

Primeira parte: Preliminares

Capítulo 1: Introdução

Mário A. Perini

1. OBJETIVOS DA PESQUISA

A pesquisa relatada neste livro tem como objetivo básico descrever a estrutura interna do sintagma nominal em português, com atenção especial à ordem dos termos.

Seguindo o uso tradicional, denominamos *sintagma nominal (SN)* qualquer construção passível de ocorrer em uma frase em função de sujeito (isto é, em relação de concordância com um verbo); o SN também ocorre em outras funções, como as tradicionalmente chamadas “objeto direto” e “regime de preposição”.

O sintagma nominal já foi objeto de diversas tentativas de descrição. Algumas foram estritamente formais, como Lemle, 1984; Perini, 1989. Outras levaram em conta fatores semânticos e/ou discursivos, como o trabalho de Waugh, 1977, para o francês. Outras, finalmente, como Lucas, 1988, estudaram o SN dentro de uma perspectiva gerativista, com intenções de depreender traços da gramática universal. Na nossa opinião, nenhuma dessas análises chega a ofere-

cer uma descrição adequada da ordem dos elementos dentro do sintagma.

Neste trabalho elaboramos uma análise que descreve a ordenação dos termos internos do SN, considerando-a resultado da interação de regras formais (“mórficas”) com regras e condições semântico-pragmáticas. Acreditamos ter argumentos que mostram ser impossível dar conta da ordenação dos termos através do uso exclusivo de regras formais sem introduzir na gramática um grau inaceitável de complexidade. Comparada às anteriores, a análise que propomos é mais adequada aos dados, e ao mesmo tempo os descreve de maneira mais simples e natural.

2. RESUMO DA ANÁLISE

Os detalhes da análise serão expostos e defendidos nos capítulos 2 a 4. Aqui daremos apenas uma idéia prévia, para facilitar a leitura.

A ordem dos termos internos do SN é resultado da ação de dois tipos de mecanismo: **regras mórficas e condições semântico-pragmáticas**, ou **CSPs**. As regras mórficas são puramente formais, e estipulam uma série de funções sucessivas dentro do SN, cada uma das quais é ocupada por um conjunto (“classe”) de itens léxicos. As funções assim estipuladas se definem pela ordem em que ocorrem, e são apenas três, a saber: **pré-determinante, determinante e termo livre**:

PDet + Det + TL^a

O termo livre pode ocorrer repetido; já o pré-determinante e o determinante são apenas um por sintagma.

Os itens léxicos são marcados quanto à função ou funções mórficas que podem ocupar no SN. Por exemplo, **todos** pode ser pré-determinante ou então termo livre; **aquele** só pode ser determinante; **amigo** só pode ser termo livre; e assim por diante.

Há algumas restrições de combinação, a saber:

(a) o PDet não ocorre sem Det (assim, temos, em português padrão, **todos os jacarés**, mas não * **todos jacarés**);

(b) o Det não ocorre sozinho no SN (assim, não ocorrem SNs compostos unicamente de artigo, nem dos itens **este**, **aquele** etc., exceto em situação anafórica);

(c) PDet + Det só ocorre se houver também TL (assim, não ocorrem SNs como * **todos os**, * **todos esses** etc. excetuando-se ainda aqui a situação anafórica).

Por outro lado, nenhum dos termos é obrigatório, tanto é que há SNs sem PDet (**os rapazes**); sem Det (**todos**; **poucos rapazes**); e sem TL (**todos**). Considerando, para simplificar, apenas SNs com um único TL, as combinações possíveis são quatro:

PDet
Det + TL
TL
PDet + Det + TL

Essa análise, naturalmente, é semelhante em princípio às análises formais já propostas. A diferença está no pequeno número de funções. Assim, um sintagma como

(1) Um engenheiro nuclear competente

se analisará da seguinte maneira:

um engenheiro nuclear competente
Det TL TL TL

Ora, sabemos que a ordenação dos três últimos termos não é indiferente:

- (2) a. * Um nuclear engenheiro competente
- b. * Um engenheiro competente nuclear
- c. * Um competente nuclear engenheiro

Essas restrições de ordenação não podem ser descritas em termos das funções mórnicas, pois os três elementos finais de (1) têm todos a mesma função, TL. Em nossa análise as restrições observadas entre os TLs se descrevem em termos de condições baseadas em traços semânticos e/ou pragmáticos dos itens em questão e do SN como um todo (essas condições compõem o que chamaremos **componente sêmico**). Nos capítulos 3 e 4 essa análise será desenvolvida e devidamente fundamentada.

3. POSTULADOS DA ANÁLISE

Quatro pontos teóricos principais norteiam nosso trabalho de descrição:

- (a) descrição superficial e conseqüente interpretação semântica das estruturas superficiais; ou seja, não aceitação de estruturas múltiplas relacionadas transformacionalmente;
- (b) consideração estritamente separada de traços de forma e traços de significado;
- (c) componente mórnico delimitado residualmente; e
- (d) descrição não distinta em componentes compactos, homogêneos e ordenados (sintaxe, semântica, pragmática). Antes, aceitação da possibilidade de que uma regra que estipula aspectos formais possa lançar mão de informação semântica; de que uma regra que interpreta semanticamente uma estrutura possa lançar mão de informação pragmática etc.

O resultado pode ser classificado, em termos gerais, como um modelo funcionalista, no sentido de que reconhece a relevância de considerações informacionais para a descrição da estrutura dos enunciados; e cognitivista, no sentido de que admite em larga escala a

relevância de traços semânticos na descrição de fenômenos tradicionalmente considerados sintáticos (em especial, a ordem dos termos). Por outro lado, igualmente admitimos a necessidade de lançar mão de recursos puramente formais; ou seja, em certos casos não se pode vislumbrar motivação semântica nem funcional para certos traços da estrutura.

Os quatro pontos acima são detalhados a seguir.

3.1. DESCRIÇÃO SUPERFICIAL

O termo **superficial** tem sido utilizado nos últimos tempos em mais de um sentido, e por isso será necessário explicitar o que se entenderá por descrição (da estrutura) superficial.

Em primeiro lugar, não se trata da estrutura-S da teoria da regência e vinculação (**government and binding**). A estrutura-S, nesse modelo, pertence a um nível bastante abstrato, incluindo, por exemplo, índices de co-referência e categorias vazias. Estamos definindo a estrutura superficial mais no espírito do nível homônimo de Chomsky, 1965.

Estrutura superficial, neste trabalho, designa um nível onde se representam, basicamente, os itens léxicos fonologicamente realizados, sua estruturação em constituintes e a ordenação destes últimos. Como se vê, não se trata de um nível composto unicamente de elementos realizados concretamente (já que a estrutura em constituintes não se manifesta na fonética). Por outro lado, é um nível extremamente concreto de análise, se comparado com os níveis “superficiais” admitidos em muitas teorias atuais.

Em particular, a estrutura superficial tal como aqui se entende não inclui os seguintes traços:

— constituintes ou categorias vazias, concebidas como sintaticamente presentes, mas sem representação fonológica;

— relações sintáticas entre estruturas formalmente diversas (isto é, relações de natureza transformacional, tais como as transforma-

ções do modelo gerativo padrão, o processo de mover-alfa da teoria GB, ou as correspondências definidas em Perini, 1989).

Os fatos usualmente capturados pela postulação de categorias vazias serão considerados parte da semântica, e tratados no componente sêmico. Assim, não figuram na estrutura superficial os elementos entre colchetes das frases seguintes:

(3) Silvinha pretende [Silvinha] ir ao concerto.

(4) Joaquim toca flautim e Anette [toca] trompete.

Os pressupostos do nosso modelo levam a analisar morficamente essas estruturas como

(3b) Silvinha pretende ir ao concerto.

(4b) Joaquim toca flautim e Anette trompete.

ou seja, atendo-nos bem mais de perto às formas observáveis. Essas estruturas sofrem a ação de regras semânticas, que atribuem ao agente “Silvinha” a ação expressa por *ir*, e a ação “*toca*” ao agente “Anette”.

Nossa opção por esse tipo de análise é motivada pela preocupação de retratar os fatos da maneira mais imediata possível; vemos como perfeitamente possível que, considerando-se os interesses da descrição completa da língua, a melhor análise seja a que postula os elementos vazios mostrados em (3) e (4). O compromisso da presente descrição é com a representação imediata dos fatos, não com a (igualmente importante) simplicidade e grau de generalidade da análise global. Este trabalho vale enquanto possibilita a representação fiel dos fatos; é um primeiro passo, indispensável mas não necessariamente suficiente, da análise.

Passando à questão das relações transformacionais, o modelo proposto não considera morficamente relevantes relações entre estruturas formalmente diferentes. Muitos modelos de análise definem

uma relação sintática entre, por exemplo, as frases de

(5) a. A cozinheira queimou o feijão.

b. O feijão, a cozinheira queimou.

Pode-se falar, nesses casos, em “transporte” do constituinte **o feijão** de uma posição pós-verbal para o início da frase. Naturalmente, falar de “transporte” só faz sentido se se admitir uma relação formal entre (5a) e (5b). Já em outros casos essa relação não se estabelece, como em

(6) a. A cozinheira desacatou o patrão.

b. O patrão desacatou a cozinheira.

Aqui nunca se diz que **o patrão**, por exemplo, foi transportado para o início da frase em (6b). Em vez disso, (6a) e (6b) são encaradas como sentenças geradas independentemente — isto é, como não tendo relação transformacional entre si.

O modelo descritivo aqui utilizado não diferencia os dois casos dentro do componente mórfico: tanto em (5) quanto em (6) temos frases independentes, que ilustram duas estruturas distintas: uma delas é exemplificada por (5b), e pode ser descrita como SN - SN - V; a outra é exemplificada por (5a), (6a) e (6b), sendo composta de SN - V - SN.

A evidente relação de significado que existe entre (5a) e (5b) fica a cargo do componente sêmico, que interpretará **o feijão** tanto em (5a) quanto em (5b) como “paciente” da ação expressa por **queimou** e praticada pela cozinheira. O fato de não se poder acrescentar um objeto a (5b):

(5c) * O feijão, a cozinheira queimou o almoço.

se explica pela saturação da grade temática de **queimar**, que comporta apenas um agente e um paciente: um dos argumentos expressos em

(5c) fica sem lugar. Trata-se do princípio expresso por Chomsky, 1982, sob o rótulo de **critério-teta**; entretanto, note-se que aqui a grade temática e suas decorrências se colocam explicitamente no componente sêmico.

Esse tipo de descrição é aqui denominado **superficial**. Mantendo os exemplos acima, diremos que em português ocorre a estrutura SN - V - SN (exemplos (5a), (6a) e (6b)); e também a estrutura SN - SN - V (exemplo (5b)).

A opção por uma descrição superficial nas linhas acima esboçadas nos leva, naturalmente, a estabelecer que a relação entre essas estruturas superficiais e suas interpretações semânticas se faz diretamente, sem passar por estruturas formais intermediárias. Por isso teremos de interpretar, por exemplo, as duas frases de (5)

(5) a. A cozinheira queimou o feijão.

b. O feijão, a cozinheira queimou.

independentemente, sem postular que **o feijão** desempenha a mesma função em ambas. Para nossos objetivos, **o feijão** tem funções diferentes em (5a) e (5b), por causa de sua diferente relação de ordem com os demais termos. Por conseguinte, teremos que relacionar, nas regras semânticas, a interpretação “paciente de **queimar**” com duas funções morficamente distintas.

3.2. O COMPONENTE MÓRFICO COMO RESÍDUO

3.2.1. Descrição fonológica, semântica e morfossintática

O estudo da organização mórfica da frase (freqüentemente, mas nem sempre, identificado com a sintaxe) difere da fonologia e da semântica por ser exclusivamente relacional. A fonologia se relaciona de maneira bastante direta com os sons (ver a respeito o capítulo 9 de Chomsky & Halle, 1968); e a semântica se relaciona com os significados. E tanto os sons quanto os significados podem ser consi-

derados entidades observáveis.

Por exemplo, é um fato inegável, independente de teorização propriamente lingüística, que a palavra **barriga** se inicia unindo os lábios e fazendo vibrar as cordas vocais; e é igualmente pré-teórica (lingüisticamente) a observação de que essa palavra designa uma parte do corpo. Pode-se dizer, então, que tanto a fonologia quanto a semântica tratam da relação da língua com o mundo extra-lingüístico; por isso mesmo não se podem considerar “puramente” lingüísticas.

Já o componente mórfico depende em grau muito maior de teorização lingüística; tipicamente, as afirmações feitas nesse componente não se relacionam diretamente com o mundo extra-lingüístico. Dizer que em **esses livros verdes** as palavras **esses e verdes** “concordam” com a palavra **livros** não pode ser considerado um fato inegável, pré-teórico; antes, é resultado do estabelecimento de diversas entidades de valor puramente lingüístico, tais como “palavra”, “concordar” etc.

Podemos exprimir isso dizendo que a semântica estuda a relação entre a estrutura da língua e os conceitos (esquemas) presentes na memória dos falantes; a fonologia estuda a relação entre a estrutura da língua e os sons. O componente mórfico (que para muitos será co-extensivo com a morfossintaxe) estuda relações entre entidades lingüísticas, sem consideração de sons ou de significados. Esse componente estuda **exclusivamente** traços da estrutura da língua; em outras palavras, a descrição mórfica é puramente relacional.

3.2.2. Traços de forma e traços de significado

Um fato da maior importância metodológica é que os falantes podem ter, pelo menos até certo ponto, acesso intuitivo à diferença entre **traços de significado e traços de forma**. Isso não significa que um falante, ou mesmo um lingüista treinado, possa dizer com segurança, em todos os casos, que determinado fenômeno se deve colocar neste ou naquele componente da gramática; essa é, novamente, uma decisão que depende de teorização. Mas significa que, em muitos casos, conseguimos identificar um ingrediente de significado evidente,

impossível de ignorar, em um fenômeno lingüístico em estudo.

Vejamos um exemplo: considerando os sintagmas

(7) a. Oração subordinada adverbial

b. * Oração adverbial subordinada

podemos observar que na versão aceitável, (7a), os dois adjetivos (que funcionam como “termos livres”, segundo a análise do Cap. 3) se ordenam de tal forma que o segundo deles é mais restritivo que o primeiro: toda oração adverbial é necessariamente subordinada (as orações adverbiais são um subconjunto das subordinadas). Essa é uma observação imediata e inegável, para quem domina a nomenclatura gramatical; não depende de teorização lingüística prévia, e faz parte da interpretação normal do sintagma, estando portanto necessariamente à disposição do falante.

Acabamos de considerar (7) do ponto de vista do significado, e depreendemos um de seus traços de significado, a saber, **a relação de inclusão existente entre os dois T_Ls**. Vamos agora considerar (7) do ponto de vista da forma.

Uma observação imediata é que (7a) difere de (7b) quanto à ordem de seus constituintes: os elementos **subordinada** e **adverbial** trocam de lugar. Essa é uma observação formal imediata que se pode fazer acerca de (7), e não depende de teorização.

Outras afirmações de interesse dependem da aceitação de alguma teoria. Mas uma boa parte da teoria é não controversa em lingüística; em particular, a imensa maioria admitiria que (7) se compõe de três “palavras”, a saber, **oração, subordinada, adverbial**. Louvando-nos nesse consenso, podemos ir adiante, dizendo que “(7a) difere de (7b) na ordem de suas palavras”. A isso chamaremos um traço de forma.

Para efeitos práticos, portanto, tanto o traço de significado “o significado de **adverbial** é mais restritivo do que o de **subordinada**” quanto o traço de forma “a palavra **adverbial**, em (7a), vem depois da palavra **subordinada**” podem ser considerados como de observa-

ção imediata. O primeiro o é por fazer parte da interpretação normal do sintagma. O segundo depende não só de observação imediata, mas também de uma noção teórica, a definição de “palavra”; mas essa noção teórica pode, na prática, ser considerada como de domínio comum, não controverso.

O que queremos dizer com esse arrazoado é que, ao considerarmos uma forma lingüística complexa, dispomos de um grupo de observações imediatas sobre as quais podemos basear uma análise; e que essas observações se dividem nitidamente em duas categorias: **traços de significado e traços de forma**. Sustentamos que é possível distinguir intuitivamente os traços de significado dos traços de forma; e que em muitos casos os traços de significado são bastante claros. E’ verdade que muitos outros traços de significado são difíceis de depreender, em parte por causa de nosso deficiente conhecimento da semântica. Baste-nos, por ora, a observação de que muitos desses traços são suficientemente claros para serem úteis na análise.

3.2.3. Sua relação com os componentes da gramática

E’ preciso distinguir cuidadosamente entre, de um lado, a identificação de um traço da língua como **de forma** ou **de significado** e, de outro lado, a colocação de um traço no componente **mórfico (CM)** ou no **sêmico (CS)** — ou, equivalentemente neste caso, na sintaxe ou na semântica.

A primeira dessas decisões pode ser feita em termos intuitivos: ninguém sustentaria que a relação entre a palavra **barriga** e uma parte do corpo seja um traço de forma; é um traço que se impõe imediatamente (intuitivamente) como de significado, ou seja, como de relação com conceitos. Já a decisão de se colocar um traço em um ou outro componente decorre de posicionamentos de caráter teórico: o componente mórfico e o sêmico são componentes da descrição gramatical, e esta é uma construção teórica. Não se pode dizer, portanto, que um falante tenha acesso intuitivo à diferença entre componente mórfico e componente sêmico (ou entre sintaxe e semântica). Conseqüentemente, a colocação de muitos traços neste ou naquele com-

ponente é controversa; e é igualmente controversa a própria necessidade de distinguir os dois componentes.

As duas questões (separação **traços de forma / traços de significado** e separação **componente mórfico / componente sêmico**) são, em princípio, independentes. A separação entre traços de forma e traços de significado é uma posição metodológica previamente assumida, motivada por uma concepção particular dos objetivos da investigação lingüística: a concepção de que a lingüística trata primariamente da relação entre traços de forma (as “imagens acústicas” de Saussure, 1916) e traços de significado (os “conceitos”). Já a separação entre os componentes mórfico e sêmico (ou entre sintaxe e semântica) como componentes da gramática é um dos aspectos dessa relação a serem investigados. Em outras palavras, uma das perguntas que o lingüista se deve fazer é:

Será necessário colocar alguns traços de forma em um componente separado da gramática (o “componente mórfico”)? Ou será que todos os traços de forma podem ser previstos a partir da descrição do significado?

A resposta, até onde podemos ver, é que o componente mórfico é necessário; as razões serão vistas no decorrer deste trabalho.

3.2.4. O princípio do “CM residual”

A linha divisória entre sintaxe e semântica é objeto de controvérsia entre os lingüistas; e não é em absoluto nosso objetivo aqui contribuir para a solução de um problema tão complexo. Por outro lado, o lingüista descritivo tem necessidade de pontos de referência que guiem sua análise, entre estes um critério de distinção entre “fatos mórficos” (ou “formais”) e “fatos sêmicos”. Vamos propor adiante um critério desse tipo, a ser entendido como instrumento metodológico — uma hipótese de trabalho, útil na sistematização dos fatos observados. A questão da separação entre os componentes da gramática continuará em aberto, naturalmente; estamos apenas tentando cons-

truir um modelo de apresentação de dados.

Voltemos aos exemplos (7)

(7) a. Oração subordinada adverbial

b. * Oração adverbial subordinada

Vimos acima que há dois traços bastante evidentes nesses sintagmas: primeiro, um dos dois termos em questão tem significado mais restritivo do que o outro (traço de significado); depois, um deles vem necessariamente em último lugar (traço de forma).

Podemos estabelecer uma relação entre esses dois traços: o termo (TL) de significado mais restritivo vem depois do de significado menos restritivo (isso será denominado, no capítulo 4, **Condição de Restritividade Crescente**). E, como veremos, essa relação é de validade geral: sempre que dois TLs se seguem, e um deles tem significado mais restritivo, a ordem aceitável é apenas o mais restritivo depois do menos restritivo. Em outras palavras, em (7) o **traço de forma observado é previsível a partir do traço de significado observado**: conhecendo a relação de restritividade dos dois TLs podemos prever sua ordenação.

Vamos confrontar essa situação com outra; sejam as frases

(8) a. Carolina gosta de Jarbas.

b. Carolina ama Jarbas.

Um dos traços de forma que distinguem as duas frases é a presença, em (8a), da palavra **de** antes de **Jarbas**; sabemos que esse **de** é obrigatório ali. Já em (8b) não se pode colocar **de**:

(9) a. * Carolina gosta Jarbas.

b. * Carolina ama de Jarbas.

No entanto, não foi possível até hoje determinar nenhum traço de significado que corresponda à exigência ou não da presença de **de** com certos verbos. No caso de (8), inclusive, pode-se dizer que as duas frases têm significados muito próximos (são quase sinônimas); pelo menos, a relação semântica entre **gosta** e **Jarbas** em (8a) é a mesma que existe entre **ama** e **Jarbas** em (8b). Temos aqui, portanto, um traço de forma ao qual não corresponde nenhum traço de significado. Isto é, em (8) o traço de forma observado **não** é previsível a partir de nenhum traço de significado — situação oposta à que se verificou com os exemplos (7).

Outro exemplo, também relacionado com a distribuição das preposições: que traço semântico poderá ajudar a prever as estranhas restrições que regem o uso das preposições com as divisões do dia? Com efeito, temos:

*de manhã / pela manhã / * à manhã*
*de dia / * pelo dia / * ao dia*
de tarde / ? pela tarde / à tarde
*de noite / * pela noite / à noite*

Aqui, novamente, teremos de dar conta dessas diferenças através de marcas formais, desprovidas de motivação semântica.

Neste trabalho chamamos **componente sêmico (CS)** o conjunto dos dispositivos gramaticais que descrevem os significados e a relação destes com as formas; e **componente mórfico (CM)** o conjunto de dispositivos gramaticais que descrevem a estruturação das formas, naquilo em que essa estruturação não decorre de fatores semânticos ou discursivos. Com isso estaremos seguindo uma tradição comum em lingüística (embora em geral seja colocada em termos de “sintaxe” e “semântica” — termos que evitamos pelas razões dadas na seção precedente).

Para efeitos da descrição superficial que objetivamos, pode-se então formular o seguinte princípio de análise:

Princípio do CM Residual

Analisa-se como fazendo parte do *componente mórfico* somente as particularidades formais que não sejam previsíveis a partir do conhecimento de seu significado (incluindo-se no significado certos traços de seu “status” discursivo).

Assim, a exigência da presença de uma preposição com o complemento de certos verbos (como **gostar**) é um traço a ser incluído na análise mórfica desses verbos, e das estruturas em que ocorrem. Mas a ordem dos TLs ilustrada nos exemplos (7) não deve ser incluída no componente mórfico; antes, fará parte da descrição sêmica das estruturas em questão. Relembro que **componente sêmico** aqui significa o sistema que relaciona formas e significados, e não a descrição pura e simples dos significados (conceitos, esquemas).

A maneira como esse princípio se manifesta na análise será amplamente exemplificada nos capítulos 3 e 4. Aqui daremos apenas uma idéia rápida.

No caso da exigência da preposição no complemento de **gostar** (traço semanticamente imprevisível, e portanto mórfico), teremos de lançar mão de marcas idiossincráticas, não motivadas fora do componente mórfico. O verbo **gostar** será marcado, no léxico, como exigindo complemento com a preposição **de**, e recusando complemento sem preposição. Essas são marcas **formais**, isto é, para o que nos interessa, **mórficas**.

Já no caso dos exemplos de (7)

(7) a. Oração subordinada adverbial

b. * Oração adverbial subordinada

analisaremos sintaticamente tanto **subordinada** quanto **adverbial** como TLs — isto é, ambos têm a mesma função, e portanto podem, morficamente falando, ocorrer em qualquer ordem: o componente mórfico não proíbe a ocorrência de (7b).

Mas, naturalmente, uma construção só é aceitável se for bem for-

mada de acordo com todos os componentes da gramática. E (7b) é semicamente mal formada, em virtude da Condição de Restritividade Crescente, mencionada acima, que estipula que se um TL tem referência necessariamente mais restritiva do que outro TL, então o mais restritivo deve ocorrer depois do menos restritivo. A ordenação oposta resultará em interpretação semântica mal formada — e é essa a razão da inaceitabilidade de (7b).

Essa proposta resulta em um **componente mórfico residual**, definível aproximadamente como o repositório dos traços formais não previsíveis a partir de correlatos de significado. O componente mórfico é encarregado de estipular as estruturas permissíveis na língua, e consta de uma série de instruções, ou regras, para se construir uma estrutura (oração, sintagma) bem formada. O resultado da aplicação do CM é um conjunto de estruturas.

Mas nem todas essas estruturas serão aceitáveis; são apenas formalmente bem construídas. O componente sêmico lhes atribuirá significados (interpretações), e poderá identificar ainda casos de má formação. Por exemplo, um sintagma como

(10) * Um rapaz grávido

é inaceitável porque a interpretação semântica que recebe é mal formada, em vista do que sabemos sobre certos aspectos do mundo: só mulheres (e fêmeas em geral) podem ficar grávidas. Mas pode-se sustentar que (10) é morficamente bem formado. Como se vê, a inaceitabilidade de uma forma não implica em má formação mórfica: implica apenas em má formação em algum nível da gramática. (10) foi gerado normalmente no componente mórfico, mas o componente sêmico o rejeitou.

O mesmo se dá com

(7b) * Oração adverbial subordinada

A interpretação semântica de (7b) é (para quem conhece algo de análise sintática) mal formada: corresponderia, *grosso modo*, a dizer-

mos de uma oração que é “adverbial, e além disso subordinada” — quando sabemos que toda adverbial é necessariamente subordinada. A má formação de (7b) decorre de uma condição semântica que interpreta sempre o último TL como mais restritivo do que o precedente — o que, no caso, resulta em má formação, pois **subordinada** não pode ser entendido, em se falando de orações, como mais restritivo que **adverbial**.

Em resumo, o componente mórfico seleciona certas seqüências como bem formadas — automaticamente marcando as demais seqüências possíveis como mal formadas.

O componente sêmico atribui interpretações às seqüências sintaticamente bem formadas, segundo regras próprias. Algumas dessas interpretações podem ser, por sua vez, consideradas mal formadas em virtude de condições semânticas de boa formação. Uma seqüência que não possa receber nenhuma interpretação bem formada será excluída por essas condições de boa formação, que atuam como uma espécie de controle de qualidade final. Como resultado, só serão aceitáveis as seqüências que forem produzidas (e não filtradas) pelo componente mórfico, e que receberem interpretação semântica não anômala.

Uma conseqüência do princípio do CM residual é que nem todos os traços de forma podem ser considerados fenômenos mórficos; por exemplo, a ordenação dos TLs em (7), embora descrita em termos sêmicos, continua presente como traço formalmente observável. Ou seja, não podemos definir o componente mórfico simplesmente como o estudo dos traços de forma. Por outro lado, todo traço de significado figurará na gramática como parte do componente sêmico (que, como se verá no capítulo 4, inclui certos aspectos do que tradicionalmente se chama “discurso” e “pragmática”).

A presença de um traço de forma não faz com que sua descrição direta seja interessante. Por exemplo, sabemos que o verbo **beber** seleciona determinados objetos; assim, aceita **água** e **refrigerante**, mas rejeita **farinha** e **empadão**. Esse é um traço de forma, e poderia ser descrito diretamente (por exemplo, poderia ser incluído nas restrições de transitividade do verbo **beber**). Não fazemos isso, contu-

do, porque o traço de forma em questão é totalmente previsível a partir da semântica de **beber**, mais a semântica do objeto direto de **beber**, mais a semântica de nomes como **água**, **farinha** etc. Por conseguinte, explicitar formalmente a impossibilidade de **farinha** ser objeto de **beber** seria no mínimo redundante.

No outro extremo temos casos como os dos exemplos (8) e (9), ou a distribuição das preposições com os itens **manhã**, **tarde**, **noite**: o traço de forma observado não é previsível a partir dos traços de significado desses itens e de suas funções. Por isso, devem ser descritos diretamente, como traços de forma, e incluídos no componente mórfico.

3.2.5. Vantagens e limitações

Um traço evidente da delimitação proposta entre componente mórfico e componente sêmico é que a qualquer momento um fato inicialmente colocado no componente mórfico poderá ser posto em correlação com novos fatos semânticos (ou discursivos) que permitam sua previsão — e, por conseguinte, esse fato deverá ser eliminado do componente mórfico. Ou seja, os limites entre os componentes serão sempre provisórios.

Esse caráter provisório, a nosso ver, não prejudica a expressividade das descrições. O que procuramos com o presente modelo descritivo é uma maneira de sistematizar fatos (formais e semânticos); sua colocação neste ou naquele componente é secundária. Assim, nos exemplos (7) e (8) acima os traços de forma estão preservados, sendo acessíveis ao investigador interessado: no caso de (8), a exigência de preposição consta das restrições mórficas a que é submetido o verbo **gostar**; e no caso de (7) a ordenação dos TLs, embora não presente no componente mórfico, pode ser derivada da interpretação semântica dos mesmos, mais a Condição de Restritividade Crescente.

3.3. SEMÂNTICA E DISCURSO COMO UM COMPONENTE ÚNICO

Ao iniciarmos este estudo, tínhamos adotado o pressuposto de que é possível descrever os aspectos centrais do significado das formas sem considerar unidades maiores que o constituinte em questão — ou, pelo menos, sem considerar traços externos à oração. Seguíamos a opinião, freqüente entre os lingüistas, de que o significado compreende um ingrediente computável a partir da estrutura sintática oracional; e que esse ingrediente interage com fatores discursivos para fornecer a interpretação final.

No entanto, a tentativa de descrever as condições que regem a ordem dos elementos no SN acabou levando-nos a questionar seriamente esse pressuposto. E, embora não consideremos a questão solucionada, fomos obrigados a abandonar a separação estrita entre os componentes semântico e discursivo, para adotar um modelo em que o significado só pode ser computado a partir de informações provenientes tanto da oração quanto de traços do contexto.

Como se verá, nem todas as restrições de ordem que funcionam entre os termos do SN podem ser explicadas em termos do componente mórfico. Em alguns casos uma ordenação é interdita em virtude de uma regra mórfica que estipula a ordenação oposta; mas em outros casos é bastante evidente que uma das ordenações possíveis é interdita porque resulta em má formação semântica. A ordenação observada é, por conseguinte, resultado da interação de regras mórficas com condições que atuam sobre aspectos do significado.

Essas condições que atuam sobre o significado por vezes levam em conta elementos do contexto extra-oracional, merecendo por isso o rótulo de **discursivas**. Já outras se formulam em termos intra-oracionais (geralmente em termos de traços semânticos de um item léxico), podendo ser consideradas simplesmente **semânticas**. Ora, concluímos ser necessário admitir que todas essas condições (semânticas e discursivas) precisam aplicar-se no final das derivações; ou seja, não existe um resultado independente das condições semânticas, sobre o qual atuariam as condições discursivas. Assim, não há,

para efeitos da aplicação dessas condições, um limite entre semântica e discurso, análogo ao limite que admitimos existir entre semântica e sintaxe (ou entre os componentes sêmico e mórfico).

Por isso, colocamos todas as condições, independentemente de sua formulação ser semântica ou discursiva, em um componente único, denominado **sêmico** — uma instância única de controle de qualidade, que inspeciona as interpretações atribuídas às estruturas mórficas, marcando, se for o caso, algumas como mal formadas.

Uma consequência importante dessa posição é termos de admitir que a interpretação semântica completa de uma estrutura só se realiza no contexto. Isso não nos impede de distinguir, por um lado, traços semânticos (definíveis intra-oracionalmente) e, por outro lado, traços discursivos; nem nos impede de falar em “condições semânticas” em oposição a “condições discursivas”. No entanto, tais termos se referem apenas à **formulação** dessas condições; do ponto de vista de seu ponto de atuação nas derivações, as condições são sempre **semântico-discursivas**, e assim serão chamadas formalmente.

Aceitamos portanto as duas hipóteses seguintes: (a) a ordem dos termos depende em parte de fatores semânticos, e (b) os fatores semânticos dependem em parte de fatores discursivos. Essas hipóteses acarretam uma consequência de natureza metodológica: só com cuidado especial se pode utilizar julgamentos de aceitabilidade de elementos fora de contexto.

Como se vê, a reunião de semântica e discurso tem consequências muito amplas dentro da teoria e da metodologia. Por isso, adotaremos aqui uma posição de prudência, afirmando apenas que **as condições extramórficas que regem a ordenação dos termos do SN são de natureza semântico-discursiva**. Trata-se, claro, de uma parcela de evidência em favor de se unificar de vez os dois componentes; mas não nos parece suficiente para liquidar a questão. A hipótese é adotada aqui por causa de suas evidentes vantagens descritivas; fica para outra ocasião a tentativa de inseri-la no contexto de um debate teórico sobre a separação dos componentes da gramática.

Há outras observações que parecem favorecer essa junção de

componentes — ou, pelo menos, que sugerem que muito do que se trata aqui como fenômenos semânticos mais ou menos “puros” tem na verdade raízes na pragmática. E’ o caso, por exemplo, da condição que denominamos **Condição de Restritividade Crescente**. Essa condição, como se viu acima, estipula que quando em um sintagma há dois TLs que denotam subconjuntos do mesmo conjunto o de significado mais restritivo deve sempre ocorrer à direita do de significado menos restritivo. O exemplo citado foi

(7) a. Oração subordinada adverbial

b. * Oração adverbial subordinada

Como **adverbial**, nesse contexto, denota um subconjunto das subordinadas, sendo portanto mais restritivo, deve ocorrer sempre depois de **subordinada**.

Mas essa condição pode ser derivada de uma das máximas do discurso propostas por Grice, 1982, a saber, a Máxima de Quantidade, assim formulada:

“1. Faça com que sua contribuição seja tão informativa quanto requerido (para o propósito corrente da conversação);

2. Não faça sua contribuição mais informativa do que é requerido.”

[Grice, 1982 (1967)]

Ora, ao dizermos * **oração adverbial subordinada**, o último termo é redundante, justamente porque as adverbiais são um subconjunto das subordinadas: ou seja, uma adverbial só pode ser subordinada, e portanto o termo **subordinada** está dando mais informação do que o necessário. O sintagma viola a Máxima de Quantidade, e é sentido como mal formado.

Pode pois valer a pena examinar as condições semânticas propostas no capítulo 4 para verificar se algumas delas não serão de na-

tureza pragmática, e não de natureza semântica no sentido tradicional.

3.4. PODE-SE SUPRIMIR O COMPONENTE MÓRFICO?

Dentro dos pressupostos da presente análise, é de se esperar que algumas das regras do componente mórfico venham a ser, eventualmente, transferidas para o componente sêmico, por se descobrirem novas condições de ordem semântica ou discursiva que as prevêm. Vale perguntar, então, se essa tendência não poderá, pelo menos em princípio, ser levada a sua conclusão lógica, eliminando-se totalmente o componente mórfico; ou seja, será possível, um dia, descrever a ordem dos termos utilizando-se exclusivamente condições semântico-pragmáticas?

Alguns autores parecem convencidos de que isso é possível. Para Reid, 1991, as regras da sintaxe são “arbitrárias” porque “não decorrem de nada mais que o falante sabe” (p. 27), ao passo que os princípios pragmáticos são motivados em outros conhecimentos. E ele propõe o “abandono da sintaxe” como uma meta alcançável (ver também García, 1979).

E’ verdade que as regras mórficas (“sintáticas”) não encontram motivação fora de seu próprio componente, e talvez possam por isso ser consideradas “arbitrárias” no sentido de Reid (esse, a nosso ver, não é o mesmo sentido dado ao termo por Saussure). No entanto, não acreditamos que tais “arbitriedades” possam vir a ser inteiramente dispensadas na análise da língua, nem em particular na análise da ordem dos termos do SN ou da oração.

Primeiro, o componente mórfico não está totalmente explicado em termos semântico-pragmáticos, e é perfeitamente possível que nunca venha a sê-lo. Por exemplo, a regra que exige que o determinante venha antes do TL não decorre de nenhum princípio semântico conhecido; e não parece ser sensível a traços do contexto (o SN * **casa** a é claramente inaceitável em qualquer contexto). Não nos sentimos autorizados a preconizar um abandono de regras, ainda que estas não tenham motivação extramórfica, em favor de princípios e regras ain-

da não descobertos.

Há muito o que discutir quanto a esse ponto, e não gostaríamos de dar a impressão de que se trata de ponto pacífico para nós. Mas a questão das dimensões a serem atribuídas aos diferentes componentes (mórfico, sêmico) é uma questão empírica, que só a pesquisa futura poderá responder; não há possibilidade de encontrar uma resposta baseada em princípios *a priori*. E, por outro lado, a evidência até agora disponível sugere que o componente mórfico é indispensável.

4. EMBLOCAMENTO

A formulação das condições que integram o componente sêmico depende em parte de uma relação semântica entre os itens do sintagma, relação essa que denominamos **emblocamento**. Vamos agora definir essa relação.

Muitas condições têm um importante traço em comum, a saber, parecem descrever situações em que determinadas seqüências podem ser facilmente interpretadas como uma unidade. Por exemplo, comparem-se os sintagmas abaixo:

(11) Os pretensos muitos namorados de Laurinha

(12) ?? Os ricos muitos namorados de Laurinha

A análise mórfica desses sintagmas é idêntica: ambos representam uma seqüência de determinante, mais quatro ocorrências de TL. No entanto, (11) é nitidamente mais aceitável do que (12).

Propomos que a razão dessa diferença de aceitabilidade tem a ver com a presença em (11), logo antes do item **muitos**, de um item (**pretensos**) que exprime a opinião do falante quanto à existência de alguma coisa, ao passo que o item correspondente de (12), **ricos**, não tem esse significado.

Note-se que “pretensos” em (11) não se predica dos namorados de Laurinha (isto é, o falante não exprime ceticismo sobre Laurinha ter namorados), mas da quantidade destes (ceticismo quanto a eles

serem muitos). Em outras palavras, a predicação expressa por **pretensos** incide não sobre o significado de **namorados**, mas sobre o da seqüência **muitos namorados**. O resultado é bem formado porque não há nada de estranho em predicar “pretensos” de “muitos namorados” — são “pretensamente muitos”.

Já em (12) a situação é diferente. Seria fácil predicar “ricos” de “namorados”, como em **os ricos namorados de Laurinha**, sintagma perfeitamente aceitável. Mas em (12) temos que predicar “ricos” de “muitos namorados”: deveria significar que eles são “ricos enquanto muitos”, ou “ricamente muitos”, o que não parece fazer sentido.

A esse fenômeno denominamos **emblocamento**. Diremos então que a presença de um item como **pretensos** (que exprime uma opinião do falante) possibilita o emblocamento de uma seqüência de **muitos** + TL. Se em vez de **pretensos** tivermos **ricos**, o emblocamento não se verifica, ou é dificultado.

O emblocamento, tal como o concebemos, é um fenômeno semântico, e não decorrência automática da estruturação de constituintes. Preferimos pensar assim porque não nos parece plausível que (11) e (12) tenham estruturas *mórficas* distintas. No entanto, o mecanismo de interpretação trata diferentemente os dois sintagmas, em virtude de certos traços *semânticos* que os distinguem.

5. DELIMITAÇÃO DO ESTUDO

5.1. “NOMINAIS”

O presente estudo considera, dentro do léxico da língua, apenas os itens “nominais” — o que inclui os “substantivos”, “adjetivos”, “artigos”, “numerais” e “pronomes” da gramática tradicional.

Essa delimitação foi estabelecida *a priori* na crença de que a classificação tradicional, nesse particular, capta algo da estrutura da língua. Por outro lado, os detalhes da classificação final serão, com toda certeza, profundamente diferentes dos da classificação tradicional, especialmente no que diz respeito à lista das subclasses em que se dividem os nominais, assim quanto aos limites entre elas. Basta con-

siderar, por exemplo, a falta de clareza das classificações tradicionais na distinção entre “adjetivos” e “substantivos”.

Já no que diz respeito ao englobamento das cinco classes tradicionais em um tipo geral de itens (os “nominais”), acreditamos que a intuição tradicional será basicamente confirmada: os nominais realmente têm todos algo em comum, que os diferencia nitidamente dos verbos, advérbios, preposições etc. Esses traços comuns são não apenas morfológicos, mas também semânticos. Ou seja, acreditamos que faz sentido, para a descrição da língua, falar em “nominais” como um grupo de itens.

5.2. OS MESMOS ITENS FORA DO SN

Muitos dos itens aqui estudados podem também ocorrer fora de sintagmas nominais. Por exemplo, o item **nosso** pode ocorrer no SN (com a função de **termo livre**, como em **nossa casa**), e também como **complemento do predicado** (“predicativo do sujeito”), como em

(13) Esse sítio é **nosso**.

onde presumivelmente não constitui SN.

Na presente pesquisa essas ocorrências fora de SNs não são consideradas. Os itens são estudados exclusivamente com referência a sua ocorrência dentro de SNs.

5.3. SINTAGMAS PREPOSICIONADOS NO SN

E’ preciso ainda notar que só incluímos na análise do SN, neste estágio inicial, termos formados de palavras individuais, deixando de lado os que são formados de sintagmas preposicionados, como **de Maria em**

(14) **A casa de Maria**

O estudo dos sintagmas preposicionados dentro do SN será even-

tualmente atacado, não só por sua óbvia importância para a compreensão da estrutura do SN, mas também porque seu comportamento parece ser em grande medida paralelo ao dos itens que podem desempenhar a função de TL. Ao que tudo indica, também os sintagmas preposicionados funcionam como TLs, sendo morficamente semelhantes a itens como **materno**, por exemplo.

5.4. EXPANSÕES

Deixamos também de lado, para pesquisa futura, as possibilidades de expansão dos diferentes termos através de elementos não nominais, como por exemplo **só**, **exatamente** e **muito** nos sintagmas seguintes:

(15) *Só um envelope / um só envelope*

(16) *Exatamente este lugar*

(17) *Café muito forte*

(18) *Muito poucos sinais*

5.5. EXCLUSÃO DO CONTEXTO ANAFÓRICO

Finalmente, é preciso observar que não levamos em conta em nossa análise a possibilidade de ocorrência de SNs reduzidos em contexto anafórico. Assim, diremos que o SN

(19) * *Algumas*

é mal formado, muito embora possa ocorrer, como em

(20) *As operárias se rebelaram; algumas se recusaram a comparecer às chamadas.*

Essa limitação é algo desconfortável, em vista de nossa opção pela não separação entre semântica e discurso em nosso modelo. No entanto, trata-se de um fenômeno relativamente bem delimitado, e há esperanças de que possa ser eventualmente incorporado à análise sem grandes dificuldades. Os SNs anaforicamente reduzidos poderão ser incorporados à análise quando se estudar o fenômeno geral da redução anafórica; e então, acreditamos, eles serão estudados como um caso especial das reduções impostas pelo contexto anafórico a estruturas plenas em geral.

Todas essas limitações são inevitáveis ao se encetar o estudo de uma área tão mal conhecida. Esperamos produzir uma análise que ofereça possibilidades imediatas de expansão, levando a descrições cada vez mais completas do comportamento dos nominais em português.

6. RESULTADOS

6.1. SÍNTESE DOS RESULTADOS

O princípio do CM residual (apresentado em 3.2.4), com sua consequência — a consideração bastante ampla de fatores semântico-discursivos na definição da ordem dos termos no SN — levou à elaboração de uma análise bastante diferente da habitual e, acreditamos, mais reveladora e interessante do que ela.

As análises prévias da estrutura interna do SN (como a de Lemle, 1984, ou a de Perini, 1989, retomada em Perini, 1995¹) lançaram mão exclusivamente de restrições de ordem sintática (isto é, formal), independentes de qualquer motivação extra-sintática. Assim, em Perini, 1989, afirma-se que o SN se compõe de sete termos estritamente ordenados. A ordenação não tem motivação extra-sintática, sendo considerada idiossincrática; assim, por exemplo, a ordem **Quantificador - Pré-núcleo** (funções definidas em Perini, 1989) é determinada por regras sintáticas, e nada mais há a dizer a respeito.

Encontramos dificuldades com análises desse tipo assim que ten-

tamos aplicá-las a um número maior de exemplos. As exceções se multiplicaram, e logo se evidenciou que o uso exclusivo de restrições formais não era o melhor caminho para descrever os fatos. Partindo dessa observação, obtivemos resultados bem melhores utilizando também restrições de base semântica e discursiva à ordenação — e, concomitantemente, simplificando consideravelmente o componente formal (aqui denominado “mórfico”). As restrições formais não foram excluídas de todo, e acreditamos que algumas são necessárias: como já foi apontado, a posição inicial do determinante não tem, que saibamos, nenhuma motivação de ordem semântica.

Por outro lado, em numerosos casos conseguimos formular condições semântico-discursivas que funcionam para restringir as possibilidades de ordenação. Por exemplo, os TLs freqüentemente têm ordenação rígida, como em

(21)a. Uma moça alta belíssima

b. * Uma moça belíssima alta

Nas análises anteriores só se poderia dizer que essa ordenação é estipulada idiossincraticamente: haveria, digamos, duas funções após o “núcleo” (Modificadores I e II); o item **alta** só poderia ter a função de Mod I, e **belíssima** só a de Mod II.

Nosso estudo desses fenômenos revelou a existência nesses casos de uma condição de base não-formal sobre a ordenação, que estabelece, em síntese, que um TL com significado superlativo nunca pode ocorrer antes de um TL com significado não superlativo. Isso nos permite simplificar o CM, admitindo apenas um tipo de TL (que pode ocorrer repetido).

Como se verá, existem várias condições semelhantes a essa; e sua interação com as restrições puramente mórficas forma um sistema muito flexível, que a nosso ver chega bem mais perto de dar conta dos fatos observados do que as propostas anteriores.

O resultado é um sistema complexo, mas de grande interesse, que funciona com base na interação de restrições formais e restrições

semântico-discursivas. Dado o estado do nosso conhecimento da estrutura semântica da língua, essas condições foram formuladas de maneira bastante rudimentar; e certamente deve haver outras que nos escaparam. Vemos aqui um campo promissor para pesquisas futuras.

6.2. INTERPRETAÇÃO TEÓRICA

O exame dos dados que nos levou aos resultados acima resumidos sugeriu a hipótese que está na base de nossa análise. E' evidente que essa hipótese não nasceu espontaneamente dos dados; mas, uma vez formulada, foi corroborada a ponto de merecer explicitação como parte dos próprios resultados.

A hipótese a que nos referimos é a seguinte:

A ordenação dos termos dentro do SN é resultado de dois fatores principais:

(a) as funções mórficas desempenhadas por esses termos;

e

(b) o efeito de condições semântico-discursivas.

Assim a ordenação de *o* e **nosso** é estrita em

(22)a. O nosso clube

b. * Nosso o clube

e isso se deve à diferença de funções dos dois elementos: *o* é determinante, e **nosso** é termo livre; e o componente mórfico estabelece que no SN o determinante vem sempre antes do termo livre. De acordo com o princípio do CM residual, esse fenômeno se coloca no componente mórfico porque não foi possível descobrir nenhum correlato semântico-discursivo para ele.

Por outro lado, a ordenação observada em

(21)a. Uma moça alta belíssima

b. * Uma moça belíssima alta

decorre, como vimos, não de diferença de função entre os termos **alta** e **belíssima** (que são ambos termos livres), mas da condição que estipula que um termo livre com significado superlativo não pode vir antes de outro termo livre com significado não superlativo. A interação entre as regras mórficas e as condições semântico-discursivas dá como resultado as relações de ordem observadas dentro do SN.

NOTAS

¹ Devemos esclarecer, em nossa defesa, que esse livro foi terminado em 1992, quando apenas se iniciava o trabalho no presente projeto.