Representação Sintáctica em HPSG

António Branco e Francisco Costa

Introdução

Este capítulo apresenta uma introdução à representação sintáctica em HPSG. Para tanto, são apresentadas nos seus aspectos essenciais as análises gramaticais para alguns fenómenos sintácticos básicos.

Cabe relembrar que na HPSG, a representação gramatical é uma descrição integrada das diversas dimensões linguísticas. A apresentação separada da componente de representação sintáctica faz-se apenas por conveniência de exposição.

Na segunda Secção, são apresentados os traços e tipos mais importantes para a organização da informação necessária para modelar os fenómenos sintácticos abrangidos neste capítulo.

De seguida, é apresentada na terceira Secção a forma de codificar a informação categorial e de subcategorização nos itens lexicais, e na Secção 4 são apresentadas as regras sintácticas principais.

As secções seguintes debruçam-se sobre o tratamento da concordância, na quinta Secção, e das dependências de longa distância, na sexta Secção.

Inclui-se ainda na sétima Secção uma breve referência ao tratamento da ordem de palavras. O presente capítulo termina com a uma apresentação da literatura HPSG produzida até ao momento sobre a língua portuguesa, na última Secção.

Tipos e Estruturas de Traços

As entidades linguísticas encontram-se organizadas num sistema de tipos, pelo que cada objecto linguístico é de um determinado tipo. Em particular, qualquer constituinte sintáctico — seja uma palavra ou um sintagma — é uma instância do tipo *sign*.

Cada tipo está associado a um conjunto de traços, os quais são pares formados por atributos e seus valores. Os atributos definidos para o tipo *sign* são PHON (PHONOLOGY)[1] e SYNSEM (SYNTAX-SEMANTICS). O valor do atributo PHON guarda a representação ortográfica do constituinte a que diz respeito, enquanto o valor de SYNSEM contém informação categorial e semântica.

[1] Em implementações computacionais, utiliza-se o atributo ORTH (por vezes chamado STEM), uma vez que o que interessa é processar texto. Na literatura é comum ver-se o atributo PHON, mas por simplificação muitas vezes o valor representado é uma representação ortográfica. É este o uso que adoptamos aqui.

Os valores destes atributos são também objectos linguísticos, tendo por isso um tipo associado. O atributo SYNSEM toma como valores instâncias do tipo *synsem*, para as quais dois atributos são apropriados: LOCAL e NON-LOCAL. O atributo NON-LOCAL contém informação necessária para acomodar as dependências de longa distância. O atributo LOCAL, por sua vez, guarda informação acerca da sintaxe e acerca da semântica, que se encontra nos valores do atributos CAT (CATEGORY) e CONT (CONTENT), respectivamente.[2]

A informação semântica é discutida no Capítulo "Representação do Significado em HPSG". Quanto à informação sintáctica, sob CAT encontram-se os traços HEAD, que codifica informação sobre a classe de palavras, e VAL (VALENCE), que guarda informação sobre a valência nos sub-traços SUBJ (SUBJECT), SPR (SPECIFIER) e COMPS (COMPLEMENTS).

Em suma, utilizando o formato MAV, as definições de adequação para os tipos *sign*, *synsem*, *loc*, *cat* e *val* são as seguintes:

$$\begin{bmatrix} sign & \\ \text{PHON} & list(str) \\ \text{SYNSEM} & synsem \end{bmatrix} \begin{bmatrix} synsem & \\ \text{LOCAL} & loc \\ \text{NON-LOCAL} & non\text{-}loc \end{bmatrix} \begin{bmatrix} loc & \\ \text{CAT} & cat \\ \text{CONT} & cont \end{bmatrix} \begin{bmatrix} cat & \\ \text{HEAD} & head \\ \text{VAL} & val \end{bmatrix} \begin{bmatrix} val & \\ \text{SUBJ} & list(synsem) \\ \text{SPR} & list(synsem) \\ \text{COMPS} & list(synsem) \end{bmatrix}$$

Em termos de estrutura de traços, a representação gramatical de um objecto do tipo *sign* toma então o seguinte aspecto:

$$\begin{bmatrix} sign & \\ \text{PHON} & list(str) \\ \text{SYNSEM} & \begin{bmatrix} synsem & \\ \text{LOCAL} & \begin{bmatrix} loc & \\ \text{CAT} & \begin{bmatrix} cat & \\ \text{HEAD} & head \\ \text{VAL} & \begin{bmatrix} val & \\ \text{SUBJ} & list(synsem) \\ \text{SPR} & list(synsem) \\ \text{COMPS} & list(synsem) \end{bmatrix} \end{bmatrix} \\ \text{CONT} & cont \end{bmatrix} \\ \text{NON-LOCAL} & non\text{-}loc \end{bmatrix} \end{bmatrix}$$

Há mais atributos sob os traços NON-LOCAL e CONT, que são apropriados para os tipos *non-loc* e *cont*, que por conveniência de exposição se encontram omitidos aqui. Estes são apresentados na sexta Secção abaixo e no Capítulo "Representação do Significado em HPSG", respectivamente.

---

[2] Adicionalmente, no traço LOCAL encontra-se também um atributo CTXT (CONTEXT) para representar informação pragmática, que não utilizaremos nestes capítulos.

O valor do atributo PHON é uma lista de representações fonológicas (ou ortográficas).

O valor dos atributos SUBJ, SPR e COMPS é uma lista de objectos do tipo *synsem*, onde se representa informação de subcategorização para o sujeito, o especificador e os complementos. Isto significa que a informação sobre o que pode ser subcategorizado não inclui informação fonológica, uma vez que, embora PHON seja um traço de *sign*, não é um traço apropriado para o tipo *synsem*.

Os tipos pertencem a uma hierarquia de herança que se organiza debaixo da relação "é-um". Por exemplo, qualquer sintagma (tipo *phrase*) e qualquer palavra (tipo *word*) são também um signo (tipo *sign*) e portanto *sign* é um supertipo de *phrase* e *word*, sendo por isso os tipos *phrase* e *word* subtipos de *sign*.

A Figura 1 apresenta uma hierarquia para estes três tipos agora referidos e ainda para os que serão apresentados nas secções seguintes. Esta hierarquia servirá de base para a exposição do material que se segue neste capítulo.[3]

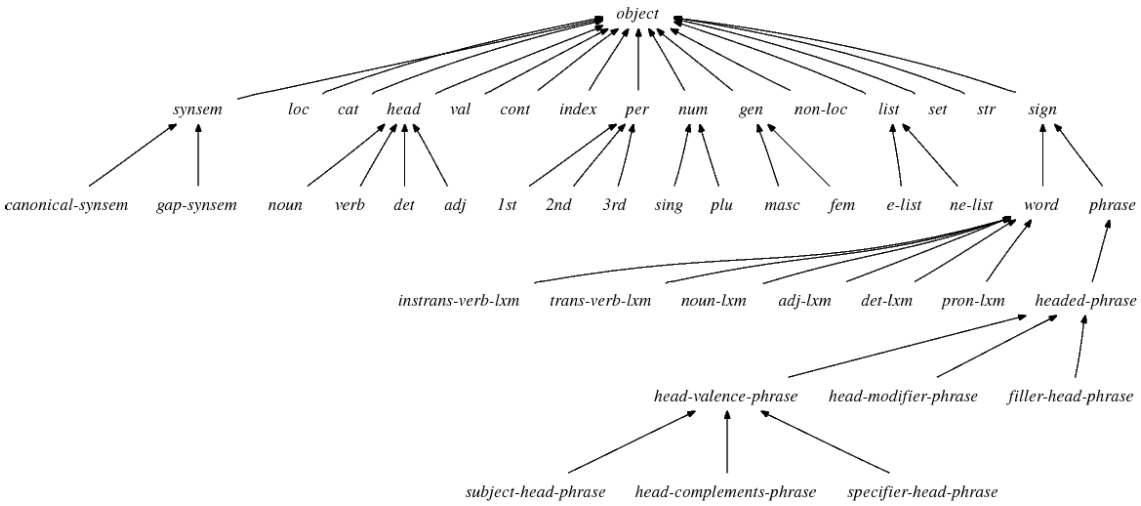


Fig. 1 Hierarquia de tipos

Um tipo herda as restrições definidas para todos os seus supertipos. Assim, todas as instâncias de *word* e *phrase*, por exemplo, têm um traço PHON e um traço SYNSEM uma vez que estes dois traços são apropriados para o seu supertipo *sign*.

Juntamente com a definição de adequação, a hierarquia de tipos ajuda a definir os valores que são possíveis para os vários traços. Por exemplo, por os tipos *noun*, *verb*, *det* (*determiner*) e *adj* (*adjective*) serem subtipos de *head*, eles constituem valores possíveis para o atributo HEAD, o qual na definição de adequação respectiva é

[3] Cabe notar que esta hierarquia se encontra simplificada relativamente a uma hierarquia mais completa para uma gramática que abranja mais fenómenos linguísticos do que aqueles que são abordados aqui.

declarado como tomando valores do tipo *head*. Em contraponto, por exemplo, o tipo *verb* não pode ser um valor válido para o atributo CAT pois não é um subtipo de *cat*.

Há atributos que tomam listas como valores. Estas podem ser vazias (tipo *e-list* (*empty-list*)) ou conter elementos (tipo *ne-list* (*non-empty-list*)). No último caso, o primeiro elemento da lista é representado no atributo FIRST, tomando o atributo REST como valor a lista que contém os restantes elementos. As listas não vazias têm portanto a seguinte definição de adequação (em que σ é o tipo dos elementos da lista):

$$\begin{bmatrix} \textit{ne-list}(\sigma) & \\ \text{FIRST} & \sigma \\ \text{REST} & \textit{list}(\sigma) \end{bmatrix}$$

Por exemplo, uma lista com "a" e "b" (com "a" à frente na lista) é uma lista *ne-list(str)* cujos elementos são do tipo *str* (*string*), representada na seguinte MAV:

$$\begin{bmatrix} \textit{ne-list(str)} & \\ \text{FIRST} & \text{“a”} \\ \text{REST} & \begin{bmatrix} \textit{ne-list(str)} & \\ \text{FIRST} & \text{“b”} \\ \text{REST} & \textit{e-list(str)} \end{bmatrix} \end{bmatrix}$$

A notação < "a", "b" > é uma simplificação notacional equivalente. A lista vazia (*e-list*) também pode ser representada por < >.

Léxico e Subcategorização

Para exemplificação de um tipo lexical, em baixo encontra-se a estrutura de traços associada a um outro dos tipos da hierarquia, o *intrans-verb-lxm* (*intransitive-verb-lexeme*). Este tipo contém as restrições apropriadas para todos os verbos intransitivos.[4]

$$\begin{bmatrix} \textit{intrans-verb-lxm} & \\ \text{SYNSEM}|\text{LOCAL}|\text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{verb} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \left\langle \begin{bmatrix} \text{LOCAL}|\text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{noun} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

[4] A notação ATRIBUTO1|ATRIBUTO2 *x* é equivalente a ATRIBUTO1 [ATRIBUTO2 *x*]. É usual omitir a referência aos tipos quando estes são recuperáveis pelo contexto. Também é usual omitir os traços que são herdados de um supertipo se não estão sujeitos a restrições adicionais (é o caso de PHON neste exemplo).

Esta MAV indica que este objecto é da categoria verbo (valor *verb* para o traço HEAD) e selecciona como sujeito um elemento com o valor *noun* para o seu atributo HEAD.[5] Para além disso, o sujeito tem de estar saturado (visto ter listas vazias para os traços de valência) isto é, tem de ser um SN completo.

A entrada lexical para um verbo intransitivo como "correr" resulta de se associar este item lexical a esse tipo e de se preencher o valor do atributo PHON com a seguinte lista singular de *strings* < "correr" >:

$$\begin{bmatrix} \textit{intrans-verb-lxm} \\ \text{PHON} \left\langle \text{“correr”} \right\rangle \end{bmatrix}$$

Como é ilustrado neste exemplo, a informação de subcategorização associada a um item lexical encontra-se codificada no valor do traço VAL.

O traço SUBJ recolhe informação sobre o sujeito. Por sua vez, os atributos SPR e COMPS contêm informação sobre o especificador e os complementos de um determinado item, respectivamente.

Por exemplo, os nomes contáveis encontram-se no léxico com o requisito de se terem de combinar com um determinante, o que resultará na seguinte estrutura de traços para o seu tipo lexical:

$$\begin{bmatrix} \textit{noun-lxm} & \\ \text{SYNSEM}|\text{LOCAL}|\text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{noun} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{SPR} & \left\langle \begin{bmatrix} \text{LOCAL}|\text{CAT}|\text{HEAD} & \textit{det} \end{bmatrix} \right\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

A informação sobre complementos é dada de forma análoga. Os verbos transitivos seleccionam um SN sujeito e um SN complemento, e o seu tipo lexical é o seguinte:

[5] Tanto os nomes como os pronomes pessoais estão definidos no léxico com o valor *noun* para este atributo.

$$\begin{bmatrix} \textit{trans-verb-lxm} & \\ \text{SYNSEM}|\text{LOCAL}|\text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{verb} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \left\langle \begin{bmatrix} \text{LOCAL}|\text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{noun} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \left\langle \begin{bmatrix} \text{LOCAL}|\text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{noun} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Os itens que não seleccionam dependentes têm os atributos dentro do valor de VAL todos com listas vazias. Para pronomes pessoais, por exemplo, o seu tipo lexical consiste na seguinte estrutura de traços:

$$\begin{bmatrix} \textit{pron-lxm} & \\ \text{SYNSEM}|\text{LOCAL}|\text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{noun} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Constituência e Funções Gramaticais

Qualquer construção sintáctica é um objecto linguístico que é um subtipo de *phrase*. Este tipo *phrase* herda os traços do seu supertipo *sign* e como tal possui os atributos PHON e SYNSEM. Adicionalmente, possui o atributo DAUGHTERS, que representa os constituintes sintácticos imediatos da construção em causa. Estes constituintes imediatos podem ser palavras ou outros sintagmas, pelo que o valor deste atributo DAUGHTERS é uma lista de objectos do tipo *sign*.

Na definição de adequação de *phrase*, o valor de PHON do nó mãe é a concatenação do valor de PHON dos nós filhos. Sendo a concatenação de listas representada pelo símbolo ⊕,[6] a estrutura de traços de *phrase* vem:

$$\begin{bmatrix} \textit{phrase} & \\ \text{PHON} & \boxed{1} \oplus \cdots \oplus \boxed{n} \\ \text{DAUGHTERS} & \textit{ne-list} \left\langle \begin{bmatrix} \textit{sign} & \\ \text{PHON} & \boxed{1} \end{bmatrix}, \cdots, \begin{bmatrix} \textit{sign} & \\ \text{PHON} & \boxed{n} \end{bmatrix} \right\rangle \end{bmatrix}$$

[6] Assumimos a definição habitual para a concatenação de listas de elementos de tipo σ:

- $\textit{e-list}(\sigma) \oplus \boxed{1}\ \textit{list}(\sigma) = \boxed{1}$
- $\begin{bmatrix} \textit{ne-list}(\sigma) & \\ \text{FIRST} & \boxed{1}\ \sigma \\ \text{REST} & \boxed{2}\ \textit{list}(\sigma) \end{bmatrix} \oplus \boxed{3}\ \textit{list}(\sigma) = \begin{bmatrix} \textit{ne-list}(\sigma) & \\ \text{FIRST} & \boxed{1} \\ \text{REST} & \boxed{2} \oplus \boxed{3} \end{bmatrix}$

A ordem dos elementos na lista de DAUGHTERS corresponde à ordem de palavras (e.g. o primeiro elemento corresponde ao filho mais à esquerda, etc).

Cabe notar que etiquetas $\boxed{i}$ idênticas denotam identidade entre valores. Neste caso, por exemplo, a identidade entre o primeiro operando da operação de concatenação que constrói o valor de PHON do nó mãe e o valor de PHON do filho mais à esquerda é representada pelas duas etiquetas $\boxed{1}$.

Um dos subtipos de *phrase* é o *headed-phrase* (vd. Figura 1). Na definição deste último, é introduzido o atributo HEAD-DTR, cujo valor corresponde à caracterização do núcleo do constituinte com esse tipo. O valor desse atributo HEAD-DTR é identificado com um dos elementos da lista em DAUGHTERS, consoante a ordem de palavras em construções específicas (e.g. com o primeiro elemento dessa lista se se tratar de uma construção em que o núcleo ocorre no início, ou seja à esquerda).

Um dos princípios sintácticos nucleares é o Princípio do Traço do Núcleo (*Head Feature Principle*), que capta o facto de os sintagmas e os seus núcleos partilharem características categoriais. Este princípio estabelece que em qualquer constituinte que possui núcleo, i.e. que seja do tipo *headed-phrase*, o valor do atributo HEAD é idêntico ao valor do atributo com esse mesmo nome no nó filho que corresponde ao núcleo. Por exemplo, numa construção para realizar ou projectar os complementos à direita de um núcleo, o valor deste atributo é idêntico ao do filho mais à esquerda.

Este princípio corresponde à seguinte restrição no tipo *headed-phrase*, que é supertipo de todas as construções em que um dos filhos é um núcleo:

Princípio do Traço do Núcleo

$$\begin{bmatrix} \textit{headed-phrase} \\ \text{SYNSEM}|\text{LOCAL}|\text{CAT}|\text{HEAD } \boxed{1} \\ \text{HEAD-DTR} \begin{bmatrix} \textit{sign} \\ \text{SYNSEM}|\text{LOCAL}|\text{CAT}|\text{HEAD } \boxed{1} \end{bmatrix} \end{bmatrix}$$

A identidade entre os dois atributos HEAD é representada pelas duas etiquetas $\boxed{1}$. Todas as construções que possuem um núcleo herdam de *headed-phrase*. Como tal, também herdam esta restrição.

Desta forma, como veremos de seguida, todo o aparato necessário para codificar regularidades relativas à estrutura sintáctica de constituência consiste na definição apropriada da hierarquia de subtipos de *phrase* e nas definições de adequação para cada um deles.

Sujeitos

A construção em que o sujeito surge projectado à esquerda do SV é de tipo *subject-head-phrase*, que é um subtipo de *headed-phrase* com as seguintes restrições adicionais:

$$\begin{bmatrix} \textit{subject-head-phrase} \\ \text{SYNSEM}|\text{LOCAL}|\text{CAT}|\text{VAL} \begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{SPR} & \boxed{1} \\ \text{COMPS} & \boxed{2} \end{bmatrix} \\ \text{HEAD-DTR } \boxed{3} \\ \text{NON-HEAD-DTR } \boxed{4} \\ \text{DAUGHTERS} \left\langle \boxed{4}\begin{bmatrix}\text{SYNSEM } \boxed{5}\end{bmatrix}, \boxed{3}\begin{bmatrix}\text{SYNSEM}|\text{LOCAL}|\text{CAT}|\text{VAL} \begin{bmatrix} \text{SUBJ} & \langle \boxed{5} \rangle \\ \text{SPR} & \boxed{1} \\ \text{COMPS} & \boxed{2}\langle\rangle \end{bmatrix}\end{bmatrix} \right\rangle \end{bmatrix}$$

Esta definição de adequação estabelece que o núcleo da construção é o nó filho da direita uma vez que através de $\boxed{3}$ o atributo HEAD-DTR está unificado com o último elemento de DAUGHTERS. O valor de SYNSEM do nó filho da esquerda (o Sujeito) é identificado através de $\boxed{5}$ com o elemento que está no atributo SUBJ do núcleo. Desta forma as propriedades do objecto linguístico que instanciar o Sujeito têm de ser compatíveis com as propriedades impostas pelo núcleo relativamente a essa posição subcategorizada por si. O nó mãe tem a lista vazia como valor do atributo SUBJ e os restantes traços de valência são idênticos aos do seu núcleo.[7]

As regras sintácticas utilizadas numa construção são os subtipos de *phrase* que não têm subtipos. Nos exemplos vistos até agora, isto só se aplica ao tipo *subject-head-phrase*. Este é um subtipo de *headed-phrase* e de *phrase* pelo que herda as restrições impostas nestes últimos. Deste modo, todas as restrições sobre *subject-head-phrase*, incluindo as herdadas dos seus supertipos, são as seguintes:[8]

$$\begin{bmatrix} \textit{subject-head-phrase} \\ \text{PHON } \boxed{1} \oplus \boxed{2} \\ \text{SYNSEM}|\text{LOCAL}|\text{CAT} \begin{bmatrix} \text{HEAD} & \boxed{3} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{SPR} & \boxed{4} \\ \text{COMPS} & \boxed{5} \end{bmatrix} \end{bmatrix} \\ \text{HEAD-DTR } \boxed{6} \\ \text{NON-HEAD-DTR } \boxed{7} \\ \text{DAUGHTERS} \left\langle \boxed{7}\begin{bmatrix}\text{PHON } \boxed{1} \\ \text{SYNSEM } \boxed{8}\end{bmatrix}, \boxed{6}\begin{bmatrix}\text{PHON } \boxed{2} \\ \text{SYNSEM}|\text{LOCAL}|\text{CAT} \begin{bmatrix} \text{HEAD} & \boxed{3} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle \boxed{8} \rangle \\ \text{SPR} & \boxed{4} \\ \text{COMPS} & \boxed{5}\langle\rangle \end{bmatrix} \end{bmatrix}\end{bmatrix} \right\rangle \end{bmatrix}$$

---

[7] A restrição de que o atributo COMPS é a lista vazia é explicada maos à frente na Subsecção Complementos.

[8] Por agora, são omitidas aqui as restrições definidas no tipo *head-valence-phrase*, que também é um supertipo de *subject-head-phrase* e que são apresentadas mais à frente na Secção Dependências de Longa Distância.

Tendo sido apresentadas as estruturas de traços lexicais para verbos intransitivos e para pronomes assim como a estrutura de traços para as construções intransitivas, apresenta-se na Figura 2 o resultado da sua unificação e da instanciação com as entradas lexicais para o exemplo "Eles correm".

$$\begin{bmatrix}
\textit{subject-head-phrase} \\
\text{PHON } \boxed{1} \oplus \boxed{2} = \langle \text{“Eles”, “correm”} \rangle \\
\text{SYNSEM|LOCAL|CAT} \begin{bmatrix} \text{HEAD} & \boxed{3}\ \textit{verb} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{SPR} & \boxed{4}\langle\rangle \\ \text{COMPS} & \boxed{5}\langle\rangle \end{bmatrix} \end{bmatrix} \\
\text{HEAD-DTR } \boxed{6} \\
\text{NON-HEAD-DTR } \boxed{7} \\
\text{DAUGHTERS} \left\langle \boxed{7} \begin{bmatrix} \textit{pron-lxm} \\ \text{PHON } \boxed{1}\langle \text{“Eles”} \rangle \\ \text{SYNSEM } \boxed{8} \begin{bmatrix} \text{LOCAL|CAT} \begin{bmatrix} \text{HEAD} & \textit{noun} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}, \boxed{6} \begin{bmatrix} \textit{intrans-verb-lxm} \\ \text{PHON } \boxed{2}\langle \text{“correm”} \rangle \\ \text{SYNSEM|LOCAL|CAT} \begin{bmatrix} \text{HEAD} & \boxed{3} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle \boxed{8} \rangle \\ \text{SPR} & \boxed{4} \\ \text{COMPS} & \boxed{5} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle
\end{bmatrix}$$

Fig. 2 Análise para a frase "Eles correm": representação em MAV.

As relações de dominância imediata e de precedência linear encontram-se codificadas no valor do atributo DAUGHTERS. Pode dar-se a estas relações uma representação alternativa em termos diagramáticos, em que a representação da Figura 2 recebe a representação equivalente em árvore na Figura 3.

$$\begin{bmatrix} \textit{subject-head-phrase} \\ \text{PHON } \boxed{1} \oplus \boxed{2} = \langle \text{“Eles”, “correm”} \rangle \\ \text{SYNSEM|LOCAL|CAT} \begin{bmatrix} \text{HEAD} & \boxed{3}\ \textit{verb} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{SPR} & \boxed{4}\ \langle\rangle \\ \text{COMPS} & \boxed{5}\ \langle\rangle \end{bmatrix} \end{bmatrix} \\ \text{HEAD-DTR } \boxed{6} \\ \text{NON-HEAD-DTR } \boxed{7} \\ \text{DAUGHTERS } \langle \boxed{7}, \boxed{6} \rangle \end{bmatrix}$$

$$\boxed{7} \begin{bmatrix} \textit{pron-lxm} \\ \text{PHON } \boxed{1} \langle \text{“Eles”} \rangle \\ \text{SYNSEM } \boxed{8} \begin{bmatrix} \text{LOCAL|CAT} \begin{bmatrix} \text{HEAD} & \textit{noun} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \qquad \boxed{6} \begin{bmatrix} \textit{intrans-verb-lxm} \\ \text{PHON } \boxed{2} \langle \text{“correm”} \rangle \\ \text{SYNSEM|LOCAL|CAT} \begin{bmatrix} \text{HEAD} & \boxed{3} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle \boxed{8} \rangle \\ \text{SPR} & \boxed{4} \\ \text{COMPS} & \boxed{5} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Fig. 3 Análise para a frase "Eles correm": representação mista em MAV e diagrama de árvore de constituintes.

Complementos

As expressões em que os complementos do núcleo se encontram projectados são caracterizadas por um outro subtipo de *headed-phrase*, o tipo *head-complements-phrase*. Este último possui as seguintes restrições adicionais:

$$\begin{bmatrix} \textit{head-complements-phrase} \\ \text{SYNSEM|LOCAL|CAT|VAL} \begin{bmatrix} \text{SUBJ} & \boxed{1} \\ \text{SPR} & \boxed{2} \\ \text{COMPS} & \langle\rangle \end{bmatrix} \\ \text{HEAD-DTR } \boxed{3} \begin{bmatrix} \text{SYNSEM|LOCAL|CAT|VAL} \begin{bmatrix} \text{SUBJ} & \boxed{1} \\ \text{SPR} & \boxed{2} \\ \text{COMPS} & \textit{ne-list} \langle \boxed{A_1}, \cdots, \boxed{A_n} \rangle \end{bmatrix} \end{bmatrix} \\ \text{NON-HEAD-DTRS } \boxed{4} \langle \begin{bmatrix} \text{SYNSEM } \boxed{A_1} \end{bmatrix}, \cdots, \begin{bmatrix} \text{SYNSEM } \boxed{A_n} \end{bmatrix} \rangle \\ \text{DAUGHTERS } \langle \boxed{3} \rangle \oplus \boxed{4} \end{bmatrix}$$

Esta regra projecta todos os complementos de um núcleo. Daí que o nó mãe apresente a lista vazia para o valor do atributo COMPS. O núcleo é o filho mais à esquerda desta construção. Os restantes filhos têm o seu SYNSEM unificado com os elementos do COMPS do núcleo, pela mesma ordem em que lá ocorrem.

As restantes listas de valência, nomeadamente SUBJ e SPR, são idênticas às listas de valência do núcleo.

A restrição nesta regra segundo a qual o COMPS do núcleo é uma lista não vazia (*ne-list*) impede que a regra se aplique sobre si própria, produzindo um ciclo infinito.

Cabe assinalar que a restrição no *subject-head-phrase* segundo a qual o COMPS do núcleo dessa construção é vazio garante que os complementos são projectados numa posição sintáctica inferior àquela em que o sujeito ocorre. Isto é, a construção *subject-head-phrase* pode aplicar-se tomando como HEAD-DTR uma construção do tipo *head-complements-phrase*, mas a HEAD-DTR de uma construção *head-complements-phrase* não pode ser uma instância de *subject-head-phrase*.

Note-se que a regra *head-complements-phrase*, tal como foi apresentada aqui, não permite a não realização de complementos opcionais, nem contempla as diferentes possibilidades de realização dos complementos que diferem quanto à ordem por que são realizados. Estas questões não são abordadas neste capítulo introdutório.

Especificadores

Como decorre do que foi apresentado anteriormente, os núcleos seleccionam o seu especificador através do atributo SPR. Por sua vez, os especificadores também têm visibilidade sobre o núcleo, nomeadamente através de um outro traço, o traço SPEC, que se encontra sob HEAD. As restrições que os especificadores impõem sobre o seu núcleo, através deste traço, são sobretudo de natureza semântica e são apresentadas no Capítulo "Representação do Significado em HPSG". Neste capítulo, vamos considerar que os determinantes se encontram na posição de especificador[9] pelo que se assume que o atributo SPEC é apropriado para o tipo *det*, o qual recebe a seguinte definição:

$$\begin{bmatrix} det & \\ \text{SPEC} & synsem \end{bmatrix}$$

Por via do valor do seu atributo SPEC, o determinante deescreve aa propriedades sintácticas do núcleo com que se pode combinar.

As restrições sintácticas nas entradas lexicais dos determinantes são as seguintes:

$$\begin{bmatrix} det\text{-}lxm & \\ \text{SYNSEM}|\text{LOCAL}|\text{CAT} & \begin{bmatrix} \text{HEAD} & det \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Por sua vez, a construção em que os especificadores se encontram projectados é do tipo *specifier-head-phrase* e tem a definição:

[9] Esta é a análise adoptada em (Pollard e Sag, 1994). Numa análise alternativa, a chamada hipótese DP (*Determiner Phrase*), os determinantes são o núcleo e o nó irmão é seleccionado como seu complemento.

$$\begin{bmatrix} \textit{specifier-head-phrase} & \\ \text{SYNSEM}|\text{LOCAL}|\text{CAT}|\text{VAL} & \begin{bmatrix} \text{SUBJ} & \boxed{1} \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \boxed{2}\,\langle\rangle \end{bmatrix} \\ \text{HEAD-DTR}\boxed{3} & \begin{bmatrix} \text{SYNSEM}\ \boxed{4} & \begin{bmatrix} \text{LOCAL}|\text{CAT}|\text{VAL} & \begin{bmatrix} \text{SUBJ} & \boxed{1} \\ \text{SPR} & \langle\boxed{5}\rangle \\ \text{COMPS} & \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{NON-HEAD-DTR}\ \boxed{6} & \begin{bmatrix} \text{SYNSEM}\ \boxed{5} & \begin{bmatrix} \text{LOCAL}|\text{CAT}|\text{HEAD}|\text{SPEC}\ \boxed{4} \end{bmatrix} \end{bmatrix} \\ \text{DAUGHTERS} & \langle\boxed{6}\,,\,\boxed{3}\rangle \end{bmatrix}$$

Os atributos SUBJ e COMPS do núcleo são percolados para o nó mãe. Esta construção também requer que o traço COMPS do núcleo tenha como valor a lista vazia, obrigando eventuais complementos do nome a serem projectados numa posição sintáctica mais baixa do que aquela em que os determinantes se realizam.

Para ilustrar a unificação das três regras sintácticas discutidas até ao momento, nomeadamente *subject-head-phrase*, *head-complements-phrase* e *specifier-head-phrase*, a Figura 4 apresenta a representação sintáctica para a frase exemplo "Eles viram os livros".

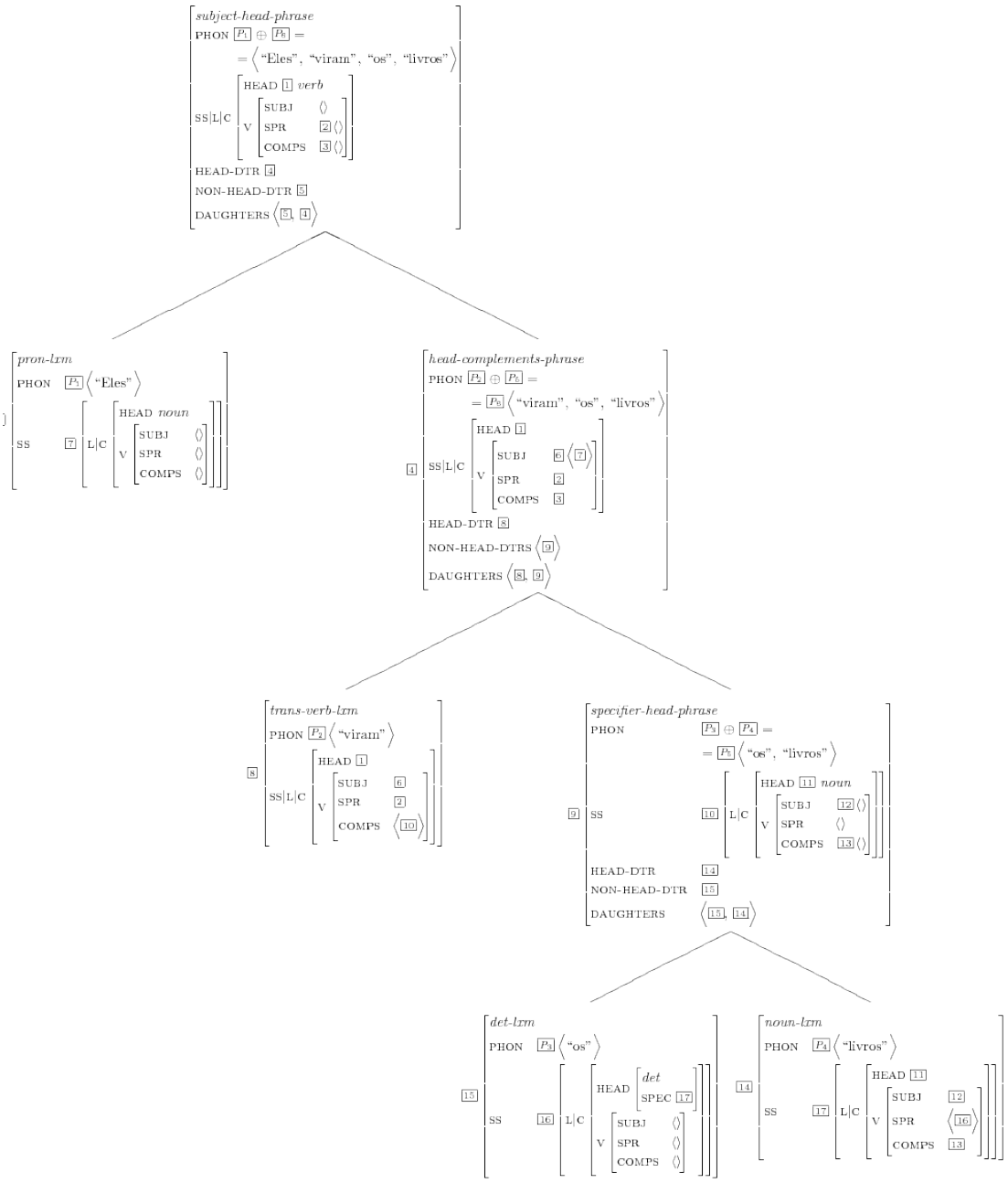


Fig. 4 Análise sintáctica da frase "Eles viram os livros". Para economia de espaço, SS abrevia SYNSEM, L abrevia LOCAL, C abrevia CAT e V abrevia VAL.

Modificadores

As estruturas de modificação vão ser apresentadas tomando os adjectivos como pretexto. Note-se porém que os dispositivos gramaticais responsáveis por dar conta das construções com modificadores são independentes da categoria sintáctica do modificador e do núcleo presentes, e que as restrições relevantes sobre os modificadores de outras categorias são semelhantes às que serão apresentadas para os adjectivos.

Os modificadores possuem um atributo MOD (MODIFIED). Este atributo encontra-se sob o traço HEAD, tem um objecto de tipo *synsem* como seu valor, e descreve as propriedades sintácticas do núcleo que uma dada categoria pode modificar.

Para os adjectivos, o atributo MOD é introduzido no tipo *adj*, o qual é o valor do atributo HEAD dos adjectivos, e constitui um subtipo de *head* :

$$\begin{bmatrix} adj & \\ \text{MOD} & synsem \end{bmatrix}$$

Uma vez que os adjectivos modificam projecções nominais que não são SNs completos, as restrições para as entradas lexicais dos adjectivos sem complementos são as seguintes:[10]

$$\begin{bmatrix} adj\text{-}lxm & \\ \text{SYNSEM}|\text{LOCAL}|\text{CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} adj & \\ \text{MOD}|\text{LOCAL}|\text{CAT} & \begin{bmatrix} \text{HEAD} & noun \\ \text{VAL}|\text{SPR} & ne\text{-}list \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Os adjectivos seleccionam constituintes nominais com a posição de especificador por projectar, visto o valor do SPR do constituinte seleccionado ser uma lista não vazia (*ne-list*). Desta forma, não podem por exemplo modificar pronomes pessoais, que vêm no léxico com a lista vazia para o atributo SPR.

As construções de modificação são do tipo *head-modifier-phrase*, um subtipo de *headed-phrase*. A par dos traços e restrições que herda deste último, o tipo *head-modifier-phrase* integra restrições adicionais, ilustradas na seguinte definição de adequação para uma construção que permite projectar um adjunto à direita do seu núcleo:

$$\begin{bmatrix} head\text{-}modifier\text{-}phrase & \\ \text{SYNSEM}|\text{LOCAL}|\text{CAT}|\text{VAL} & \boxed{1} \\ \text{HEAD-DTR} & \boxed{2}\begin{bmatrix} \text{SYNSEM} & \boxed{3}\begin{bmatrix} \text{LOCAL}|\text{CAT}|\text{VAL} & \boxed{1} \end{bmatrix} \end{bmatrix} \\ \text{NON-HEAD-DTR} & \boxed{4}\begin{bmatrix} \text{SYNSEM}|\text{LOCAL}|\text{CAT}|\text{HEAD}|\text{MOD} & \boxed{3} \end{bmatrix} \\ \text{DAUGHTERS} & \left\langle \boxed{2}, \boxed{4} \right\rangle \end{bmatrix}$$

Apresentamos a análise para o SN "o carro azul pequeno" na Figura 5. Como se pode ver pelo exemplo, este tipo *head-modifier-phrase* permite recursão.

[10] Por simplificação, assumimos que os adjectivos não têm sujeito. Isto pode ter de ser alterado para se abranger adjectivos em posição predicativa.

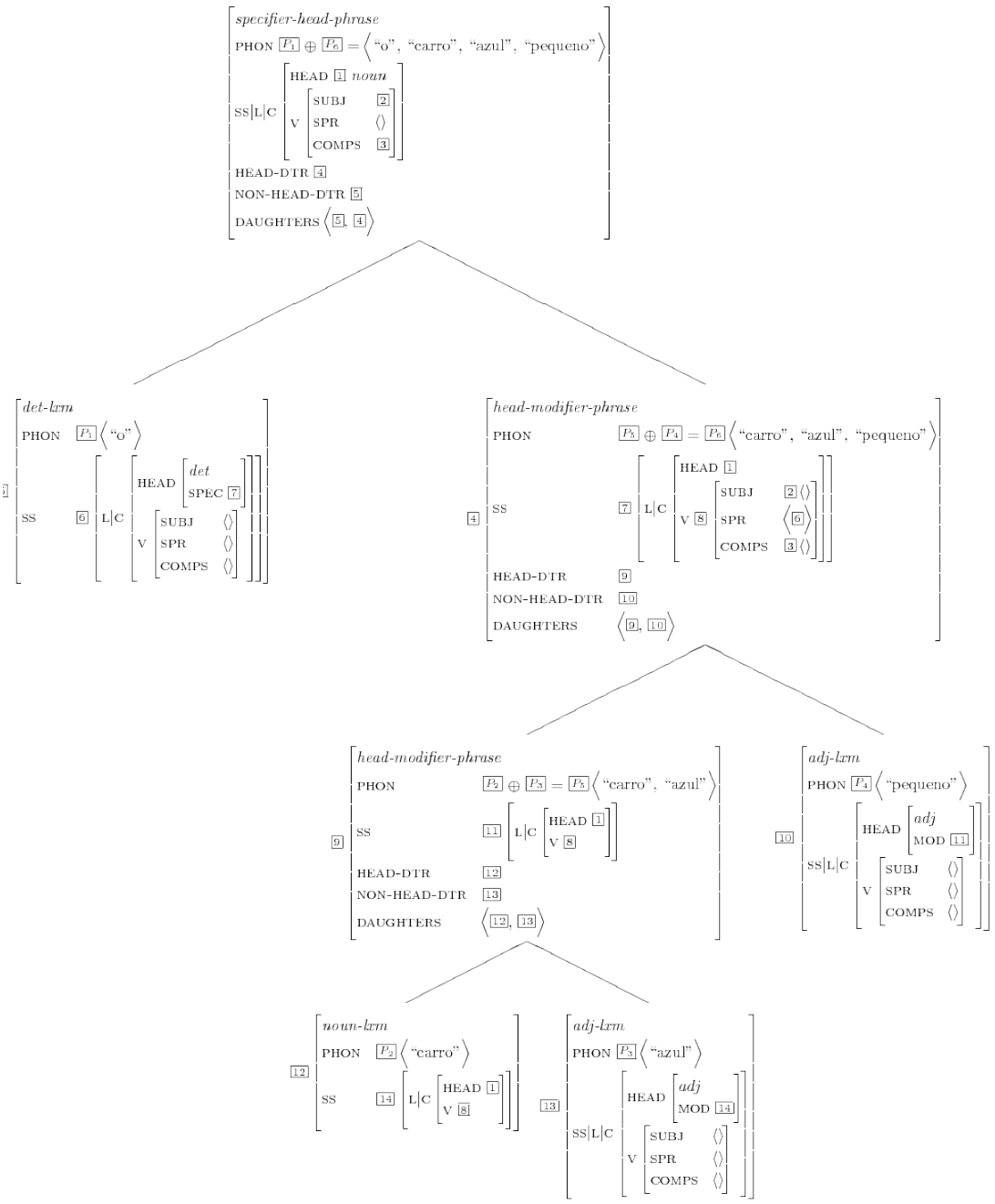


Fig. 5 Análise para o SN "o carro azul pequeno". Por economia de espaço, SS abrevia SYNSEM, L abrevia LOCAL, C abrevia CAT e V abrevia VAL.

Concordância

O estabelecimento de concordância morfo-sintáctica resulta da unificação entre atributos que representam a informação relevante.

Essa informação (sobre a flexão em pessoa, género e número) é colocada nos atributos PERSON, NUMBER, GENDER, os quais são agrupados no tipo INDEX. Este tipo é o valor do atributo AGR (AGREEMENT), que se encontra sob o atributo HEAD.

$$\begin{bmatrix} noun & \\ \text{AGR} & index \end{bmatrix} \qquad \begin{bmatrix} index & \\ \text{PERSON} & per \\ \text{NUMBER} & num \\ \text{GENDER} & gen \end{bmatrix}$$

Na mini-gramática que está a ser apresentada ao longo deste capítulo só precisamos do traço AGR para nomes e pronomes, por isso definimos o traço AGR como apropriado para *noun*.

Os subtipos dos três tipos *per*, *num* e *gen* apresentados na hierarquia na Figura 1 são os valores possíveis para estes atributos.

A informação sobre o valor destes três atributos vem codificada nas entradas lexicais. Por exemplo, as estruturas de traços associadas aos nomes contêm informação de género. Eis um exemplo de entrada lexical, para o nome "lápis":

$$\begin{bmatrix} noun\text{-}lxm & \\ \text{PHON} & \langle \text{“lápis”} \rangle \\ \text{SYNSEM}|\text{LOCAL}|\text{CAT}|\text{HEAD} & \begin{bmatrix} noun & \\ \text{AGR} & \begin{bmatrix} \text{PERSON} & 3rd \\ \text{NUMBER} & num \\ \text{GENDER} & masc \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Este exemplo tem o valor do traço NUMBER sub-especificado, uma vez que o nome em questão não apresenta morfologia diferente em virtude de diferente número gramatical. De forma análoga, os nomes comuns de dois, como "pianista", têm o traço GENDER sub-especificado, apresentando este o valor mais geral que lhe é apropriado, *gen*.

Para simplificar a exposição, neste capítulo de introdução não se apresentam as regras lexicais para a flexão, pelo que se assume um léxico com todas as formas flexionadas.

Através do seu traço MOD, os adjectivos restringem os atributos apropriados sob o traço AGR da categoria que modificam. Os determinantes, por sua vez, restringem o AGR do seu nó irmão através do atributo SPEC. Tal é ilustrado, respectivamente, com os exemplos do adjectivo "amarelos" e do determinante "bastantes":

$$\begin{bmatrix} \textit{adj-lxm} \\ \text{PHON} \left\langle \text{“amarelos”} \right\rangle \\ \text{SYNSEM}|\text{LOCAL}|\text{CAT}|\text{HEAD} \begin{bmatrix} \textit{adj} \\ \text{MOD}|\text{LOCAL}|\text{CAT}|\text{HEAD} \begin{bmatrix} \textit{noun} \\ \text{AGR} \begin{bmatrix} \text{NUMBER} & \textit{plu} \\ \text{GENDER} & \textit{masc} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

$$\begin{bmatrix} \textit{det-lxm} \\ \text{PHON} \left\langle \text{“bastantes”} \right\rangle \\ \text{SYNSEM}|\text{LOCAL}|\text{CAT}|\text{HEAD} \begin{bmatrix} \textit{det} \\ \text{SPEC}|\text{LOCAL}|\text{CAT}|\text{HEAD}|\text{AGR}|\text{NUMBER} & \textit{plu} \end{bmatrix} \end{bmatrix}$$

Para ilustrar uma estrutura em que é estabelecida uma relação de concordância, no exemplo a seguir apresenta-se a representação sintáctica para o sintagma "lápis amarelos":

$$\begin{bmatrix} \textit{head-modifier-phrase} \\ \text{PHON } \boxed{P_1} \oplus \boxed{P_2} = \left\langle \text{“lápis”, “amarelos”} \right\rangle \\ \text{SYNSEM}|\text{LOCAL}|\text{CAT} \begin{bmatrix} \text{HEAD} & \boxed{1} \begin{bmatrix} \textit{noun} \\ \text{AGR} \begin{bmatrix} \text{PERSON} & \textit{3rd} \\ \text{NUMBER} & \textit{plu} \\ \text{GENDER} & \textit{masc} \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \boxed{2} \begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{SPR} & \left\langle \begin{bmatrix} \text{LOCAL}|\text{CAT}|\text{HEAD} & \textit{det} \end{bmatrix} \right\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \\ \text{HEAD-DTR } \boxed{3} \\ \text{NON-HEAD-DTR } \boxed{4} \\ \text{DAUGHTERS} \left\langle \boxed{3} \begin{bmatrix} \textit{noun-lxm} \\ \text{PHON} & \boxed{P_1} \left\langle \text{“lápis”} \right\rangle \\ \text{SYNSEM} & \boxed{5} \begin{bmatrix} \text{LOCAL}|\text{CAT} \begin{bmatrix} \text{HEAD} & \boxed{1} \\ \text{VAL} & \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix}, \boxed{4} \begin{bmatrix} \textit{adj-lxm} \\ \text{PHON } \boxed{P_2} \left\langle \text{“amarelos”} \right\rangle \\ \text{SYNSEM}|\text{LOCAL}|\text{CAT} \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{adj} \\ \text{MOD} & \boxed{5} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

Uma vez que o *synsem* do nome é unificado com o valor do traço MOD do adjectivo e o atributo HEAD do nome é unificado com o mesmo traço do nó mãe, este último vai ter um AGR que é a unificação das duas estruturas seguintes:

$$\begin{bmatrix} \text{PERSON} & \textit{3rd} \\ \text{NUMBER} & \textit{num} \\ \text{GENDER} & \textit{masc} \end{bmatrix} \sqcap \begin{bmatrix} \text{PERSON} & \textit{per} \\ \text{NUMBER} & \textit{plu} \\ \text{GENDER} & \textit{masc} \end{bmatrix} = \begin{bmatrix} \text{PERSON} & \textit{3rd} \\ \text{NUMBER} & \textit{plu} \\ \text{GENDER} & \textit{masc} \end{bmatrix}$$

A primeira corresponde ao AGR do nome, com o valor de número sub-especificado e portanto com o valor mais geral que lhe é apropriado, *num*; a

segunda corresponde ao AGR seleccionado pelo adjectivo, com o valor de pessoa sub-especificado.

A unificação opera de forma recursiva e todos os subtraços são também unificados — assim, a unificação de valores de SYNSEMs implica a unificação do traço LOCAL|CAT|HEAD|AGR. Para além disso, a unificação de dois tipos é o tipo mais específico dos dois: por exemplo, no caso do valor de número, o resultado da unificação de *num* e *plu* é *plu* porque *plu* é um subtipo de *num*.

Cabe notar que o sintagma resultante não se poderia combinar com a forma do artigo definido "as" por esta requer como nó irmão um constituinte com o valor *fem* para o traço NUMBER: *fem* e *masc* são incompatíveis, porque nenhum é um subtipo do outro nem têm nenhum subtipo comum na hierarquia de tipos assumida. Este sintagma pode contudo combinar-se com a forma "bastantes", apresentada acima, pois os valores de pessoa, número e género são unificáveis.

A concordância sujeito-verbo, por sua vez, é modelada através de restrições nas formas verbais relativas ao traço AGR do sujeito. Eis um exemplo para uma forma de terceira pessoa singular:

$$\begin{bmatrix} \textit{trans-verb-lxm} \\ \text{PHON} \quad \langle \text{“viu”} \rangle \\ \text{SYNSEM|LOCAL|CAT|VAL|SUBJ} \quad \left\langle \begin{bmatrix} \text{LOCAL|CAT|HEAD|AGR} \begin{bmatrix} \text{PERSON} & \textit{3rd} \\ \text{NUMBER} & \textit{sing} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

Dependências de Longa Distância

Em alguns tipos de construções, como as orações interrogativas ou relativas, por exemplo, um constituinte pode não ocorrer na posição canónica em que ocorreria numa construção declarativa análoga não marcada. Apesar de se conformar às restantes dependências sintácticas e semânticas a que se conformaria caso ocorresse nessa posição canónica, esse constituinte pode ocorrer numa posição irrestritamente distante.

Para ilustrar a análise deste tipo de dependências, ditas de longa-distância, a seguir apresenta-se a análise de uma construção de topicalização, seguindo a abordagem proposta em (Bouma *et al.* 2001). A frase que vai servir de exemplo de trabalho será "Isso, eles viram".

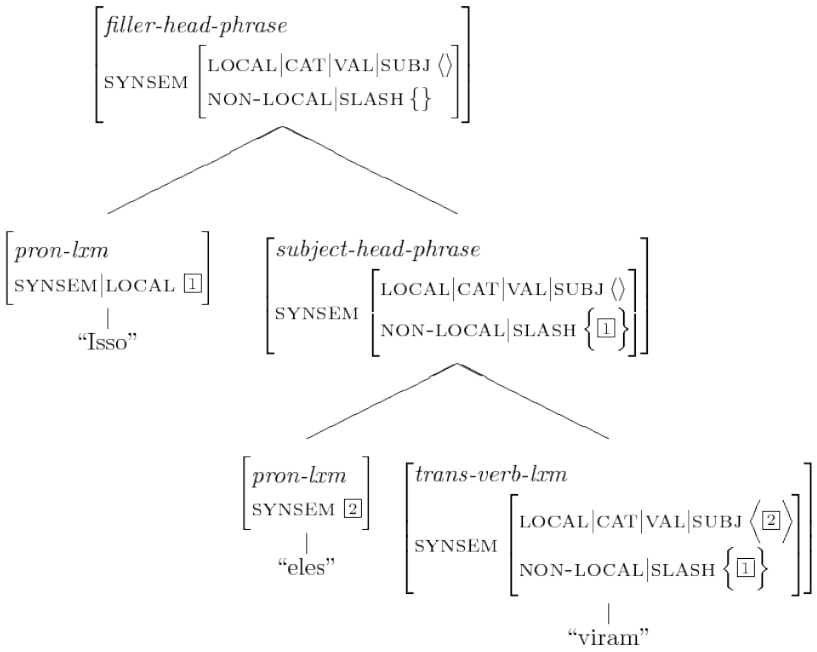
A metáfora subjacente para se entender os pormenores da análise da topicalização apresentados a seguir é a de que, relativamente a um dado predicador, um dos seus argumentos não ocorre localmente na posição canónica: em vez disso, a informação quanto às restrições de subcategorização desse argumento é "posta a circular" pela árvore de constituência até encontrar um nó em que um dos seus nós filhos se conforme a essas restrições. Esse nó filho é o argumento topicalizado.

Esta abordagem envolve assim três componentes principais: (i) um tipo de construção em que o nó filho da esquerda é um constituinte topicalizado; (ii) um dispositivo de percolação da informação de subcategorização acerca do argumento topicalizado; (iii) um mecanismo que, relativamente a um predicador, remove a informação sobre o argumento topicalizado da sua lista de subcategorização e a "coloca a circular" ao longo da árvore de constituência.

Estes três componentes têm em comum integrarem um traço que guarda e percola a informação sobre o constituinte subcategorizado. Esse traço é SLASH, que se encontra sob o atributo NON-LOCAL e que toma como valores conjuntos de objectos do tipo *loc*:[11]

$$\begin{bmatrix} \textit{non-loc} & \\ \text{SLASH} & \textit{set(loc)} \end{bmatrix}$$

A seguinte árvore ilustra o papel do atributo SLASH na percolação da informação de subcategorização entre o predicador e a periferia esquerda da frase, onde se encontra o constituinte topicalizado:



Esta análise é melhor apreendida com a descrição de cada um dos três componentes principais das dependências de longa distância mencionados acima.

[11] Na análise de outras construções com dependências de longa distância, faz-se uso de mais dois outros traços sob NON-LOCAL, os traços REL e QUE, para dependências de longa distância que envolvem elementos-QU. Por limitação de espaço, estes traços não são apresentados neste capítulo introdutório.

Tópico

A construção responsável por projectar um constituinte na periferia esquerda da frase é do tipo *filler-head-phrase*, com a seguinte definição:[12]

$$\begin{bmatrix} \textit{filler-head-phrase} & \\ \text{SYNSEM} & \begin{bmatrix} \text{LOCAL}|\text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{verb} \\ \text{VAL} & \boxed{1}\begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \\ \text{NON-LOCAL}|\text{SLASH} & \boxed{2} \end{bmatrix} \\ \text{HEAD-DTR} & \boxed{3}\begin{bmatrix} \text{SYNSEM} & \begin{bmatrix} \text{LOCAL}|\text{CAT}|\text{VAL} & \boxed{1} \\ \text{NON-LOCAL}|\text{SLASH} & \boxed{2} \cup \left\{\boxed{4}\right\} \end{bmatrix} \end{bmatrix} \\ \text{NON-HEAD-DTR} & \boxed{5}\begin{bmatrix} \text{SYNSEM}|\text{LOCAL} & \boxed{4} \end{bmatrix} \\ \text{DAUGHTERS} & \left\langle \boxed{5}, \boxed{3} \right\rangle \end{bmatrix} \quad \textit{tal que } \boxed{2} \cap \left\{\boxed{4}\right\} = \{\}$$

Note-se que o constituinte da direita, com a etiqueta $\boxed{3}$, tem um SLASH não vazio, fazendo com que este sintagma seja bem formado quando algures em nós inferiores existe um predicador sem um dos seus argumentos a ocorrer localmente e que, por isso, tenha posto a circular informação através de SLASH. O constituinte da esquerda, por sua vez, tem o valor de LOCAL unificado com um dos elementos do SLASH do constituinte direito.

Em virtude de o argumento subcategorizado a longa distância ter já ocorrido, a informação acerca desse elemento não é percolada para nós superiores, pelo que na construção *filler-head-phrase* o SLASH do nó mãe é um conjunto sem essa informação.

Percolação

A segunda componente da análise das dependências de longa distância consiste na percolação do valor de SLASH.

As restrições que unificam os valor de SLASH no nó mãe de uma construção e nos nós filhos estão associadas ao tipo *head-valence-phrase*:

$$\begin{bmatrix} \textit{head-valence-phrase} \\ \text{SYNSEM}|\text{NON-LOCAL}|\text{SLASH } \boxed{1} \\ \text{HEAD-DTR}|\text{SYNSEM}|\text{NON-LOCAL}|\text{SLASH } \boxed{1} \end{bmatrix}$$

Este é um supertipo de várias das construções *subject-head-phrase*, *head-complements-phrase* e *specifier-head-phrase* já apresentadas (vd. Figura 1).

[12] Para efeitos da presente exposição introdutória, trata-se de uma simplificação relativamente ao tipo *filler-head-phrase* definido em (Bouma *et al.* 2001).

Lista de Subcategorização

Resta considerar o terceiro componente da análise de dependências de longa distância, relativo à definição do tipo do predicador, e como nesta definição o valor de SLASH é devidamente instanciado.

O tipo *synsem* tem dois subtipos, nomeadamente *canonical-synsem* e *gap-synsem*. O valor de *synsem* para os constituintes, sejam palavras ou sintagmas, é do tipo *canonical-synsem*. Por conseguinte, o tipo *sign* tem a seguinte restrição adicional, a qual recebe a designação de Princípio de Canonicidade (*Principle of Canonicality*):

$$\begin{bmatrix} \textit{sign} & \\ \text{SYNSEM} & \textit{canonical-synsem} \end{bmatrix}$$

O tipo *canonical-synsem* é também o subtipo de *synsem* dos elementos que se encontram nas listas de subcategorização de um predicador e que, em relação a esse predicador, ocorrem localmente.

O tipo *gap-synsem*, por sua vez, é o subtipo de *synsem* dos elementos que se encontram nas listas de subcategorização de um predicador e que ocorrem a longa distância deste.

É de notar que nenhuma regra sintáctica é capaz de projectar um elemento de SUBJ ou de COMPS que seja do tipo *gap-synsem*, uma vez que estes elementos são unificados com o SYNSEM de outro constituinte, o qual em virtude do Princípio de Canonicidade tem de ter sempre o tipo *canonical-synsem*.

Para além desta extensão em termos dos subtipos de *synsem*, os traços em que a informação de subcategorização de um predicador é codificada vão também ser alargados. É utilizado um traço adicional, o traço DEPS (DEPENDENTS), que tal como os traços VAL|SUBJ e VAL|COMPS, também se encontra sob CAT. O seu valor é a concatenação das listas de valência SUBJ e COMPS. O valor deste atributo DEPS só é relevante para palavras, pelo que o valor de DEPS é definido no tipo *word* através das seguintes restrições adicionais, as quais recebem a designação de Princípio de Realização do Argumento (*Argument Realization Principle*):

Princípio de Realização do Argumento

$$\begin{bmatrix} \textit{word} & \\ \text{SYNSEM}|\text{LOCAL}|\text{CAT} & \begin{bmatrix} \text{VAL} & \begin{bmatrix} \text{SUBJ} & \boxed{1} \\ \text{COMPS} & \boxed{2} \ominus \textit{list(gap-synsem)} \end{bmatrix} \\ \text{DEPS} & \boxed{1} \oplus \boxed{2} \end{bmatrix} \end{bmatrix}$$

Sendo o seu valor uma lista de elementos de tipo *synsem*, e podendo por isso conter elementos do tipo *gap-synsem*, o valor do traço DEPS inclui a concatenação dos elementos de SUBJ com os elementos de COMPS que não são do tipo *gap-synsem*.

Para tirar partido deste alargamento, e em vista da análise das dependências de longa distância, os verbos transitivos recebem a seguinte definição:

$$\begin{bmatrix} \textit{trans-verb-lxm} \\ \text{SYNSEM}|\text{LOCAL}|\text{CAT} \begin{bmatrix} \text{HEAD} & \textit{verb} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle \textit{synsem} \rangle \\ \text{SPR} & \langle\rangle \end{bmatrix} \\ \text{DEPS} & \left\langle \begin{bmatrix} \textit{synsem} \\ \text{L}|\text{C} \begin{bmatrix} \text{HEAD } \textit{noun} \\ \text{V} \begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \textit{synsem} \\ \text{L}|\text{C} \begin{bmatrix} \text{HEAD } \textit{noun} \\ \text{V} \begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

Nesta definição não se especifica o subtipo de synsem dos SNs seleccionados pelo verbo. Consequentemente, estas restrições impostas na entrada lexical de um verbo transitivo podem ser satisfeitas por duas estruturas distintas, consoante o segundo elemento de DEPS, relativo ao Objecto Directo, acontece ser resolvido para um objecto do subtipo *canonical-synsem*, em (a), ou do subtipo *gap-synsem*, em (b):

$$\text{(a)} \begin{bmatrix} \textit{trans-verb-lxm} \\ \text{S}|\text{L}|\text{C} \begin{bmatrix} \text{V} \begin{bmatrix} \text{SUBJ} & \left\langle \boxed{1} \begin{bmatrix} \text{L}|\text{C} \begin{bmatrix} \text{HEAD } \textit{noun} \\ \text{V} \begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \\ \text{COMPS} & \left\langle \boxed{2} \begin{bmatrix} \text{L}|\text{C} \begin{bmatrix} \text{HEAD } \textit{noun} \\ \text{V} \begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix} \\ \text{DEPS} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \end{bmatrix}$$

$$\text{(b)} \begin{bmatrix} \textit{trans-verb-lxm} \\ \text{S}|\text{L}|\text{C} \begin{bmatrix} \text{V} \begin{bmatrix} \text{SUBJ} & \left\langle \boxed{1} \begin{bmatrix} \text{L}|\text{C} \begin{bmatrix} \text{HEAD } \textit{noun} \\ \text{V} \begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \\ \text{DEPS} & \left\langle \boxed{1}, \begin{bmatrix} \textit{gap-synsem} \\ \text{L}|\text{C} \begin{bmatrix} \text{HEAD } \textit{noun} \\ \text{V} \begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

Finalmente, há que estabelecer que quando um elemento de DEPS é do tipo *gap-synsem*, como em (b), a lista no SLASH do verbo tem como elemento o LOCAL desse *gap-synsem*. Isto é obtido através da seguinte definição para o tipo *gap-synsem*:

$$\begin{bmatrix} \textit{gap-synsem} \\ \text{LOCAL } \boxed{1} \\ \text{NON-LOCAL}|\text{SLASH} \left\{ \boxed{1} \right\} \end{bmatrix}$$

Esta definição faz que no nosso exemplo (b) o valor de SLASH do SN complemento contenha como membro o LOCAL desse SN.

Para além desta definição de *gap-synsem*, há que impor uma restrição adicional sobre o tipo *word*, chamada Amálgama SLASH (SLASH *Amalgamation*):[13]

Amálgama SLASH

$$\begin{bmatrix} \textit{word} & \\ \text{SYNSEM} & \begin{bmatrix} \text{LOCAL}|\text{CAT}|\text{DEPS} & \left\langle \begin{bmatrix} \text{NON-LOCAL}|\text{SLASH} & \boxed{S_1} \end{bmatrix}, \cdots, \begin{bmatrix} \text{NON-LOCAL}|\text{SLASH} & \boxed{S_n} \end{bmatrix} \right\rangle \\ \text{NON-LOCAL}|\text{SLASH} & \boxed{S_1} \cup \cdots \cup \boxed{S_n} \end{bmatrix} \end{bmatrix}$$

$$n \geq 0$$

Esta restrição estabelece que o SLASH de uma palavra é a união dos traços SLASH de todos os seus dependentes sintácticos.

Desta forma, o LOCAL do complemento de tipo *gap-synsem* de um verbo é um membro do SLASH desse verbo.

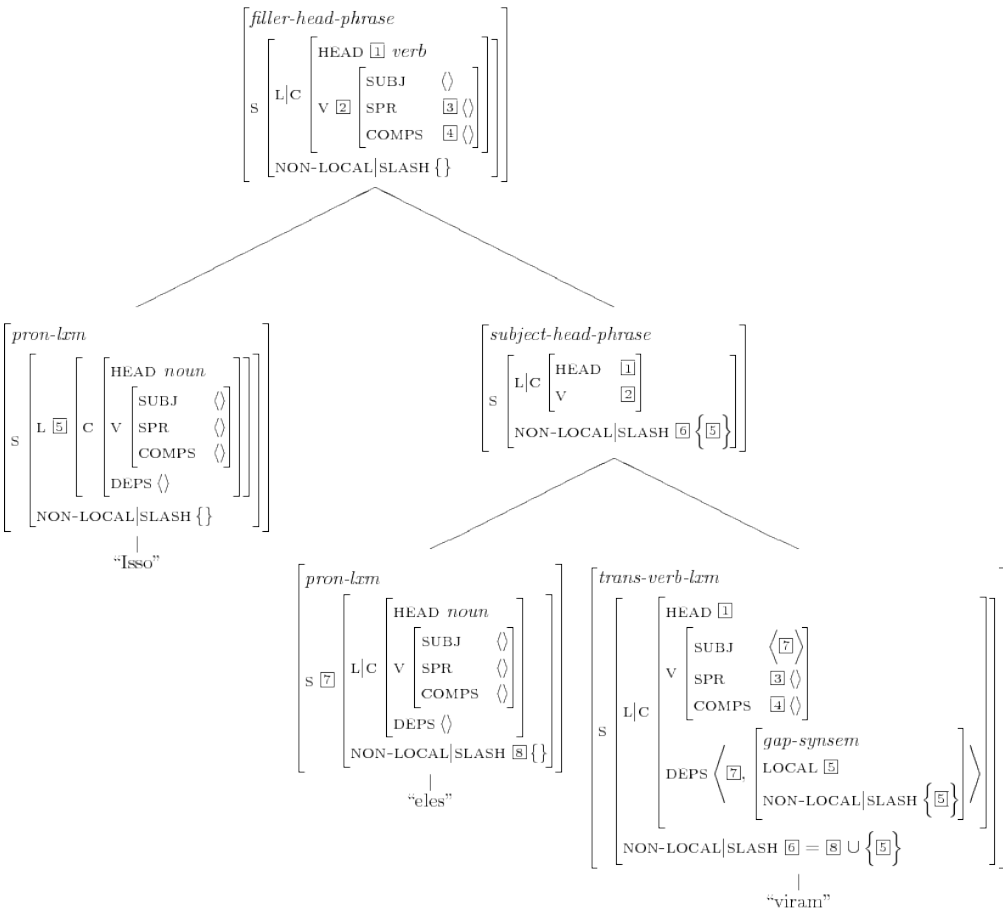
Com estas restrições acrescentadas à gramática, o caso (b) contém a seguinte informação:

$$\begin{bmatrix} \textit{trans-verb-lxm} & \\ \text{SYNSEM} & \begin{bmatrix} \text{LOCAL}|\text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{verb} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \left\langle \boxed{1} \begin{bmatrix} \text{LOCAL} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{noun} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \\ \text{NON-LOCAL}|\text{SLASH} & \boxed{A} \end{bmatrix} \right\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \\ \text{DEPS} & \left\langle \boxed{1}, \begin{bmatrix} \textit{gap-synsem} & \\ \text{LOCAL} & \boxed{2} \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{noun} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \\ \text{NON-LOCAL}|\text{SLASH} & \left\{\boxed{2}\right\} \end{bmatrix} \right\rangle \end{bmatrix} \\ \text{NON-LOCAL}|\text{SLASH} & \boxed{A} \cup \left\{\boxed{2}\right\} \end{bmatrix} \end{bmatrix}$$

É de notar que o valor de SLASH com a etiqueta $\boxed{A}$ toma o conjunto vazio quando o verbo é combinado com o Sujeito pela construção *subject-head-phrase*.

Após a apresentação dos dispositivos acima, é possível retomar o exemplo de trabalho, "Isso, eles viram", apresentando a sua análise de forma mais completa no que toca aos dispositivos chave envolvidos nas dependências de longa distância:

[13] A definição que apresentamos da Amálgama SLASH é uma simplificação relativamente à apresentada em (Bouma *et al*. 2001), uma vez que, dentro das dependências de longa distância, neste capítulo só estamos interessados em abranger topicalizações.



Configuracionalidade

Para efeitos da exposição neste capítulo, estabeleceu-se que a ordem superficial de palavras, representada na ordem dos elementos de PHON, é replicada pela ordem dos elementos de DAUGHTERS, através da seguinte definição do tipo *phrase*:

$$\begin{bmatrix} \textit{phrase} & \\ \text{PHON} & \boxed{1} \oplus \cdots \oplus \boxed{n} \\ \text{DAUGHTERS} & \textit{ne-list} \left\langle \begin{bmatrix} \textit{sign} & \\ \text{PHON} & \boxed{1} \end{bmatrix}, \ldots, \begin{bmatrix} \textit{sign} & \\ \text{PHON} & \boxed{n} \end{bmatrix} \right\rangle \end{bmatrix}$$

A relação entre os elementos de PHON e os de DAUGHTERS pode porém não ser esta. É possível estabelecer as relações de constituência dissociando as restrições de precedência linear das restrições de dominância. Esta estratégia é sobretudo utilizada para línguas a que alguns autores se referem como não configuracionais, as quais apresentam uma ordem de palavras bastante livre.

Devido a limitação de espaço, não é possível apresentar estas abordagens neste capítulo. O leitor interessado pode consultar as obras mais proeminentes sobre este tópico, como (Reape 1994) e (Kathol 1995).

Leituras Subsequentes

O estudo do material introdutório apresentado neste capítulo pode ser continuado com o auxílio dos dois volumes fundacionais de (Pollard e Sag 1987, 1994). Para o aprofundamento da discussão e análise dos diversos fenómenos sintácticos, serve como apoio fundamental o repositório bibliográfico online de HPSG, com acesso livre à maior parte das publicações aí listadas, que se encontra em (HPSG Página Web).

Têm sido desenvolvidos vários trabalhos de investigação que tomam em consideração dados empíricos do Português e em que são focados aspectos diversos da gramática do Português ou da gramática universal.

As estruturas nominais com elipse foram estudadas por (Branco e Costa 2006), tendo a estrutura do sintagma nominal sido aprofundada por (Costa 2007).

A coordenação recebeu atenção de (Villavicencio *et al.* 2005); (Chaves 2005, 2006, 2007).

Os clíticos foram alvos de análise por (Crysmann 1998, 1999, 2000b,a, 2002). A anáfora e a teoria da ligação foram exploradas por (Branco 1996); (Branco e Marrafa 1997, 1999); (Branco 2000, 2005).

O fenómeno da concordância negativa foi estudado por (Branco e Crysmann 2001). A acomodação de diferenças dialectais, em particular entre a variante americana e europeia do Português, foi trabalhada por (Branco e Costa 2007a,b,c).

Há resultados de investigação sobre outras línguas românicas, de fácil acesso e eventualmente relevante para o Português. A compilação em (Balari e Dini 1997) constitui leitura inicial interessante.

Referências

BALARI, Sergio e DINI, Luca, orgs., (1997). *Romance in HPSG*. Stanford: CSLI Publications.

BOUMA, Gosse, MALOUF, Rob e SAG, Ivan (2001). "Satisfying constraints on extraction and adjunction", *Natural Language and Linguistic Theory*, vol. 1, nº 19, pp. 1-65.

BRANCO, António (1996). "Branching split obliqueness at the syntax-semantics interface", *in Proceedings of COLING 96*. Copenhaga: Center for Sprogteknologi, pp. 149-156.

_____ (2000). *Reference processing and its universal constraints*. Lisboa: Edições Colibri.

_____ (2005). "Reflexives: escaping exemption via domain reshuffling", i*n* MÜLLER, Stefan (org.), *The proceedings of the 12th International Conference on Head-Driven Phrase Structure Grammar*. Stanford: CSLI Publications, pp. 467-481.

BRANCO, António e COSTA, Francisco (2006). "Noun ellipsis without empty categories", *in* MÜLLER, Stefan (org.), *The proceedings of the 13th International Conference on Head-Driven Phrase Structure Grammar*. Stanford: CSLI Publications, pp. 81-101.

_____ (2007a). "Accommodating language variation in deep processing", *in* KING, Tracy Holloway e BENDER, Emily M. (orgs.), *Proceedings of the Grammar Engineering Across Frameworks workshop (GEAF07)*. Stanford: CSLI Publications, pp. 67-86.

_____ (2007b). "Identification and handling of dialectal variation with a single grammar", *in* DIRIX, Peter, SHUURMAN, Ineke, VANDEGHINSTE, Vincent e VAN EYNDE, Frank (orgs.), *Proceedings of the 17th meeting of Computational Linguistics in the Netherlands (CLIN17)*. Utrecht: LOT, pp. 5-19.

_____ (2007c). "Self- or pre-tuning? Deep linguistic processing of language variants", *in ACL 2007 Workshop on Deep Linguistic Processing*. Praga: Association for Computational Linguistics, pp. 57-64.

BRANCO, António e CRYSMANN, Berthold (2001). "Negative concord and the distribution of quantifiers", *in* D'HULST, Yves, ROORYCK, Johan e SCHROTEN, Jan (orgs.), *Romance languages and linguistic theory*. Amsterdão: John Benjamins.

BRANCO, António e MARRAFA, Palmira (1997). "Subject-oriented and non Subject-oriented long-distance anaphora: an integrated approach", *in PARK,* Byung-Soo e KIM, Jong-Bok (orgs.), *Proceedings of the 11th Pacific-Asia Conference on Language, Information and Computation (PACLIC'96).* Seul: Language Education and Research Institute, Kyung Hee University, pp. 21-31.

_____ (1999). "Long-distance reflexives and the binding square of opposition", *in* WEBELHUTH, Gert, KÖNIG, Jean-Pierre e KATHOL, Andreas (orgs.), *Lexical and constructional aspects of linguistic explanation*. Stanford: CSLI Publications, pp. 111-123.

CHAVES, Rui Pedro (2005). "A linearization-based approach to gapping", *in* JÄGER; G, MONACHESI, P., PENN, G. e WINTNER, S. (orgs.), *FG-MOL 2005: The 10th conference on Formal Grammar and The 9th meeting on Mathematics of Language*. Stanford: CSLI Publications.

_____ (2006). "Coordination of unlikes without unlike categories", *in* MÜLLER, Stefan (org.), *The proceedings of the 13th International Conference on Head-Driven Phrase Structure Grammar*. Stanford: CSLI Publications, pp. 102-122.

_____ (2007). *Coordinate structures – constraint-based syntax-semantics processing*. Lisboa: Universidade de Lisboa, tese de doutoramento.

COSTA, Francisco (2007). *Deep linguistic processing of Portuguese noun phrases*. Lisboa: Universidade de Lisboa, tese de mestrado.

CRYSMANN, Berthold (1998). "The conspiracy of quantification and linear precedence in European Portuguese proclisis", *in Proceedings of the Eastern States Conference on Linguistics (ESCOL '97)*. Ithaca: CLC Publications.

_____ (1999). "Licensing proclisis in European Portuguese", *in* CORBLIN, F., MARANDIN, J.-M. e DOBROVIE-SORIN, C. (orgs.), *Empirical Issues in Formal Syntax and Semantics*. *Selected papers from the Colloque de Syntaxe et de Sémantique de Paris (CSSP 1997)*. Haia: Thesus, pp. 255-276.

_____ (2000a). "Clitics and coordination in linear structure", *in* GERLACH, Birgit e GRIJZENHOUT, Janet (orgs.), *Clitics in phonology, morphology, and syntax*. Amsterdão: John Benjamins, pp. 121-159.

_____ (2000b). "Syntactic transparency of pronominal affixes", *in* CANN, Ronnie, GROVER, Claire e MILLER, Philip (orgs.), *Grammatical interfaces in* HPSG. Stanford: CSLI Publications, pp. 77-96.

_____ (2002). "Constraint-based coanalysis: Portuguese cliticisation and morphology-syntax interaction in HPSG". Saarbrücken Dissertations in Computational Linguistics and Language Technology. Saarbrücken: Deutsches Forschungszentrum für Künstliche Intelligenz e Universität des Saarlandes.

HPSG (Página Web). *The HPSG bibliography*. Repositório de referências a publicações de trabalhos de investigação realizados no quadro da HPSG. http://hpsg.fuberlin.de/HPSG-Bib.

KATHOL, Andreas (1995). *Linearization-based German syntax*. Ohio State University, Tese de doutoramento.

POLLARD, Carl e SAG, Ivan (1987). *Information-based syntax and semantics*. Stanford: CSLI Publications, vol. 1.

_____ (1994). *Head-Driven Phrase Structure Grammar*. Stanford: Chicago University Press e CSLI Publications.

REAPE, Mike (1994). "Domain union and word order variation in German", *in* NERBONNE, John, NETTER, Klaus e POLLARD, Carl J. (orgs.), *German in Head-Driven Phrase Structure Grammar*. Stanford: CLSI Publications, n. 46 de CSLI Lecture Notes, pp. 151-197.

VILLAVICENCIO, Aline, SADLER, Louisa e ARNOLD, Doug (2005). "An HPSG account of closest conjunct agreement in NP coordination in Portuguese", *in* MÜLLER, Stefan (org.), *The proceedings of the 12th International Conference on Head-Driven Phrase Structure* Grammar. Standford: CSLI Publications, pp. 427-447.