



HAL
open science

Linguagem & Informação

Jorge Baptista

► **To cite this version:**

| Jorge Baptista. Linguagem & Informação. 2017. hal-04267580

HAL Id: hal-04267580

<https://hal.science/hal-04267580v1>

Preprint submitted on 6 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

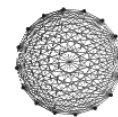
Linguagem & Informação

*Uma síntese da teoria da linguagem
desenvolvida por Harris S. Harris
ilustrada com exemplos do Português
Jorge Baptista, U. Algarve, Faro, Portugal
6 de Outubro de 2017*

Conteúdo

<i>Apresentação</i>	3
<i>Panorâmica</i>	5
<i>Desvios a partir da equiprobabilidade combinatória dos elementos discretos do sistema</i>	5
<i>As linguagens naturais não têm uma metalinguagem externa</i>	7
<i>Linguagem natural vs. linguagem artificial</i>	7
<i>Exemplos de metalinguagem</i>	8
<i>Restrições fundamentais</i>	14
<i>Ordem parcial de entrada das palavras nas frases</i>	14
<i>Probabilidade de co-ocorrência operador-argumento(s)</i>	21
<i>Redução</i>	29
<i>Redução correferencial</i>	29
<i>Redução apropriada</i>	33
<i>Linearização</i>	35
<i>Restrições complementares</i>	44
<i>1ª restrição complementar</i>	44
<i>2ª restrição complementar</i>	49
<i>Questões teóricas e metodológicas</i>	53
<i>Natureza do problema</i>	53
<i>Baixa redundância e relevância do método</i>	55
<i>Simplicidade</i>	61
<i>Pode o significado ser usado?</i>	65
<i>Propriedades universais da linguagem</i>	68

Este documento foi composto pelo autor com base no *template* "Sprint Beyond the Book" desenvolvido pela Society for Scholarly Publishing, e disponível na plataforma Overleaf.



Society for
Scholarly Publishing
Innovative People Advancing Scholarly Communication

Overleaf

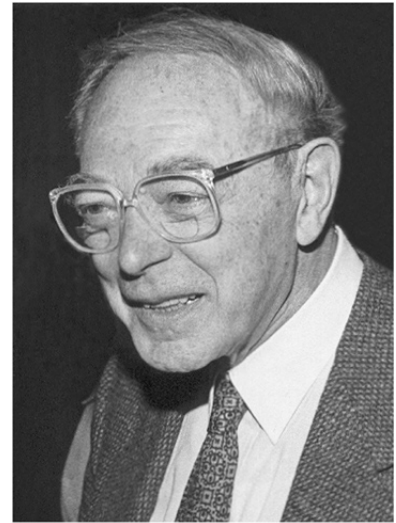
Apresentação

Poucas pessoas deverão ter influenciado tanto a história do pensamento linguístico ocidental como o linguista norte-americano *Zellig Sabettai Harris* (1909 — 1992), figura que, por um conjunto de razões históricas e científicas complexas, assinala, de uma forma indelével, a transição da primeira para a segunda geração do estruturalismo, justamente na passagem da primeira para a segunda metade do séc. XX. Não é nosso propósito fazer aqui nem sequer um esboço do papel de Harris, quer como homem, quer como figura política e intelectual comprometida com o seu tempo, quer ainda como cientista que profundamente marcou várias gerações de linguistas. Para uma perspetiva global do trabalho de Zellig S. Harris ao longo, sobretudo, da segunda metade do séc. XX, veja-se a biografia que dele fez Robert F. Barsky (2011). Para os contributos do pensamento de Harris em vários domínios científicos e na investigação sobre a linguagem, veja-se o volume de contribuições editado por Bruce E. Nevin e Stephen B. Johnson (2002).

Este documento pretende ser apenas fornecer uma perspetiva panorâmica, necessariamente sucinta, da teoria harrissiana da linguagem, tendo sido elaborado sobretudo com base na síntese que o autor dela fez na sua *magna opus: A Theory of Language and Information - A Mathematical Approach*, publicada em 1991, perto do final da sua vida. Aqui se irá, pois, apresentar essencialmente e de forma muito sucinta: (i) o ponto de partida da teoria, uma teoria baseada em restrições, fundada e inspirada na teoria matemática da informação; (ii) as restrições fundamentais e a forma como elas contribuem para a informação que é veiculada pela linguagem; (iii) a metodologia a que necessariamente se chega pela adopção desta perspetiva sobre o funcionamento e o estudo da linguagem; (iv) a natureza das descrições que é possível elaborar a partir da teoria e apoiadas na metodologia assim definida; e, finalmente, (v), alguns dos resultados já obtidos na descrição de largas secções de línguas naturais, com especial referência ao português.

A motivação principal por detrás deste texto é, antes de mais, a falta de materiais disponíveis em português sobre as contribuições do linguista e estruturalista norte-americano para a teoria da linguagem, em geral, e da descrição do inglês, em particular. Esta lacuna é tanto mais lamentável quanto, há mais de três décadas, um conjunto de linguistas com trabalho sobre a língua portuguesa, de ambos os lados do Atlântico, vêm reivindicando a sua filiação científica neste grande personagem da história do pensamento linguístico.

Em segundo lugar, pretende-se constituir aqui uma espécie de roteiro de leitura, um texto introdutório, numa linguagem acessível



Zellig S. Harris

Figura 1: Zellig S. Harris (1909 — 1992)



Figura 2: A obra-síntese de Zellig S. Harris, publicada em 1991, um ano antes da sua morte.

ao não especialista, capaz de suscitar a curiosidade e desenvolver o espírito crítico necessário para melhor compreender os textos de Harris. Este guião poderá, talvez, servir de apoio a um pequeno curso de introdução ao pensamento e método harrissianos. Por essa razão, o documento apresenta uma extensa bibliografia de obras do autor ou outras sínteses e elaborações mais recentes da perspetiva harrissiana.

Naturalmente, não se pretende substituir a leitura das fontes primárias, os textos originais de Harris: para deveras se conhecer em profundidade o seu pensamento, continuará a ser imprescindível ler os seus textos, seguir paciente e laboriosamente as suas demonstrações, refletir profundamente sobre os exemplos que apresenta, atentar na sua sequência e compreender as suas consequências, deixar-se surpreender com a subtilidade das suas observações, bem como admirar a perspicácia das suas conclusões.

Finalmente, ao preparar este documento, tive diante de mim os meus alunos, a quem dedico estas páginas, num desejo sincero de que, como eu, encontrem na visão de Harris uma compreensão mais profunda do funcionamento da linguagem.

Jorge Baptista Faro, Portugal
6 de Outubro de 2017

Panorâmica

Nesta secção, pretende-se apresentar uma visão panorâmica da teoria harrissiana, tal como formulada na síntese de Harris (1991).

A teoria harrissiana parte de duas observações de base relativamente ao funcionamento da linguagem:

- (1) O sistema linguístico é constituído por elementos discretos (sons/fonemas, letras, palavras/morfemas, ...) que não se combinam de forma indiferenciada uns com os outros. Pelo contrário, esses elementos apresentam importantes *desvios a partir da sua equiprobabilidade combinatória*. Estes desvios representam aquilo que, em rigor, se pode chamar a *informação* veiculada pelo sistema linguístico, entendendo aqui *informação* nos termos da teoria matemática da informação.
- (2) *A linguagem natural não apresentam uma meta-linguagem externa*, que possa ser utilizada para se lhe referir. Dito de outro modo, a linguagem natural é auto-referencial e não há um “lado de fora” da linguagem em que esta se pudesse colocar para se referir a si mesma.

Analisemos, então, com mais pormenor cada uma destas observações.

Desvios a partir da equiprobabilidade combinatória dos elementos discretos do sistema

Efetivamente, ao invés de se encontrarem num estado amorfo ou *entrópico*, em que qualquer elemento se poderia combinar indiferentemente com qualquer outro, os elementos discretos da linguagem apresentam diversas restrições quanto às combinações que estabelecem uns com os outros. Por *elementos discretos* entende-se a possibilidade de considerar, em diferentes níveis de análise, objetos individualizáveis e percebidos como semelhantes ou como distintos entre si. Em última análise, estes elementos são os próprios sons da linguagem, os *fonemas* (ou no caso da sua representação na escrita, os *grafemas*).

Esta primeira observação é relativamente óbvia e fácil de compreender e de verificar. Assim, por exemplo, considere-se o conjunto dos *grafemas* do português. Se tivermos um dicionário suficientemente abrangente, é possível dizer que este representa de algum modo o léxico da língua. Nesse caso, é possível calcular, a partir da lista das entradas a *probabilidade* de, a seguir a um dado *grafema*, se seguir outro *grafema*, a que chamaremos a *probabilidade sequencial de grafemas*. Verifica-se, nesse caso, quem para cada combinação potencial,

Dois observações de partida da teoria harrissiana sobre a linguagem, em particular as linguagens naturais:

- desvios a partir da equiprobabilidade combinatória dos elementos discretos do sistema linguístico
- ausência de uma metalinguagem externa.

entropia

elementos discretos

fonemas

grafemas

sistema de grafemas do Português

léxico

probabilidade sequencial de grafemas

há diferentes probabilidades de cada uma ocorrer. Por exemplo, em posição de início de palavra, o <a> ou o <c> têm uma probabilidade muito alta, já que há muitas palavras que começam com estas letras. Pelo contrário, o número de palavras começada com <y> ou <z> é bastante mais pequeno, o que reduz a probabilidade de estes grafemas ocorrerem nesta posição. Quando passamos para as combinações com dois caracteres, o número de palavras com dois <aa> justapostos já é bastante pequeno (resumindo-se basicamente às *abreviaturas* de *aná* ou de *autores*) e rapidamente cai para apenas um (*aal*) no caso de palavras com três caracteres. Não há nenhuma palavra com quatro caracteres começada por <aa>. No caso das palavras começadas com , apenas se encontrou uma palavra com dois <bb> (abreviatura de *bombordo*) e outra começada assim e com três caracteres (*bbl*, abreviatura de *barril*). É fácil verificar que, sendo dada uma combinação de caracteres em posição inicial, rapidamente o número de palavras abrangidas decai para zero à medida que se considera um conjunto progressivamente mais longo de caracteres que podem ocorrer à direita dessa combinação particular.

abreviaturas

Um exemplo semelhante pode ser dado no caso da posição de palavras em frases. Neste caso, considere-se um *corpus* suficientemente extenso para poder ser representativo da distribuição das palavras nas frases de uma língua, e nele detenhamo-nos nas sequências de palavras que ocorrem em posição inicial de frase. É fácil verificar que certas palavras (*a, aquilo, as, este, estes, isto, o, os, um, uns, etc.*) tem uma muito alta probabilidade de ocorrer em início de frase, ao passo que certas palavras raramente ocorrem nessa posição.

probabilidade sequencial de palavras em frases
corpus
representatividade
distribuição

O que importa aqui reter é que, quando entendida deste modo, a probabilidade é um conceito estatístico, matematicamente preciso, quantificável. Embora neste curso não tenhamos de nos preocupar especificamente com aspectos relacionados com este tipo de quantificação, ela está na base de numerosas aplicações em *processamento (computacional) de linguagem natural* (PLN), entre outras, por exemplo, a anotação morfossintática e sua desambiguação automáticas (ing. *part-of-speech tagging*).

estatística

Processamento de Linguagem Natural

Part-of-speech tagging

Os desvios à equiprobabilidade combinatória (*entropia*) dos elementos discretos do sistema linguístico, independentemente do nível de análise em que nos situemos, constituem a informação (*negentropia*) veiculada pelo sistema. As línguas naturais são isso mesmo: sistemas que usamos para veicular informação. O estudo da linguagem, independentemente de outras perspectivas que se possa adoptar relativamente ao complexo fenómeno das linguagens humanas, consiste, essencialmente, na descrição da forma como a linguagem se estrutura para servir essa função de veículo de informação. Nesse sentido, importa compreender a forma como a linguagem representa essa

negentropia
língua natural

informação. Iremos ver, mais adiante, que a linguagem se estrutura com base num conjunto de restrições às combinações que os seus elementos podem apresentar. Essas restrições podem ser determinadas a partir dos desvios à equiprobabilidade combinatória dos elementos linguísticos. Mas não nos adiantemos demais neste momento. Vejamos agora, com maior profundidade, a segunda observação que é ponto de partida da teoria harrissiana.

As linguagens naturais não têm uma metalinguagem externa

Qualquer linguagem natural tem um conjunto de elementos *que fazem parte dela própria* e que são usados para se referirem à própria linguagem – aquilo a que chamamos, justamente, *metalinguagem* (do gr. *μετα meta-*, ‘para além de’). Esta observação é mais difícil de explicar do que a anterior e também é mais difícil de compreender, em especial quando é considerada em relação ao profundo impacto que tem, quer sobre a natureza e funcionamento da linguagem, quer sobre a natureza, o método e a forma da teoria da linguagem que é possível elaborar. Como todos os conceitos teóricos fundamentais (elementares), o alcance deste princípio é vasto e as implicações complexas.

metalinguagem

Normalmente, esta propriedade das linguagens naturais é apresentada *pela positiva*, como, por exemplo, o fazem Mihalilicek & Wilson (2011) e Burton *et al.* (2016) a partir dos chamados *design features* alistados por Charles Hockett (1966), indicando-se, nomeadamente, como uma das suas características fundamentais o facto de serem *autorreflexivas*.

autorreflexividade

Harris, contudo, faz questão de enunciar esta propriedade *pela negativa*, já que tal é determinante para a forma da teoria e a metodologia que esta pressupõe. Ora, o facto de as linguagens naturais *não* terem uma metalinguagem externa quer dizer que não é possível alguém referir-se à linguagem (às suas partes, aos seus elementos, às propriedades dos seus elementos, às relações entre os seus elementos, às relações entre os seus elementos e os objetos/estados/processos do mundo extralinguístico, etc.) sem utilizar forçosamente a própria linguagem. Por outras palavras, *não há um lado de fora da linguagem*, em que nos pudéssemos situar para, a partir daí, falar sobre a própria linguagem. Vamos tentar explicar e aprofundar esta difícil noção com alguns exemplos.

Linguagem natural vs. linguagem artificial

Antes de mais, comecemos por comparar uma língua natural com uma língua artificial, como uma linguagem de programação, por exemplo, o PERL. Aliás, trata-se de uma das linguagens mais populares em PLN.

Língua natural vs. linguagem artificial. Para uma definição mais precisa do que distingue uma língua natural de outras formas de linguagem, veja-se a síntese de Burton *et al.* (2016).

Perl

É possível usar uma *linguagem de programação* para construir *programas* que *executam* diversas *tarefas*. De forma muito grosseira, esses programas podem ser entendidos como um conjunto de *proposições* (frases), em que se faz a *definição* de certos *objetos* (variáveis) e *condições*, as quais determinam um conjunto de *ações* a executar.

Ora, é possível considerá-la uma *linguagem* na medida em que é formada por um conjunto de elementos discretos (os seus *operadores* e o conjunto de objetos que se define como as variáveis que esses operadores manipulam), assim como um conjunto de regras para a boa formação das proposições que constituem, no seu conjunto, as instruções do programa. A este conjunto de regras podemos chamar a *sintaxe* da *linguagem* de programação. Por exemplo, a definição de uma condição é constituída por dois termos (*if ... then / se ... então*): um *antecedente*, que enuncia a condição propriamente dita, e um *consequente*, que indica a ação a executar.

Para nos podermos referir ao programa ou às suas partes, podemos usar qualquer linguagem natural (por exemplo, o português). Como muitas outras linguagens de programação, esta “re-usa” elementos lexicais do inglês para os seus operadores fundamentais (IF ‘se’).

Exemplos de metalinguagem

Ora, a situação é muito diferente quando se trata de um texto numa língua natural. Considere o pequeno anúncio (exemplo real) da Figura 3, abaixo:

***Aluga-se apartamento
a estudantes universitárias
em bom estado.***

linguagem de programação

Figura 3: Pequeno anúncio

Como seria possível explicar *sem usar uma língua natural* o que está acontecendo nesta frase? Pelo contrário, é perfeitamente possível, usando uma língua natural, aliás a própria língua em que a frase se encontra escrita, descrever e, de algum modo, explicar (e desfazer!) a graça do anúncio. Por exemplo, seria possível dizer:

- (i) *O que está em bom estado é o apartamento e não as estudantes universitárias;*
- (ii) *Aluga-se apartamento que está em bom estado a estudantes universitárias;*

- (iii) *Aluga-se apartamento a estudantes universitárias. Esse apartamento está em bom estado..*

Se se usasse uma linguagem mais técnica (a negrito), poderia dizer-se:

(iv) O **complemento preposicional** *em bom estado* encontra-se separado do **nome** que **modifica** , *apartamento*, pelo **complemento indireto** do **verbo** *alugar*, *a estudantes universitárias*. Tal conduz a uma leitura em que a **expressão** *em bom estado* se encontraria antes a **modificar** o **constituente** *a estudantes universitárias*. Esta segunda leitura produz um efeito de *ironia* , pois a **expressão** *em bom estado não se costuma aplicar* a **nomes humanos** (*universitários*).

complemento preposicional
modificador

constituente
X aplicar-se a Y: restrições de seleção

Claro que uma frase assim perdia a graça toda! Mas note-se que o facto de se ter utilizado uma linguagem mais técnica (uma *terminologia*) para descrever o fenómeno é equivalente a utilizar as explicações mais singelas que se apresentou acima.

terminologia

No primeiro exemplo (i), estabeleceu-se um *distinguo* por meio de uma extração *ser . . . que*; no segundo caso (ii), reestabeleceu-se a dependência entre o modificador e o nome *apartamento* por meio de uma relativa; no terceiro (iii), dá-se conta de que o texto é formado por uma *frase complexa* , desdobrando-a nas *frases de base* ou *frases elementares* que lhe deram origem. Em rigor, os exemplos (i) a (iii) não são *expressões metalinguísticas*, mas sim *paráfrases* mais ou menos complexas da frase de partida.

frase complexa
frases de base ou *elementares*
paráfrases

Por outro lado, o discurso, bastante técnico, apresentado em (iv) inclui um conjunto de *termos* cuja definição (complexa) implica um conjunto de conhecimentos (neste caso gramaticais) sobre o funcionamento da linguagem. Contudo, esses conhecimentos, conceitos e definições correspondem necessariamente a frases da língua. É por um processo de progressiva abstração a partir de proposições elementares que construímos essas definições, conceitos e conhecimentos. (Iremos discutir melhor o conceito de *sublinguagem* , aplicado às terminologias científicas e técnicas, mais adiante).

termos

sublinguagem

Há, no entanto, outras expressões de uso mais comum do que as que vimos em (iv), isto é, frases que não são tão marcadas do ponto de vista da sua natureza terminológica. Podemos também usar esse tipo de expressões para compreender melhor o que é um *discurso metalinguístico*. Assim, por exemplo, considere as frases:

Terminologia: Entende-se aqui *terminologia* no sentido de uma sublinguagem especializada, limitada tanto do ponto de vista do objeto que nomeia como do grupo sócio-profissional que a emprega. Sobre o conceito de sublinguagem, no sentido harrissiano, falar-se-á mais adiante.

- (1) *A frase* O João é muito esperto *tem cinco palavras*.
- (2) *A palavra* João *tem o som* [ãw̃].
- (3) *A expressão* à-vontade *escreve-se com um hífen quando é usada como um nome*.

Estas frases (ou discursos) são certamente parte da linguagem natural, neste caso de uma língua a que chamamos português, mas

elas têm uma natureza ou um estatuto nesta língua que é diferente do de frases como:

O Pedro gosta de chocolates

Trata-se, efetivamente de *frases metalinguísticas*, na medida em que os elementos que nelas se encontram não têm referente no universo extralinguístico (o mundo real, fora da língua), mas se referem antes a elementos do próprio sistema linguístico. Encontramos elementos especiais (ou *especializados*) para nos referirmos aos restantes elementos da linguagem: termos como *frase* ou *palavra*, mas também termos de uso menos preciso/mais geral como *expressão* (qualquer coisa que se possa dizer é uma *expressão*) ou *som*; outros elementos referem-se especificamente à (orto)grafia da língua, como os termos *letra* ou *hífen*.

frases metalinguísticas

Um outro mecanismo está ainda presente nestas frases. Repare-se que os termos metalinguísticos *frase*, *palavra*, *som* e *expressão* introduzem elementos que fazem parte dos discursos da linguagem propriamente dita. Estes elementos não estão referidos entre aspas (usou-se em vez disso o contra-italico), mas poderiam perfeitamente estar. Este mecanismo de *citação* é, igualmente, um processo de natureza metalinguística, em que os termos metalinguísticos funcionam como *classificadores* de elementos de natureza mais puramente linguística.

citação

classificadores

Noutros casos, outro tipo de mecanismos metalinguísticos são empregues, como por exemplo os que estabelecem relações (por exemplo, de tipo posicional) entre diferentes palavras num discurso. Considere-se o seguinte discurso:

informação posicional

(4) *Na frase* Antes de o comprar, tenho de verificar se o livro está em boas condições, *a palavra* o, *que aparece imediatamente antes da palavra* comprar *refere-se ao mesmo objeto a que se refere a palavra* livro, *que aparece a seguir a se*.

Neste caso, além da informação posicional (uma palavra ou expressão aparece num discurso antes ou depois de outra palavra ou expressão), está também presente uma informação que relaciona as palavras no discurso (isto é, na frase) com objetos fora do universo do discurso, isto é, com objetos do mundo real – aquilo que, numa linguagem mais técnica, se costuma chamar um processo de *referência*. No caso concreto, ao dizer-se que o pronome *o* se refere ao mesmo objeto extralinguístico a que outra palavra, o nome *livro*, também se refere, este discurso metalinguístico está a estabelecer uma relação de *correferência*. Podia-se usar uma notação mais ou menos artificiosa para representar esta relação, como fazem, por exemplo, certos autores quando analisam linguisticamente a frase. Neste sentido, é frequente usar-se *índices de correferência*, como no exemplo

referência

correferência

índices de correferência

abaixo, que retoma o conteúdo do discurso apresentado na frase (4) e usa estes índices para relacionar entre si elementos desse discurso:

(4a) *Antes de o_i comprar, tenho de verificar se o livro $_i$ está em boas condições*

Note-se que o uso de índices de correferência, como o que se ilustra no exemplo acima, constitui uma *convenção* e que esta implica necessariamente um processo de *definição explícita*, para que o leitor a compreenda. Obviamente, a explicitação dessa convenção tem de ser sempre formulada numa língua natural.

convenção

No caso de um sistema de representação mais formalizada, em que se procede à anotação deste tipo de relações e que é destinada a processamento computacional, a mesma relação poderia ser representada assim:

(4b) *Antes de <anaph ID=123>o</anaph> comprar, tenho de verificar se <NP ID=123>o livro</NP> está em boas condições*

Neste caso, apresenta-se uma notação com etiquetas “embebidas” no texto, num formato semelhante ao do XML, em que um par de etiquetas delimita o elemento anafórico (<anaph> ... </anaph>), enquanto o outro par indica o constituinte (<NP> ... </NP>) que é antecedente desse elemento. A relação anafórica é estabelecida pelos identificadores (<ID=123>).

XML: eXtended Markup Language

Ora, é importante salientar que os exemplos (4a) e (4b) *não são* frases da língua. São diferentes meios de notação de um tipo particular de relação linguística observável num dado discurso. Essa notação, para poder ser corretamente interpretada, tem de ser previamente definida, e a única forma de o fazer é usando a própria linguagem. Em rigor, a informação que estes dois exemplos veiculam é exatamente a mesma que a frase (4), com a diferença de que esta é uma frase da língua enquanto (4a) e (4b) são uma representação formalizada convencional do conteúdo – exatamente o mesmo conteúdo – da frase (4).

Em suma, ninguém negará que as frases (1) a (4) são frases da língua. Assim, é preciso não só assumir que a língua contém em si mesma a sua própria metalinguagem e, mais importante, não há nenhuma outra forma de a linguagem se referir a si mesma que não seja usando a própria linguagem.

Conquanto o conjunto de tipos de processos metalinguísticos seja em grande medida semelhante para todas as línguas, cada língua desenvolve os seus recursos metalinguísticos próprios em função das categorias linguística que, nessa língua, são descritivamente relevantes. As diferentes teorias gramaticais, com todo o aparato terminológico que lhe está associado ou pelo emprego de um conjunto mais ou menos sofisticado de notações para representar aos objetos

e as relações gramaticais, não mais são do que um substituto conveniente, para efeitos de descrição/explicação, da própria linguagem que lhes subjaz, ou, para ser mais exato, da metalinguagem que é parte integrante da linguagem a descrever.

A principal consequência desta propriedade fundamental das línguas naturais consiste no facto de que, não havendo um ponto de vista externo a partir do qual se pudesse fazer referência à própria linguagem, a única metodologia possível para descrever o seu funcionamento se baseia em *métodos probabilísticos* (ou *estocásticos*). Dito de uma maneira muito crua, a única forma de descrever a linguagem é descrever o modo como os seus elementos se combinam uns com os outros, o que se expressa em termos de restrições à combinação desses elementos, em rigor, em termos de desvios a partir da equiprobabilidade combinatória desses elementos. Ou seja, a construção de uma gramática é essencialmente um projeto que descreve a linguagem com base na *distribuição* diferencial dos elementos discretos do sistema linguístico. Esses desvios a partir da equiprobabilidade combinatória constituem, pois, a informação veiculada pela linguagem, apresentando os elevados níveis de redundância que permitem a comunicação verbal humana e que são estudados na perspectiva da teoria matemática da informação.

métodos probabilísticos (ou estocásticos)

distribuição

Chegamos assim à formulação essencial que define o método distribucional e que é a base da teoria harrissiana: uma teoria baseada em restrições combinatórias entre os elementos do sistema.

Outra consequência das duas propriedades fundamentais que aqui vimos tem a ver com a natureza do objeto de estudo e a base material que lhe serve de suporte. Em rigor, a gramática harrissiana assume uma postura *descritivista*, presumindo que se dispõe de um *corpus* suficientemente extenso para ser representativo dos discursos de uma língua natural. Na medida em que as expressões metalinguísticas fazem parte da linguagem natural, deverá haver, nesse *corpus* ideal, instâncias suficientes para determinar com rigor o seu comportamento, quer dizer, a sua distribuição.

descritivismo

O que é importante reter aqui é que, em última instância, não parece necessário (e não seria epistemologicamente adequado) recorrer a mecanismos externos à linguagem, notações complicadas, nem esquemas esotéricos para descrever o comportamento das línguas naturais. Na ausência de uma metalinguagem externa, a linguagem natural provê todos os meios necessários para representar a informação que os seus discursos veiculam. Quaisquer terminologias, notações, esquemas e outras estratégias, por mais complexos que sejam (e podem mesmo ser muito complexos!), são, em última instância, meros substitutos de – e têm a sua origem em – as propriedades

distribucionais que os elementos discretos da linguagem apresentam nos discursos.

Esta ênfase num *corpus* de discursos como base empírica para a investigação sobre o funcionamento da linguagem não significa que Harris negue a famosa dicotomia saussuriana (de Saussure, 1916) entre, por um lado a *língua* (fr. *langue*), enquanto abstração teórica, um conhecimento de algum modo partilhado por uma comunidade de falantes, e a *fala* (fr. *parole*), enquanto manifestação concreta desse comportamento humano, nas diferentes *modalidades*: fala, escrita e gesto (para me referir, no resto deste documento, a esta manifestação da língua, utilizarei sobretudo o termo *linguagem* (por oposição ao termo *língua*), reservando o termo *fala* para a modalidade oral-auditiva da linguagem). Contudo, Harris foca-se numa abordagem descritiva, isto é, que parte de instâncias concretas para determinar o conjunto de elementos discretos da linguagem e as suas propriedades, com base nas diferenças que estes apresentam entre si quanto à sua distribuição. Os métodos estocásticos utilizados para tal são, de resto, os únicos que é possível utilizar no caso de *línguas mortas*, isto é, sem falantes nativos vivos, baseando-se nos diferentes registros que chegaram até à atualidade. Neste caso, apenas a evidência empírica que nos é dada pelos discursos que sobreviveram à usura do tempo podem ser empregues para descrever essas línguas.

No entanto, no caso das *línguas vivas*, é possível complementar a informação distribucional, obtida a partir de *corpora* dos discursos dessas línguas, recorrendo a dados obtidos por *introspeção* a partir de falantes nativos, mediante certas cautelas metodológicas. Sobre tal, veja-se, entre outros, a discussão de Laporte (2015).

Antes de terminar esta secção, deixa-se apenas uma nota sobre o aparente “superficialismo” (não confundir com superficialidade!) da abordagem harrissiana à descrição da linguagem, conceito que é entendido aqui no sentido de uma análise muito próxima à materialidade das formas linguísticas efetivamente observáveis nos discursos. Esta perspectiva não põe em causa a existência de uma metalinguagem interna, ou seja, de frases como as que ilustrámos mais acima. As frases metalinguísticas veiculam sobretudo uma informação que se encontra, frequentemente, reduzida porque redundante. É essa *redundância* que permite reconstituir, com elevado grau de *precisão* e *reprodutibilidade* os elementos que foram reduzidos dos discursos. O funcionamento dos discursos efetivamente observáveis (em superfície) não poderia ser justificado sem o recurso a essas expressões, elas próprias parte integrante da linguagem, que veiculam a informação que não se encontra numa forma aberta (ing. *overt*). O objetivo da análise é, pois, reconstituir a partir dos discursos observados os ele-

de Saussure: *langue/parole*

modalidades da linguagem humana:
fala, escrita e gesto

línguas mortas

línguas vivas

introspeção
falante nativo

superficialismo

redundância

mentos reduzidos e, com eles, explicitar a informação veiculada por esses discursos.

A teoria linguística deve, portanto, desenvolver os métodos adequados para que essa análise se possa realizar de forma controlada, sistemática e reproduzível. Tais métodos dependem crucialmente do objeto a descrever. Ora, enquanto objeto de estudo, a linguagem estrutura-se com base num conjunto de restrições à combinatória dos seus elementos discretos. É dessas restrições fundamentais que trata o capítulo seguinte.

Restrições fundamentais

A linguagem natural estrutura-se a partir de um conjunto de quatro restrições fundamentais e duas restrições complementares. As quatro restrições fundamentais são as seguintes:

restrições fundamentais

- (1) ordem parcial de entrada das palavras nas frases;
- (2) probabilidade de co-ocorrência operador-argumento(s);
- (3) redução; e
- (4) linearização.

As duas restrições complementares operam apenas entre sequências de frases concatenadas no discurso:

restrições complementares

- (1) a 1ª restrição complementar impõe a repetição de elementos ao longo de sequências de frases concatenadas no discurso; enquanto que
- (2) a 2ª restrição complementar impõe a repetição de relações operador-argumento(s) ao longo de sequências de frases concatenadas no discurso.

Por uma questão de clareza iremos apresentar cada uma destas restrições sequencialmente, embora seja por vezes difícil isolá-las, já que todas operam simultaneamente na construção do discurso.

Ordem parcial de entrada das palavras nas frases

A 1ª restrição fundamental é igualmente a restrição mais importante, dela derivando todas as outras. Efetivamente, uma única relação, a ordem parcial de entrada das palavras na frase, é responsável por toda a estrutura da informação na linguagem. Esta restrição consiste em verificar que, para uma dada palavra poder entrar numa frase e formar um discurso, é necessário que já estejam presentes nessa frase uma ou mais outras palavras. Assim, por exemplo, para uma palavra

como *cair* entrar numa frase e formar um discurso é necessário que já lá esteja presente outra palavra, por exemplo, *jarra*, o que dá origem ao discurso:

A jarra caiu

De modo semelhante, para uma palavra como *ler* são necessárias duas outras palavras, por exemplo, *Pedro* e *livro*, para formar o discurso

O Pedro leu o livro

Já com a palavra *emprestar* são precisos três outros elementos, por exemplo, *Pedro*, *livro* e *João*, a fim de construir a frase

O Pedro emprestou o livro ao João

Esta restrição cria assim uma assimetria fundamental entre os elementos do léxico, organizando-os a todos em dois conjuntos distintos:

- | | |
|---|------------|
| (i) os <i>operadores</i> , ou seja, as palavras (ou morfemas) que, para poderem entrar numa frase requerem a entrada prévia de outras palavras antes delas; e | operadores |
| (ii) os <i>argumentos</i> , ou seja, as palavras (ou morfemas) sobre as quais os operadores atuam. | argumentos |

Por uma questão de simplicidade, não iremos falar agora de <i>morfemas</i> , que, de resto, representam uma definição mais rigorosa destes conceitos, e apresentaremos os conceitos apenas com base em <i>palavras</i> , como se estas fossem formadas por um único morfema. Mais adiante veremos como esta distinção é extremamente relevante na descrição de alguns fenómenos linguísticos.	morfemas/palavras
---	-------------------

Assim, na sequência dos exemplos acima, diremos que *cair* é um operador sobre *jarra*; *ler* um operador sobre o par ordenado de argumentos constituído por *Pedro* e *livro*; e *emprestar* um operador sobre o conjunto ordenado de três elementos, a saber, *Pedro*, *livro* e *João*.

par ordenado

Por sua vez, a classe dos operadores pode ser organizada em dois subconjuntos. Por um lado, encontramos os *operadores de primeira ordem*, isto é, aquelas palavras cujos argumentos não podem ser eles mesmos operadores; por outro lado, os *operadores de segunda ordem*, isto é, aquelas palavras cujos argumentos incluem pelo menos um operador.

operadores de primeira ordem
operadores de segunda ordem

Note-se, em primeiro lugar, que os operadores de primeira ordem determinam, assim, um conjunto particular do léxico, os *argumentos elementares*, ou seja, as palavras que entram sempre em primeiro lugar na frase, não requerendo a presença prévia de nenhuma palavra

argumentos elementares

antes delas. Trata-se, em geral, de nomes comuns que designam objetos concretos (e.g. *livro*) ou nomes próprios que designam diferentes tipos de entidades (e.g. *Pedro*).

Em segundo lugar, os operadores de segunda ordem requerem que pelo menos um argumento do seu domínio de argumentos seja ele mesmo um operador. No caso de um domínio de argumentos com aridade superior a 1, os outros argumentos podem ser tanto argumentos elementares como operadores.

Entre os operadores de segunda ordem incluem-se todos os elementos predicativos que selecionam uma oração subordinada completa, como, por exemplo, *importar*. Numa frase como:

Importa ler este livro

dizemos que *importar* é um operador de segunda ordem que opera sobre *ler*; e que este último é um operador de primeira ordem, que opera sobre um argumento, por exemplo, *as pessoas*, e *livro*; o sujeito de *ler* designa um elemento indefinido, que se encontra aqui reduzido (ver adiante, a secção sobre a 3ª restrição, *redução*), e tal é responsável pela interpretação genérica da construção.

Constitui-se, assim, uma *ordem parcial de entrada das palavras na frase*, na medida em que cada operador entra na frase operando diretamente sobre o seu respetivo domínio de argumentos; no caso de um desses argumentos ser ele próprio um operador, o operador superveniente opera sobre o *resultado* da operação do operador anterior sobre o seu respetivo domínio de argumentos. É possível representar esta ordem parcial dizendo que uma palavra entra *antes* '<' ou *depois* '>' de outra palavra na frase. Uma palavra pode também entrar *ao mesmo tempo* que outra palavra, caso em que ambas são argumentos do operador que sobrevenha. Assim, para o exemplo que demos acima, poderíamos representar a ordem parcial de entrada das palavras do seguinte modo:

importar>ler>(alguém, livro)

Note-se que *alguém* e *livro* entram simultaneamente na frase, sob *ler*, daí se encontrarem ligados por vírgula entre parênteses.

A tarefa da análise linguística consiste, justamente, em reconstituir esta ordem parcial de entrada, já que ela representa uma parte substancial da informação veiculada pelos discursos. A tarefa não é fácil já que uma dada expressão superficial pode, devido a diversos fenómenos de redução, implicar a reconstituição desses elementos reduzidos. Por exemplo, a frase:

O Pedro gosta de chocolates

domínio argumental

aridade: número de elementos de um conjunto

Os operadores de segunda ordem operam sobre o *resultado* da operação prévia dos seus argumentos que são eles mesmos operadores

parece poder autorizar, pelo menos provisoriamente, a classificação do verbo *gostar* como um operador com dois argumentos elementares, v.g. *Pedro* e *chocolates*. Contudo, uma análise mais cuidadosa leva a considerar igualmente a seguinte construção deste verbo:

O Pedro gosta de fazer isso.

em que ele aparece construído com uma completiva-objeto (*fazer isso*), o que levaria a classificá-lo como um operador de segunda ordem, em que o primeiro argumento (*Pedro*¹) é um argumento elementar e o segundo (*fazer isso*) é o resultado de um operador (*fazer*) sobre o seu domínio de argumentos (*Pedro*² e *isso*). Ora, ao considerar que o significado do verbo *gostar* é o mesmo em ambas as frases, somos levados a reavaliar a classificação provisória que atribuímos ao verbo *gostar de*, à luz desta segunda construção. Como é possível reconstituir um elemento (*comer*) que opera entre o verbo *gostar* e o nome *chocolates*, tal como exemplificado na frase seguinte:

O Pedro gosta de (comer) chocolates

somos levados a classificar *gostar* como um operador *Ono*.

Repare-se que isto só é possível porque a frase acima é uma paráfrase da frase de partida, ou seja, verifica-se uma invariância de significado. Por outras palavras, diremos que um operador opera imediatamente sobre um determinado domínio de argumentos *sse* não houver nenhum outro operador que possa ter entrado na frase antes desse operador e depois dos elementos que se pretende analisar como argumentos do operador de topo. Como não é esse o caso aqui, consideraremos que a ordem parcial da frase acima será :

gostar de > (*Pedro*¹, *comer* > (*Pedro*², *chocolates*))

ou seja, *gostar de* opera sobre dois argumentos: um argumento elementar, o nome *Pedro*¹, e o operador *comer*; por sua vez, este segundo verbo opera sobre dois argumentos elementares, o nome *Pedro*² e *chocolates*.

Tanto quanto se sabe, a maioria dos operadores das línguas naturais têm apenas 1 ou 2 argumentos (e.g. *cair* ou *ler*, respetivamente), e há algumas classes interessantes (quer pelo número de elementos, quer pelo tipo de predicado que expressam) de operadores com três argumentos (e.g. *emprestar*, *dizer*, *colocar*). Os casos de operadores com mais do que 3 argumentos são raros ou podem por vezes ser analisados como o resultado da concatenação de dois ou mais operadores (por uma questão de clareza deixaremos essa discussão para outro momento).

Inversamente, muitas línguas naturais têm um pequeno conjunto de operadores que exprimem predicados denotando *fenómenos meteorológicos* (e.g. *chover*, *nevar*, *trovejar*) ou *partes do dia* (e.g. *amanhecer*,

completiva-objeto

Distinguimos as duas instâncias do nome *Pedro*, o sujeito de *gostar de* e o de *fazer*, com índices numéricos; o facto de estes nomes serem iguais e se referirem à mesma entidade extralinguística não é aqui considerado (ver adiante o capítulo sobre *redução*).

sse = se e só se

Os elementos lexicais repetidos foram aqui numerados sequencialmente para mais fácil referência. Não explicamos agora a questão da correferência entre o sujeito de *gostar de* (*Pedro*¹) e de *comer* (*Pedro*²), que explica porque este último também sofre uma redução. Voltaremos a falar deste assunto quando tratarmos da 3ª restrição fundamental, justamente a *redução*.

Sobre este tipo de predicados, veja-se Ruwet (1986).
fenómenos meteorológicos
partes do dia

anoitecer, entardecer), que apresentam aquilo a que, em gramática tradicional se chama *construções impessoais*, sem qualquer argumento (sujeito). Em português, há ainda o caso absolutamente excepcional da construção impessoal do verbo *tratar-se de*, e.g.:

Iremos falar deste assunto. Trata-se de um tema muito importante

Trata-se, de qualquer modo, de fenómenos perfeitamente marginais no conjunto das construções verbais do Português, que estudaremos com mais pormenor mais adiante.

Nas línguas como o português, o fenómeno gramatical das construções impessoais (sem sujeito) não é problemático, já esta língua não exige a expressão sintática (em superfície) do sujeito gramatical. Já noutras línguas, como o Inglês ou o Francês, o preenchimento desta posição sintática de sujeito é obrigatório, não pode ser deixado vazio, pelo que esta posição é preenchida por um pronome de terceira pessoa singular neutro; v.g. ing.: *it*, fr.: *il*):

It rains; Il pleut

Note-se que este pronome, que constrói as expressões meteorológicas ou de parte de dia, não se confunde com a forma do pronome pessoal de terceira pessoa, já que, neste caso, não se observa variação em género, sendo as formas masculina e feminina inaceitáveis:

**He/She rains; *Elle pleut*

É possível, assim, construir a seguinte *tipologia de operadores* (Tabela 1), na qual representamos os respetivos domínios de argumentos com índices *o* ou *n*, consoante se trate, respetivamente, de operadores ou de argumentos elementares. Falaremos, pois, de operadores do tipo *O_o*, *O_n*, *O_{oo}*, *O_{on}*, *O_{no}*, *O_{nn}*, etc.

		nº de argumentos		
		1 argumentos	2 argumentos	3 argumentos
tipo de operador	2ª ordem	<i>Oo</i> : <i>importar</i>	<i>Ooo</i> : <i>causar</i>	<i>Onoo</i> : <i>preferir</i>
	1ª ordem		<i>Ono</i> : <i>pensar</i>	<i>Onon</i> : <i>dizer</i>
			<i>Oon</i> : <i>alegrar</i>	<i>Onno</i> : <i>obrigar</i>
		<i>On</i> : <i>cair, zumbir</i>	<i>Onn</i> : <i>amar</i>	<i>Onnn</i> : <i>emprestar</i>
argumentos elementares (n)				

Tenha-se em atenção que quase todos os verbos indicados nesta tabela podem apresentar mais do que uma *construção* léxico-sintática-semântica. Por essa razão, não é o verbo (unidade morfológica) mas sim a construção do verbo que pode ser classificada como operador

construções impessoais

Este tipo de sujeito impessoal *ele* também está atestado em variantes dialetais do Português europeu, e.g. *Ele/*Ela* *choveu* muito esta noite.

O símbolo */** assinala a inaceitabilidade da expressão.

Tabela 1: Tipologia de operadores (verbos).

construção

de um determinado tipo. No caso de um verbo ter mais do que uma construção – e, por consequência, diferentes significados (*homonímia*), suficientemente distintos uns dos outros para poderem ser determinados de forma clara e reproduzível – é necessário desdobrar esse verbo, enquanto unidade morfológica, em diferentes entradas lexicais distintas, em rigor, entradas léxico-sintático-semânticas.

Assim, por exemplo, o operador representado pelo verbo *importar*¹ / *Oo* corresponde ao emprego:

*Oo: importar*¹: *Importa que isso se faça assim*
(= *ser importante*)

Esta construção não se confunde com o operador do verbo *importar*² / *Ono*:

*Ono: importar*²: *O Pedro importa-se com que o João faça isso*
(= *incomodar-se*)

nem com o operador do verbo *importar*³ / *Onnnn*:

*Onnnn: importar*³: *O Pedro importou bananas da Madeira para o continente*
(= *comprar, transportar; antônimo = exportar*)

Em relação às construções dos restantes verbos representadas na Tabela 1, a maioria é relativamente evidente, mas algumas merecem um comentário suplementar. Trata-se das construções ilustradas nos exemplos a seguir.

Em primeiro lugar, a par da construção de *importar*¹ / *Oo*, encontramos também neste grupo de operadores todas as construções de verbo auxiliar, como *estar a Vinf*:

Oo: estar a Vinf: O Pedro está a ler o livro

A construção de *verbo auxiliar* (aspetual) *estar a Vinf* – em que *Vinf* representa qualquer verbo no infinito não flexionado ou impessoal – é analisada como um operador do tipo *Oo*, cujo argumento é o verbo que auxilia.

Em segundo lugar, o verbo *preferir* exige, além do seu sujeito humano, duas completivas, em contraponto, daí a sua classificação como *Onoo*.

Onoo: preferir: O Pedro prefere ler um livro a ver televisão

Saliente-se que esta 1ª restrição fundamental *não* determina que um operador de segunda ordem exija que o operador inferior (seu argumento) seja ele mesmo um operador *de um tipo particular*. Por outras palavras, a 1ª restrição não permite dizer que, para um operador qualquer, por exemplo do tipo *Ono* (e.g. *adorar*), o seu argumento

homonímia

desdobramento de entradas lexicais

verbo auxiliar: Na sequência de Baptista et al. (2010), consideramos que os verbos auxiliares são transparentes às restrições de seleção que o verbo principal (o verbo principal, que é auxiliado por estes verbos) impõe ao preenchimento lexical das respetivas posições argumentais. Nesse sentido, os verbos auxiliares introduzem na construção do verbo sobre o qual operam diferentes valores gramaticais, tipicamente valores temporais, aspetuais e modais. Nesta perspetiva, os verbos auxiliares podem ser vistos como meros instrumentos gramaticais, cuja função, em termos da informação veiculada, é semelhante à dos morfemas de flexão nas formas simples dos verbos plenos (ou distribucionais).

o tenha de ser de um tipo específico qualquer, por exemplo, um *On* (e.g. *bocejar*). Dito de outro modo, quando um operador seleciona para seu argumento um outro operador, qualquer palavra que pertença à classe dos operadores pode satisfazer a 1ª restrição e ocupar essa posição argumental.

Finalmente, o leitor mais atento terá reparado que apenas se deram até aqui exemplos de operadores com verbos. Ora, a noção de operador define uma restrição fundamental que está para além das chamadas *partes do discurso* (ou *classes de palavras* ou, ainda, *categorias morfossintáticas*) tradicionais, já que apenas se baseia na ordem parcial de entrada das palavras na frase. Por essa razão, há que considerar a existência operadores em praticamente todas as classes de palavras, todas as categorias gramaticais tradicionais, quando se trata de elementos predicativos. Em contraponto, palavras de certas categorias, como alguns adjetivos derivados de nomes, por exemplo, o adjetivo *lunar* (<*lua*), não deverão ser considerados como formas da base mas sim o resultado de operações sobre nomes (argumentos elementares) quando se encontram sob outro operador. É o que acontece em:

a gravidade lunar
 ← *a gravidade da lua*
 ← *A lua tem/exerce (uma certa) gravidade*

o que explica as suas propriedades sintáticas, nomeadamente, neste exemplo, o facto de *lunar* não ser um adjetivo predicativo, isto é, não aceitar um contexto claramente predicativo, como a construção com um *verbo copulativo* (como *ser* ou *estar*):

**a gravidade é/está lunar*

Enquanto as categorias gramaticais tradicionais são definidas com base numa confluência de diversos critérios (morfológicos, sintáticos, semânticos e até pragmáticos), estes critérios não se traduzem numa única propriedade fundamental comum. Pelo contrário, a relação operador-argumento(s), que instaura a ordem parcial de entrada das palavras na frase, é uma única relação fundamental, que abrange, pois, todas as categorias gramaticais, não só as categorias ditas *lexicais* (e.g. os verbos, os nomes e os adjetivos predicativos, e os advérbios), mas inclui também outros elementos, geralmente integrados nas chamadas classes de *palavras funcionais* (as preposições e as conjunções).

Assim, por exemplo, a par dos verbos, encontramos operadores de outras classes gramaticais. Note-se que o facto de uma palavra ser *simples* ou *composta* (*locução*) é irrelevante para efeitos da classificação gramatical.

- **nomes - Oo** : *importância* (*ter*), *urgência* (*ter - de/em*), *costume* (*ser*);

partes do discurso, classes de palavras,
categorias morfossintáticas

adjetivo predicativo
contexto predicativo
verbo copulativo

categorias "lexicais"

categorias "funcionais"
Não trataremos agora das classes de
determinantes e das interjeições.

palavra simples/palavra composta
(locução)

No caso dos nomes predicativos,
indicamos ao lado o *verbo-suporte* da
respetiva construção de base.

Ooo : causa (ser Det – de); *Ono* : pressa (ter), questão (fazer – de/em);

- **adjetivos** - *Oo* : evidente (ser), importante (ser), urgente (ser), provável (ser); *On* : gordo (estar, ser), alto (estar, ser), forte (estar, ser); *Onn* : alérgico (ser); *Ono* : relutante (estar); *Oon* : irritante (ser);
- **advérbios** - *Oo* : provavelmente; *Ooo* : portanto, no entanto, consequentemente;
- **conjunções** - *Ooo* : e, mas, para, porque, a fim de;
- **preposições** - *Onn* : em, em cima de, por trás de.

Esta gramática sem partes-do-discurso não significa que se ponha em causa o conceito, já que, como se percebe rapidamente, a 1ª restrição permite recortar algumas das categorias gramaticais. Por exemplo, todas as construções de verbos auxiliares são do tipo *Oo*; a classe dos advérbios é formada, na sua maioria, por operadores do tipo *Oo* e *Ooo*; todas as conjunções são do tipo *Ooo*; a maioria dos adjetivos e nomes predicativos são do tipo *On* (*azul*, *alto*; *sono* (ter), *zelo* (ter)) e *Onn* (*subordinado* (estar/ser – a); *comando* (estar ao – de)); no caso dos adjetivos, há um importante número de construções completivas, do tipo *Oon* (*irritante*) e mais raramente *Ono* (*ansioso*), etc., embora o fenómeno seja muito mais raro no caso dos nomes predicativos (*expectativa* (ter – de)).

Qualquer palavra na língua tem necessariamente de estar integrada ou na classe dos operadores ou na dos argumentos elementares. Isto é particularmente notório no caso de *neologismos*, isto é, de palavras novas que, quando “cunhadas”, são imediata e necessariamente construídas ou como um operador ou como um argumento elementar.

Na medida em que o operador exerce a sua ação sobre o seu domínio de argumentos, ele é um predicado sobre esses argumentos, ou seja, o operador “diz qualquer coisa sobre” esses argumentos. Neste sentido, as relações operador-argumento(s) são responsáveis pela maior parte da informação veiculada na frase, a qual se traduz, em grande medida, no significado das expressões linguísticas.

A 1ª restrição cria todas as frases do *kernel* da língua, as *frases da base da gramática*, estruturando todos os seus elementos discretos com base numa única relação fundamental, a ordem de entrada das palavras na frase.

Probabilidade de co-ocorrência operador-argumento(s)

Esta 2ª restrição fundamental está diretamente ligada à observação inicial, isto é, a constatação de que os elementos discretos do sistema linguístico apresentam importantes desvios a partir da equiprobabilidade combinatória. Na secção anterior, a 1ª restrição traduzia essa

No caso dos adjetivos predicativos, indicamos ao lado o *verbo copulativo* que auxilia a respetiva construção de base.

uma gramática sem partes-do-discurso

neologismos

frases do *kernel* da língua ou frases da base da gramática

observação em termos de relações entre operadores e de relações entre estes e os argumentos elementares. Nesta secção, essa observação é agora enquadrada nas relações entre o operador o seu domínio de argumentos. Neste sentido, é possível verificar que, para um operador de um determinado tipo, certos elementos que satisfazem a 1ª restrição têm uma maior probabilidade do que outros de ocorrer como seus argumentos.

Assim, por exemplo, para o operador *On*: *zumbir*, alguns argumentos elementares têm uma probabilidade maior do que outros de ocorrer como argumentos desse verbo. É possível, ainda que de uma forma grosseira, ordenar intuitivamente esses nomes relativamente à probabilidade de co-ocorrência como argumentos do verbo. Num extremo, com uma probabilidade relativamente normal-alta, encontraríamos nomes como *abelha*, *mosca*, *mosquito*, etc., e outros nomes que designem *insetos* (alados):

A abelha/mosca zumbia

A maioria, porém, dos nomes da língua não ocorre como argumento deste verbo.

Acrescente-se que é relativamente a este *emprego* dito *literal* que se definem os *empregos figurados* ou *metafóricos*, incluindo os que são usados no sentido de *forma de dizer*, ou seja, aquilo a que se convencionou chamar *verbum dicendi*. Por outras palavras, frases como:

A impressora zumbia <suavemente>
 “–Não quero fazer isso!”, *zumbiu o Pedro*

deverão ser consideradas como derivando, através de processos que envolvem, nomeadamente, um processo de comparação, da construção literal, acima ilustrada.

Repare-se que, na definição da 2ª restrição, se determina a probabilidade de co-ocorrência operador-argumentos relativamente ao *tuplo ordenado* de palavras que constituem o *domínio argumental* desse operador. Assim, por exemplo, para um operador do tipo *Onn* como *comer*, tanto o par ordenado (*Pedro, bife*) como o par ordenado (*bife, Pedro*) satisfazem ambos a 1ª restrição, já que são ambos argumentos elementares, mas só o primeiro par ordenado satisfaz a 2ª restrição:

O Pedro comeu um bife
 **O bife comeu o Pedro*

Do mesmo modo, ambos os pares ordenados (*vaca, erva^{PE}/capim^{PB}*) e (*erva^{PE}/capim^{PB}, vaca*) podem satisfazer a 1ª restrição mas também aqui apenas o primeiro par ordenado satisfaz a 2ª restrição:

emprego literal vs. figurado

verbum dicendi (do lat. ‘verbo de dizer’)

tuplo ordenado de argumentos
 domínio argumental

A vaca comeu a erva/o capim
 **A erva/O capim comeu a vaca*

Isto acontece porque a 2ª restrição está diretamente ligada à 4ª restrição fundamental – *linearização* (ver adiante), a qual determina que cada argumento desempenhe, relativamente ao operador, um determinado *papel semântico* (AGENTE, OBJETO, etc.). Ora, justamente, a forma como certas línguas, como o português, procedem à atribuição do papel semântico aos argumentos de um predicado está em grande medida ligada à ordem desses elementos relativamente o operador de que dependem. Em concreto, trata-se da atribuição das posições que correspondem às funções sintáticas de sujeito ou complemento numa frase de base na voz ativa. Na ordem básica dos constituintes determinada pela língua, o sujeito ocorre normalmente antes do verbo e o complemento depois. A natureza da relação semântica expressa nesse papel está, naturalmente, muito estreitamente relacionada com a *classe distribucional* a que pertencem os elementos que podem ocupar essa posição argumental (nome humano/não-humano, por exemplo).

Portanto, a 2ª restrição, ao determinar uma probabilidade de ocorrência operador-argumentos, não o faz relativamente a cada argumento, separadamente, mas sim em relação ao *tuplo* ordenado de palavras que constituem o domínio argumental de um dado operador. Por outras palavras, esta 2ª restrição não se limita a dizer que *Pedro* e *bife* são, *separadamente*, bons argumentos de *comer*; ou *vaca* e *erva^{PE}/capim^{PB}*, relativamente a *comer*. De facto, trata-se, de uma restrição bastante mais complexa, pois determina que o grau de probabilidade de co-ocorrência do operador com o seu domínio de argumentos resulta ainda do facto de *Pedro* ou *vaca* desempenharem o papel semântico de AGENTE do processo descrito por *comer*; e *bife* e *erva^{PE}/capim^{PB}* desempenharem o papel de OBJETO. Tal traduz-se, em línguas como o português, pelo facto de estes argumentos ocuparem diferentes posições na frase, que correspondem a diferentes funções sintáticas, nomeadamente as funções de sujeito e de complemento.

Nos exemplos acima, os exemplos estavam na voz ativa. Note-se, também, que uma transformação, como a *apassivação* (ou *Passiva*), produz frases aceitáveis (naturais), desde que estas alterações da forma das frases operem sobre os tuplos adequados:

O bife foi comido pelo Pedro.
A erva/o capim foi comida pela vaca.

e expressões inaceitáveis com os tuplos inadequados:

**O Pedro foi comido pelo bife.*
 **A vaca foi comida pela erva/pelo capim.*

linearização

papel semântico

classe distribucional

nome humano/não-humano

transformação

Passiva

Aliás, o conceito de *transformação* é, justamente, definido com base nesta propriedade: a conservação do grau de (in-)aceitabilidade de uma dada distribuição operador-argumento(s) através das diferentes transformações a que uma dada construção pode ser sujeita. Na esteira de Harris (1955, 1961), consideramos aqui transformações como as relações não orientadas, de equivalência parafrástica, entre pares frases, que não alteram as relações semânticas entre os elementos das frases, nomeadamente os papéis semânticos dos argumentos relativamente ao operador de que dependem, mas sim a sua ordem linear e, eventualmente a estrutura sintática da construção.

O que se quer enfatizar aqui, neste momento, é que as restrições de selecção se determinam relativamente aos *tuplos* e não para cada posição argumental separadamente.

Repare-se, portanto, que tanto *Pedro* como *vaca* são bons primeiros argumentos do operador *comer*; e que tanto *bife* como *erva* são bons segundos argumentos deste operador. Contudo, dada esta propriedade de a 2ª restrição se definir relativamente aos *tuplos* e não para cada posição independentemente, tal leva a que as expressões como:

?,*O Pedro comeu a erva/o capim

?,*A vaca comeu o bife

transformação

conservação da (in-)aceitabilidade das combinações operador-argumento(s)

Os símbolos '?' e '*' assinalam que a expressão é de aceitabilidade duvidosa ou muito duvidosa, respetivamente.

se encontrem abaixo do nível de probabilidade normal de co-ocorrência definível para um verbo como *comer*.

Esta 2ª restrição corresponde, pois, àquilo que, em sintaxe contemporânea, se chama *restrições de selecção*. Diz-se, por exemplo, que o verbo "restringe os elementos que seleciona" para ocupar as diferentes posições argumentais desse verbo. Neste caso, com *zumbir*, poderíamos falar do conjunto de nomes que designam *insetos alados*.

restrições de selecção

Note-se que muitos traços distribucionais mais não são do que formas de simplificar o processo de descrever as complexas restrições que se observam entre um elemento predicativo e o seu domínio de argumentos. Assim, por exemplo, um verbo como *calçar* seleciona praticamente qualquer peça de <calçado> (*botas, chinelos, pantufas, sapatos, etc.*) mas também *meias* (para os pés) e, surpreendentemente, *luvas* (para a mãos).

Esta forma de conceber as restrições de selecção, no quadro harriano, também pode ser interpretada no sentido inverso. Na medida em que se pode caracterizar (e quantificar) a probabilidade de um dado operador ocorrer com um determinado conjunto de elementos como seus argumentos, é igualmente possível caracterizar (e quantificar) a probabilidade de um dado elemento poder ocorrer como argumento de um conjunto de operadores. Assim, por exemplo, é possível dizer que, para um nome como *mosquito*, certos verbos têm uma probabilidade muito mais alta (e.g. *picar, pousar, transmitir, voar,*

zumbir, etc.) do que outros (e.g. *nidificar*, *respirar*, etc.), enquanto que alguns terão mesmo uma probabilidade muito baixa (e.g. *cantar*).

Na perspectiva mais habitual (*zumbir* vs. *insetos*), falamos então de uma seleção “para baixo” (do ing. *downward selection*), enquanto que, no sentido inverso (*mosquito* vs. *voar*, *respirar*, *nidificar*), falaríamos de seleção “para cima” (do ing. *upward selection*).

Apesar de a termos definido acima de forma intuitiva, grosseira, a probabilidade de co-ocorrência operador-argumento(s) pode, contudo, ser quantificada com rigor, com base nas ocorrências desse operador num *corpus* de dimensões apreciáveis.

Em muitas situações, uma descrição grosseira do grau de probabilidade de co-ocorrência é suficiente. Harris considera cinco níveis de probabilidade: normal; superior ou inferior ao normal; e probabilidade excepcionalmente alta ou excepcionalmente baixa. Vejamos, separadamente, exemplos de cada um destes níveis de probabilidade de co-ocorrência.

- casos com uma probabilidade de co-ocorrência *normal*: trata-se da situação em que uma palavra ocorre sem que nenhum efeito de estranheza/naturalidade seja especialmente percebido;
 – normal
- casos com uma probabilidade de co-ocorrência *excepcionalmente alta*: são situações em que apenas um elemento ou um número muito pequeno de elementos pode ocorrer numa dada posição argumental, razão por que esse elemento pode ser reduzido sem perda de informação (fenómeno a que Harris chamou *redução apropriada*, e de que voltaremos a falar mais adiante);
 – excepcionalmente alta
- casos com uma probabilidade de co-ocorrência *superior à normal*: consiste na situação em que é relativamente simples enumerar, por vezes de forma extensional, o conjunto de elementos que constituem os argumentos preferenciais desse operador (o exemplo de *zumbir* e dos < *insetos* > alados vêm de novo à mente);
 – superior à normal
- casos com uma probabilidade de co-ocorrência *inferior à normal*: trata-se de combinações operador-argumento(s) pouco usuais e que, em grande medida, são motivados por fenómenos de natureza tropológica (metáforas), envolvendo diversos tipos de redução;
 – inferior à normal
- casos com uma probabilidade de co-ocorrência *excepcionalmente baixa*: são as situações em que a maioria dos falantes consideraria a expressão como semanticamente incongruente.
 – excepcionalmente baixa

Repare-se que, no quadro harrissiano nenhuma combinação de palavras pode ser descrita como tendo uma probabilidade *zero*. Desde que a 1ª restrição não seja violada, qualquer combinação de um elemento lexical (operador) com outros elementos lexicais (seus argu-

selecção “para baixo”
(*downward selection*)

selecção “para cima”
(*upward selection*)

graus de probabilidade
de co-ocorrência:

mentos) tem uma determinada probabilidade de ocorrência – que pode ser extremamente baixa, mas é sempre diferente de zero.

Nesse sentido, não é exacto considerar que mesmo a famosa frase

As ideias verdes dormem furiosamente

não possa ter uma probabilidade associada .

Ora, ao conceber a 2ª restrição desta forma, o facto de poder haver combinações com graus de probabilidade variados permite que o sistema linguístico, como um todo, possa evoluir, historicamente, na medida em que certas combinatórias vão, ao longo do tempo, vindo variar a verosimilhança com que podem ocorrer. De facto, muitas expressões modernas e que são hoje, de algum modo, excêntricas, tiveram provavelmente origem em combinatórias com um grau de probabilidade normal. Por exemplo, a preposição *trás* apenas sobrevive no topónimo composto *Trás-os-Montes* (região nordeste de Portugal continental) e nas locuções prepositivas (*por*) *detrás de*, *atrás de*. Contudo, não só a sobrevivência no topónimo, como o mecanismo de formação das locuções, ambos atestam a natureza dinâmica da língua, em particular quanto à probabilidade de co-ocorrência dos seus elementos.

Em contraponto, ao considerar diferentes graus de probabilidade de co-ocorrência de certas palavras como argumentos de um dado operador, encontramos um conjunto relativamente pequeno de palavras que tem uma probabilidade particularmente alta de ocorrer como argumento de praticamente qualquer operador. Trata-se das palavras e expressões a que chamamos *proformas*, na medida em que a sua principal função na língua é a de poder substituir outros elementos sem perda substancial de informação. O termo *proformas*, decalcado de “pro-nomes” (que ocorrem *em vez de* nomes), pretende, justamente, dar conta deste seu estatuto particular.

As *proformas* incluem, entre outras expressões:

- os pronomes indefinidos, e.g. *alguém, ninguém, algo, nada, tudo*;
- um conjunto de advérbios de valor genérico, e.g. *algures, assim*; ou com valor referencial, v.g. *aí, aqui, ali, cá, lá, acolá*;
- ou expressões nominais com valor genérico/indefinido, como *pessoa, coisa* /^{PB}*negócio, lugar/local* /^{PE}*sítio*, etc.; a estes nomes genéricos chamamos *nomes não-específicos* (Nne).

Os pronomes indefinidos e os nomes genéricos têm, além disso, uma importante função na arquitetura da gramática de operadores harrissiana: eles servem para satisfazer a 1ª restrição quando não é preciso, por qualquer razão, fornecer uma informação específica sobre um dos argumentos de um dado operador. Por exemplo, numa frase como:

Podemos considerar que a probabilidade p de uma palavra w é a razão do número de vezes c que ela ocorre num *corpus* pelo número total de palavras nesse *corpus* N , ou seja $p(w) = c/N$; e, por sua vez, considerando que a probabilidade de uma frase S é dada pelo produto da probabilidade individual de cada uma das palavras por que essa frase é formada, ou seja, $p(S) = p(w^1) * p(w^2) ... p(w^n)$; então, num *corpus* como o CETEMPúblico, que tem cerca de 190 milhões de palavras, verifica-se que a probabilidade desta famosa frase é aproximadamente de $1,929^{-24}$, ou, dito de outro modo, $p(S) = 0,000\ 000\ 000\ 000\ 000\ 000\ 000\ 002$. É uma probabilidade muito baixa, certamente, mas não é igual a zero. v. Pereira (2002).

Proformas:

- pronomes indefinidos
- advérbios (proformas advérbias)
- nomes genéricos/indefinidos (nomes não-específicos)

Função de pronomes indefinidos e nomes genéricos na gramática: satisfazer a 1ª restrição.

É importante ^{PT}lavar/^{PB}escovar bem os dentes depois de comer.

o caráter *genérico* da expressão é dado pelo facto de os sujeitos de *lavar* e de *comer* serem preenchidos por indefinidos como *as pessoas* (*em geral*).

Certos valores locativos estão diretamente associados à situação comunicativa: o locutor fala sempre de um lugar (*aquí, cá*) e num determinado momento (*agora*), razão por que estes adverbais estão implícitos em frases como:

O Pedro ainda não chegou (*aquí*).

Note-se que, no caso dos advérbios, especialmente *cá, lá* e *aí*, estes têm outros empregos não exatamente locativos mas com diferentes valores, sobretudo de natureza pragmática, em que não se comportam como proformas, e.g.:

Deixa-me *cá* pensar nisto

Lá estás tu a fazer isso outra vez!

E *aí* ele disse que *sim*

Neste último exemplo, o advérbio *sim* tem, igualmente, um valor anafórico, na medida em que substitui um discurso anterior, como pode ser demonstrado expandindo a frase com o seu conteúdo:

E *aí* ele disse que *sim*, que tinha feito isso.

Os pronomes pessoais (*eu, tu, ele, ela, você, me, te, se, mim, ti, si*, etc.) são um tipo particular deste tipo de expressões, já que têm um valor referencial bem marcado, mas que podem igualmente combinar-se com praticamente qualquer operador, desde que respeitadas as restrições distribucionais genéricas (humano/não-humano) e as condições referenciais a que estão sujeitos (ver adiante). Note-se que tal conduz a que se considerem igualmente como proformas outras classes de pronomes e advérbios, nomeadamente os pronomes e advérbios relativos (*quem, que, o qual, cujo; onde*), os interrogativos (*quem, que* (vários empregos); *quando, (Prep) onde, como, quanto/quão, qual*), etc.

Na gramática harrissiana, a 1ª restrição – a ordem parcial de entrada das palavras na frase – cria todas as frases do *kernel* da língua, mas nada determina sobre a natureza distribucional dos elementos que constroem com os operadores essas frases da base. Na medida em que o operador “é dito” em relação ao seu domínio de argumentos, podemos igualmente dizer que esse operador “predica” os seus argumentos. Nesse sentido, os operadores constroem, por meio da 1ª restrição, uma parte significativa do significado das frases.

Já a 2ª restrição é responsável, em grande medida, pelo restante significado que é possível atribuir aos elementos lexicais. Tal é particularmente óbvio nas definições de muitos dicionários, que mais não

genericidade

proformas adverbiais locativas e situação comunicativa

expressões adverbiais: outros empregos

pronomes pessoais

pronomes e advérbios relativos
pronomes e advérbios interrogativos

1ª restrição e significado

2ª restrição e significado

são do que séries de sinónimos, uma para cada um dos sentidos que essa palavra pode ter. Na prática, o que o dicionário dá como definição, ainda que de forma implícita e aproximativa, são as distribuições (previamente conhecidas) de outras palavras em cujos contextos de ocorrência poderia, com maior ou menor grau de probabilidade, ocorrer igualmente a palavra a definir.

Em rigor, excepto em circunstâncias excepcionais, nenhuma palavra da língua tem exatamente a mesma distribuição, pelo que praticamente não há sinónimos perfeitos. Contudo, palavras com distribuições semelhantes tendem a ter significados semelhantes. Conhecendo a distribuição de uma palavra *A*, se se souber que ela pode ser praticamente sempre substituída pela palavra *B* nos contextos em que ocorre, é muito provável que *B* tenha o mesmo significado (ou, pelo menos, um significado bastante semelhante) que a palavra *A*.

Um exercício simples serve para ilustrar este ponto. Nas minhas aulas, costumo dar como exemplo um nome real do português europeu, mas que é de uso dialetal. Trata-se do nome *aiola*. Como ensino numa região diferente daquela em que cresci e fui criado, e que é onde esta palavra é muito usual, só por acaso é que algum aluno poderá saber à partida o que essa palavra significa. Ora, após ter assegurado que a palavra existe em Português, vou apresentando sucessivamente diferentes frases que servem para contextualizar (isto é, para *caracterizar a distribuição*) de forma progressivamente mais específica esse nome, até que a sua natureza (genérica) se torna óbvia para os alunos, ainda que nenhum tenha alguma vez visto uma *aiola* na vida. Por exemplo:

- informação insuficiente:
Esta aiola é azul
Esta aiola é grande
- alguma informação, mas ainda insuficiente (será uma peça de mobiliário?)
Esta aiola é feita de madeira de pinho
Esta aiola leva 3 pessoas
- informação suficiente:
Esta aiola flutua muito bem
Esta aiola teve de ancorar aqui

O emprego de verbos como *flutuar* (menos específica) ou *ancorar* (mais específica) implicam, sobretudo quando usadas em conjunto, uma interpretação não ambígua ou, pelo menos, muito menos ambígua do que as *pistas* dadas pelas frases anteriores. Estas pistas são apenas o facto de o nome desconhecido se encontrar como argumento de diferentes operadores com diferentes restrições de seleção.

definições de dicionários, sinonímia e distribuição

sinónimos perfeitos

O *jogo das perguntas* baseia-se no mesmo princípio: o desafio é uma pessoa adivinhar um nome (animal, objeto, etc.) pensado por uma pessoa e ir fazendo perguntas sucessivas até delimitar a distribuição do nome que designa esse objeto. As perguntas servem o mesmo objetivo de contextualizar, isto é, determinar a distribuição do nome-alvo.

É, assim, possível dizer que *aiola* deverá ser, provavelmente, um *meio de transporte aquático, grosso modo*, um tipo de *embarcação/barco*, pois estes nomes (ou um dos seus *hipónimos*) pode substituir o nome desconhecido nesses contextos. Porém, o significado preciso do que é uma *aiola*, a sua forma e função, etc. continua por definir.

hipónimos

Ainda assim, já é possível, por exemplo, prever que a frase seguinte será muito provavelmente aceitável:

Esta aiola navegava junto à costa

Este tipo de exercício demonstra que o significado pode ser representado com base na distribuição de uma palavra. O uso de *elementos* (nomes, neste caso) *classificadores*, tais como *embarcação/barco* é apenas uma forma de *substituir* a indicação *direta* de uma distribuição (ver exemplos acima) por uma indicação *indireta*: se uma *aiola* é um tipo de *embarcação*, deverá poder ocorrer nos mesmos contextos em que *embarcação* também ocorre.

nomes classificadores

Por fim, repare-se que tanto a primeira como a 2ª restrição, porém, são conceitos que assentam na probabilidade de co-ocorrência entre elementos lexicais. Como essas probabilidades são estimadas por cada falante de forma diferenciada, a (subtil) variação entre falantes que se observa no uso das expressões linguísticas é um dos ingredientes fundamentais da *mudança linguística*. Não iremos, no entanto, falar desse tipo de processos neste momento.

mudança linguística

Redução

A 3ª restrição é responsável pela maioria das alterações da forma das frases da base (do *kernel*), dando origem aos discursos que efetivamente observamos. Estas reduções ocorrem pelo fenómeno de redundância da informação que se dá quando as frases da base se concatenam em sequências mais ou menos longas.

Consideram-se dois tipos principais de redução:

- redução por efeito da repetição de elementos correferentes;
- redução apropriada de elementos.

Redução correferencial

A *redução correferencial* ocorre quando, em duas frases concatenadas, duas instâncias do mesmo elemento lexical se referem à mesma entidade extralinguística. Por exemplo, na frase complexa, formada por coordenação:

redução correferencial

^t*O Pedro_i leu o jornal e o Pedro_i leu a revista*

as duas instâncias do nome *Pedro* referem-se à mesma entidade extralinguística, *grosso modo* à mesma pessoa, o que é indicado através dos

índices de correferência i . Tal repetição torna a frase extremamente redundante, o que leva à redução a zero (ing. *zeroing*) dos elementos repetidos na segunda frase da coordenação, incluindo o verbo:

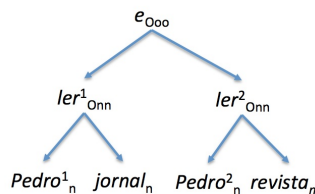
- O Pedro_{*i*} leu o jornal e leu a revista
- O Pedro_{*i*} leu o jornal e a revista

Repare-se que a frase de partida é uma frase da língua. Trata-se, concretamente, uma frase do *kernel*, o que é indicado pelo índice t . A frase resulta da operação da conjunção *e* sobre duas frases, cada uma com uma instância diferente do verbo *ler*, e cada uma destas instâncias deste verbo tem um domínio de argumentos próprio e distinto: *Pedro* e *jornal* na primeira instância e *Pedro* e *jornal* na segunda. A reduzida aceitabilidade/naturalidade desta frase vem da redundância que a repetição de elementos produz e que a 3ª restrição, justamente, elimina, reduzindo-os (neste caso, a zero).

Podemos representar a ordem parcial de entrada das palavras nesta frase, por exemplo, através de uma notação simplificada, como a que já referimos atrás, usando $x > y$ para indicar que x entra na frase *depois de* y , e indicando os elementos que entram simultaneamente na frase dentro de parênteses. Para mais fácil referência, os elementos repetidos encontram-se numerados com expoentes e em índice indicamos o tipo de operador (O_{oo} , O_{nn}) ou se se trata de um argumento elementar n . Assim, e ignorando os determinantes e outros elementos gramaticais como o tempo verbal, podíamos produzir uma *representação parentética* como a seguinte:

$$e_{O_{oo}} > (ler^1_{O_{nn}} > (Pedro^1_n, jornal_n), ler^2_{O_{nn}} > (Pedro^2_n, revista_n))$$

Contudo, como esta forma de representação não é dos formalismos mais fáceis de interpretar (por humanos), podemos, em vez dela, usar uma estrutura arborescente, que designaremos por *lattice*, para representar exatamente a mesma informação (Fig. 4).



Em rigor, a estrutura da Fig. 4 é apenas uma *semillattice* da *lattice* mais completa que representaria toda a frase

O Pedro leu o jornal e a revista

redução a zero (ing. *zeroing*)

representação parentética

representação por (semi)lattice

Figura 4: Exemplo de uma (semi)semillattice, representando a ordem parcial de entrada das palavras na frase.

lattice vs. *semillattice*

já que só parte da informação se encontra aqui representada, muito embora uma parte significativa da informação já esteja presente nesta representação. Repare-se também que esta semillattice não representa a *estrutura* da frase, mas sim a *estrutura da informação* que esta veicula, definida em termos da ordem parcial de entrada das palavras na frase. Assim, por exemplo, embora se indique que *ler* é um operador *Onn*, nada na lattice dá conta de que se trata de uma construção transitiva direta (com complemento direto). Não se confunda, pois, estas (*semi*)lattices com outras formas de representação, habituais em sintaxe formal, em que o que está em jogo é a estrutura sintática das expressões.

estrutura da informação

A redução de elementos repetidos abrange diferentes níveis de análise, não apenas as palavras. Assim, por exemplo, na coordenação de advérbios terminados em *-mente*, uma muito bem conhecida regra de redução determina que tanto a conjunção coordenativa como o sufixo se reduzem em todas as frases coordenadas menos na frase do último advérbio:

redução em diferentes níveis de análise

^tO Pedro leu o livro *atentamente* (e o Pedro leu o livro) *interessadamente* (e o Pedro leu o livro) *pausadamente* (...) e (o Pedro leu o livro) *paulatinamente*

→ O Pedro leu o livro *atenta, interessada, pausada* (...) e *paulatinamente*

Repare-se que, neste caso, se considera ainda que não é apenas a conjunção e o sufixo que se reduzem, mas todas as instâncias da primeira frase *O Pedro leu o livro*, que se encontram repetidas sob várias instâncias de *e*. A Fig. 5 representa, de forma simplificada, esta coordenação de frases com advérbios terminados em *-mente*. Voltaremos a analisar esta frase, com mais profundidade, mais adiante.

Outras formas de redução envolvem casos de subordinação. Assim, quando o sujeito da oração subordinada é correferente do sujeito da oração principal, aquele reduz-se a fim de evitar a repetição.

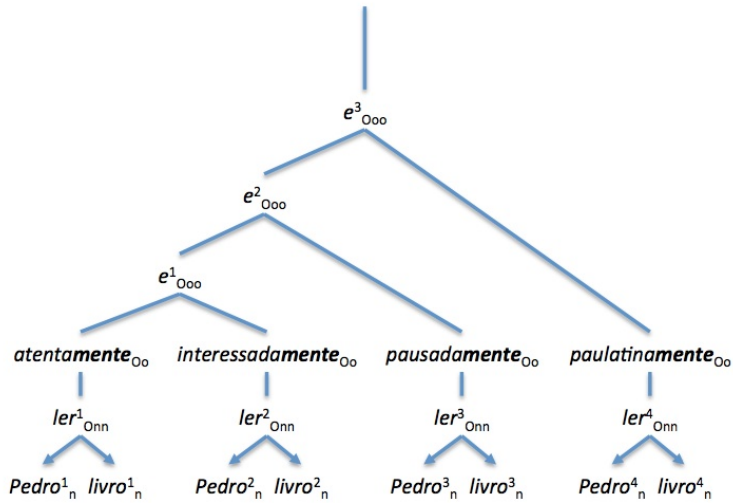
^tO Pedro_i leu o jornal enquanto o Pedro_i tomava o pequeno-almoço^{PT}/café-da-manhã^{PB}

→ O Pedro_i leu o jornal enquanto tomava o pequeno-almoço/café-da-manhã

No caso de certas orações subordinadas, o verbo da oração subordinada também se reduz a uma infinitiva, já que as marcas de tempo-modo-aspeto e pessoa-número da flexão são supridas pela estrutura sintática de subordinação (completiva integrante):

^tO Pedro_i quer que o Pedro_i vá a Lisboa

→ O Pedro_i quer ir_i a Lisboa



O Pedro leu o livro atenta, interessada, pausada, e paulatinamente

Repare-se que, em todos os exemplos acima, a redução a zero dos elementos correferentes repetidos é *obrigatória*, já que a sua manutenção na frase implicaria, justamente, que não se referissem à mesma entidade. Por outras palavras, as frases de base, em que um elemento correferente se repete, têm um estatuto teórico (o que se indica pelo expoente “t”), a redução é o mecanismo que transforma essas frases com elementos repetidos correferentes (de outro modo inaceitáveis) nas expressões efetivamente observadas nos discursos. Caso contrário, se se aceitar como boas as frases marcadas com “t”, então, os elementos repetidos *não* podem ser correferentes e as reduções a zero (ou outras) não podem ter lugar. Veremos, mais adiante, que outras restrições operam no complexo fenómeno das reduções.

A par das reduções a zero, observa-se igualmente a redução de palavras (ou morfemas) a proformas, nomeadamente a pronomes. Assim, por exemplo, uma frase complexa em que um ou mais elementos aparecem repetidos, alguns deles podem reduzir-se a zero e outros se reduzem a proformas. Considerando as duas formas de base (separadas por ‘#’):

O Pedro_i ofereceu um livro_j; ao João_k # O Pedro_i tinha falado desse livro_j; ao João_k

a formação da oração relativa a partir da segunda frase dá origem à frase complexa (ignoramos, por agora, os problemas de deslocação de constituintes):

Figura 5: Coordenação de advérbios em *-mente*

redução a *proformas*: pronomes

→ *O Pedro_i ofereceu ao João_k um livro_j de que_j lhe_k tinha falado_i*

Ou seja, enquanto o sujeito (correferente) da oração subordinada relativa *o Pedro_i* é reduzido a zero (tendo nós passado a indicar o índice de correferência junto ao verbo da oração secundária); tanto o complemento direto *o livro_j* como o complemento indireto *o João_k* se reduzem a proformas, respetivamente, ao pronome relativo *que_j* e ao pronome pessoal dativo *lhe_k*.

Importa reter, portanto, que as reduções não alteram a quantidade de informação presente nas frases da base da gramática. Estas operações apenas alteram a forma das frases mantendo a respetiva informação, razão por que é possível interpretá-las de forma precisa, reproduzível e não ambígua. Apenas nas frases de base a informação está presente de forma visível/explicita, enquanto que nas frases reduzidas é necessário levar em conta os processos de redução que nelas tiveram ou puderam ter tido lugar.

Um dos principais objetivos da análise gramatical é reconstituir as frases de base que deram origem às frases. Cada língua determina o conjunto de reduções (a zero ou a proformas) e as condições de correferência (ou não) em que estas operações podem ocorrer. Por exemplo, a redução das frases coordenadas com advérbios terminados em *-mente* é um fenómeno característico do português.

Redução apropriada

Um segundo tipo de redução tem uma natureza diferente, já que não implica a repetição de elementos lexicais. Pelo contrário, trata-se de um fenómeno de natureza lexical, resultando, sobretudo, da 2^a restrição. Como já vimos atrás, há situações em que apenas um elemento ou um número muito pequeno de elementos pode ocorrer numa dada posição argumental, razão por que pode esse elemento ser reduzido sem perda de informação. A este fenómeno Harris (1976:113-115) chamou *redução apropriada* e, dada a sua natureza lexical, ou seja, o facto de depender de combinações específicas de alguns elementos lexicais, é um domínio de difícil descrição, desde logo pela dificuldade em inventariar todas as situações em que ocorre.

Contudo, embora ocorra em numerosas outras situações, há um conjunto de casos em que o fenómeno da redução apropriada é particularmente evidente. Trata-se do caso das construções que envolvem um nome parte-do-corpo. Este nomes implicam, necessariamente, uma *relação metonímica* (parte-todo) entre eles e um elemento de que fazem parte. A este tipo de relação também é costume chamar-se de *posse inalienável*. Num exemplo como:

O Pedro penteou o cabelo_i do João_i

nome parte-do-corpo

relação metonímica

posse inalienável

dizemos que há entre *João* e *cabelo* uma relação metonímica (o cabelo é parte integrante do João, o que é assinalado no exemplo pelos índices de correferência metonímica “_i”).

Ora, acontece que entre o verbo *pentear* e o nome *cabelo* há uma relação de *apropriação*, na medida em que este nome é o que tem a mais alta probabilidade de ocorrer como segundo argumento deste verbo. Isso não significa que outros nomes não possam igualmente ocorrer nesta posição, tendo em conta, naturalmente, as restrições de seleção impostas pelo verbo *pentear* ao preenchimento lexical seu complemento direto. Efetivamente, nomes como *barba*, *bigode*, *madeixas*, *patilhas*, *sobrancelhas*, etc. poderiam ocupar aquela posição com um grau de probabilidade normal. Porém, só *cabelo* estabelece essa relação de *apropriação*, já que só ele pode reduzir-se sem que haja alteração do significado global da frase:

→ *O Pedro penteou o João*

Numa frase como esta, a parte do corpo que, com maior verosimilhança, terá sido penteada foi, obviamente, o *cabelo*.

No caso em que a entidade a que o nome parte-do-corpo pertence é também o sujeito da construção, diferentes reduções se podem observar, desde logo a simples redução a zero da segunda instância do nome repetido (a), bem como a redução apropriada e posterior redução a pronome reflexo da segunda instância do nome repetido (b):

O Pedro_i penteou o cabelo_i do Pedro_i

→ (a) *O Pedro_i penteou o cabelo_i*

→ (b) *O Pedro_i penteou-se_i*

Repare-se que não está em jogo o grau de especialização de um dado verbo relativamente ao nome parte-do-corpo com que se constrói. Verbos com construções extremamente especializadas (isto é, com uma distribuição muito restrita), como *frisar*^{PT} (*cabelo*), que significa ‘tornar crespo, encrespar, encararolar’, podem não admitir esta redução apropriada:

O Pedro_i frisou o cabelo_j da Ana_j.

→ **O Pedro frisou a Ana*

embora admitam a redução do elemento repetido correferente ao sujeito, desde que se mantenha o nome parte-do-corpo:

O Pedro_i frisou o cabelo_i do Pedro_i

→ *O Pedro_i frisou o cabelo_i*

→ **O Pedro_i frisou-se_i*

É, pois, necessário um laborioso e paciente trabalho de descrição linguística, caso a caso, destas situações a fim de se poder determinar em que condições um nome se encontra em contexto apropriado.

Outro exemplo, mais complexo, de redução apropriada implica, como já vimos atrás, a determinação dessa relação entre um verbo numa oração principal e uma relação verbo-objeto na oração secundária. Na frase:

O Pedro adora livros de banda desenhada^{PB}/história em quadrinhos^{PB}

dizemos que adorar não opera diretamente sobre o nome *livros de banda desenhada^{PB}/história em quadrinhos^{PB}* já que se trata de um operador do tipo *Ono*, como o prova a expressão:

← *O Pedro adora ler livros de banda desenhada^{PB}/história em quadrinhos^{PB}*

em que o verbo *ler* é o elemento lexical com a mais alta probabilidade de ocorrer como argumento de *adorar* quando tem como complemento direto um nome como *livros de banda desenhada^{PB}/história em quadrinhos^{PB}*. Dizemos, assim, que *ler* se encontra num contexto ou posição apropriada sob *adorar* e sobre *livros de banda desenhada^{PB}/história em quadrinhos^{PB}*.

Aliás, esta propriedade permite inclusive distinguir duas construções diferentes de *adorar*. A par da que acima se ilustrou, é necessário considerar igualmente a construção de *adorar* como um operador do tipo *Onn*:

O Pedro adora a Ana/uma divindade egípcia

já que neste caso não é possível inserir nenhum outro operador que entrasse na frase *imediatamente antes de adorar e imediatamente depois dos nomes* que se encontram em posição de complemento:

O Pedro adora [???] a Ana/uma divindade egípcia

Linearização

Como o nome indica, a 4^a restrição fundamental – a *linearização* – tem a ver com a forma como os elementos lexicais se dispõem linearmente no discurso. Dada a sua natureza, a fala exige que os elementos linguísticos se disponham linearmente, uns após os outros (com a excepção de alguns morfemas que são expressos por processos fonológicos suprasegmentais – a entoação).

Até aqui, a 1^a restrição determinada a natureza dos elementos lexicais, distinguindo os operadores dos argumentos elementares e ordenando a entrada dos elementos das frases. Contudo, esta restrição nada determina em relação à ordem linear das palavras nas frases.

ordem linear dos elementos no discurso

processos fonológicos suprasegmentais: entoação

Por seu turno, a 2ª restrição, ao determinar diferenças de probabilidade de co-ocorrência operador-argumento(s), embora o faça relativamente a um tuplo ordenado de elementos, por si só não determina a ordem linear destes, apenas implica que as restrições distribucionais que o operador impõe se determinam relativamente a cada uma das posições argumentais, tendo em conta simultaneamente o preenchimento das restantes.

Esta restrição tem, assim, como efeito, o arranjo linear das palavras (ou morfemas) nas frases do discurso. A linearização pode ser concebida, num primeiro momento, no quadro da frase elementar. Neste sentido, esta restrição é responsável, entre outros aspectos, pela interpretação associada à relação entre cada argumento e o operador de que depende diretamente, por outras palavras, o seu papel semântico.

linearização e papel semântico

Verifica-se não só que cada língua determina uma ordem linear básica dos argumentos relativamente ao operador de que dependem diretamente, como também recorre a mecanismos diversos para a expressão desta restrição. Certas línguas naturais, como, por exemplo, o latim, são *a priori* mais flexíveis quanto à ordem por que os elementos se apresentam dispostos linearmente nas frases, na medida em que desenvolveram uma morfologia flexional que inclui um sistema de *casos*.

ordem básica

O *caso gramatical*, ao indicar a função sintática do constituinte, determina pelo menos em parte, a interpretação que o elemento lexical adquire relativamente ao operador de que depende. Assim, por exemplo, as frases seguintes tem os mesmos elementos lexicais e estes estão dispostos exatamente na mesma ordem, mas como o caso gramatical dos nomes *Ioanes* (João) e *Maria* é diferente, a sua interpretação também é distinta:

caso gramatical

*Ioanes*_{nominativo} *Mariam*_{acusativo} *amat* 'João ama Maria'
*Ioanem*_{acusativo} *Mariã*_{nominativo} *amat* 'Maria ama João'

Por outro lado, qualquer uma das frases abaixo tem exatamente o mesmo conteúdo:

*Ioanes*_{nominativo} *Mariam*_{acusativo} *amat*
*Ioanes*_{nominativo} *amat* *Mariam*_{acusativo}
*Mariam*_{acusativo} *Ioanes*_{nominativo} *amat*
*Mariam*_{acusativo} *amat* *Ioanes*_{nominativo}
Amat *Ioanes*_{nominativo} *Mariam*_{acusativo}
Amat *Mariam*_{acusativo} *Ioanes*_{nominativo}
 'João ama Maria'

Quando os mesmos elementos se apresentam numa ordem alternativas, tais alterações à ordem básica são portadoras de informação. Nomeadamente, os elementos que ocorrem em posição inicial de

frase são, geralmente, o foco/tópico da proposição (tema), enquanto os restantes elementos são o conteúdo da proposição, ou seja, o que se diz sobre esse elemento (rema).

Assim, uma transformação como a passiva tem, entre outros efeitos, a função de trazer para a posição de tópico um dos argumentos do operador que, na frase ativa correspondente, não teria o mesmo destaque. Pelo contrário, o sujeito da frase ativa, ao ocupar uma posição sintática não destacada pode, frequentemente, reduzir-se a zero.

O número de casos depende de cada língua particular, sendo frequente encontrar línguas em que relações semânticas distintas que ou são expressas por diferentes casos ou são enquadradas sob o mesmo caso.

A par dos casos, as línguas naturais também podem recorrer a outros mecanismos formais, tais como as *preposições* (ou *posposições*, como no caso do coreano, japonês, síndi, turco, urdu e outras línguas), isto é, certos elementos morfológicos com função idêntica à dos casos, isto é, que funcionam como *marcadores de argumentos*, assinalando a relação que estes elementos estabelecem com o operador de que dependem.

Finalmente, a própria disposição na frase dos argumentos relativamente ao operador de que dependem pode ser usada como um mecanismo equivalente ao dos casos ou das preposições. Tal sucede em línguas em que a ordem básica das palavras é ou se tornou menos flexível.

Como é natural – até pela sua natureza histórica e necessariamente contingente, poucas línguas naturais expressam esta restrição de um modo “quimicamente puro”, ou seja, quase todas as línguas naturais apresentam uma complexa combinação destes diferentes processos. É o caso do Português: embora o caso gramatical tenha praticamente desaparecido, o que levou a uma maior rigidez da ordem linear dos argumentos de um operador na frase, este processo combina-se com o emprego de preposições em todas as posições argumentais menos no sujeito e no complemento direto dos verbos.

As preposições básicas, ou seja, os marcadores de estatuto argumental são geralmente em número reduzido. Em Português, estas preposições básicas (não locativas) são *a*, *com*, *de* e *em*, a que se juntam ainda, mas de forma menos relevante, *para*, *por* e *sobre*. No entanto, nem sempre estas preposições expressam a mesma relação semântica, a qual depende de forma crucial do operador a que estão associadas.

Assim, por exemplo, a preposição *a* introduz muitas vezes o complemento indireto, com valor de INTERLOCUTOR, nas construções dos verbos de comunicação:

O Pedro disse isso a alguém

preposições
posposições

marcadores de argumentos

Expressão da linearização em Português: um sistema híbrido

preposições básicas (não locativas) em Português: *a*, *com*, *de* e *em*

valores semânticos de complementos essenciais introduzidos a preposição *a*

Já o mesmo tipo de constituinte apresenta o valor de *destinatário*, nas construções que exprimem uma *transferência de objeto*:

O Pedro emprestou o livro ao João

Porém, neste último tipo de construções, a direção da transferência não pode ser calculada diretamente a partir da preposição e implica levar simultaneamente em consideração o próprio verbo. Em predicados como:

O Pedro roubou o livro ao João

o complement indireto não é o DESTINATÁRIO/BENEFICIÁRIO, mas a FONTE/VÍTIMA.

Por último, refira-se o caso das construções do mesmo tipo com o verbo *alugar* e seus sinónimos, que são ambíguas quanto à orientação do processo. Em:

O Pedro alugou o apartamento ao João

o *João* tanto pode ser o proprietário como o inquilino, sendo esta ambiguidade impossível de decidir sem informação adicional.

Um exemplo curioso é o caso da preposição *de* nas construções com verbos locativos de movimento/dinâmicos. De um modo geral, esta preposição introduz o complemento locativo que indica ORIGEM:

O Pedro veio de Lisboa

Contudo, com o verbo *aproximar-se*, ela indica o DESTINO/ALVO:

O Pedro aproximou-se de Lisboa

Um segundo aspecto da linearização observa-se não ao nível da frase elementar (ou da base) mas já ao nível das frases complexas. Certas alterações da ordem da básica resultam do processo de concatenação de formas de base num discurso complexo. Este aspecto interage, por sua vez, com a redução.

Assim, por exemplo, nas construções com construção completiva-sujeito (1), a expressão da completiva na posição básica perturba a curva entoacional típica de uma frase declarativa em português, o que leva a deslocar esse constituinte para o final da frase (a), razão porque se chamou a esta operação *permuta de comprimento*. Repare-se que a redução da completiva-sujeito a uma proforma (um indefinido como *isso*) conduz a que a ordem básica resulte totalmente natural (b):

(1) ?*Que o Pedro faça isso assim importa-me imenso*

→ (a) *Importa-me imenso que o Pedro faça isso assim*

→ (b) *Isso importa-me imenso*

permuta de comprimento

Outro aspecto prende-se com o comprimento relativo dos constituintes que são complementos de um verbo. Perante dois complementos de um verbo dispostos na ordem básica dos constituintes (2), a progressiva extensão do primeiro complemento com novos modificadores, ainda que seja um processo perfeitamente “gramatical”, acaba por tornar a frase progressivamente menos natural (c)-(e) até se tornar praticamente inaceitável (e), sobretudo porque parece perder-se a relação de dependência entre o segundo complemento verbal, o complemento indireto *ao João*, e o verbo *emprestar* de que este depende:

- (2)
- (a) *O Pedro emprestou o livro ao João*
 - (b) *O Pedro emprestou o livro de matemática ao João*
 - (c) *?O Pedro emprestou o livro de matemática de capa amarela ao João*
 - (d) *??O Pedro emprestou o raríssimo livro de matemática de capa amarela ao João*
 - (e) *?*O Pedro emprestou o raríssimo livro de matemática de capa amarela e lombada lilás ao João*

Neste caso, o constituinte mais longo é deslocado para o final da frase. Compare-se a fraca aceitabilidade das frases (2)(c)-(e) com a muito maior aceitabilidade das frases correspondentes (3)(c)-(e), em que o complemento indireto é deslocado para junto do verbo:

- (3)
- (a) *?O Pedro emprestou ao João o livro*
 - (b) *O Pedro emprestou ao João o livro de matemática*
 - (c) *O Pedro emprestou ao João o livro de matemática de capa amarela*
 - (d) *O Pedro emprestou ao João o raríssimo livro de matemática de capa amarela*
 - (e) *O Pedro emprestou ao João o raríssimo livro de matemática de capa amarela e lombada lilás*

Repare-se ainda que, no caso da frase de base (3)(a), a permuta correspondente produz um resultado menos natural, já que há uma alteração da ordem básica. Essa transformação, contudo, pode ser usada para conferir maior saliência na frase ao complemento anteposto (e, inversamente, reduzir a saliência do complemento deslocado para o fim da frase. Note-se ainda no facto de que tanto a frase (2)(b) como a frase correspondente com a permuta do complemento *o livro de matemática*, ou seja, a frase (3)(b), parecerem igualmente naturais, ou, pelo menos, ser muito difícil determinar com precisão uma diferença de aceitabilidade.

A redução de um constituinte complemento a uma *proforma* (prónome) e a sua colocação em posição clítica (ligado ao verbo) pode ser enquadrada neste fenómeno mais geral, como se mostra nos exemplos (4)(a)-(b):

- (4) *O Pedro emprestou o livro ao João*
 → (a) *O Pedro emprestou-o ao João*
 → (b) *O Pedro emprestou-lhe o livro*

A alteração da ordem dos pronomes na frase, resultante da presença de certos adverbiais, nomeadamente os *advérbios de negação* ou os *advérbios conjuntivos*, ilustrados nos exemplos (5)(a)-(b)), pode ser vista como assinalando a natureza complexa da predicação expressa nesses discursos, já que estes advérbios são, por definição, operadores de segunda ordem. Em Português Europeu, a aceitabilidade destas frases é francamente má quando o pronome se encontra em posição de ênclise, como se vê em (5)(c)-(d):

- (5) *O livro # o Pedro emprestou-o ao João*
 (a) *O livro # o Pedro não o emprestou ao João*
 (b) *O livro # o Pedro também o emprestou ao João*
 (c) **O livro # o Pedro não emprestou-o ao João*
 (d) **O livro # o Pedro também emprestou-o ao João*

Compare-se estas frases com as dos exemplos (6)-(6a), em que a oração com os constituintes pronominalizados se encontra subordinada a outro verbo:

- (6) *O Pedro disse que emprestara o livro ao João*
 → (a) *O Pedro disse que o emprestara ao João*

Repare como a linearização é entendida como uma restrição, neste contexto. O leitor atento já viu que nem todas estas mesmas frases tem uma aceitabilidade idêntica se se tratasse da variante do Português falado no Brasil. Cada língua (e neste caso, cada variante) impõe diferenças na expressão da linearização.

Por último, refiramos o caso da interação entre a expressão de elementos repetidos em orações complexas formadas por subordinação. Como se sabe, na subordinação adverbial, as orações secundárias podem deslocar-se para o início da oração principal, fenómeno que é assinalado, na oralidade, com uma breve pausa ou, na escrita, com uma vírgula. Como já vimos, quando há repetição elementos correferentes, a segunda ocorrência pode reduzir-se. O tipo de redução (a zero ou a pronome) pode depender de vários fatores, mas é igualmente função da interação entre estas duas restrições fundamentais: a linearização e a redução. Assim, na frase (7), em que não houve permuta da subordinada, escolha da forma adequada de redução, a zero ou a pronome, de elementos repetidos depende da existência ou não de correferência entre o sujeito da oração principal e o sujeito da subordinada. Em Português Europeu, no caso de haver correferência, a redução a zero é obrigatória, caso contrário um sujeito da subordinada reduzido a pronome seria interpretado como não correferente ao sujeito da oração principal.

advérbio de negação

advérbio de conjuntivo

Repare-se, no entanto, que restrições diferentes se aplicam no caso de uma subordinada infinitiva: *O Pedro preferia emprestar o livro ao João* → *O Pedro preferia emprestar-lo ao João*, vs. *O Pedro preferia o emprestar ao João*.

- (7) ^tO Pedro_i lê o jornal
 enquanto o Pedro_i/João_j toma o pequeno-almoço/café da manhã
 → (a) O Pedro_i lê o jornal
 enquanto 0_i/ele_j toma_{i/j} o pequeno-almoço/café da manhã

Percebe-se, assim, em que medida a redução deve ser considerada uma restrição fundamental: cada língua (e variante) determina as condições que permitem a uma dada palavra reduzir-se a uma proforma ou a zero.

Já no caso das frases com permuta da oração subordinada, é possível considerar duas situações: ou a redução tem lugar na oração principal (8a), ou pode ocorrer na subordinada deslocada para o início da frase (8b).

- (8) ^tO Pedro_i lê o jornal
 enquanto o Pedro_i/o João_j toma o pequeno-almoço/café da manhã
 → (a) Enquanto o Pedro_i toma_i o pequeno-almoço/café da manhã,
 0_{i/*j}/ele_{2i/j} lê o jornal
 → (b) Enquanto 0_{i/*j}/ele_{2*i/j} toma_i o pequeno-almoço/café da manhã,
 o Pedro_i lê o jornal

Como se pode ver pelos exemplos, a situação é, neste caso, muito menos simples do que no exemplo anterior. No caso de a redução ocorrer na oração principal (7a), a redução a zero só permite a interpretação correferencial. Já a redução a zero, embora privilegie a interpretação de sujeito não-correferente não exclui totalmente a interpretação correferencial. Caso a redução ocorra na oração subordinada (7b), o grau de aceitabilidade é semelhante mas a redução a pronome correferente parece muito menos natural.

Uma nota final sobre as relações de correferência aqui envolvidas e o conceito de *anáfora* / *catáfora*. Com base na disposição dos elementos da frase, chama-se *anáfora* a redução de um elemento *depois* do seu antecedente na ordem linear do discurso. O *antecedente* é o elemento ao qual um pronome (ou o lugar estrutural, deixado vazio pela sua redução) se refere. É o caso quer do exemplo (7a) e (8a), em que as formas reduzidas (a zero ou a pronome), também chamadas de expressões anafóricas, aparecem linearmente depois do respetivo antecedente. Em contrapartida, chama-se *catáfora* quando o elemento reduzido (a zero ou a pronome) se encontra *antes* do seu antecedente na ordem linear do discurso. É o que parece acontecer no caso do exemplo (8b).

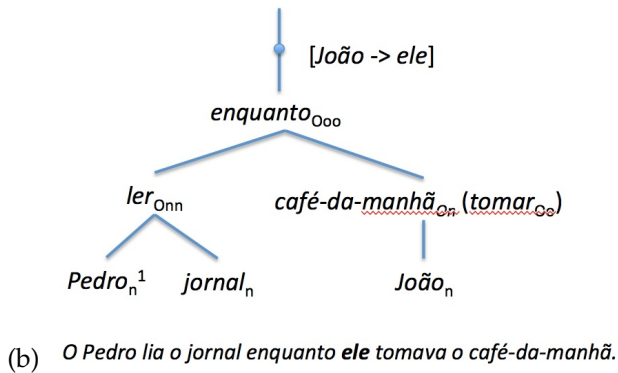
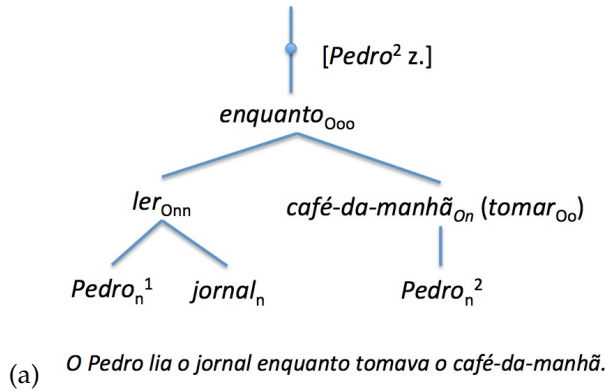
A representação sob a forma de semilattices destas reduções é relativamente transparente. No caso, a frase com a subordinada na posição básica (7)(a-b) pode ser ilustrada pelas lattices da Fig.6.

Repare-se que as duas representações são idênticas, mas que as frases de base se distinguem apenas pelo facto de o sujeito da oração subordinada (*Pedro* / *João*) ser ou não o mesmo que o sujeito da oração

anáfora/catáfora

antecedente

expressões anafóricas, ing. *anaphor*
 catáfora



principal (*Pedro*). Contudo, esta diferença tem consequências para a escolha da redução adequada (a zero ou a pronome). (Trata-se, saliente-se, de uma semillatice, em que falta ainda representar os mecanismos de correferência, de que falaremos mais adiante.)

No caso do exemplo (8)(a-b), a permuta da subordinada é apenas outra operação formal (transformação) que tem lugar na frase, a par da redução. Contudo, havendo duas operações, é necessário determinar a sua ordem de aplicação. Por outras palavras, no caso de haver uma permuta da oração subordinada e uma redução de elementos repetidos, levanta-se a questão do momento em que se pode dar cada uma das operações. Nesse sentido, a gramática harrissiana determina o seguinte princípio de ordenação das operações na derivação representada pelas lattices:

*As operações sobre as frases têm lugar **assim que** estiverem reunidas condições para ocorrerem.*

(Retenha este princípio, já que iremos voltar a ele mais adiante.) Obviamente, um elemento só pode encontrar-se repetido se já tiver ocorrido *antes* no discurso.

Ora, no caso da frase (8a), uma vez que a redução ocorre na oração principal e o antecedente ocorre primeiro na ordem linear das

Figura 6: Redução de elementos repetidos em orações subordinadas: (a) redução a zero com correferência entre o sujeito da oração principal e o da subordinada; (b) redução a pronome se não houver correferência entre o sujeito da oração principal e o da subordinada.

princípio de ordenação das operações

palavras no discurso, a permuta tem lugar *antes* da redução. Pelo contrário, no caso da frase (8b), a permuta tem lugar *depois* da redução.

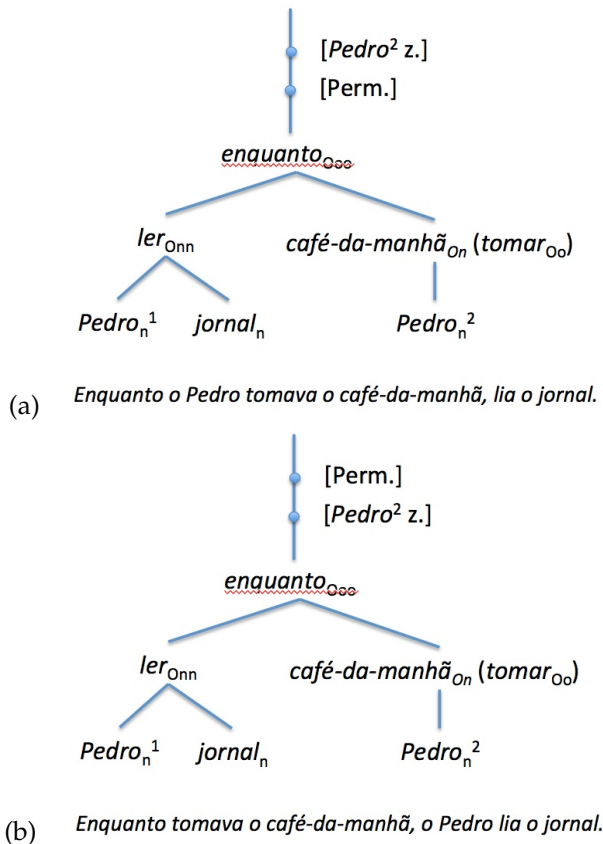


Figura 7: Redução de elementos repetidos, isto é, com correferência entre o sujeito da oração principal e o da subordinada, em orações subordinadas permutadas: (a) redução a zero do sujeito da subordinada; (b) redução a zero do sujeito da oração principal.

Por outras palavras, o próprio conceito de *catáfora* parece violar um princípio básico de ordenação das operações sobre frases. Efetivamente, a *catáfora*, entendida nesta perspetiva, parece corresponder a uma situação bizarra, em que a redução de elementos tem lugar em frases apesar de o respetivo “antecedente” (à falta de melhor termo) *ainda não* ter aparecido no discurso. Assim, esses elementos não deveriam poder reduzir-se.

O facto de se considerar duas operações distintas (a permuta e a redução) e um princípio geral de ordenação das operações regulariza a descrição gramatical, simplificando-a e tornando-a consistente.

Nas secções anteriores, tratámos das quatro restrições fundamentais, ou seja, as restrições responsáveis pela construção das frases da língua. Nas secções seguintes, iremos tratar das restrições complementares, que dão conta de uma outra importante dimensão das línguas naturais, nomeadamente, da sua metalinguagem.

Restrições complementares

Além das restrições fundamentais, que vimos na secção anterior, e que operam, sobretudo, no quadro das frases da base (ou *kernel*) da gramática; as línguas naturais apresentam ainda duas outras restrições complementares, de que trataremos agora. Estas restrições, porém, só se verificam em sequências de frases concatenadas, ou seja, quando duas ou mais frases da base se juntam para formar um discurso. Por outro lado, não há, propriamente, um nome para cada uma delas, pelo que muito prosaicamente as distinguiremos como *primeira* e *segunda* restrições complementares. Estas restrições constituem uma segunda cadeia de frases que está para além do discurso efetivamente observável, fazendo parte do aparato metalinguístico da linguagem, na medida em que se referem a elementos presentes no discurso.

1ª restrição complementar

A 1ª restrição complementar determina que certas palavras (ou morfemas) se repitam em sequências de frases concatenadas.

Tomemos como exemplo a formação de uma oração subordinada relativa (processo a que também se poderia chamar de *relativização*) e que assinalamos a negrito na frase seguinte:

*O Pedro emprestou ao João o livro **de que lhe falara***

Para que seja possível formar este tipo de subordinada, é necessário que em cada uma das duas frases de base haja um elemento lexical que se repita e que se refira à mesma entidade extralinguística (por outras palavras, que as duas instâncias de um nome repetido sejam correferentes). Tomemos, assim, as seguintes frases de base, considerando que a primeira vai constituir a oração principal e ignorando para já a repetição dos nomes *Pedro* e *João* bem como outros aspectos relevantes atinentes ao processo de subordinação:

O Pedro emprestou um livro; ao João # O Pedro falara de um livro; ao João

Ora, nestes exemplos, a correferência das duas instâncias do nome *livro* foi assinalada com índices de correferência 'i'. Trata-se de um processo notacional que, como já vimos, é perfeitamente artificial. Na verdade, se se incorporar na prática da descrição gramatical a observação de que as línguas naturais não têm uma metalinguagem externa, temos forçosamente de chegar à conclusão de que a língua deve dispor de um mecanismo para veicular a informação de que as duas instâncias do nome *livro* nestas duas frases se referem ao *mesmo* objeto extralinguístico.

Ora, o leitor mais perspicaz já percebeu que a frase acima contém esse elemento – *mesmo* – não sendo preciso nenhuma notação esotérica e artificial como aquele índice de correferência que, por mera comodidade didática, usamos nas duas frases acima.

O que a 1ª restrição complementar faz é, justamente, explicitar essa informação, ao assumir que, na base, há uma frase metalinguística que determina a correferência de dois nomes que ocorrem em frases distintas e sequenciais. Embora a formulação exata dessa frase possa variar (como o próprio exemplo acima ilustra), vamos aqui considerar que se trata de uma frase como:

O [objeto a que se refere o nome] livro da primeira frase é o mesmo [objeto a que se refere o nome] livro da segunda frase

Esta 1ª restrição complementar consiste, pois, num mecanismo de referenciação das instâncias dos elementos repetidos em cada uma das duas frases concatenadas e o operador *mesmo*, que estabelece a correferência entre estes elementos. Tanto o mecanismo de referenciação como o operador *mesmo* são parte integrante da língua. Contudo, as frases metalinguísticas como a que ilustramos acima constituem uma *segunda cadeia de frases*, que está *para além* das frases efetivamente observadas, mas cuja existência é necessária se se assumir a ausência de uma metalinguagem externa à linguagem natural e se se quiser dar conta da informação efetivamente veiculada pelas frases num discurso concatenado.

segunda cadeia de frases

Indo, então, mais longe, e considerando agora todos os elementos repetidos nas duas frases da base, podemos construir a seguinte representação da informação veiculada pela frase com a oração relativa (por uma questão de comodidade didática e maior simplicidade, voltamos a anotar as diferentes instâncias dos nomes repetidos com índices, desta vez numéricos).

*O Pedro₁ emprestou um livro₁ ao João₁ #
O Pedro₂ falara de um livro₂ ao João₂*

*O Pedro₁ é o mesmo que o Pedro₂
O livro₁ é o mesmo que o livro₂
O João₁ é o mesmo que o João₂*

Naturalmente, só a existência destas frases metalinguísticas pode dar conta da informação (e, logo, do significado) da frase complexa com a oração relativa, em que várias reduções tiveram lugar sem que, contudo, haja qualquer perda da informação.

Retomando essa frase:

O Pedro emprestou ao João o livro de que lhe falara

verificamos que:

- (i) o sujeito de *falara* na oração relativa (*Pedro*₂) se reduziu a zero, pois é correferente do elemento *que*, na oração principal, desempenha a mesma função sintática;
- (ii) o complemento preposicional de *falara* (*de um livro*₂) se desloca para o início da oração relativa; a preposição que introduz o complemento (*de*) acompanha-o para o início da relativa e o nome que é a cabeça desse complemento reduz-se à forma do pronome relativo apropriado (*que*); e
- (iii) o complemento indireto de *falara* (*ao João*₂) se reduz à forma dativa do pronome pessoal (*lhe*), a fim de evitar a repetição do nome *João*₂.

A Fig. 8 mostra a semillatice que representa (parcialmente) a oração relativa que acabámos de analisar.

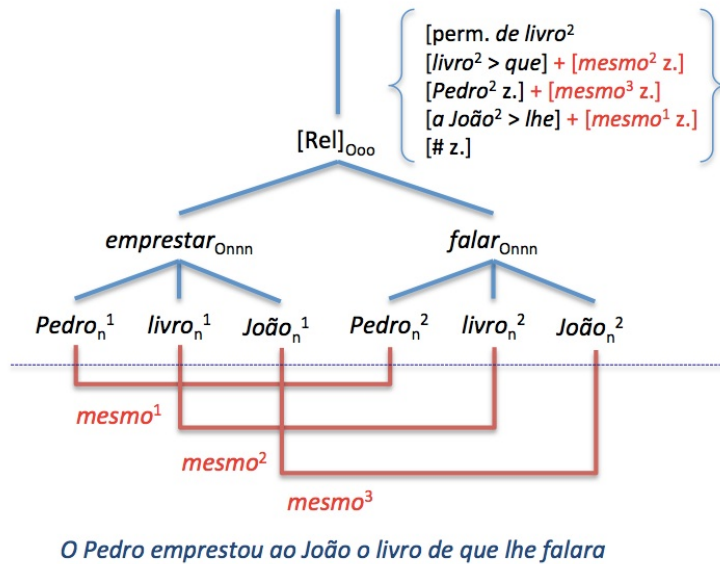


Figura 8: Formação de oração relativa.

Tratando-se de uma subordinação, ou seja, de uma “união de orações”, postula-se a existência de um operador *Ooo*, de tipo conjuntivo, a que, na falta de melhor designação, chamaremos de *operador de relativização* e que foi notado [*Rel*]. Os operadores metalinguísticos de correferência (*mesmo*) estão representados na parte inferior, constituindo, como dissemos, uma espécie de segunda cadeia de frases que se referem às expressões linguísticas que efetivamente ocorrem na superfície do discurso. As várias operações envolvidas no processo de relativização e descritas acima são representadas explicitamente, embora se trate de um único processo complexo. Recorde-se que [*perm.*] indica a *permuta* do constituinte repetido e que é o *pivot* da

operador de relativização

operação de relativização, bem como a notação [X z.] que representa a redução a zero (do ing. *zeroing*) de um elemento correferente. No caso das reduções a proforma (pronomes), usou-se o símbolo '>' e estes pronomes estão explicitamente indicados.

Repare-se que não há redução a zero da segunda ocorrência de *ao João*, ainda que haja identidade referencial (é o mesmo *João*) bem como identidade de função sintática: são ambos complementos indiretos (ou dativos). Tal poderá ser explicado pelo facto de o papel semântico não ser idêntico, ainda que a função sintática de *ao João* seja a mesma (complemento indireto/dativo): com *emprestar* (transferência de objeto) está-se perante um RECIPIENTE OU DESTINATÁRIO, enquanto que com *falar* (ato comunicativo) se trata de um INTERLOCUTOR. Vemos aqui uma relação subtil (e provavelmente pouco estudada) entre a 3ª e a 4ª restrições fundamentais. Pelo contrário, a redução a zero de um sujeito da subordinada correferente ao da principal parece uma operação muito menos dependente do papel semântico – que também é subtilmente diferente! – desse constituinte.

Finalmente, repare-se na última operação formal, indicada por [# z.]. Como o leitor atento reparou, a frase complexa que estamos a analisar é aquilo a que se chama uma *relativa restritiva*. Ora, o operador de relativização tem como função apenas unir as duas frases de base. Nesse processo, a oração a subordinar desloca-se para a posição *imediatamente a seguir* ao nome antecedente do elemento que é *pivot* da relativa, mas mantendo-se a autonomia relativa entre elas:

relativa restritiva

^tO Pedro emprestou um *livro* ao João # [Rel]_{Ooo} # O Pedro falou desse *livro* ao João
 → O Pedro emprestou um *livro* # [Rel]_{Ooo} # O Pedro falou desse *livro* ao João # ao João

Têm então lugar as restantes operações, representadas na lattice, e que conduzem à frase:

→ O Pedro emprestou ao João um *livro* # de *que* lhe falara #

Ora, como se percebe, esta expressão é aquilo a que se chama uma oração *relativa explicativa*. A sua posição na frase, obrigatoriamente a seguir ao antecedente, resulta do processo de inserção por meio do operador de relativização, que referimos acima. Esta oração caracteriza-se ainda pelo facto funcionar como uma espécie de inserção na frase matriz, mantendo-se, contudo, a autonomia das duas orações. Tal é assinalado na escrita, pelo facto de a relativa explicativa se destacar da frase matriz, encontrando-se delimitada por vírgulas (ou assinalada na fala por uma pausa nítida).

relativa explicativa

Porém, a relativa restritiva não é uma mera união de frases apostas uma outra, mas funciona como parte integrante da determinação do

seu antecedente, restringindo a referência do nome que é cabeça do grupo nominal em que se encontra. É por isso que numa frase com uma oração relativa restritiva não há vírgulas a separá-la da frase matriz.

Assim, esta última operação, [#z.], remove a fronteira de frase # e insere a oração subordinada como um modificador do nome antecedente do pronome relativo, fazendo parte do respetivo grupo nominal.

Por outro lado, o pronome relativo *que* não era a única opção disponível no processo de derivação, sendo igualmente possível em português escolher-se a proforma *o qual*. Tal daria origem à frase:

O Pedro emprestou ao João um livro # do qual lhe falara

Como se sabe, trata-se neste caso de uma forma alternativa da *relativa explicativa* a que nos referimos acima. Contudo, a escolha desse pronome bloqueia a progressão do processo de formação da relativa restritiva, já que, perante *o qual* a operação de remoção de fronteira de frase não pode ocorrer.

Refiramos, por último, a deslocação do complemento indireto *ao João*. Trata-se, como já vimos, de uma *permuta de comprimento*, resultado de este constituinte ser de comprimento menor do que o comprimento do complemento direto, quando este integra a oração subordinada relativa, tanto a explicativa como a restritiva. No caso da mera junção de orações – que pode ser vista como um processo de *aposição* de frases, o facto de a oração incisa ter de se deslocar obrigatoriamente para a posição imediatamente a seguir ao nome antecedente (e que será pivot da relativa), estabelece uma ligação entre este nome e a oração incisa. Tal leva igualmente à permuta de comprimento do complemento indireto. Compare-se:

*O Pedro emprestou um livro ao João*¹
 # *O Pedro falara desse livro ao João*²
 → **O Pedro emprestou um livro*
 # *O Pedro falara desse livro ao João*² # *ao João*¹
 [perm. *ao João*¹] → *O Pedro emprestou ao João*¹ ***um livro***
 # *O Pedro falara desse livro ao João*²

A derivação das orações relativas que acima apresentamos demonstra que, através de um mesmo conjunto de operações formais (transformações) extremamente regulares e produtivas, é possível dar conta, de forma integrada e consistente, tanto das relativas explicativas como das restritivas.

Nesta derivação, fica evidente a necessidade na arquitetura de uma gramática das línguas naturais de mecanismos, de natureza metalinguística, que explicitem as relações de correferência entre palavras que ocorrem em sequências de frases concatenadas num

discurso. Estes mecanismos, na ausência de uma metalinguagem externa à linguagem, tomam forma em expressões da própria linguagem, no caso, o operador *mesmo*. É difícil imaginar uma estrutura da linguagem que desse conta da informação veiculada pelos discursos sem que incluisse este tipo de mecanismos (qualquer que seja o formalismo que se adotasse para os descrever).

Contudo, apesar do poder explicativo desta 1ª restrição complementar, a descrição dos discursos da língua carece ainda de um segundo conjunto de mecanismos metalinguísticos a fim de dar conta de uma parte importante da redundância da informação presente nas frases concatenadas. É dessa 2ª restrição complementar que trata a secção seguinte.

2ª restrição complementar

A 2ª restrição complementar é definida de forma subtilmente diferente da primeira: Esta restrição complementar determina que relações operador-argumento(s) se repetem em sequências de frases concatenadas. Note-se que, no caso da 1ª restrição, se tratava apenas da repetição de *elementos*. Nesta 2ª restrição, trata-se da repetição de *relações* operador-argumento(s).

Considere-se a seguinte frase complexa:

A Ana perguntou ao Pedro se tinha ido ao cinema e ele respondeu-lhe que não.

Para simplificar a análise, é possível considerar que as frases de base na origem desta frase complexa são as que se representa a seguir (separando cada uma por #):

*A Ana¹ perguntou algo¹ ao Pedro²
se_{Ooo} # O Pedro¹ tinha ido ao cinema¹
e_{Ooo} # O Pedro³ respondeu algo² à Ana²
que_{Ooo} # não_{Oo} # O Pedro⁴ tinha ido ao cinema²*

*O Pedro¹ é o mesmo que o Pedro²
O Pedro¹ é o mesmo que o Pedro³
A Ana¹ é a mesma que a Ana²
O Pedro³ é o mesmo que o Pedro⁴
O cinema¹ é o mesmo que o cinema²*

Contudo, a aplicação da 1ª restrição complementar não dá conta da redução da oração completiva dependente de *responder* (e que deixa o advérbio de negação aparentemente sem nenhum argumento, o que violaria a 1ª restrição). Para que isso não aconteça, intervem a 2ª restrição complementar, que determina que o conteúdo da completiva-objeto de *responder* é (em parte) o mesmo que o conteúdo da completiva-objeto de *perguntar*. Por outras palavras, a frase de base

A simplificação aqui realizada envolve igualmente a relação entre as proformas *algo¹ algo²* e as orações completivas, bem como a completa representação do domínio argumental da conjunção subordinativa integrante *se/Ooo*. Trataremos destes aspetos mais adiante.

O Pedro tinha ido ao cinema

está repetida no conjunto das frases de base que deram origem à frase complexa que estamos analisando.

A frase complexa apresenta, pois, uma repetição das relações operador-argumento, v.g. $ir_{Omn}(Pedro_n, cinema_n)$, e é essa repetição que autoriza a redução da segunda ocorrência quando as frases com os verbos *perguntar* e *responder* se coordenam, sob a conjunção coordenativa e_{Ooo} .

Assim, ao conjunto das frases metalinguísticas determinadas pela primeira restrição complementar e já apresentadas acima, é preciso acrescentar pelo menos uma outra frase metalinguística, determinada pela 2ª restrição complementar: trata-se de uma que explicita, justamente, que o conteúdo da completiva de *responder* é o mesmo que o conteúdo da completiva de *perguntar*. Por uma questão de simplificação da notação, representaremos essa frase do seguinte modo, indicando o operador da 2ª restrição complementar em maiúsculas (MESMO), a fim de o distinguir dos operadores da 1ª restrição complementar (*mesmo*):

$ir_{Omn}^1(Pedro_n^2, cinema_n^1)$ é o MESMO que $ir_{Omn}^2(Pedro_n^4, cinema_n^2)$

Contudo, é preciso deixar bem claro que esta informação é exatamente a mesma que já referimos – basta ver a nossa frase anterior:

*o conteúdo da completiva de perguntar é o mesmo
que o conteúdo da completiva de responder*

Neste processo de redução, todos os elementos repetidos sob um operador MESMO se reduzem, apenas restando os elementos diferentes sob a coordenação. Em rigor, e na medida em que nas duas completivas coordenadas o verbo *ir* se encontra construído com o verbo auxiliar *ter* V_{pp} (*ter ido*), é sobre o resultado das frases com o verbo auxiliar que este operador metalinguístico opera.

Falta ainda referir um segundo operador metalinguístico resultante da 2ª restrição complementar. O leitor mais atento verificará que o sujeito da completiva dependente de *perguntar* tem uma alta probabilidade de se referir à mesma pessoa que o complemento indireto deste mesmo verbo. De facto, a regra geral de redução do sujeito da completiva

O Pedro tinha ido ao cinema

conduziria apenas à redução a pronome:

→ *Ele tinha ido ao cinema*

já que o seu antecedente é complemento indireto de *perguntar*:

A Ana perguntou algo ao Pedro

Contudo, verifica-se não só a redução a pronome:

A Ana perguntou ao Pedro_i se ele_i tinha ido ao cinema

mas igualmente a possibilidade de esta redução ir mais longe, reduzindo a zero esse sujeito:

A Ana perguntou ao Pedro_i se o_i tinha ido ao cinema

sem que haja qualquer perda de informação. Por outras palavras, continuamos a saber que *quem tinha ido ao cinema* é a mesma pessoa a quem se está a *perguntar* isso. Este tipo de restrição, em que um verbo impõe que um dos argumentos do argumento-operador que dele depende seja correferente de um dos seus argumentos elementares, é igualmente o resultado da 2ª restrição.

A semillatice da Fig. 9 representa a frase que temos vindo a analisar.

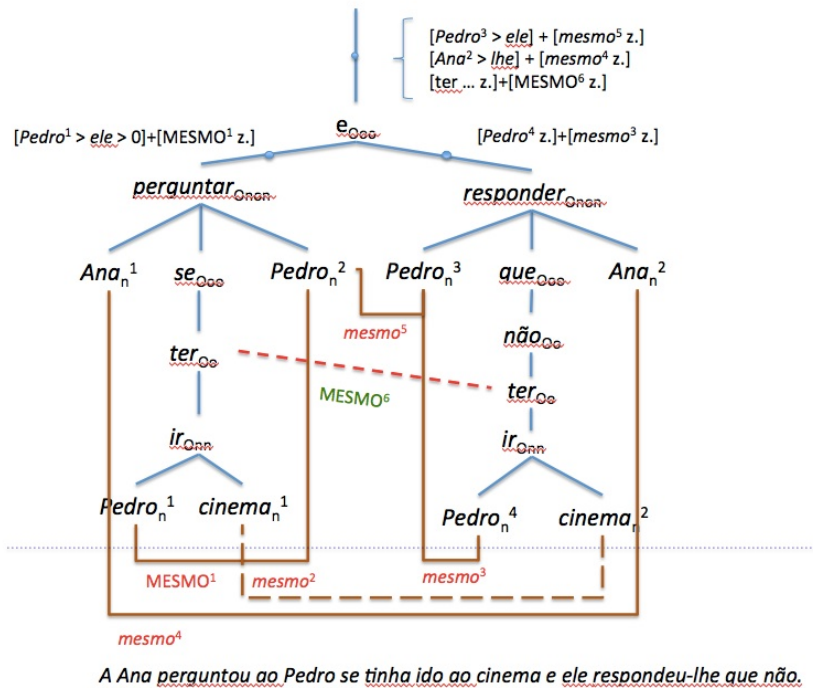


Figura 9: A 2.ª restrição complementar:
 1. Redução de elementos resultante da repetição de relações operadores-argumento em frase coordenadas.
 2. Redução de argumentos de um operador com elevada probabilidade de correferência com argumentos do operador imediatamente superior.

Finalmente, há ainda que considerar uma dependência mais subtil. À semelhança do que se descreveu acima, note-se que a redução a zero do complemento indireto de *responder* (à Ana²) também não

pode ser justificada pelas regras gerais de redução a zero, já que o seu antecedente (*a Ana*¹) desempenha uma função sintática diferente, sendo sujeito de *perguntar*. Tal situação deveria, pois, dar origem a:

O Pedro respondeu-lhe que não

Ora o que provavelmente se verifica é que, sendo dadas as várias condições de correferência atrás descritas, quando os dois elementos *perguntar* e *responder* se encontram coordenados sob a conjunção *e*, há uma alta probabilidade de o sujeito do primeiro operador se referir ao mesmo elemento do complemento indireto do segundo operador; e de o complemento indireto do primeiro operador ser correferente do sujeito do segundo. Tal restrição deveria ser veiculada por um operador *MESMO*⁴, numa representação mais complexa do que a que adoptámos até aqui.

A 2ª restrição complementar é, igualmente, o mecanismo que dá conta da coordenação e redução dos advérbios terminados em *-mente*, que vimos mais acima, em que a mesma frase se repete, coordenando-se com outras de igual conteúdo, embora cada uma seja modificada por um advérbio diferente:

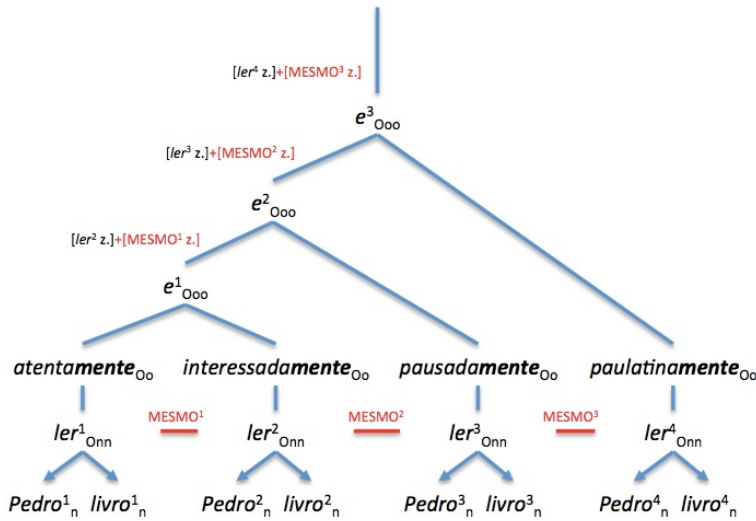
^t*O Pedro leu o livro atentamente (e o Pedro leu o livro) interessadamente (e o Pedro leu o livro) pausadamente (...) e (o Pedro leu o livro) paulatinamente*

→ *O Pedro leu o livro atenta, interessada, pausada (...) e paulatinamente*

Podemos agora completar um pouco mais a semillatice que descreve o processo de formação desta frase, representando desta vez os operadores da 2ª restrição complementar, como se ilustra na Fig. 10.

Trata-se, no entanto, de uma restrição que merece maior investigação, razão por que ainda não a representamos

Figura 10: A 2ª restrição complementar: Redução de advérbios terminados em *-mente* sob coordenação.



O Pedro leu o livro atenta, interessada, pausada, e paulatinamente

Questões teóricas e metodológicas

Nesta secção, tentamos sintetizar os aspectos fundamentais do método harrissiano de descrição das línguas naturais e a as relações com uma teoria geral da linguagem humana a que é possível chegar através da adopção desse método.

Como nos alerta Harris (1991, p.30), de um modo geral, em linguística, as questões metodológicas relativamente à relevância e a adequação das afirmações feitas sobre o fenómeno da linguagem humana nem sempre têm sido objeto de métodos controlados de observação e de argumentação. Nesse sentido, é necessário determinar que métodos são efetivamente relevantes em linguística, a fim de se estabelecerem critérios sólidos que orientem a investigação e a análise. Naturalmente, a escolha do método depende na natureza do objecto e dos problemas que os dados levantam. Por outro lado, o método adoptado deve privilegiar a descoberta de regularidades e a interdependência dos dados observados.

Natureza do problema

A escolha do método de investigação deve ser, acima de tudo, um consequência da natureza do problema a investigar. Quando se fala

Esta secção é, em grande parte uma tradução-adaptação do segundo capítulo de Harris (1991:pp. 30-49). Procurou-se apresentar, numa linguagem talvez mais simples, os conceitos teóricos e metodológicos fundamentais, com exemplos retirados do Português. Naturalmente, uma compreensão mais profunda não dispensa a leitura do texto original.

da linguagem, existem à partida vários tipos de dados e diferentes aspectos e dificuldades que é possível considerar (Fig. 11).

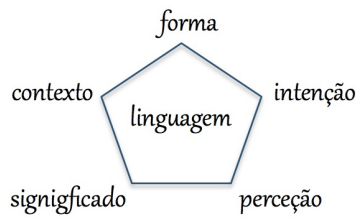


Figura 11: Aspectos da linguagem e natureza do problema: forma, intenção, perceção, significado e contexto.

Um tipo de dados sobre a linguagem resulta da relação entre o falante e aquilo que é dito, isto é, entre a sua *intenção* e a forma como a frase é construída. Em contraponto, encontra-se a *perceção* do ouvinte (ou leitor) relativamente àquilo que é dito (escrito), como é que é decodificado e o que é que o ouvinte/leitor percebe a partir disso. Um terceiro aspecto prende-se com o *significado* do que é dito. O quarto tem a ver com o *contexto* situacional e comportamental que rodeia o processo comunicativo, isto é, que situações e eventos se (co)relacionam com o que é efetivamente dito e quais os efeitos, tanto no contexto imediato, como no contexto cultural mais geral, produzidos pelo que é dito. Finalmente, temos a *forma* da linguagem, considerada aqui em todas as suas modalidades: tanto a cadeia sonora que constitui a fala, e que ocorre em cada ato comunicativo; como, em alternativa, qualquer outro meio adequado de representação da fala (gravação, espectrograma, etc.) ou outro processo equivalente à fala, tal como a *escrita*; como, ainda, a modalidade gestual-visual, empregue nas diferentes línguas gestuais.

Este último aspecto, a *forma* da linguagem, seja como uma sequência de sons, seja como uma sequência de caracteres escritos, ou ainda um sequência de gestos, tem a propriedade de poder ser observado *independentemente* da intenção, da perceção, do significado ou do contexto. Em particular, a escrita é suficiente para veicular a maior parte da informação expressa pela linguagem, como o demonstra a sua eficácia e adequação, independentemente do apoio dos sons da fala, do contacto interpessoal ou mesmo do contexto extralinguístico.

Há, pois, boas razões para abordar em primeiro lugar a *forma* das expressões e, assim, determinar os métodos mais adequados para as estudar. O objetivo de tais métodos são a elaboração de uma teoria precisa, capaz de analisar qualquer amostra de discursos de uma dada língua, produzidos num determinado período, e ser assim capaz de prever a estrutura desses discursos nessa língua (sem, contudo, prever mais do que os que são efetivamente possíveis). Como

independência das observações sobre a forma da linguagem

objetivo do método: elaboração de uma teoria precisa para analisar discursos de uma língua; registo da informação presente em cada expressão analisada.

se verá, ao atingir-se tal teoria, obtém-se igualmente um registo da informação presente em cada expressão que tiver sido analisada.

Esta posição não significa que não seja possível obter conhecimento adicional a partir dos outros aspectos da linguagem, embora tal conhecimento só possa ser obtido de forma relativamente vaga e de maneira pontual. Por outro lado, a condição de só se poder usar a informação obtida a partir da análise da forma das expressões *nas situações em que estas ocorrem* impossibilita que se atinja uma capacidade de análise sistemática da *forma* e da *informação* que a linguagem veicula, e que essa análise seja generalizável para outras situações comunicativas.

Apesar do que se disse, importa salientar que os dados sobre a *forma* das expressões da linguagem podem ser relativamente imprecisos em casos a que poderíamos chamar de *marginais*. Por outras palavras, há expressões para as quais não é claro se fazem ou não parte de uma língua, casos em que os falantes discordam ou não estão completamente seguros sobre o estatuto dessas expressões. Há igualmente muitas expressões ambíguas, isto é, com a mesma forma mas com significados diferentes, que tornam mais complexa a sua descrição. Como noutras situações, a teoria a que aqui se pretende chegar deverá, primeiro, ser capaz de descrever todas as expressões que não apresentem esses problemas; para, só em seguida, tentar dar conta desses casos de expressões com um estatuto marginal e problemático.

casos marginais

Baixa redundância e relevância do método

Como se disse acima, a análise da forma das frases em linguagem natural, independentemente das entidades que se considere (sons, caracteres; morfemas, palavras; gestos), consiste em considerar essas entidades e a forma como se modificam ou como se combinam entre si. Nesse sentido, verifica-se que só algumas combinações dessas entidades ocorrem efetivamente nos discursos de uma linguagem natural, o que resulta em importantes desvios a partir da equiprobabilidade combinatória desses elementos. Estes desvios são universais e resultam do facto de a linguagem natural não ter uma metalinguagem externa. Este aspecto é crucial: para se descrever a forma da linguagem (ou de qualquer outro sistema) é necessário identificar as suas partes, as suas modificações e as relações entre elas. Ora, como não há nenhum sistema *distinto da* (ou *exterior à*) linguagem natural, no qual fosse possível identificar os seus elementos e as suas combinatórias, não é possível descrever as estruturas das frases a não ser com um sistema que é constituído ele mesmo por frases, predicados e

diferenças de verosimilhança de co-ocorrência entre os elementos que constituem a linguagem.

Parece evidente que só uma fração muito pequena das sequências possíveis de sons ocorre efetivamente nas frases de uma linguagem natural, pelo que é possível identificar assim as fronteiras das palavras bem como a probabilidade da sua co-ocorrência com base nos seus contextos frásicos. Isto pode ser conseguido sem que seja necessário que tal nos seja dito por palavras ou mesmo sem conhecimento do que possam ser esses objetos a que chamamos *palavras*.

Uma vez identificadas as palavras, verifica-se que estas se combinam entre si através de um número relativamente reduzido de formas, que são utilizadas ao longo de todo o discurso, pelo que se torna assim possível identificar as frases, tanto as mais curtas como as mais extensas, distinguindo-as das sequências de palavras que não ocorrem no discurso. Por sua vez, dentro desses discursos mais extensos, é possível identificar segmentos (frases) que são idênticos a discursos mais curtos, observados independentemente. Assim, é possível proceder desta forma, de modo progressivamente mais complexo e independentemente se se saber sequer o que sejam (ou mesmo que existam) esses objetos a que chamamos *frases*.

Neste sentido, uma tarefa fundamental do que se tem de dizer *sobre a forma da linguagem* consiste em descrever os desvios a partir da equiprobabilidade combinatória das sequências de sons e de palavras. Tal, porém, levanta alguns problemas. Não parece ser possível descrever esses desvios determinando *probabilidades de transição* fixas entre palavras sucessivas nas frases que formam os discursos de uma língua. Tal deriva do facto de que uma caracterização geral das configurações de palavras consiste numa *ordem parcial* das palavras (e não na sua mera *sequência linear*). Além disso, o conjunto de todas as possíveis combinações de palavras é, em princípio, infinito, ao passo que a gramática de uma língua tem de ser necessariamente finita: uma gramática que fosse apenas um catálogo das combinações de palavras que ocorrem nas frases não constituiria um programa com um objetivo prático. Por último, um número muito grande de expressões são marginais (expressões ambíguas ou de aceitabilidade duvidosa), não sendo claro se se encontram ou não dentro da língua.

Todas estas dificuldades podem ser ultrapassadas se se considerar que a tarefa da teoria não é elaborar uma lista das combinações atestadas, mas sim um conjunto de restrições às combinações de palavras. Um conjunto relativamente pequeno de restrições determina um conjunto não-limitado (ing. *unbound*) de sequências de palavras que *não* podem existir; ao mesmo tempo que permite a construção de todas as outras expressões. Por outro lado, as restrições que são enunciadas sobre diferentes conjuntos de palavras ou sobre conjuntos

de palavras com fronteiras relativamente indefinidas podem criar o conjunto das formas marginais que suscitam dúvidas sobre se satisfazem ou não essas mesmas restrições.

Se se conseguir obter as expressões que efetivamente ocorrem na linguagem ao explicitar as restrições que lhes deram origem, também se estará a especificar por meio dessas restrições as várias contribuições que estas fazem para a não-equiprobabilidade combinatória total dos elementos do sistema. É, pois, necessário que essas restrições sejam formuladas de um modo que não introduza nenhuma redundância adicional além da que é a redundância mínima necessária para dar conta das expressões que são efetivamente observadas.

Há diversas formas de obter essa redundância mínima. Por exemplo, é possível considerar cada forma flexionada de um verbo (e.g. *cantar, canto, cantava, cantei, cantarei, etc.*; ou *ir, vou, ia, fui, irei, etc.*) como se se tratasse de um operador distinto, a fim de dar conta das diferenças morfológicas entre o paradigma de flexão de um verbo regular, como *cantar* por oposição à flexão idiosincrática de *ir*. Tal abordagem poderia ser justificada considerando que as formas *cantei* e *fui* co-ocorrem com um grande conjunto de operadores, como *ontem*, mas não com outros, como *amanhã*; e, inversamente, que as formas *cantarei* e *irei* apresentam uma distribuição complementar à das primeiras. Seríamos, então, levados a constituir duas classes de operadores: por um lado, a das formas que se combinam com *ontem* mas não com *amanhã*; e, por outro lado, a das formas que se combinam com *amanhã* mas não com *ontem*.

Ora, esta solução descritiva não parece ser a mais adequada, pois as propriedades definitórias destas duas hipotéticas classes de operadores (nomeadamente, a probabilidade de combinação com advérbios de tempo do tipo de *ontem* e *amanhã*) não se correlacionam com quaisquer outras propriedades dos elementos dessas classes, ou seja, com as diferentes propriedades das construções de *cantar* e de *ir* (nomeadamente, por exemplo, o facto de uma ser uma construção com complemento-objeto e a segunda envolver um complemento locativo). A redundância introduzida na descrição linguística pela abordagem acima delineada é, como se percebe rapidamente, muito elevada.

Em contrapartida, é possível dizer que todos os verbos se combinam como advérbios de tempo, mas que, em diferentes contextos, o conteúdo fonémico de certas palavras é (bastante) diferente – o que acontece com um número relativamente pequeno de palavras, de resto. Ao fazê-lo, a teoria tem de considerar um nível de análise intermédio (morfofonémico) entre os fonemas e as palavras. A adopção deste nível de análise poderia pois permitir considerar que algumas sequências de fonemas (e.g. *cant-*) formam palavras, sendo delimi-

redundância mínima: aplicação do princípio da navalha de Occam.

distribuição complementar

morfofonémica: nível de análise intermédio entre a fonologia e a morfologia

tadas por fronteiras de morfema ou de palavra; outras palavras são formadas por dois ou mais segmentos (fonemas) que “alternam” entre si, e.g. *v-* (*vou*), *fu-* (*fui*), *i-* (*ia,irei*) em condições especificáveis. Essas condições são homólogas às combinatórias de morfemas (de tempo-modo e pessoa-número) que se observam com os verbos ditos *regulares*.

Ora, este nível de análise (morfofonémico), que distingue entre palavras-morfemas e palavras com alternância de fonemas, é, de facto, necessário em muitas situações. Contudo, a restrição necessária para o determinar para uma palavra como *ir* afeta um domínio muito mais reduzido, do que aquele que teria de ser definido se se considerasse uma restrição cujo objetivo seria distinguir as diferentes formas de *ir* – e que constituem uma mera alternância morfofonémica – a fim de caracterizar a construção do verbo (por exemplo, em relação à co-ocorrência com advérbios de tempo), já que a grande maioria dos verbos apresentam essa oposição (temporal).

Assim, se se tiver diferentes descrições de uma língua, todas adequadas para caracterizar as suas expressões, mas com diferentes quantidades de desvios a partir da equiprobabilidade combinatória, atribuídos em diferentes pontos ao longo da descrição, dever-se-á optar pela descrição que apresenta a menor quantidade desses desvios, uma vez que essa descrição terá claramente acrescentado menos desvios àqueles que são inerentes à estrutura da língua que está a ser descrita.

O efeito da aplicação do *princípio da baixa redundância* a cada ponto da descrição linguística conduz a:

- uma gramática com o menor número possível de objetos, tanto os elementos como as respetivas construções, que são igualmente o mais independentes possível uns dos outros;
- uma gramática com o menor número de classes de objetos, as quais apresentam o menor grau de interseção possível;
- uma gramática com o menor número possível de regras (relações, transformações, etc.) sobre esses objetos, as quais são o mais independentes umas das outras que possível for;
- uma gramática o menor conjunto de diferenças em termos de domínio para essas regras; e, por último,
- uma gramática com o menor conjunto de conceitos abstratos (construtos).

Uma gramática assim estruturada será o menos redundante possível na medida em que cada objeto e cada regra (relação ou classificação), bem como a determinação do domínio de aplicação de cada

princípio da baixa redundância e estrutura da gramática

objeto ou regra, é uma restrição sobre a equiprobabilidade da partes da expressão que estiver a ser descrita.

Nesse sentido, na análise de uma expressão, deve acrescentar-se o mínimo possível ao sistema que descreve as formas analisadas até aí. Obviamente, deve evitar-se repetir informação que já tenha sido dada.

Por exemplo, se já se tiver dito que uma dada sequência de fonemas/letras (e.g. *invadir*) tem um determinado significado e coocorre com elevada probabilidade com um conjunto de palavras como seus argumentos (e que constituem, pois, a sua *selecção*), deve então evitar-se dar esta informação duas vezes: uma vez quando ocorre como um verbo:

Os árabes invadiram a Península Ibérica em 711 d.C.

e outra vez quando ocorre como um nome:

A invasão da Península Ibérica pelos árabes em 711 d.C.<foi rápida>

É preferível, em alternativa, considerar que a segunda forma consiste na mesma frase “encaixada” sob um outro operador posterior (v.g. *foi rápida*). Por outras palavras, é preferível evitar dar a duas palavras que ocorrem em contextos distintos diferentes classificações (neste caso, verbo ou nome) se houver alguma regularidade nas diferenças entre os contextos em que cada uma delas ocorre (as respectivas restrições de selecção, por exemplo), ou seja, se se encontrar outros casos de formas que variam do mesmo modo em contextos com diferenças semelhantes.

Este princípio aplica-se em numerosas situações: no exemplo acima, considerando que uma frase (com o verbo *invadir*) se torna argumento de um operador superior (*rápido*), e que o verbo toma a forma de um nome (*invasão*), mantendo todos os seus argumentos (os quais adquirem um estatuto adjetival ou de *complemento de nome*). Isto quer dizer que a *nominalização*, entendida aqui como o processo transformacional que “encaixa” uma frase como argumento de um operador superior, não ocorre com o verbo (ou com qualquer operador, em geral), mas sim com as frases em que este se encontra construído. Por outras palavras, uma frase nominalizada é a forma que toma uma frase quando encaixada sob um operador subsequente.

Vejamos outro exemplo: para muitos verbos, em vez de se dizer que eles apresentam duas construções distintas, uma intransitiva e outra transitiva:

O Pedro fuma vs. O Pedro fuma cigarros

pode dizer-se que, nestas construções transitivas, o objeto se pode reduzir quando é preenchido por um indefinido:

Veremos mais adiante que é possível considerar que a relação entre o grupo nominal construído com o nome *invasão* não deriva *diretamente* da frase construída com o verbo *invadir*, mas sim de uma frase da base (ou kernel) da gramática, que é uma paráfrase daquela, e que envolve um verbo-suporte do nome predicativo.

nominalização

O Pedro fuma certas coisas → *O Pedro fuma*

redução de indefinido

redução que se encontra abundantemente atestada noutros lugares da língua.

Naturalmente, há situações em que não há outra solução adequada que não seja reconhecer as diferenças que certas palavras apresentam quanto à selecção dos elementos com alta probabilidade de co-ocorrência, como sucede no caso de *palavras homónimas*, e.g.:

homónimos

São em São Francisco

são em Eles são espertos

são em Ele está são (= saudável)

geralmente com etimologias distintas (v.g. *são* < (Lat.) SANCTUS, SUNT, SANUS).

Isto acontece mesmo quando estas palavras homónimas resultam de um mesmo étimo e até ocorrem em contextos relativamente semelhantes. Por exemplo, considere-se a suas construções sintáticas transitivas-diretas do verbo *ralar*:

O Pedro rala o queijo

O Pedro rala a Ana

mas que apresentam diferenças evidentes quanto à sua selecção: o complemento da primeira é obrigatoriamente um nome não-humano, concreto, de um tipo particular (e.g. *queijo*, *pão*, certos tipos <*fruta*>, como *cenoura*, etc.; enquanto o complemento da segunda construção é necessariamente preenchido por um nome humano. Logo, estas duas construções apresentam diferenças nítidas quanto ao seu significado, isto é, a primeira designa um processo físico de manipulação/transformação de um OBJETO, ao passo que a segunda expressa um processo que produz um estado psicológico num humano com o papel de EXPERIMENTADOR. A homonímia é, nestes casos, a análise mais eficiente.

Esta preocupação em atingir uma gramática com o mínimo de redundância possível norteia tanto a identificação dos elementos como a determinação das relações entre eles (o que, em muitos casos, é equivalente). A análise de uma forma da língua deverá, pois, fazer o máximo uso das análises usadas para outras formas, evitando tanto quanto possível o recurso a novas classes, regras e domínios de aplicação, que teriam de ser definidos *ad hoc*. Explicar uma dada forma linguística consiste, pois, em encontrar a máxima semelhança que esta apresenta com outras formas, tornando-se toda a gramática o conjunto que melhor se adapta aos dados. Isto implica analisar cada forma em relação a toda a gramática relevante.

Por outro lado, entende-se aqui este conceito de *gramática* em relação a toda uma comunidade linguística reconhecível como tal, e

não em relação a um falante qualquer (real ou ficcional), uma vez que a linguagem é essencialmente uma estrutura pública (partilhada por uma comunidade).

A construção da gramática com a menor redundância, que se aplique/descreva um conjunto ilimitado de frases, implica que muitas palavras e frases (e os seus significados) sejam *derivados* a partir de outras, através de sucessivas alterações e adições, preservando, no entanto, tudo o que entrou previamente na construção dessas outras palavras e frases. Esta condição sobre a forma de gramática também acarreta que as *derivações* se processem em *pequenos passos*, os quais se repetem em diferentes combinações – em vez de passos muito grandes, cada um dos quais dando conta de todo o conjunto de diferenças entre o ponto de partida e o ponto de chegada. É o que sucede, por exemplo, no caso da derivação das orações relativas (que apresentámos acima e que dá conta simultaneamente de fenómenos tão díspares como a aposição, orações incisivas, orações reduzidas, relativas explicativas, relativas restritivas, modificadores adnominais e, até, comportamento de algumas interrogativas). Em contrapartida, considerar apenas uma única operação o fenómeno a que se chama *relativa* faz com que essa derivação se aplique apenas a um ponto da gramática – a formação da relativa. Pelo contrário, se se considerar que esta forma de frase deriva de uma série de operações menores e independentes, em que os resultados de uma operação estão no domínio de aplicação de outra operação subsequente, é possível verificar que cada um desses pequenos passos ocorre em numerosos outros lugares da gramática, em diferentes combinações. Nesse sentido, pretende-se, neste processo de elaboração de uma gramática da língua, descobrir o menor número de passos derivacionais que, pela sua máxima repetição em diferentes domínios de aplicação (independentemente definidos), dão conta de todas as formas das expressões linguísticas.

Simplicidade

A redução ou minimização da redundância na descrição gramatical produz o que se poderia informalmente chamar de *simplicidade* da gramática, o que é um aspecto não apenas de elegância ou de procedimento científico mas, sobretudo, um requisito metodológico. Dada a natureza construtiva da estrutura da linguagem, quanto mais simples for a gramática, menos processos serão necessários *in toto* para descrever todas as frases da linguagem. A simplicidade global da teoria é, em larga medida, o resultado de contribuições específicas para essa propriedade, cada uma resultando de condições específicas dos dados a descrever. Por exemplo, se uma frase sofrer vários

conservação da informação através da derivação

derivações em pequenos passos

simplicidade

processos de expansão (adições), reduções ou ainda outras transformações, não é necessário assumir que todas elas têm de ter lugar numa determinada ordem. Algumas dessas operações podem ser definidas com base na mesma forma de frase e, portanto, poderão ocorrer simultânea e independentemente umas das outras. Outro exemplo: como a língua contém frases longas e curtas, uma vez que as possibilidades de combinação de elementos é menor nas frases curtas, é possível alcançar um sistema mais simples se se considerar que as frases mais longas são formadas de combinações de frases mais curtas, ou através dos mesmos processos de análise das frases mais curtas.

Outra simplificação resulta do facto de se considerar que as línguas mudam ao longo do tempo, pelo que haverá sempre uma fímbria das formas da língua que não se encaixam perfeitamente na gramática atual. Nesse sentido, a gramática (teoria) corrente é mais simples do que a gramática da língua (sistema) na medida em que minimiza o impacto dessas formas devidas a mudanças que estão em curso. Por outro lado, isto sugere que as regras que é necessário formular para dar conta das combinações efetivamente observadas e que são relevantes no passado recente ou no futuro imediato se limitam a alterações menores nas condições ou no domínio de aplicação dessas regras. Efetivamente, um dos principais métodos de simplificação resulta de se considerar não apenas a forma mas também o seu 'comportamento', ou seja, o contexto e outros fatores que caracterizam as ocorrências dessa forma. Nessa descrição do comportamento de uma forma, é mais fácil indicar as restrições que condicionam certo tipo de combinações do que que explicitar todos os tipos de combinações de palavras que podem aparecer numa língua.

Questões de simplicidade do sistema podem determinar a análise de uma dada forma. Neste sentido, por exemplo, a classificação das distinção fonémicas em fonemas; ou de segmentos formados por sequências de fonemas em palavras; ou a estruturação de conjuntos de palavras em classes de operadores; todos estes procedimentos são levados a cabo não apenas por uma questão taxonómica, com as óbvias vantagens que qualquer abordagem taxonómica oferece; mas, antes de mais e de forma explícita, porque estas classificações, em conjunto com as restrições formuladas sobre as classes assim constituídas (de fonemas, palavras ou operadores), são muito mais simples e em menor número do que as restrições que seria necessário determinar se estas tivessem de ser explicitadas para cada distinção fonémica individual, para cada palavra separadamente ou para cada operador, individualmente. Os métodos de descrição aqui apresentados baseiam-se, assim, nas relações entres os membros das classes assim constituídas.

Contudo, a simplificação numa parte da gramática tem de ser aferida levando em consideração os seus efeitos noutras partes da mesma gramática. A simplicidade do sistema é maior se qualquer propriedade ou análise que tiver sido determinada num certo ponto da descrição puder ser preservada em pontos posteriores do desenvolvimento da descrição. Por exemplo, a frase

(1) *Um golfinho não é um peixe*

não deverá ser analisada como o resultado da asserção (com a respectiva entoação) da frase

(2) *Um golfinho é um peixe*

à qual foi, então, acrescentado um *não* negando a asserção inicial. Isto não significa que se ignore a relação (1)-(2). Mas é preferível obter a frase (1) a partir de uma frase (3') com o operador *nego*, operando sobre a frase não proferida (sem entoação) (2):

(3') *Eu nego_{O_{no}} # que_{O_{oo}} # (2) Um golfinho é um peixe*
 → (1) *Um golfinho não é um peixe*

Constitui, pois, uma contribuição para a simplicidade do sistema conferir um carácter acrecional/cumulativo a cada passo do processo de construção, da análise e da interpretação de cada estrutura da linguagem, evitando, assim, ter de reconsiderar o que tiver sido construído até esse ponto.

Esta parcimônia de análise implica, ainda, não se considerar na análise de uma dada forma distinções que a língua ela própria não faz. Por exemplo, no caso de um pronome como *ele*, não é possível em geral determinar de forma única a partir do que é dito a que indivíduo humano masculino (ou objeto) este pronome se refere. Uma gramática com preocupações de simplicidade não tentará, por isso, suplantiar a própria linguagem, criando um conjunto de leituras determinativas, uma para cada possível antecedente; um conjunto tal que pudesse, então, ser progressivamente reduzido com base em material contextual posterior (cada leitura teria, assim, um antecedente definido, de uma forma que a própria linguagem não estabelece). O facto de tais alternativas nos levarem para além daquilo que constitui a estrutura da gramática torna-se evidente pelo facto de se verificar que a gramática apresenta mecanismos específicos para localizar o antecedente de um pronome, por exemplo, ao colocá-lo em posição imediatamente após o antecedente:

O Pedro ofereceu ao João um livro; que; falava deste assunto

ou numa posição específica relativamente ao operador de que depende diretamente e deste operador em relação ao operador em cujo domínio se encontra o antecedente do pronome:

Note-se que não se diz aqui que (2) não pode ser afirmada, mas sim que (1) não deverá ser descrita como uma asserção resultante da modificação (por *não*) de uma outra asserção prévia (2).

referência de um pronome

leitura determinativas

mecanismos gramaticais específicos para localização do antecedente de um pronome

O Pedro ofereceu ao João; **um livro de que** lhe_i falara
 cp. O Pedro_j ofereceu ao João_i; **um livro de que** este_i lhe_j falara.

Finalmente, além da sua simplicidade interna, a gramática deve ter de suportar uma metalinguagem o mais simples ou reduzida possível. Isto não se refere à simplicidade da *estrutura* da metalinguagem, já que esta é uma sublinguagem de uma língua natural, cuja gramática é (apenas) parcialmente diferente da gramática da linguagem natural que a metalinguagem descreve. Não se trata, pois, da estrutura da metalinguagem mas sim dos discursos dentro dela: quanto é que tem de ser dito dentro da metalinguagem sobre a linguagem que ela descreve.

simplicidade da metalinguagem

É possível evitar certos comentários metalinguísticos sobre eventos especiais (formas e seus 'comportamentos') na linguagem se se puder obter esses eventos a partir da gramática da linguagem (que está na metalinguagem) sem nenhuma adição suplementares, ou apenas com a adição de elementos que podem ser úteis para descrever outras situações. Assim, por exemplo, ainda que se pudesse criar diferentes categorias gramaticais, correspondendo cada uma a diferentes distinções, para descrever os diversos valores semânticos das orações subordinadas adverbiais (temporais, condicionais, consecutivas, causais, etc.), não se deverá criar nenhuma categoria especial em situações que resultam de distinções que já são estabelecidas de outro modo.

Concretamente, a par do valor temporal da conjunção subordinativa TEMPORAL *quando*:

(1) O Pedro fez isto **quando** o João fez aquilo

não parece necessário postular um valor CONDICIONAL para esta conjunção, como o que aparentemente ocorre em

(2) O Pedro fará isto **quando** o João fizer aquilo

já que o mesmo pode ser derivado da combinatória de tempo-modo dos verbos das orações principal e subordinada (v.g. *futuro do indicativo* e *futuro do conjuntivo/subjuntivo*). Bastaria para tal:

- (i) deixar por definir na base da gramática a relação TEMPORAL entre as orações ligadas por *quando*, ao contrário do que sucede normalmente com este tipo de conjunção subordinativa, por exemplo:

antes de (ANTES), *depois de* (DEPOIS), *enquanto* (SIMULTÂNEO), etc.).

- (ii) fazer decorrer a aparente interpretação CONDICIONAL, e.g.

(2) \approx (2') *O Pedro fará isto se o João fizer aquilo*

da modalidade particular (irreal) do futuro do conjuntivo/subjuntivo na oração subordinada com a modalidade (irreal) do futuro do indicativo na oração principal.

Pode o significado ser usado?

A linguagem é claramente um veículo de significado, seja nas palavras seja nas formas particulares em que as palavras se combinam umas com as outras. Na determinação dos métodos de análise da linguagem, poder-se-ia pensar que as formas (as palavras e as construções) e significados que estas veiculam poderiam ser analisadas independentemente uns dos outros. Contudo, o facto de a linguagem não ter uma metalinguagem externa requer que estes dois aspectos estejam interligados. Como se viu atrás, a ausência de uma metalinguagem externa leva a que seja impossível identificar todos os elementos na cadeia da fala sem se levar em consideração as restrições que se verificam quanto à sua combinação. Do mesmo modo, é também necessário referimo-nos às restrições quanto às combinações de palavras na especificação do respetivo significado.

Isto sucede porque não há forma de definir o significado de todas as palavras, nem o significado das construções criadas pelas hierarquias de restrições à sua combinação, excepto se essa definição for feita na própria linguagem. Embora possa ser, sem dúvida, adequado definir uma palavra apontando para o seu referente; ou compreender o significado de uma palavra ou uma frase pela situação em que as mesmas são ditas – o que de facto é essencial quando se trata de capturar certas palavras e frases curtas (as interjeições, por exemplo); tais processos não são, porém suficientes para todas as palavras nem, sobretudo, para os seus diferentes significados quando em diferentes combinações. Palavras como *tempo*, *considerar*, *o* ou *de*; ou o significado de 'frase subordinada' de uma oração relativa; não podem ser adequadamente obtidos a partir de situações do mundo real em que são usados.

Por outro lado, qualquer que fosse a metalinguagem em que tais significados pudessem ser dados, esta teria de conter em si mesma, como palavras suas, o tipo de significados que está a ser explicado. Assim, os significados menos óbvios destas palavras e construções só podem ser aprendidos pelos falantes e pelos analistas da língua depois de muita experiência a partir das palavras e frases vizinhas com as quais eles ocorrem.

Efetivamente, quando procuramos o significado de uma palavra na respetiva definição num dicionário, estamos na realidade a aprendê-lo a partir dos contextos frásicos em que esta ocorre, ou seja, a partir

da distribuição (previamente conhecida) das palavras que aparecem na definição, muitas vezes uma mera lista de (quase) sinónimos; e a partir das frases de exemplo, e não por qualquer experiência extralinguística da palavra.

Além disso, muitas palavras ocorrem em diferentes combinações e, com estas diferentes distribuições, verifica-se uma variação no seu significado. De um modo geral, não podemos inferir quais as mudanças que se verificam nos coocorrentes de uma palavra a partir de uma diferença prévia de significado, mas podemos compreender a mudança de significado ao observar as mudanças nas palavras com que uma dada expressão coocorre; por exemplo, o uso conjuncional de *visto (que)*, a partir das construções gerundivas *tendo sido visto/tendo alguém visto que*; ou a oposição *porco/vaca* como tipo de *carne* a partir dos nomes designativos destas espécies de animais.

Por outro lado, os significados que podem ser expressos pela linguagem – e, de forma ainda mais relevante, os que não podem ser expressos por meio da linguagem – em geral não são objetos discretos, que pudessem ser mapeados nas palavras das construções linguísticas. Há, de facto, alguns conjuntos de significados que podem ser conhecidos independentemente da linguagem, tais como, por exemplo, os números, as relações de parentesco, as relações espaciais assim como os humanos as experienciam. Nestes casos, é possível partir dos significados e determinar como é que a linguagem os exprime.

Contudo, este método não pode ser estendido à grande massa de significados que correspondem às palavras e às construções de uma língua. Na língua, não há um conjunto definível *a priori* de significados que cubra o vasto conjunto de experiências relevantes partilhadas por uma comunidade. É necessário verificar, primeiro, qual o vocabulário dessa língua e, depois, como ele é usado, em que combinações e em que construções. Não é possível determinar *a priori* quais os objetos, ações ou relações que estarão incluídos num significado por oposição a outro. É preciso, por exemplo, ver primeiro que nomes se combinarão com *bater*, por oposição a *ferir*; ou os que ocorrem com *sob* por oposição a *embaixo de/debaixo de*.

Além de não ser possível conhecer o significado das expressões independentemente das combinações das palavras na linguagem, os significados não são suficientemente precisos para especificar/predizer todas as combinações de uma palavra. Assim, por exemplo, ainda que *negro/preto* sejam em larga medida sinónimos, apenas certas combinações são efetivamente aceitáveis na língua, tais como *chá preto/*negro*, *nota preta^{PB}/*negra*; *caixa negra/*preta*, *magia negra/*preta*.

Não só é insustentável utilizar o significado como um quadro de referência geral para a análise da linguagem, como também o significado não pode ser utilizado como critério ocasional para decidir como analisar a estrutura linguística das frases; por exemplo, quando é difícil determinar algumas restrições combinatórias. Por outro lado, o facto de se apelar ao significado quando a análise formal é difícil impede outras tentativas de encontrar uma ‘explicação’ puramente formal. Ao insistir numa abordagem estritamente formal, consegue-se chegar a explicações perfeitamente satisfatórias, mesmo nas situações em que o significado pareceria uma saída mais fácil (e.g. a análise da metáfora).

Mas o mais importante é o facto de que, quando toda a linguagem é descrita em termos de restrições combinatórias, sem que o investigador se desvie do método, fazendo apelo ao significado, se verifica que a linguagem é constituída a partir de um pequeno conjunto de regras, praticamente sem exceções, e cuja força não é completamente evidente quando apenas se dispõe de análises parciais da linguagem, ou quando apenas se fazem referências ocasionais às restrições combinatórias das formas linguísticas.

Além disso, o conjunto das construções gramaticais que é obtido através de uma análise puramente formal acaba por ter uma correlação muito mais precisa com o significado das expressões linguísticas do que a correlação a que se chegaria através de uma gramática parcialmente semântica (veja-se, por exemplo, a dificuldade em atingir algum tipo de consistência formal a partir da classificação semântica dos tipos de conjunção subordinativa/oração subordinada).

Os significados na linguagem são, pois, os significados das palavras e os das construções que estão na linguagem e que são conhecidos *antes* de tentarmos determinar os seus significados. Efetivamente, as fronteiras e a identidade das palavras e das construções pode ser estabelecida por meios estatísticos, sem qualquer referência ao conhecimento dos seus significados. A partir daí, verifica-se que os significados não são propriedades adicionais das formas, não relacionadas com as formas sintáticas, mas que são estreitamente concomitantes com as restrições quanto à escolha de palavras na relação operador-argumento(s) e quanto à possibilidade de as palavras entrarem em certas construções ou sofrerem certas reduções.

Todas estas observações acima não significam que se esteja a negar a utilidade do significado numa investigação formal, e muito menos em estudos sobre o uso da linguagem ou as formas de arte baseadas na linguagem. O significado, e em especial a percepção da *diferença de significado*, quer entre duas palavras quer entre duas ocorrências da mesma palavra, pode ser usado como uma ferramenta preliminar para tentar encontrar eventuais construções. É claro, a partir do que

Veja-se, a este respeito, o uso do conceito de APARIÇÃO na classificação das construções verbais em Baptista (2012,2013), bem como muitas outras situações homólogas.

se disse atrás, que de modo nenhum está aqui o significado a ser excluído do estudo da linguagem. Ele apenas não está a ser utilizado como critério para determinar as formas da linguagem (nem os métodos a aplicar para a estudar). Efetivamente, considerações gerais sobre o significado podem ser derivadas a partir de considerações sobre as formas da linguagem.

Finalmente, uma última observação sobre a relação entre o significado e a investigação das formas da linguagem. Recorde-se que a descrição das combinações de sons ou de palavras se deparava com duas fontes de redundância: (i) a redundância que resulta do processo de descrição e classificação, bem como do modo como são formuladas as regras que presidem às combinatórias destas formas; e (ii) a redundância que resulta das restrições residuais, derivadas da própria linguagem. Na medida em que o significado está relacionado com as restrições à combinação das palavras, ele está, pois, relacionado com o segundo tipo de redundância e não com a que resulta do processo de descrição das formas. Por essa razão, torna-se necessário atingir uma descrição com o menor grau de redundância antes de se considerar o significado e a relação interpretativa entre ele e as restrições de forma.

Propriedades universais da linguagem

A linguagem apresenta um conjunto de propriedades universais, além das restrições sobre as combinações de palavras em frases, que determinam igualmente os métodos relevantes para a sua análise.

Linguagem: um conjunto infinito de formas, um conjunto finito de falantes

O conjunto das frases de uma língua (tal como podem ser descritos por uma gramática) é, potencialmente, infinito (em ing. *unbounded*): não há nenhuma forma de determinar o comprimento máximo de uma frase, já que, potencialmente, a qualquer frase pode ser acrescentado um elemento suplementar; enquanto os falantes de uma língua são, necessariamente, seres finitos. Muitos aspectos da linguagem (nomeadamente, que sons são fonologicamente distinguidos numa dada língua; que sequências de fonemas constituem as palavras e os morfemas de uma língua; que sequências de palavras formam expressões aceitáveis, isto é, frases) dependem de um corpo necessariamente finito de experiências públicas (dos falantes). Isto implica que qualquer língua tem de poder ser descrita através de um sistema construtivo (recursivamente enumerável); um sistema que é formado por um núcleo, que pode ser razoavelmente aprendido, e sobre o qual podem ser determinados processos de repetição ou de recursividade, processos, esses sim, cuja aplicação será *a priori* ilimitada (possivelmente

Note-se que, conquanto o comprimento máximo de uma frase é, em potência, ilimitado, a capacidade real dos falantes para produzir ou compreender frases muito longas é limitada pelo seu próprio sistema cognitivo. Concretamente, limitações naturais na estrutura da memória de curta duração dificultam ou impedem mesmo que um falante possa ser capaz de “processar” frases demasiado longas, ainda que esse processamento se faça, em grande parte, sequencialmente, já que só é possível atribuir uma estrutura completa (e, conseqüentemente, uma interpretação) a tais frases quando se chega ao seu término.

com um conjunto finito de material 'idiomático' que não se encontra sujeito a estas operações de aplicabilidade ilimitada).

Elementos discretos O sucesso dos diversos sistemas de escrita, os quais usam todos marcas ou sinais discretos como veículos da linguagem, sugere que a linguagem pode ser representada por meio de elementos discretos. Por outro lado, verifica-se que é possível segmentar a cadeia da fala ou a sequência de palavras no discurso e constituir assim os seus elementos discretos. Assim que os elementos discretos da linguagem são atingidos, limita-se em grande medida a escolha dos métodos relevantes para a investigação subsequente.

Linearidade A fala tem apenas uma extensão no eixo do tempo e a sua representação na escrita é essencialmente uma sequência de sinais (embora em alguns casos possa haver outros sinais que são simultâneos ou que coocorrem com os primeiros). Isto sugere que a linguagem pode ser descrita de forma linear – ou, quanto muito, de forma multilinear, em que um dado evento é resultado de uma sobreposição simultânea de vários eventos lineares.

Contudo, a representação linear não significa que os elementos da linguagem sejam ordenados linearmente em relação à estrutura: os elementos da linguagem podem ser ordenados parcialmente, com uma projeção linear subsequente, tal como se propõe aqui. Em concreto, as relações gramaticais (sujeito, complemento, modificador, determinante, etc.) que se determina entre os elementos da linguagem são definidas não de forma linear, em relação à sequência de palavras no discurso, mas sim em relação à ordem parcial das palavras na estrutura.

Contiguidade A fala ou a escrita não se desenvolvem num qualquer espaço mensurável. A única distância entre duas palavras de uma frase é a sequência de palavras que se encontra entre elas. Assim, a única relação elementar diretamente experimentável entre duas palavras ou entidades numa sequência de palavras consiste no facto de essas duas entidades serem vizinhas uma da outra (ou serem simultâneas, como no caso da entoação).

Por essa razão, a boa formação das estruturas frásicas exige que na base exista uma sequência contígua de objetos discretos, ainda que permutações, reduções e outras operações posteriores possam intervir e modificar essa sequência inicial. A única propriedade que torna essa sequência uma construção da gramática é o facto de esses objectos não serem palavras arbitrárias, mas sim membros de classes específicas de palavras. Mas a sequência tem de ser contígua, não pode encontrar-se dispersa, com “espaços” entre os

elementos, já que não há qualquer forma de identificar os espaços ou de os medir.

Do mesmo modo, o efeito de qualquer operação que é definida na estrutura da linguagem (a mudança ou adição de elementos que essa operação introduz no respetivo operando) tem de ocorrer no seu operando ou de forma contígua a esse operando. Naturalmente, operadores posteriores que ajam no resultado de uma dada operação podem figurar entre o operador anterior e o seu operando, separando-os na sequência linear de palavras da frase, tal como sucede, por exemplo, com a coordenação e redução de advérbios terminados em *-mente*. No entanto, na identificação da ação do operador prévio sobre o seu respetivo operando, essa interrupção não pode ser encontrada, já que se deve apenas à acção de um operador posterior.

Portanto, se a linguagem deve ter uma gramática construtiva, então deve também haver uma maneira de descrever apenas por meio de relações puramente contíguas as frases da base. Essa contiguidade das palavras *do discurso* não satisfaz, porém, este requisito já que se verifica que as frases são caracterizadas por uma ordem parcial dos seus elementos, dos quais a ordem linear é uma mera projeção. Apesar disso, esta contiguidade pode ser utilizada para conceber programas informáticos para analisar a estrutura das frases.

Varição dentro de uma classe Vimos acima que a procura de uma descrição da linguagem com a menor redundância possível conduz à constituição de diferentes classes de palavras, de modo a que uma propriedade possa ser definida de maneira mais simples para a classe como um todo do que para cada membro da classe, individualmente. Tal pode suceder quando os elementos de uma classe têm todos essa propriedade e não desejamos que se tenha de repetir essa propriedade para cada um deles. Mas também sucede que apenas alguns membros de uma classe têm determinada propriedade, enquanto outros não a apresentam. Sucede ainda que certos elementos de uma classe têm umas propriedades, enquanto outros têm outras propriedades, sendo os dois conjuntos complementares, pelo que se pode considerar que pares (compósitos ou disjuntos) de elementos formam efetivamente um todo, o qual é, então, o verdadeiro membro da classe. Em todos estes casos, pode ser revelador definir uma propriedade de classe.

Assim, a formação de classes não é processo de agrupamento dos dados com base na “similaridade” entre eles, mas sim um processo de descoberta do modo mais regular de atribuir propriedades a entidades, ou seja, de definir propriedades e os seus respetivos domínios. Portanto, na constituição de classes, deve

similaridade

levar-se em consideração a sua variedade, a degenerescência da informação e outras dificuldades como que se pode deparar. Em particular, nas classes de palavras é determinante perceber em que medida é que todos os membros de uma dada classe podem ser argumentos de um conjunto de operações ou ocorrer em determinados contextos. Ou, inversamente, se todos podem ter como argumentos as mesmas classes de palavras.

Ora, tal como há muitas classes em que o conjunto dos seus membros inclui elementos cuja pertença à classe não é muito precisa (em ing. *fuzzy*), também há muitas operações e construções com domínios de aplicação relativamente imprecisos. É, pois, possível considerar dois aspectos desta imprecisão:

classes com fronteiras imprecisas

- (i) algumas operações apresentam apenas relativamente poucas diferenças em relação a outras quanto aos respetivos domínios de aplicação;
- (ii) enquanto as fronteiras dos domínios de aplicação de outras operações e construções são relativamente indefinidas.

Quanto ao primeiro aspecto, tal dá origem a *frases ambíguas* porque a sua análise pode resultar de diferentes processos de derivação aplicados sobre diferentes formas mas que, por um processo de *degenerescência da informação*, conduziram a um mesmo resultado.

degenerescência da informação

Por exemplo, o grupo nominal *a invasão de Portugal* é uma expressão ambígua pois tanto poderá derivar de uma frase equivalente a:

Alguém invadiu Portugal

como poderá ter origem numa outra frase, em que teria havido uma redução metonímica de um indefinido associado a *Portugal*

redução metonímica

(As pessoas que formam o país chamado) Portugal invadiu algum lugar

Quando ao segundo aspecto, as pessoas não têm a certeza de que uma dada operação possa ter lugar sobre uma determinada palavra, que se encontra provavelmente no seu domínio de aplicação. O resultado é uma *frase marginal*, como, por exemplo:

frase marginal

?*O gato soluçava.*

Coexistência sincrónica de formas numa mesma língua Considerando o facto não só de que há numerosas línguas naturais no mundo; mas também o facto de que as línguas apresentam internamente alguma variação dialetal; conjugado com a evidência de que as línguas mudam ao longo do tempo; tal leva-nos a considerar que há diferentes objetos que são passíveis de análise e descrição linguística.

Seria, assim, possível, descrever todas as línguas no seu conjunto, isto é, analisar os aspectos universais, comuns a todas línguas naturais; ou descrever apenas uma língua ou uma família de línguas ao longo do tempo, dando conta das formas sucessivas de certos elementos ou construções particulares.

Contudo, é particularmente importante a abordagem que se limita a *uma única comunidade linguística* (ignorando as diferenças que não afetam a intercompreensão mútua, ou as que se devem a fatores de ordem geográfica ou social) durante *um único período "sincrónico"*, no qual as mudanças são demasiado lentas para serem vistas, ou seja, num dado intervalo de tempo. Dentro destes limites, todas as formas, significados e distinções coexistem para os falantes dessa língua. É nestas condições que é possível desenvolver uma gramática que descreva a língua como um sistema completamente integrado, pois só assim é que é possível obter uma visão panorâmica das formas e das restrições que se verificam quando estas são usadas em conjunto.

análise de uma sincronia linguística

Variação diacrónica As línguas naturais são objetos abertos: é permanentemente possível criar novas frases e novos discursos. Além disso, as línguas mudam ao longo do tempo e, tanto quanto se percebe, sem que, em qualquer ponto desse processo histórico, deixem de ter uma estrutura gramatical fundamentalmente coerente.

Assim, ao longo do processo de mudança de um momento arbitrário da história t_1 de uma língua para outro momento t_2 , igualmente arbitrário, dessa história, um elemento passa de uma análise que é adequada à gramática do momento inicial para uma nova análise adequada à do momento final, sendo necessário considerar um momento intermédio t_i , no qual ambas as referidas descrições desse elemento são simultaneamente possíveis: as mudanças verificadas no elemento já permitem considerá-lo adequado à gramática de t_2 mas ainda não foram tão grandes que o tornassem inadequado à gramática de t_1 . Isto implica que algumas formas da língua possam ter múltiplas análises (ou, pelo menos, análises não únicas). Tal pode observar-se, portanto, na análise de mudanças que ainda estão em curso.

Por exemplo, a expressão *com vista a* (em Português Europeu), hoje com um valor conjuncional:

(A) *alguém fez isso com vista a fazer qualquer coisa,*

e impondo condições de correferência particulares (que por ora ignoramos); poderá (ter de) ser analisada diacronicamente a partir de múltiplas alterações de uma construção inicial, hoje inexistente:

(B) **fazer qualquer coisa está na vista de alguém*

Uma análise possível da expressão *com vista a* (do Português Europeu).

expressão que é concebida à semelhança da construção, ainda hoje usual:

(b) *fazer qualquer coisa está na intenção de alguém*

Esta última construção pode ser associada a uma outra frase com *estar* e preposição *com*:

(c) *alguém está com a intenção fazer qualquer coisa*

em que os argumentos se apresentam numa ordem linear inversa. As construções com *estar com* deste tipo dão origem a grupos preposicionais introduzidos por *com*, numa sequência de frases concatenadas (provavelmente por um processo de *gerundivização* que ainda não tratámos aqui):

alguém fez isso # alguém está com a intenção de fazer qualquer coisa
 → (C) *alguém fez isso com a intenção fazer qualquer coisa*

gerundivização

Nada obsta a que idêntico processo derivacional tenha tido lugar no caso de *com vista a*, ainda que a presença da preposição *a* requeira alguma argumentação adicional. Repare-se que (A) e (C) são idênticos, ainda que os correspondentes passos intermédios da derivação do segundo, ilustrados por (b) e (c), tenham de procurar-se em sincronias pretéritas da língua, a saber:

(b') *fazer qualquer coisa está na vista de alguém*
 (c') *alguém está com vista a fazer qualquer coisa*

O que importa reter aqui é que duas formas de frase (duas análises) coexistem para (C) num dado momento, ou seja, na presente sincronia.

A possibilidade de uma gramática estrutural estática em cada sincronia (ou corte temporal) num sistema em mudança resulta do facto de haver semelhanças e outras relações entre os vários domínios de aplicação determinados na gramática e entre as várias operações que nela foram definidas. Ao nível elevado de análise destas relações sistémicas, pode ser demonstrado que o carácter excepcional dos domínios de aplicação de certas operações resulta de extensões ou modificações de algum dos seus domínios regulares.

Estas considerações implicam que a descrição de uma língua tem de garantir a existência de itens que não se “encaixam” de modo completamente adequado nas regras do resto da gramática dessa língua, mas podem, não obstante, ser postos em relação com essas regras como extensões do seu domínio ou como modificações menores dessas operações.

Referências

- Baptista, J. (2010). Verba dicendi: a structure looking for verbs. In T. Nakamura, É. Laporte, A. Dister, and C. Fairon (Eds.), *Les Tables. La grammaire du français par le menu. Mélanges en hommage à Christian Leclère*, Number 6 in Cahiers du CENTAL, pp. 11–20. Louvain-la-Neuve: CENTAL/Presses Universitaires de Louvain.
- Baptista, J. (2012). ViPER: A Lexicon-Grammar of European Portuguese Verbs. In J. Radimsky (Ed.), *Proceedings of the 31st International Conference on Lexis and Grammar*, Nové Hradý (Czech Republic), pp. 10–16. Università degli Studi di Salerno (Italy)/University of South Bohemia in Nové Hradý (Czech Republic).
- Baptista, J. (2013). ViPER: uma base de dados de construções léxico-sintáticas de verbos do Português Europeu. In F. Silva, I. Falé, and I. Pereira (Eds.), *Actas do XXVIII Encontro da APL - Textos Seleccionados*, Faro, Portugal, pp. 111–129. XXVIII Encontro Nacional da Associação Portuguesa de Linguística, Coimbra, 2012: APL.
- Baptista, J., N. Mamede, and F. Gomes (2010). Auxiliary verbs and verbal chains in European Portuguese. In T. Pardo, A. Branco, A. Klautau, R. Vieira, and V. L. S. de Lima (Eds.), *Computational Processing of the Portuguese Language*, Number 6001 in Lecture Notes in Computer Science / Lecture Notes in Artificial Intelligence, Berlin. PROPOR 2010: Springer.
- Barsky, R. F. (2011). *Zellig Harris: From American Linguistics to Socialist Zionism*. Cambridge, MA: MIT Press.
- Burton, S., R.-M. Déchaine, and E. Vatikiotis-Bateson (2016). *Linguistics for Dummies*. Wiley.
- Daladier, A. (1990). Aspects constructifs des grammaires de Harris. *Langages* 99, 57–84.
- Harris, Z. (1964). *Papers on Syntax*, Chapter The elementary transformations, pp. 211–235. D. Reidel Publishing Company.
- Harris, Z. (1978). Operator-grammar of English. *Linguisticae Investigationes II*, 55–92.
- Harris, Z. S. (1946). From morpheme to utterance. *Langages* 22(3), 161–183.
- Harris, Z. S. (1951). *[Methods in] Structural Linguistics*. Univ. Chicago Press.
- Harris, Z. S. (1952). *Discourse Analysis*. D. Reidel Publishing Company.

- Harris, Z. S. (1955). *Co-occurrence and transformation in linguistic structure*, pp. 143–210. D. Reidel Publishing Company.
- Harris, Z. S. (1961). Strings and transformations in language description. *Papers on formal linguistics 1*.
- Harris, Z. S. (1968). *Mathematical Structures of Language*. Pennsylvania: Interscience publishers.
- Harris, Z. S. (1970). *Papers in Structural and Transformational Linguistics*, Volume 1 of *Formal Linguistics Series*. Dordrecht, Holland/New York: D. Reidel Pub. Co./Humanities Press.
- Harris, Z. S. (1976). *Notes du Cours de Syntaxe*. Paris: Seuil.
- Harris, Z. S. (1981). *Papers on Syntax*, Volume 14 of *Synthese Language Library - Texts and Studies in Linguistics and Philosophy*. Dordrecht/Boston/London: D. Reidel Pub. Co.
- Harris, Z. S. (1982). *A Grammar of English on Mathematical Principles*. New York: John Wiley & Sons, Wiley-Interscience Pub.
- Harris, Z. S. (1988). *Language and Information*. New York: Columbia University Press.
- Harris, Z. S. (1991). *A Theory of Language and Information. A Mathematical Approach*. Oxford: Clarendon Press.
- Hockett, C. (1966). The problem of universals in language. In J. Greenberg (Ed.), *Universals of language*. Cambridge, MA: MIT Press.
- Jurafsky, D. and J. Martin (2009). *Speech and Language Processing: an Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (2nd ed.). Prentice Hall series in Artificial Intelligence. Pearson/Prentice Hall.
- Laporte, E. (2015). The science of linguistics. *Inference: International Review of Science*, 2015 1(2), s/p.
- Manning, Chris; Schütze, H. (1999, May). *Foundations of Statistical Natural Language Processing* (1st ed.). Cambridge, MA: MIT Press.
- Martinot, C., C. Marque-Pucheu, and S. Gerolimich (Eds.) (2016). *Perspectives Harrisiennes*, Paris. Proceedings of the International Conference “L’héritage de Zellig Sabbetai Harris”, Duino-Aurisina, June 20-22, 2014: CLR - Cellule de Recherche en Linguistique.
- Mihalilicek, V. and C. Wilson (2011). *Language Files. Materials for an Introduction to Language and Linguistics* (11 ed.). Ohio State University Press.

Nevin, B. (1992, Spring-Summer). Zellig S. Harris: An appreciation. *California Linguistic Notes* 23(2), 60–64.

Nevin, B. E. and S. B. Johnson (Eds.) (2002). *The Legacy of Zellig Harris: Language and information into the 21st century*. Current Issues in Linguistic Theory. Amsterdam/Philadelphia: Publisher: John Benjamins Publishing Company.

Saussure, F. (1916). *Cours de Linguistique Générale*. Lausanne: Payot.