

Noções Básicas de Análise Sintáctica

1. Introdução

A análise linguística tem como objectivo descrever os mecanismos de produção e compreensão da fala, determinar, pois, a natureza da linguagem humana. A abordagem empírica ao estudo da linguagem consiste em detectar as **regularidades** presentes nas produções verbais humanas. Num primeiro momento, essas produções parecem ser extremamente numerosas e de uma grande heterogeneidade, uma vez que elas variam consideravelmente de um locutor para outro (devido a factores sociais, dialectais, etc.). Além disso, a observação das produções linguísticas de um dado locutor mostra que ele modifica a sua linguagem em função dos seus interlocutores, da sua disposição momentânea, de modas temporárias, etc.

Contudo, fazendo abstracção dessas variações, os gramáticos reconheceram desde muito cedo a presença de elementos recorrentes (estáveis) nas sequências fónicas produzidas por todos os locutores. Observaram que existiam:

- **unidades de significado constantes**, o que os conduziu à identificação de palavras e à sua representação por sinais (por exemplo, os hieróglifos do egípcio, os ideogramas do chinês);
- **elementos fónicos estáveis**, que puderam ser representados por outros sinais, que deram origem, em primeiro lugar, à constituição de listas silábicas, que depois se foram transformando em listas de sons elementares, transcritos em alfabetos.

Uma vez que a matéria fónica é, de facto, a única que é concreta e natural, é lógico que tenham sido estas regularidades as primeiras a serem observadas. Mas, apesar de tudo, foi necessária uma etapa de abstracção considerável para detectar essas regularidades a partir dessa massa enorme de dados constituídos pelas variações acidentais de pronúncias e de significados.

Em seguida, foi necessário fixar essas observações por meio de representações gráficas. Tais esforços culminaram na invenção de sistemas de escrita, alfabeticos ou outros. E só depois de terem sido elaboradas tais representações foi possível ensiná-las às elites sociais e, mais

tarde, a um maior número de pessoas. No início, a principal função das representações escritas era a de conservarem a oralidade (a palavra), nomeadamente as leis e as trocas comerciais (vendas, doações, testamentos), mas as utilizações culturais apareceram muito depressa.

Os sistemas de escrita foram-se aperfeiçoando lentamente, até chegarem, no ocidente, à forma que hoje conhecemos, em que as palavras estão separadas umas das outras por espaços. Neste processo, um dos progressos mais importantes foi o da fixação de normas ortográficas. Em relação ao português, a norma é recente, só foi estabelecida em pleno século XX. Tal como acontece para os produtos industriais, em que a normalização permite a sua rápida difusão, a adopção de uma norma ortográfica vai igualmente contribuir de forma significativa para a difusão da escrita e da leitura. Além disso, cria condições para a elaboração de dicionários. De facto, a procura de uma palavra numa lista alfabética só pode ser feita rápida e eficazmente se a ortografia da palavra for única e se o utilizador do dicionário tiver aprendido essa ortografia (ou for capaz de a reconstituir a partir da pronúncia). Em relação ao português antigo e clássico, ainda hoje não dispomos de dicionários completos, pela simples razão de o número de variantes ortográficas nos textos de época ser tão elevado, que torna os recenseamentos muito difíceis. Aliás, se esses dados tivessem sido recenseados de forma completa, surgiriam enormes dificuldades práticas para representar todas as variantes ortográficas de uma dada palavra num dicionário em papel (estas dificuldades podem actualmente ser ultrapassadas com o registo das variantes em bases de dados informatizadas ou em dicionários electrónicos)¹.

A representação material dos discursos através de sequências de palavras é o ponto de partida para estudar a organização dos textos em termos de unidades mais vastas e complexas do que as palavras: os estudos sintácticos. Estes estudos põem também ênfase em evidência uma variedade considerável de formas. Em particular, conduzem à identificação de

¹ Exemplos de variantes ortográficas num mesmo texto do século XIV (*Crónica Geral de Espanha de 1344*). Atente-se, em particular, na representação ortográfica dos sons nasais:

«E, veendo os grandes fidalgos que nom eram desses bandos e outrossi os poboos o grande mal e destruiçon que por esto viinha aa terra, veeronsse a acordar de fazerem cortes e fazerem ē ellal tal regimento per que se a terra nō perdesse».

«E, por esta razon, tragia sempre muy grāde casa e muy honrrada de muitos fidalgos e outrossi sua molher acompanhada de muitas ricas donas e donzelas de grande guisa».

E ella avya por amiga hūa muy fremosa donzella que avia nome Alquifa.

blocos de palavras de diversos tipos, bem como à determinação de diferentes níveis de composição desses blocos. Por exemplo, a partir de frases como:

O vendedor atende o cliente; mostra-lhe o casaco

a observação de que existe uma semelhança entre as sequências *o vendedor*, *o cliente*, *o casaco*, conduz à definição da categoria **Artigo** para o elemento *o* (e ao seu rápido alargamento às formas *a*, *os*, *as*)². Simultaneamente, a categoria **Nome** (ou **Substantivo**), a que pertencem *vendedor*, *cliente* e *casaco*, deve poder ser identificada. A observação de outras produções como:

O vendedor solícito atende o cliente impaciente; mostra-lhe o casaco de caxemira

permite (e exige) que, a um certo nível de análise, se considerem como equivalentes combinações de palavras como:

<i>vendedor solícito</i>	=	<i>vendedor</i>
<i>cliente impaciente</i>	=	<i>cliente</i>
<i>casaco de caxemira</i>	=	<i>casaco</i>

A existência de tais regularidades conduz então à definição de noções abstractas, que permitem generalizar as observações. Essas abstracções são essencialmente:

- designações de classes de **palavras**: os nomes das partes do discurso ou classes gramaticais, como *Artigo*, *Nome* (ou *Substantivo*), *Adjectivo*, etc. (ou, dito de outro modo, classes de equivalência de palavras que se definem por listas);
- designações para diferentes tipos de **sequências de palavras** que não se definem por listas (mais tarde, será estudada em pormenor a noção de sintagma (ou grupo nominal)).

Estes resultados, obtidos no início do tempo histórico (cf. gramáticos hindus, alexandrinos, latinos), foram utilizados sem grandes alterações no decorrer dos séculos. Com efeito, até à segunda metade do século XX, a Linguística desenvolveu-se muito lentamente. Porém, a partir daí, tem-se assistido a uma rápida e constante evolução. A Linguística adoptou (ou melhor, adaptou) modelos matemáticos, informáticos e lógicos para analisar as línguas, e introduziu noções complexas para descrever a organização das frases e dos discursos.

² Em contrapartida é menos fácil reconhecer um *o* ou um *a* subjacente às formas contraídas *ao*, *à*, etc. mas isso consegue-se pelo mesmo método de comparação e de substituição de sequências equivalentes.

No âmbito da disciplina *Linguística Portuguesa IV*, utilizaremos um **modelo geral de descrição** formal do Português, baseado num **conjunto mínimo de abstracções** gramaticais. Veremos que as funções e categorias propostas nas gramáticas escolares se revelam incompletas e mesmo pouco pertinentes num quadro de descrição formal assente sobre as combinatórias de palavras.

2. Fundamentos da Análise Sintáctica

2.1. Noção de Frase

A primeira abstracção que é indispensável introduzir é a noção de frase. Este conceito não tem que ser definido; ele identifica-se com o próprio postulado da existência da Sintaxe³. Este postulado encontra várias motivações: por um lado, possui base empírica (falaremos em seguida da noção de aceitabilidade frásica), por outro, resulta da observação da existência de diversas formas recorrentes. Vejamos o exemplo de produções linguísticas como:

- (1) *O Zé veio à faculdade mas a Ana foi ao cinema*
- (2) *Ontem os miúdos não apareceram, mas as mães vieram cá hoje*

É natural fazer a observação de que se trata de frases complexas, isto é de frases formadas por frases mais simples, a saber:

- (a) *O Zé veio à faculdade*
- (b) *A Ana foi ao cinema*
- (c) *Ontem os miúdos não apareceram*
- (d) *As mães vieram cá hoje*

Em (1) e (2), estas frases simples (ou elementares) estavam ligadas, relacionadas, duas a duas pela palavra *mas* (uma conjunção).

Uma tal observação pode ser transformada em regra geral se for formulada sem ter em conta as características (forma e conteúdo) das frases mais simples. A formulação dessa regra poderá ser então:

³ Para uma discussão da noção de frase ver Ranchhod, Elisabete M. 1997. «Algumas Considerações sobre a Noção de Frase Elementar e sua Aplicação à Análise do Português». In Brito, Ana Maria et al. (orgs.), *Sentido que a Vida Faz. Estudos para Óscar Lopes*, pp. 787-795, Porto: Campo das Letras .

$$(3) \quad F = F \text{ mas } F \quad [\text{outro formalismo: } F \rightarrow F \text{ mas } F]$$

Esta regra utiliza apenas a noção de frase (representada por **F**), que, contrariamente às partes do discurso (categorias gramaticais), não pode ser objecto de listas completas. Este exemplo de regra é extensível a algumas outras palavras como *mas*, extensão que deu origem à noção abstracta de **conjunção de coordenação** (que representaremos por **Conjc**), a categoria gramatical definida pela conhecida lista:

$$(4) \quad \text{Conjc} =: e, \text{nem}, \text{mas}, \text{porém}, \text{todavia}, \text{contudo}, \text{ou}, \text{ora}, \text{quer}, \text{seja}, \text{logo}, \text{pois}, \text{portanto}, \text{por conseguinte}, \text{por consequência}, \dots$$

Assim, é possível formular uma regra mais geral do que (3), ou seja:

$$(5) \quad F = F \text{ Conjc } F$$

Vemos, assim, através destes exemplos simples, que as abstracções necessárias à descrição sintáctica já não são directamente observáveis, como o são as palavras e os sons. Trata-se de noções definidas pela sua comodidade de utilização ou pela sua validade no quadro da descrição completa e coerente de uma dada língua, ou no quadro universal de descrição de todas as línguas, ou da descrição da faculdade (biológica ou cognitiva) da linguagem.

A descrição da noção de frase identifica-se, pois, com o programa geral da investigação em Sintaxe. Mas a Sintaxe dos nossos dias teve de limitar a sua abordagem a certas características da frase, que nem sempre coincidem com as que as gramáticas escolares (pedagógicas) procuram cobrir. Em particular, a maior parte das questões de estilo são consideradas exteriores à Sintaxe, que não se preocupará, pois, com fenómenos como:

- a escolha lexical das produções linguísticas (calão *vs* léxico cuidado: *o pessoal bazou todo – a malta pirou-se toda – toda a gente foi embora – toda a gente saiu*; *está um frio do caraças – está um frio de rachar – está muito frio*; *passei-me – saltou-me a tampa – perdi o controlo*;
- empréstimos *vs* léxico vernáculo (*uma pequena nuance – uma pequena diferença – um pequeno matiz*; *lanche – merenda*; *constatar – verificar*);
- o comprimento das frases, a adjectivação excessiva, etc. (Cf. normas estilísticas).

A Sintaxe tornou-se uma disciplina autónoma e bem separada da Semântica; muitas das correlações entre significado e forma, traduzidas em regras pedagógicas, revelaram-se ingénias e mesmo falsas (Cf. definição das categorias gramaticais nas gramáticas escolares). As razões da autonomia da Sintaxe são tanto de ordem teórica como empírica. A

Sintaxe utiliza formalismos e conceitos baseados em teorias matemáticas (álgebras booleanas, análise combinatória, teoria dos autómatos, ...). A aparelhagem conceptual que essas teorias produziram revela-se adequada para descrever formalmente a fonologia e a sintaxe das línguas, mas não parece pertinente para tratar fenómenos pouco combinatórios que envolvem noções de significado e questões estilísticas. As questões de estilo são frequentemente difíceis de objectivar, pois fazem apelo a valores estéticos, muitas vezes exclusivamente individuais. Pode dizer-se o mesmo do significado, uma vez que a interpretação do significado das palavras e das frases, fora do contexto concreto da comunicação corrente, pode variar consideravelmente de indivíduo para indivíduo.

Para resumir: os sintacticistas procuram limitar-se à descrição de situações reprodutíveis, ou, dito de outro modo, a situações que são objecto de consenso entre os especialistas do estudo da língua. Com esta atitude, mais não fazem do que adoptar o método científico, mesmo que a sua aplicação seja mais difícil nas Ciências Humanas do que nas Ciências Naturais, de tal modo são fortes as interferências entre o estudioso e o objecto de estudo. Deste ponto de vista, o domínio da Sintaxe parece ser privilegiado, uma vez que a noção de aceitabilidade sintáctica, base da detecção e da avaliação de todos os fenómenos sintáticos, revelou ser bem reprodutível.

2.2. *Enquadramento Formal*

A Sintaxe está ligada de um modo óbvio às operações combinatórias que permitem a construção de sequências naturais de palavras: **concatenação, apagamento, inserção, permuta** e **comutação**. Estas operações delimitam com precisão o quadro lógico-matemático da aparelhagem descritiva da Sintaxe. Ilustremos estas operações combinatórias através de exemplos simples:

Concatenação – as duas palavras *chegou* e *ele* podem aparecer concatenadas nas sequências: *ele chegou* e *chegou ele*, mas também em sequências que não são frases: **ele chegou ele*; **ele ele chegou*⁴;

Apagamento – expressões como : *um inteiro*, *uma maiúscula*, *um microondas*, são interpretadas como abreviações de expressões mais longas, respectivamente: *um número inteiro*, *uma letra maiúscula*, *um forno microondas*. Trata-se de expressões resultantes do apagamento apropriado (isto é da elipse de elementos com uma probabilidade de ocorrência extremamente alta) dos substantivos *número*, *letra* e *forno*;

⁴ O asterisco é utilizado para marcar formalmente uma sequência (em geral, uma frase) mal formada (inaceitável).

Inserção – na sequência: *uma ideia* a inserção de *brilhante* conduz às sequências: *uma ideia brilhante, uma brilhante ideia*;

Permuta – os pares de sequências (frases): *Ele chegou e chegou ele; Ele gosta de gelados e de gelados ele gosta*, diferem entre si por uma permuta;

Comutação – o sintagma nominal *os miúdos que conheço* integrado na frase: *Todos os miúdos que conheço gostam de gelados*, comuta com a forma nominativa de pronome pessoal *eles*: *Todos eles gostam de gelados*; de igual modo, a sequência *em breve*, integrada em: *Ele fará isso em breve*, comuta com um advérbio: *Ele fará isso brevemente*, o que é uma indicação clara quanto à sua natureza adverbial.

Numa abordagem geral, despojada da preocupação de qualquer aplicação, a criação de abstracções deverá ser submetida às regras metodológicas da Ciência. Ora, muitas vezes as categorias gramaticais usuais foram estabelecidas no quadro de aplicações da linguística: pedagogia e didáctica das línguas (materna e estrangeiras) e, recentemente, tratamento informático das línguas. Diversas noções necessitam de ser rediscutidas e revistas, o que constitui o próprio objecto da linguística teórica.

O formalismo elementar que usámos antes na configuração (5):

$$(5) \quad F \quad = \quad F \text{ Conj } F$$

mostra imediatamente que a lista escolar (4) é incoerente. Ela levaria a que uma sequência como:

$$(6) \quad A \text{ } Ana \text{ } trabalha \text{ } nem \text{ } o \text{ } Zé \text{ } dorme$$

fosse considerada gramatical (ou bem formada), o que é de todo inadequado: tal sequência é inaceitável como frase. Mesmo que se tenha em conta a restrição de que *nem* deve coordenar frases negativas, é difícil aceitar uma forma como:

$$(7) \quad A \text{ } Ana \text{ } não \text{ } trabalha \text{ } nem \text{ } o \text{ } Zé \text{ } dorme$$

Embora (7) possa ser interpretada, do ponto de vista sintáctico está ainda mal formada⁵, sendo a forma aceitável:

$$(8) \quad Nem \text{ } a \text{ } Ana \text{ } trabalha \text{ } nem \text{ } o \text{ } Zé \text{ } dorme$$

⁵ A situação é diferente se o sujeito das orações coordenadas for co-referente:

A Ana não trabalha nem dorme

A Ana nem trabalha nem dorme

Ana é sujeito de *trabalhar* e de *dormir*

Ora esta sequência não corresponde à forma (5) mas a uma forma diferente, iniciada pela conjunção:

$$(9) \quad F = \text{Conjc } F \text{ Conjc } F$$

A posição *Conjc* pode ser preenchida por outras conjunções, como:

$$(10) \quad \text{Ou a Ana trabalha ou não faz a cadeira}$$

A regra (5) também se aplica mal, entre outras, à conjunção *ora*, pois dá origem à sequência:

$$(11) \quad \text{A Ana trabalha, ora o Zé dorme}$$

que também não é uma frase bem formada. Contudo, se esta forma for prolongada:

$$(12) \quad \text{A Ana trabalha, ora o Zé dorme, logo não pode fazer barulho}$$

a subsequência (11) tem agora melhor aceitabilidade. Contudo, para a sua formação é necessário formular outra regra mais complexa (uma regra ternária):

$$(13) \quad F = \text{Fora } F \text{ logo } F$$

que é de facto uma forma de um silogismo.

Estes exemplos mostram que o interesse científico e pedagógico da enumeração ilustrada por (4) é bastante limitado, uma vez que a categoria gramatical em causa não permite formular nenhuma regra baseada sobre a combinatória dos elementos. Os exemplos ilustram também que, para tornar coerentes as categorias habituais da gramática, é necessário proceder a várias correcções, introduzir, na análise sintáctica, um nível de descrição mais preciso.

2.3. Aceitabilidade Sintáctica

O fundamento empírico dos estudos sintáticos reside assim, e antes demais, num processo intuitivo que tem como base o reconhecimento de que certas sequências de palavras correspondem à noção de frase. O programa científico da Sintaxe consiste assim em construir (ou em descobrir) as regras combinatórias que permitem formar sequências que sejam aceites como frases **e só estas**. Deste modo, a sequência:

Essa solução convém a muitas pessoas

é, sem qualquer dúvida, considerada como uma frase. Em contrapartida, a sequência:

A solução convém essa muitas pessoas

não é considerada como frase, igualmente sem qualquer hesitação. No entanto, ela só difere da precedente pela permuta de *essa* e *a* (respectivamente, um determinante demonstrativo e uma preposição). Este método de avaliação da boa formação de frases é igualmente pertinente em relação a exemplos de natureza diferente, como os seguintes (adaptados de, respectivamente, Lucien Tesnière, 1953 e Noam Chomsky, 1957)⁶:

(1) *O silêncio vertebral indispõe a vela lícita*

(2) *As ideias verdes incolores dormem furiosamente*

As combinações de palavras (1) e (2) não têm qualquer interpretação (ou, pelo menos, não têm uma interpretação óbvia e imediata). No entanto, possuem uma organização tal que podem ser consideradas gramaticais, no sentido de estarem sintaticamente bem formadas. Se, contudo, as palavras que as constituem estiverem dispostas de um modo diferente, como em:

(3) *O indispõe silêncio vertebral a vela lícita*

(4) *Silêncio a lícita vertebral o indispõe vela*

já não podem ser consideradas frases. Para descrever a diferença entre, por um lado, (1) – (2) e, por outro (3) – (4), diremos que as primeiras são **sequências estruturadas**, sendo a sua estrutura uma estrutura de frase, enquanto as últimas não têm qualquer organização interna. As dificuldades que se encontram para ler (3) e, sobretudo, (4) — e muito mais ainda para as memorizar —, põem bem em evidência o papel fundamental das estruturas sintácticas, papel claramente independente do significado. As dificuldades de leitura e de memorização indicam também que existem relações evidentes entre a sintaxe e a prosódia: só as sequências sintaticamente bem formadas podem receber entoação, e, por sua vez, a existência de uma linha prosódica facilita a memorização.

Poder-se-iam fazer as mesmas observações a propósito de exemplos como:

(5) *Os batões fatáveis grunem dinamente*

⁶ «Le silence vertébral indispose la voile licite», in Tesnière, L. 1953. *Esquisse de syntaxe structurale*, Paris: Klincksieck.

«Colorless green ideas sleep furiously», in Chomsky, N. 1957. *Syntactic Structures*, Haia: Mouton.

(6) Batões os fatáveis dinamente grunem

Estes exemplos podem ser mais elucidativos, uma vez que eles são constituídas por léxico artificial (embora a estrutura morfológica dos elementos lexicais artificiais imite a das palavras do português). Apesar de constituídos por léxico artificial, é claro que, de um ponto de vista estritamente sintáctico, (5) corresponde uma frase bem formada, enquanto (6) não tem tais características. Nenhuma das sequências tem qualquer significado, mas é possível dizer a que categoria gramatical pertence cada uma das «palavras» por que são constituídas. Em (5), contrariamente ao que se passa em relação a (6), essas categorias estão estruturadas, mantendo entre si relações gramaticais (não é difícil identificar o sujeito, por exemplo).

A avaliação da boa formação dos exemplos apresentados antes levava a que eles pudessem ser considerados, sem qualquer hesitação, frases ou não frases. Se considerarmos agora as sequências seguintes:

muitos dos seus livros
surdo como uma porta
num abrir e fechar de olhos

verificamos que, embora bem formadas, não podem ser consideradas frases. A nossa cultura gramatical permite-nos designá-las na metalinguagem da gramática como construções (sintagmas) nominais, adjetivais e adverbiais, respectivamente. Mas este conhecimento não tem já uma relação directa com a intuição de gramaticalidade, que é largamente independente da formação escolar.

Nem todas as combinações de palavras são claramente aceitáveis ou inaceitáveis (como acontecia com os exemplos anteriores). Algumas, causam um certo desconforto que não permite afirmar com segurança se são ou não são frases. As causas que dificultam a avaliação da aceitabilidade sintáctica prendem-se fundamentalmente com o contexto distribucional (restrições de selecção) ou com fenómenos de natureza variada, que, por comodidade, consideraremos como estilísticos.

2.3.1. Distribuição

Os testes de aceitabilidade nem sempre conduzem imediatamente à avaliação que se pretende fazer das sequências de palavras, sendo em muitos casos difícil separar as interdições sintácticas dos bloqueios semânticos ou estilísticos. É o que se observa, com bastante frequência, quando se quer determinar a distribuição dos nomes em relação aos

verbos. Consideremos o verbo *comer*. Sabemos intuitivamente que, se usado literalmente, só aceita na posição de complemento (à sua direita) nomes de alimentos:

(1) *O Zé comeu uma maçã*

Assim sendo, avaliação da boa formação de uma expressão como:

(2) *O Zé comeu um tijolo*

não é simples: o *Zé* pode ser um doente mental (ou estar momentaneamente perturbado) e, neste universo de referência, poder-se-iam encontrar, na posição de complemento directo de *comer*, numerosos substantivos concretos.

Um processo diferente de interpretação sistemática dos objectos concretos é sugerida pelo exemplo:

(3) *O Zé comeu um tijolo de chocolate*

Existem outros processos, como a intervenção da negação, que facilitam a aceitabilidade dos complementos concretos, como mostram os exemplos:

(4) *O Zé não comeu o tijolo, apenas o aproximou da boca*

(5) *Não se deve comer a terra dos legumes*

Estas extensões da aceitabilidade não parecem contudo aplicar-se aos substantivos abstractos. Formas como:

?* *O Zé comeu uma incredulidade*

?* *O Zé comeu uma concatenação*

?* *O Zé não comeu uma incredulidade*

?* *O Zé não comeu uma concatenação*

não correspondem a estruturas bem formadas e exigem grande imaginação para poderem ser interpretadas. A construção de contextos onde estas formas pudessem ter uma interpretação coerente passaria certamente por uma redefinição do verbo *comer* e/ou dos substantivos abstractos (*incredulidade* e *concatenação*), mas isso levaria à avaliação da boa ou má formação de dados diferentes dos inicialmente testados.

Mas o que é de salientar neste ponto é que todos estes exemplos com o verbo *comer*, são generalizáveis, no sentido de ilustrarem processos gerais baseados na estrutura formal das expressões:

- A frase (iii) está bem formada devido à adjunção a (ii) de um modificador interno ao complemento directo do verbo, o que altera a sua interpretação, mas não impede que (iii) continue a ter uma estrutura idêntica à da forma canónica (i);
- O exemplo (ii) só é aceitável se o sujeito da frase puder ser interpretado de forma específica; o complemento directo, não tendo qualquer complemento que o modifique, continua a não ser «comestível»;
- A negação do verbo *comer* faz com que a expressão (iv) se torne aceitável; para isso, contribui também o facto de, através da oração justaposta, se ter posto em contraste a interpretação do verbo;
- O exemplo (v) é genérico, tem a interpretação de um conselho, ou de um provérbio, frases que têm formas particulares.

É igualmente de salientar que a natureza semântica destas variações não é simples, ao passo que as alterações formais dizem apenas respeito à introdução da negação e a uma modificação do complemento directo da mesma estrutura sintáctica: *sujeito - verbo - complemento*.

Em todas estas observações sobre a aceitabilidade, seguimos a tradição e utilizámos um suporte sintáctico, gramatical, para tentar localizar o significado. Mas isso nem sempre poderá ser feito. De facto, existem numerosas expressões linguísticas em que é impossível estabelecer uma relação entre a análise léxico-sintáctica e a sua interpretação. O caso mais óbvio é o das expressões idiomáticas (expressões não compostionais tanto do ponto de vista sintáctico como semântico). Pode considerar-se que, do ponto de vista do significado, as frases seguintes estão próximas:

- (6) *O Zé comeu o pão que o diabo amassou*
 (7) *O Zé passou por momentos difíceis*

Contudo é evidente que o significado da frase idiomática (6) não está directamente relacionado nem com a sua estrutura sintáctica nem com o léxico por que é constituída.

2.3.2. Factores Estilísticos

Discutiremos seguidamente alguns exemplos que levantam questões que são por vezes consideradas como estilísticas. Observemos a frase:

- (1) *O gato bebe o leite*

e a sequência (mal formada):

(2) **Gato bebe o leite*

Esta é inaceitável devido à ausência de determinante na posição de sujeito. Esta restrição é muito geral em português⁷. Nos casos em que isso acontece, observa-se frequentemente uma inversão do sujeito, como acontece nas seguintes construções existenciais:

**Ainda pessoas simpáticas existem*

Ainda existem pessoas simpáticas

Mas, voltando ao exemplo (2), verifica-se que, se for modificado do seguinte modo:

(3) *Gato esfomeado bebe o leite frio ou quente*

passa a estar bem formado, ainda que o sujeito continue a não ser precedido por qualquer determinante. A razão parece ser a seguinte: ao acrescentar modificadores adjetivais ao sujeito e ao complemento (*esfomeado*, *frio* e *quente*) criou-se uma forma que, tanto do ponto de vista sintáctico como lexical e rítmico, está próxima do provérbio:

Gato escaldado de água fria tem medo

O recurso a um processo analógico particular tornou possível alterar a aceitabilidade, embora a estrutura, estritamente falando, permaneça agramatical.

Consideremos agora a seguinte frase exclamativa (com valor de palavra de ordem):

(4) *Abaixo o ministro!*

Comparemo-la com a seguinte, que não difere desta a não ser pelo tipo de determinante que precede *ministro*:

(5) **Abaixo um ministro!*

Verifica-se que, contrariamente à anterior, (5) não é aceitável. A razão da inaceitabilidade prende-se obviamente com os determinantes (artigo definido e artigo indefinido). Poder-se-á acrescentar uma explicação semântica banal, que consiste em observar que nas frases

⁷ A ausência de determinante na posição de sujeito é frequente nos (talvez mesmo característica dos) títulos da imprensa escrita:

Acidente de viação faz três vítimas mortais

Greve de fome de reclusos gera contradições

que, por essa e outras razões, se constituem como discursos específicos.

imperativas se faz apelo a entidades bem identificadas, ou bem determinadas, ou seja a sintagmas nominais definidos. Em todo o caso a forma:

(6) *Abaixo um ministro que reduz(a) o orçamento das Universidades!*

tem melhor aceitabilidade, embora o determinante de ministro seja um artigo indefinido. Deixando de lado as considerações semânticas quanto à referência definida e indefinida das expressões nominais (a adjunção de uma oração relativa torna o sintagma nominal indefinido menos indefinido), descreveremos a diferença entre (4) e (5) em termos sintácticos: neste tipo de construção, o artigo indefinido obriga à co-ocorrência de um modificador (em (6), uma oração relativa). Estas restrições sintácticas apresentam uma certa regularidade. Observam-se de forma sistemática noutro tipo de construções (frases com verbos-suporte e nomes predicativos⁸):

?* *O Zé está numa agitação*

O Zé está numa agitação!

O Zé está numa agitação espantosa

?* *O Zé é de uma generosidade*

O Zé é de uma generosidade!

O Zé é de uma generosidade extraordinária

Os exemplo bem e mal formados indicam que, em certos, casos as expressões exclamativas são equivalentes a um *Modificador*, sintacticamente realizado como adjetivo ou como oração relativa⁹

Para concluir, faremos algumas referências a aspectos, exclusivamente lexicais, que se prendem com questões de registo ou nível de língua que podem levantar dificuldades à avaliação da boa ou má formação sintáctica das frases. De facto, a presença de calão ou de palavras obscenas numa dada frase pode levar a considerar essa frase como mal formada. Esta rejeição nada tem que ver com má formação sintáctica. A noção de aceitabilidade que temos estado a discutir é independente do nível de língua (culto, informal, familiar, grosseiro, etc.). Assim, ela aplica-se de igual modo às frases (7) e (8), por um lado, e (9) e (10), por outro:

(7) *Não gosto desse indivíduo*

(8) *A cara desse indivíduo não me diz nada*

⁸ Ranchhod, Elisabete Marques. 1990. *Sintaxe dos Predicados Nominais com Estar*, Lisboa: INIC.

⁹ Harris, Z. S. 1981. *Papers on Syntax*, Dordrecht: D. Reidel.

- (9) *Não gramo esse gajo*
- (10) *A tromba desse gajo não me diz nada*
- (7') **Não gosto esse indivíduo*
- (8') **A cara desse indivíduo não me escreve nada*
- (9') **Não gramo desse gajo*
- (10') **A tromba desse gajo não me escreve nada*

A observação que há a fazer em relação a exemplos deste tipo é o conhecimento do léxico (a competência lexical) ou a atitude cultural perante certas palavras varia de indivíduo para indivíduo.

Salvas as devidas distâncias, pode-se estabelecer um paralelismo entre este tipo de situações e o que se observa em relação ao vocabulário técnico e científico, que é em larga medida desconhecido dos falantes não especialistas. Nem todos os falantes do português sabem que *cefaleia* e *parotidite* são termos equivalentes a, respectivamente, *dor de cabeça* e *papeira* (ou *trasorelho*).

A mistura de níveis de língua pode levar à criação de expressões linguísticas, que se encontram, por essa razão, próximas da não aceitabilidade:

O ilustre mancebo tinha uma lábia do caneco
 (vs *O ilustre mancebo tinha um porte altivo*)

ilustre mancebo pertence a um nível de língua cuidado, uma *lábia* e *do caneco* correspondem a um registo popular. No entanto, se bem utilizada, a mistura de diferentes tipos de léxico permite obter efeitos humorísticos ou sarcásticos¹⁰.

Em síntese: verifica-se que mesmo se a intuição de aceitabilidade de uma dada frase é clara, todas as frases submetidas a avaliação devem ser cuidadosamente elaboradas: é preciso assegurar que, aquando da aplicação dos testes formais, a escolha do *vocabulário*, as *formas sintácticas* ou as *condições de enunciação* das frases não introduzem factores exteriores ao fenómeno que se pretende estudar. Só assim poderão ser emitidos juízos de aceitabilidade de boa qualidade e, portanto, reproduutíveis e formalizáveis. Tais precauções exigem conhecimentos aprofundados acerca a estrutura da(s) língua (s). É por esta razão que, tal

¹⁰ Veja-se a este propósito o efeito irónico conseguido por Vitorino Nemésio (*Carta a Raul Brandão*, 8 Ago. 1929. In http://bnd.bn.pt/ed/vitorino_nemesio/itinerario/313.html):

«Da história de Portugal é esse o período que sinto, – tão agitado, tão ansioso, e cujos gérmenes ainda revolvem o nosso coração século XX. Hoje temos mais civilização – ou mais lábia, coisa a que chamam cultura;»

como noutras domínios científicos, estas experiências só podem ser feitas por especialistas. Há que distinguir, então, esta situação da utilização de inquéritos sócio-linguísticos, destinados esses a falantes não especialistas.

2.3.3. Formalização da aceitabilidade

Quando se utilizam quadros teóricos baseados nas combinatórias de elementos e de sequências de elementos, os juízos de aceitabilidade deveriam ser formalizados em termos binários ('+' e '-'); deste modo os exemplos que não são aceitáveis (que não existem na língua) seriam marcados com asterisco ' $*$ '; os exemplos bem formados não seriam marcados. Contudo, do ponto de vista empírico, uma tal dicotomia afigura-se muito redutora. Em numerosas situações, hesitamos, com base em argumentos adequados, entre um e outro valor. De facto, atendendo ao carácter experimental destas avaliações, parece cómodo introduzir a noção de 'forma duvidosa', precedida de um ponto de interrogação '?', e a noção de 'forma muito duvidosa', precedida de interrogação e asterisco '?*'. Estas duas últimas notações não têm lugar nos quadros teóricos que se baseiam nas combinatórias. Com efeito, de um ponto de vista teórico, os juízos de aceitabilidade devem ser necessariamente binários (o modelo dos autómatos e das gramáticas formais não autoriza outras possibilidades). Contudo podem-se imaginar - e tirar partido de - outros quadros teóricos que autorizam mais do que dois valores de aceitabilidade. É o caso, por exemplo, das probabilidades, que permitem que a aceitabilidade possa ser ponderada.

Elisabete Marques Ranchhod

Lisboa, Fevereiro de 2005