

Les structures de traits

B. Habert

LIMSI & Université Paris X – habert@limsi.fr

15 février 2005

Les structures de traits ¹ permettent de noter sous une forme claire et manipulable des informations linguistiques, lexicales aussi bien que syntaxiques et sémantiques. Elles ont été popularisées par les formalismes d'unification depuis le début des années 90 (Shieber, 1990)(Abeillé, 1993). On trouvera dans (Ligozat, 1994) une présentation plus formelle.

1 Les structures de traits : caractéristiques générales

1.1 « Grammaire » des structures de traits

Une structure de traits se définit de la manière suivante :

```
<structure de traits> ::= <association trait / valeur>+
<association trait / valeur>+ ::= <nom de trait> = <valeur>
<nom de trait> ::= <symbole>
<valeur> ::= <symbole> | <structure de traits>
```

1.2 Facettes des structures de traits

Dans les exemples ci-dessous, on utilise les conventions employées pour le langage PATR-II (Shieber, 1990) :

- une structure de traits est une suite d'*équations* terminée par un point ;
- un *chemin* est un symbole ou une suite de symboles entre chevrons et il permet de désigner un trait ;
- une équation se compose d'un chemin à gauche du signe = et à droite soit d'un autre chemin (partage de valeur entre traits) soit d'une valeur atomique.

Les informations sont nommées Elles sont dès lors plus facilement compréhensibles et utilisables. On comparera :

nom commun masculin singulier

et :

¹La terminologie anglo-saxonne, rappelée par S. Shieber (? , p. 12) est : *feature structures*, *categories*, *f-structures* (LFG), *feature bundles*, *feature matrices* (GPSG), DAGs (PATR-II) . . En linguistique française, *traits* est plutôt utilisé pour la phonétique et la phonologie, voire la sémantique. Il a un sens plus restreint et correspond plutôt à une liste plate d'association <trait, valeur>. *Catégorie* est également plus limité, puisque ce terme renvoie surtout à la partie du discours.

<catégorie> = nom
 <type> = commun
 <genre> = masculin
 <nombre> = singulier .

ce qui peut se noter aussi, sous la forme de *matrice trait-valeur* :

catégorie	nom
type	commun
genre	masculin
nombre	singulier

L'ordre des traits n'importe pas Dans la mesure où les traits sont nommés de manière non ambiguë, leur ordre n'entre pas en ligne de compte. Les deux structures suivantes sont alors équivalentes :

<catégorie> = nom
 <type> = commun
 <genre> = masculin
 <nombre> = singulier .

<type> = commun
 <catégorie> = nom
 <nombre> = singulier .
 <genre> = masculin

Une valeur peut être une structure de traits Au lieu de disposer d'une liste plate de couples <trait, valeur>, on peut regrouper les informations relevant du même domaine (morphologie, syntaxe, sémantique...). Ainsi, la structure :

<catégorie> = nom
 <genre> = masculin
 <nombre> = singulier .

pourra être réorganisée de la manière suivante :

<catégorie> = nom
 <accord genre> = masculin
 <accord nombre> = singulier .

Le trait <accord> est un trait dont la valeur est *complexe*, par opposition à <catégorie> dont la valeur est *atomique*. Ce regroupement évite d'émettre les informations. Elle les rend plus claires. Elle facilite aussi leur manipulation. Une autre notation est :

catégorie	nom				
type	commun				
accord	<table style="border-collapse: collapse;"> <tr> <td style="border-right: 1px solid black; padding-right: 5px;">genre</td> <td style="padding-left: 5px;">masculin</td> </tr> <tr> <td style="border-right: 1px solid black; padding-right: 5px;">nombre</td> <td style="padding-left: 5px;">singulier</td> </tr> </table>	genre	masculin	nombre	singulier
genre	masculin				
nombre	singulier				

Les structures de traits sont aussi complexes que souhaitable Elles peuvent être aussi « profondes » et aussi « larges » qu'il est nécessaire pour rendre compte du domaine traité.

Certains informations peuvent rester non-spécifiées L'absence d'une information peut correspondre au fait que la structure en question est compatible avec toutes les valeurs que peut prendre le trait absent. Par exemple, donner au mot *secrétaires* la structure :

<catégorie> = nom
<type> = commun
<accord nombre> = pluriel .

est équivalent à indiquer :

<catégorie> = nom
<type> = commun
<accord nombre> = pluriel
<accord genre> = {masculin féminin} .

Une seule valeur pour un trait Il n'est pas possible pour un même trait de recevoir deux valeurs différents dans une même structure de traits. Ainsi, la structure suivante est mal formée :

<catégorie> = nom
<genre> = masculin
<genre> = féminin .

Il est par ailleurs possible d'indiquer qu'un trait (atomique ou complexe) correspond à une disjonction (non exclusive) de valeurs. Par exemple, au mot *secrétaire*, on pourra associer la structure suivante :

<catégorie> = nom
<type> = commun
<accord genre> = {masculin féminin} .

Deux traits peuvent partager une valeur Ils ont la même *occurrence* de valeur et non le même *type* de valeur. Par exemple, dans :

<constituant catégorie> = GN
<constituant accord> = <fils1 accord>
<constituant accord> = <fils2 accord>
<fils1 catégorie> = déterminant
<fils1 type> = article_défini
<fils1 accord genre> = féminin
<fils1 accord nombre> = singulier
<fils2 lemme> = le
<fils2 forme> = la
<fils2 catégorie> = nom
<fils2 type> = commun
<fils2 lemme> = secrétaire
<fils2 forme> = secrétaire .

trois traits partagent la *même valeur* : <constituant> <accord>, <fils1 accord> et <fils2 accord>. La matrice correspondante est :

constituant	<table style="border-collapse: collapse;"> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">catégorie</td> <td style="padding: 2px 5px;">GN</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">accord</td> <td style="padding: 2px 5px;">[1]</td> </tr> </table>	catégorie	GN	accord	[1]										
catégorie	GN														
accord	[1]														
fils1	<table style="border-collapse: collapse;"> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">catégorie</td> <td style="padding: 2px 5px;">déterminant</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">type</td> <td style="padding: 2px 5px;">article_défini</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">accord [1]</td> <td style="padding: 2px 5px;"> <table style="border-collapse: collapse;"> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">genre</td> <td style="padding: 2px 5px;">féminin</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">nombre</td> <td style="padding: 2px 5px;">singulier</td> </tr> </table> </td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">lemme</td> <td style="padding: 2px 5px;">le</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">forme</td> <td style="padding: 2px 5px;">la</td> </tr> </table>	catégorie	déterminant	type	article_défini	accord [1]	<table style="border-collapse: collapse;"> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">genre</td> <td style="padding: 2px 5px;">féminin</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">nombre</td> <td style="padding: 2px 5px;">singulier</td> </tr> </table>	genre	féminin	nombre	singulier	lemme	le	forme	la
catégorie	déterminant														
type	article_défini														
accord [1]	<table style="border-collapse: collapse;"> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">genre</td> <td style="padding: 2px 5px;">féminin</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">nombre</td> <td style="padding: 2px 5px;">singulier</td> </tr> </table>	genre	féminin	nombre	singulier										
genre	féminin														
nombre	singulier														
lemme	le														
forme	la														
fils2	<table style="border-collapse: collapse;"> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">catégorie</td> <td style="padding: 2px 5px;">nom</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">type</td> <td style="padding: 2px 5px;">commun</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">accord [1]</td> <td style="padding: 2px 5px;"> <table style="border-collapse: collapse;"> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">nombre</td> <td style="padding: 2px 5px;">singulier</td> </tr> </table> </td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">lemme</td> <td style="padding: 2px 5px;">secrétaire</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">forme</td> <td style="padding: 2px 5px;">secrétaire</td> </tr> </table>	catégorie	nom	type	commun	accord [1]	<table style="border-collapse: collapse;"> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">nombre</td> <td style="padding: 2px 5px;">singulier</td> </tr> </table>	nombre	singulier	lemme	secrétaire	forme	secrétaire		
catégorie	nom														
type	commun														
accord [1]	<table style="border-collapse: collapse;"> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">nombre</td> <td style="padding: 2px 5px;">singulier</td> </tr> </table>	nombre	singulier												
nombre	singulier														
lemme	secrétaire														
forme	secrétaire														

2 Relations entre structures de traits

2.1 Subsumption

(Shieber, 1990, p. 37) Les structures de traits ont par nature une structure de treillis basée sur la relation de *subsumption*. Il s'agit d'une relation d'ordre sur les structures de traits qui, pour résumer sommairement, mesure la compatibilité et la spécificité relative de l'information qu'elles contiennent.

(Shieber, 1990, p. 38) ...une structure de traits D subsume une structure D' (noté $D \sqsubseteq D'$), si D contient un sous-ensemble de l'information de D' .

[]	D_{var}
[catégorie SN]	D_{SN}
[catégorie SN accord [nombre singulier]]	D_{SNsg}
[catégorie SN accord [personne troisième nombre singulier]]	D_{SN3sg}
[catégorie SN accord [personne troisième nombre singulier] sujet [personne troisième nombre singulier]]	$D_{SN3sgSuj}$

$$\left[\begin{array}{l} \text{catégorie} \quad \text{SN} \\ \text{accord} \left[\begin{array}{l} \text{personne} \quad \text{troisième} \\ \text{nombre} \quad \text{singulier} \end{array} \right] \\ \text{sujet} \quad \left[\text{accord} \quad \text{I} \right] \end{array} \right]_{D'_{SN3sgSuj}}$$

$$D_{var} \sqsubseteq D_{SN} \sqsubseteq D_{SNsg} \sqsubseteq D_{SN3sg} \sqsubseteq D_{SN3sgSuj} \sqsubseteq D'_{SN3sgSuj}$$

3 Combinaison de structures de traits

(Shieber, 1990, p. 39) La subsomption n'est qu'une relation d'ordre partiel. En effet, deux structures de traits quelconques ne sont pas nécessairement dans une telle relation l'une par rapport à l'autre. Ainsi, deux structures de traits portant des informations différentes, mais compatibles, ne se subsument ni l'une ni l'autre.

$$\left[\begin{array}{l} \text{catégorie} \quad \text{SN} \\ \text{accord} \quad \left[\begin{array}{l} \text{nombre} \quad \text{singulier} \end{array} \right] \end{array} \right]_{D_{SNsg}}$$

$$\left[\begin{array}{l} \text{catégorie} \quad \text{SN} \\ \text{accord} \quad \left[\begin{array}{l} \text{personne} \quad \text{troisième} \end{array} \right] \end{array} \right]_{D_{SN3}}$$

(Shieber, 1990, p. 39) Il en est de même pour des structures qui portent des informations incompatibles :

$$\left[\begin{array}{l} \text{catégorie} \quad \text{SN} \\ \text{accord} \quad \left[\begin{array}{l} \text{nombre} \quad \text{singulier} \end{array} \right] \end{array} \right]_{D_{SNsg}}$$

$$\left[\begin{array}{l} \text{catégorie} \quad \text{SN} \\ \text{accord} \quad \left[\begin{array}{l} \text{nombre} \quad \text{pluriel} \end{array} \right] \end{array} \right]_{D_{SNpl}}$$

(Shieber, 1990, p. 40) La différence entre ces deux cas est que, dans le premier, il existe une structure plus spécifique qui est subsumée par les deux structures de traits, soit :

$$\left[\begin{array}{l} \text{catégorie} \quad \text{SN} \\ \text{accord} \quad \left[\begin{array}{l} \text{personne} \quad \text{troisième} \\ \text{nombre} \quad \text{singulier} \end{array} \right] \end{array} \right]_{D_{SN3sg}}$$

tandis que dans le second une telle structure n'existe pas.

Recherche de la structure qui contient toute l'information des structures combinées, mais pas plus, à la différence de :

$$\left[\begin{array}{l} \text{catégorie} \quad \text{SN} \\ \text{accord} \quad \left[\begin{array}{l} \text{personne} \quad \text{troisième} \\ \text{nombre} \quad \text{singulier genre} \quad \text{masculin} \end{array} \right] \end{array} \right]_{D_{SN3sgm}}$$

(Shieber, 1990, p. 40) ... nous définissons l'*unification* de deux structures de traits D' et D'' comme la structure de traits la plus générale, D , telle que $D' \sqsubseteq D$ et $D'' \sqsubseteq D$. Nous adoptons la notation $D = D' \sqcup D''$.

$$D_{SN3} \sqcup D_{SNsg} = D_{SN3sg}$$

$$\left[\begin{array}{l} \text{catégorie SN} \\ \text{accord} \left[\begin{array}{l} \text{personne troisième} \\ \text{nombre singulier} \end{array} \right] \end{array} \right]_{D_{SN3sg}}$$

(Shieber, 1990, p. 40) ... l'unification est parfois impossible pour certaines paires de structures de traits qui contiennent des informations incompatibles. Dans ce cas, on dit que l'unification *échoue*.

$$D_{SNpl} \sqcup D_{SNsg} = \perp$$

Propriétés de l'unification

Ajout d'information

$$\left[\begin{array}{l} \text{catégorie SN} \end{array} \right]_{D_{SN}} \sqcup \left[\begin{array}{l} \text{accord} \left[\begin{array}{l} \text{nombre singulier} \end{array} \right] \end{array} \right]_{D_{sg}}$$

$$= \left[\begin{array}{l} \text{catégorie SN} \\ \text{accord} \left[\begin{array}{l} \text{nombre singulier} \end{array} \right] \end{array} \right]_{D_{SNsg}}$$

Idempotence

$$\left[\begin{array}{l} \text{catégorie SN} \end{array} \right]_{D_{SN}} \sqcup \left[\begin{array}{l} \text{catégorie SN} \\ \text{accord} \left[\begin{array}{l} \text{nombre singulier} \end{array} \right] \end{array} \right]_{D_{SNsg}}$$

$$= \left[\begin{array}{l} \text{catégorie SN} \\ \text{accord} \left[\begin{array}{l} \text{nombre singulier} \end{array} \right] \end{array} \right]_{D_{SNsg}}$$

Élément neutre : les variables

$$\left[\right]_{D_{var}} \sqcup \left[\begin{array}{l} \text{catégorie SN} \\ \text{accord} \left[\begin{array}{l} \text{nombre singulier} \end{array} \right] \end{array} \right]_{D_{SNsg}}$$

$$= \left[\begin{array}{l} \text{catégorie SN} \\ \text{accord} \left[\begin{array}{l} \text{nombre singulier} \end{array} \right] \end{array} \right]_{D_{SNsg}}$$

Identité versus partage de valeur Les effets de l'unification diffèrent selon que les valeurs sont semblables ou identiques :

$$\left[\begin{array}{l} \text{accord} \left[\begin{array}{l} \text{nombre singulier} \end{array} \right] \\ \text{sujet} \left[\begin{array}{l} \text{accord} \left[\begin{array}{l} \text{nombre singulier} \end{array} \right] \end{array} \right] \end{array} \right]_{D_{Sujsg}}$$

$$\sqcup$$

$$\left[\begin{array}{l} \text{sujet} \left[\begin{array}{l} \text{accord} \left[\begin{array}{l} \text{personne troisième} \end{array} \right] \end{array} \right] \end{array} \right]_{D_{Suj3}}$$

$$\begin{aligned}
&= \left[\begin{array}{l} \text{accord} \left[\begin{array}{l} \text{nombre} \quad \text{singulier} \end{array} \right] \\ \text{sujet} \left[\begin{array}{l} \text{accord} \left[\begin{array}{l} \text{personne} \quad \text{troisième} \\ \text{nombre} \quad \text{singulier} \end{array} \right] \end{array} \right] \end{array} \right]_{D_{\text{Suj}3\text{sg}}} \\
\text{versus} \\
&\left[\begin{array}{l} \text{accord}_{\boxed{1}} \left[\begin{array}{l} \text{nombre} \quad \text{singulier} \end{array} \right] \\ \text{sujet} \left[\begin{array}{l} \text{accord} \quad \boxed{1} \end{array} \right] \end{array} \right]_{D'_{\text{Suj}3\text{sg}}} \sqcup \left[\begin{array}{l} \text{sujet} \left[\begin{array}{l} \text{accord} \left[\begin{array}{l} \text{personne} \quad \text{troisième} \end{array} \right] \end{array} \right] \end{array} \right]_{D_{\text{Suj}3}} \\
= \\
&\left[\begin{array}{l} \text{accord}_{\boxed{1}} \left[\begin{array}{l} \text{personne} \quad \text{troisième} \\ \text{nombre} \quad \text{singulier} \end{array} \right] \\ \text{sujet} \left[\begin{array}{l} \text{accord} \quad \boxed{1} \end{array} \right] \end{array} \right]_{D'_{\text{Suj}3\text{sg}}}
\end{aligned}$$

Références

- ABEILLÉ, A. (1993). *Les nouvelles syntaxes : grammaires d'unification et analyse du français*. Paris : Armand Colin.
- LIGOZAT, G. (1994). *Représentation des connaissances et linguistique*. Acquis Avancés de l'Informatique. Paris : Armand Colin.
- SHIEBER, S. (1990). Les grammaires basées sur l'unification. In P. MILLER & T. TORRIS, Eds., *Formalismes syntaxiques pour le traitement automatique du langage naturel*, Langue, raisonnement, calcul, chapitre 1, pp. 27–85. Paris : Hermès.