

1994

Extraction of *De*-Phrases from the French NP

Ivan A. Sag
Stanford University

Danièle Godard
CNRS, Université Paris 7

Follow this and additional works at: <https://scholarworks.umass.edu/nels>



Part of the [Linguistics Commons](#)

Recommended Citation

Sag, Ivan A. and Godard, Danièle (1994) "Extraction of *De*-Phrases from the French NP," *North East Linguistics Society*. Vol. 24 : Iss. 2 , Article 12.

Available at: <https://scholarworks.umass.edu/nels/vol24/iss2/12>

This Article is brought to you for free and open access by the Graduate Linguistics Students Association (GLSA) at ScholarWorks@UMass Amherst. It has been accepted for inclusion in North East Linguistics Society by an authorized editor of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

Extraction of *De*-Phrases from the French NP*

Ivan A. Sag
Stanford University

Danièle Godard
CNRS, Université Paris 7

1 Introduction

This paper addresses a number of empirical problems surrounding the analysis of 'extraction' from French nominal phrases. The lexically based analysis that we present expresses a fundamental generalization in this domain, namely, the correlation between the potential for extraction from NPs and the possibility of 'pied piping'. As observed by Godard (1992), this generalization is left unexplained by existing accounts of extraction.

We will also sketch how our treatment of extraction and pied piping fits into a broader analysis of the core syntactic phenomena of modern French. Our lexical treatment of cliticization, itself a consequence of the strict lexicalism that we embrace, allows us to unify the analysis of unbounded extraction phenomena with that of cliticization, in the process explaining their common properties.

We are able to derive the facts in question through the interaction of independently motivated constraints on representations. The relevant generalizations are naturally cast as constraints in the framework of **Head-Driven Phrase Structure Grammar (HPSG)**, whose relevant constructs we explain in section 3 below.

2 The Basic Data

NPs are in general extraction islands in French, as shown by the following familiar data.¹

- (1) a. *le sport auquel [l'aptitude _] n'est pas très répandue en France
(‘the sport to which the aptitude is not very widespread in France’)

*This paper is part of a larger investigation of French grammar undertaken together with Anne Abeillé and Philip Miller, both of whom we thank for their continued contributions to our thinking. Thanks also to Susanne Riehemann and Emma Pease for their help. We also gratefully acknowledge the support of the National Science Foundation (Grant SBR-930958), CNRS, CSLI and OTS.

¹There are **apparent** exceptions to this generalization that we do not discuss here, e.g. *A et B, entre lesquels la différence est importante.*

- b. *la conférence à laquelle [la participation –] est surtout américaine
(‘the conference to which the participation is particularly american’)

But *de*-phrase arguments may be extracted in certain cases, yielding, e.g. relative clauses introduced by *dont* and cliticization with *en*:

- (2) a. Marie lit la fin du livre. (‘Marie is reading the end of the book’)
 b. le livre dont Marie lit la fin (‘the book of which Marie is reading the end’)
 c. Marie en lit la fin. (‘Marie is reading the end of it.’)

However, as many have noted, there are a variety of constraints on which *de*-phrase arguments are extractable. For example, Godard (1992) observes that extraction is impossible in examples like the following:

- (3) ??le jeu dont la passion a perdu les aristocrates russes paraît aujourd’hui désuet (‘gambling of-which (= for which) the passion has ruined Russian aristocrats seems today really outdated’)
 *les serpents dont on mesure la peur par les représentations qu’on en fait (‘the snakes of-which one measures the fear (= the fear of which one measures) by the representations that are made of them’)

Similarly, if an agentive *de*-phrase is expressed, then a theme *de*-phrase cannot be extracted (Ruwet 1972):

- (4) a. la jeune fille dont le portrait est à la Fondation Barnes
(‘the young girl of-which the portrait (= whose portrait) is at the BF’)
 b. *la jeune fille dont le portrait de Corot est à la Fondation Barnes
(‘the young girl of-which the portrait of Corot (= painted by Corot) is at the Barnes Foundation’)
 c. la jeune fille dont le portrait par Corot est à la Fondation Barnes
(‘the young girl of-which the portrait by Corot is at the Barnes Foundation’)
 d. Le portrait en serait à la Fondation Barnes.
 e. *Le portrait de Corot en serait à la Fondation Barnes.

And if a possessive determiner or *de*-phrase possessor is expressed, then no other complement can be extracted (Milner 1978, 1982):

- (5) a. la neuvième, dont j’ai beaucoup aimé l’interprétation par Karajan
(‘the Ninth, of-which I have very much loved the interpretation by K.’)
 b. *la neuvième, dont j’ai beaucoup aimé son interprétation
(‘the Ninth, of-which I have very much loved his interpretation’)

EXTRACTION FROM THE FRENCH NP

- (6) a. le Corbusier dont les maisons ne sont pas très confortables
 ('Le C. of-which the houses (= whose houses) are not very pleasant to live in')
- b. *le Corbusier dont la maison de M. X n'est guère confortable
 ('Le C. of-which the house of Mr. X. is not really pleasant to live in')

Previous analyses (Giorgi & Longobardi 1991, Pollock 1989, Sportiche 1989, Stowell 1989) have assumed that extractability from NP should follow from constraints like the ECP or Subjacency, exploiting structural differences between complements. But, as shown by Godard (1992), no structural property cleanly separates the extractable complements from their unextractable counterparts. Moreover, any appeal to conditions on extraction encounters difficulties with the last and most surprising set of data that we discuss. The constraints on extraction from NP also govern the possibilities for pied piping:

- (7) a. la table à un pied de laquelle nous nous accrochons toujours
 ('the table into a leg of which we keep bumping')
- b. les romans de Balzac sur la moitié desquels j'ai travaillé
 ('the novels of B. on the half of which I have worked')
- (8) a. *le sport à l'aptitude auquel je ne m'intéresse pas
 ('the sport in the aptitude to which I am not interested')
- b. *la conférence sur la participation à laquelle on s'interroge
 ('the conference on the participation to which one wonders')
- c. ??le jeu à la passion duquel les aristocrates russes ne renoncèrent pas
 ('gambling from the passion of which Russian aristocrats did not go away')
- d. *les serpents à la peur desquels certains mythologues ont consacré leur œuvre
 ('the snakes to the fear of which some mythologists devoted their work')
- e. *la neuvième, à son interprétation de laquelle je me suis intéressé
 ('the Ninth, in his interpretation of which I'm interested')

That is, a given *de*-phrase can contain a relative word only if that *de*-phrase is extractable. This correlation is particularly curious given the fact that there is no general correlation between the constraints on pied piping and those governing extraction, as illustrated by the well-known data in (9):

- (9) a. Pierre dont j'ai rencontré l'ami ('P. of-which I met the friend')
- b. *Pierre dont j'ai voté contre l'ami ('P. of-which I voted against the friend')
- c. *Pierre, dont j'ai parlé au frère ('P. of-which I talked to the brother')

3 HPSG – The Basic Framework

Many of the central constructs of HPSG are motivated by its adherence to strict lexicalism, a thesis that entails that syntactic operations cannot operate on or make reference to internal properties of lexical items. Any lexically based theory necessarily employs rich lexical representations and HPSG's UG is a small set of principles that allow the grammar of phrases to be projected from the particular information encoded in lexical heads. One might think of the core of HPSG theory as an attempt to simplify both grammatical structures and their grammar, deriving the effects of head movement, functional categories and the projection principle all from the interaction of X' theory and strict lexicalism.

All X' theories embody some variant of the following principle, whose specific formulation presumes that HEAD is a feature taking a feature structure complex as its value:

(10) **The Head Feature Principle (HFP):**

The HEAD value of a headed phrase is identified with that of its head-daughter.

Whereas most X' theories are formulated in terms of hierarchical bar levels, all essentially adapted from the systems of exponential categories developed by Zellig Harris in the 1940s (e.g. Harris 1946), the HPSG X' theory replaces this component with one based on argument cancellation, an idea derived directly from the tradition of categorial grammar (starting with Ajdukiewicz 1935):

(11) **The Valence Principle (VALP):**

For each valence feature F, the F value of a headed phrase is the head-daughter's F value minus the realized non-head-daughters (e.g. Subj-Dtr, Complement-Dtrs, Spec-Dtr).

These X' principles, which again presuppose a particular feature geometry, serve to constrain the structures described by a small inventory of universally available schemata, e.g. the following:

(12) a. **Schema 1:**

$$X \longrightarrow \text{Subj-Dtr, Head-Dtr} \\ [\text{COMPS } \langle \rangle]$$

b. **Schema 2:**

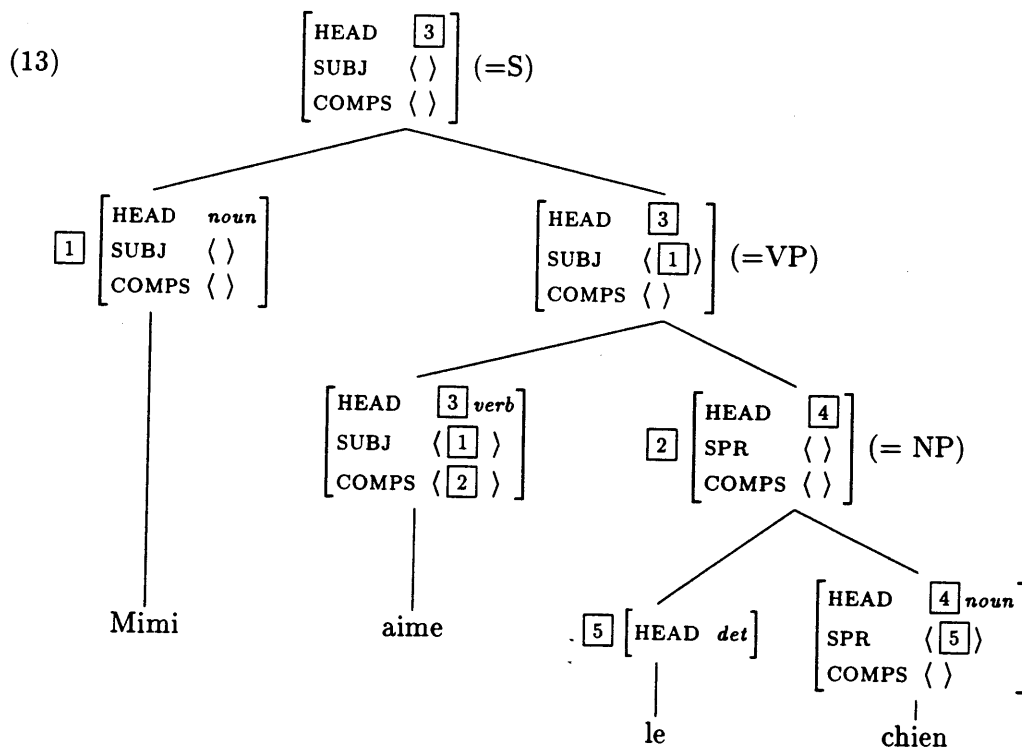
$$X \longrightarrow \text{Head-Dtr, Complement-Dtrs} \\ [\text{COMPS } \langle \rangle]$$

c. **Schema 3:**

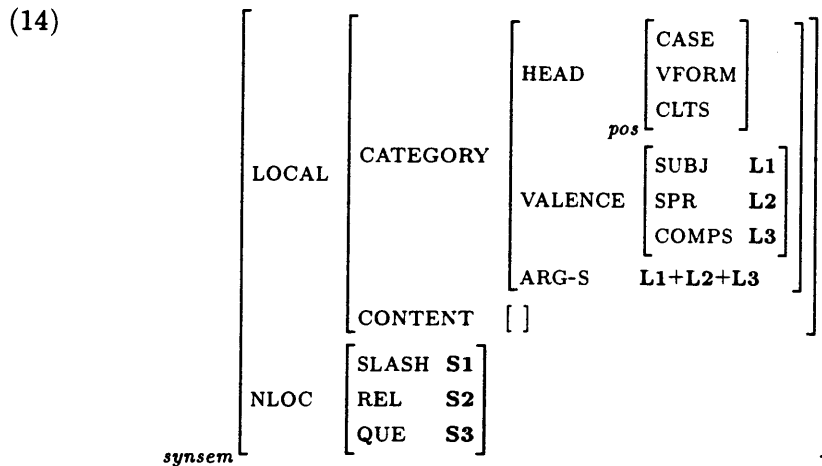
$$X \longrightarrow \text{Spec-Dtr, Head-Dtr} \\ [\text{COMPS } \langle \rangle]$$

EXTRACTION FROM THE FRENCH NP

These schemata describe phrases that consist of, respectively: (1) a subject daughter and a head daughter that already contains its complements (if any): (2) a lexical head daughter and all of its (non-subject) complements; and (3) a head daughter and its specifier. Because of X' theory, the head daughter's HEAD information is maximally projected in any given phrase (by the **HFP**) and the head's valence information determines the elements that the maximal projection contains (in accordance with the **VALP**). Thus each subtree in the following structure satisfies one of the schemata and all of the principles of UG:



The boxed integers in these tree diagrams are variables used to 'tag' certain feature values within the structure as being token identical, as required by the **HFP** or the **VALP**. The feature geometry here and throughout is simplified in irrelevant respects. For example, the version of HPSG theory outlined in Pollard and Sag (1994: chap. 9) would include the following hierarchical relations among feature values:



Note that canonically the values of the various valence features (L1+ L2+L3 in (14)) 'add up' to the argument structure of a given word. The argument structure, coded in terms of the attribute ARG-S, will be the locus of the binding theory's constraints, *inter alia*.²

The feature geometry here is in no way arbitrary. For example, *synsems* are the objects that appear as members of the lists that serve as values of the valence features and hence contain all the information that can be selected by a head daughter in any given construction. Since *synsem* objects contain information about part of speech, agreement classes, case, semantic content, and valence, it follows that such information can be selected by heads. But since *synsem* objects contain only information associated with the specifier, complement, or subject that the head combines with directly, it follows that all head-dependent selections are local. The very geometry of linguistic information in HPSG thus allows us to provide a precise and unified account of such fundamental issues as the locality of subcategorization, case and role assignment, and nonanaphoric agreement.

In order for the schemata and principles stated above to function, lexical items must be richly specified, as they are in any theory that embraces strict lexicalism. The structure in (13), for example, presupposes a set of lexical elements like the following:

- (15) a. aime ('loves') b. chien ('dog') c. le ('the')
- $$\begin{array}{ccc}
 \left[\begin{array}{c} \text{HEAD } \textit{verb}[\textit{fin}] \\ \text{SUBJ } \langle [1] \text{NP}[\textit{nom}]_{3s} \rangle \\ \text{COMPS } \langle [2] \text{NP}[\textit{acc}] \rangle \\ \text{ARG-S } \langle [1], [2] \rangle \end{array} \right] &
 \left[\begin{array}{c} \text{HEAD } \textit{noun} \\ \text{SUBJ } \langle \rangle \\ \text{COMPS } \langle \rangle \\ \text{SPR } (\textit{Det}_{3s}) \end{array} \right] &
 \left[\begin{array}{c} \text{HEAD } \textit{det} \\ \text{CONTENT } \left[\text{INDEX } [3sgm] \right] \end{array} \right]
 \end{array}$$

²This enables a treatment of 'pro-drop' phenomena, for instance, in terms of verbs whose valence features take reduced values, but whose argument structures contain unexpressed pronominal elements.

EXTRACTION FROM THE FRENCH NP

525

<p>d. <i>portrait</i> ('picture')</p> <table style="border-left: 1px solid black; border-right: 1px solid black; border-bottom: 1px solid black; padding: 5px;"> <tr><td style="padding: 2px 5px;">HEAD</td><td style="padding: 2px 5px;"><i>noun</i></td></tr> <tr><td style="padding: 2px 5px;">SUBJ</td><td style="padding: 2px 5px;">⟨ ⟩</td></tr> <tr><td style="padding: 2px 5px;">COMPS</td><td style="padding: 2px 5px;">⟨(NP[de]_i),(NP[de]_j)⟩</td></tr> <tr><td style="padding: 2px 5px;">SPR</td><td style="padding: 2px 5px;">⟨Det_{3s}⟩</td></tr> <tr><td style="padding: 2px 5px;">CONTENT</td><td style="padding: 2px 5px;"><i>portrait</i>(ag:j,th:i)</td></tr> </table>	HEAD	<i>noun</i>	SUBJ	⟨ ⟩	COMPS	⟨(NP[de] _i),(NP[de] _j)⟩	SPR	⟨Det _{3s} ⟩	CONTENT	<i>portrait</i> (ag:j,th:i)	<p>e. <i>de</i> ('of')</p> <table style="border-left: 1px solid black; border-right: 1px solid black; border-bottom: 1px solid black; padding: 5px;"> <tr><td style="padding: 2px 5px;">HEAD</td><td style="padding: 2px 5px;"><i>mark</i>[de]</td></tr> <tr><td style="padding: 2px 5px;">SUBJ</td><td style="padding: 2px 5px;">⟨ ⟩</td></tr> <tr><td style="padding: 2px 5px;">COMPS</td><td style="padding: 2px 5px;">⟨ ⟩</td></tr> <tr><td style="padding: 2px 5px;">SPR</td><td style="padding: 2px 5px;">⟨NP[unm]⟩</td></tr> </table>	HEAD	<i>mark</i> [de]	SUBJ	⟨ ⟩	COMPS	⟨ ⟩	SPR	⟨NP[unm]⟩
HEAD	<i>noun</i>																		
SUBJ	⟨ ⟩																		
COMPS	⟨(NP[de] _i),(NP[de] _j)⟩																		
SPR	⟨Det _{3s} ⟩																		
CONTENT	<i>portrait</i> (ag:j,th:i)																		
HEAD	<i>mark</i> [de]																		
SUBJ	⟨ ⟩																		
COMPS	⟨ ⟩																		
SPR	⟨NP[unm]⟩																		

Of course such lexical items are neither arbitrary nor stipulated in HPSG. An HPSG lexicon has a rich, deductive structure, one that expresses certain kinds of generalizations in terms of cross-classifying word sorts and multiple inheritance hierarchies³ and others in terms of 'lexical rules' that participate in an inductive definition of the lexical entries of a given language.

One such lexical rule, formulated in (16), systematically expands the lexicon by adding lexical items that resemble basic lexical items in every way except for the effect of moving one element from the COMPS list of the input word to the SLASH value of the output word.

(16) Complement Extraction Lexical Rule (CELR):

$$\left[\begin{array}{l} \text{HEAD} \quad \textit{verb} \\ \text{COMPS} \quad \langle \dots, \left[\begin{array}{l} \text{LOC} \quad \boxed{1} \\ \text{NLOC|SLASH} \quad \{ \boxed{1} \} \end{array} \right], \dots \rangle \\ \text{NLOC|SLASH} \quad S \end{array} \right] \Rightarrow \left[\begin{array}{l} \text{COMPS} \quad \langle \dots \dots \rangle \\ \text{NLOC|SLASH} \quad S \cup \{ \boxed{1} \} \end{array} \right]$$

Intuitively, SLASH encodes missing constituents – if a phrase's value for the feature SLASH contains a category X, then there is a constituent of category X missing from that phrase.⁴ A *synsem* object that appears on a feature list and whose NLOC|SLASH value is identified with its LOCAL value is functioning analytically as a trace, though not as a phonetically empty constituent.

It is a noteworthy property of this lexical rule that it applies not just to lexical items that might be antecedently specified as taking 'slashed' complements, but also to those whose complements are simply *unspecified* in this respect. Thus CELR maps the lexical entry in (17a) into its counterpart in (17b), 'unifying in' to the verb's argument structure the relevant information about the element being slashed.

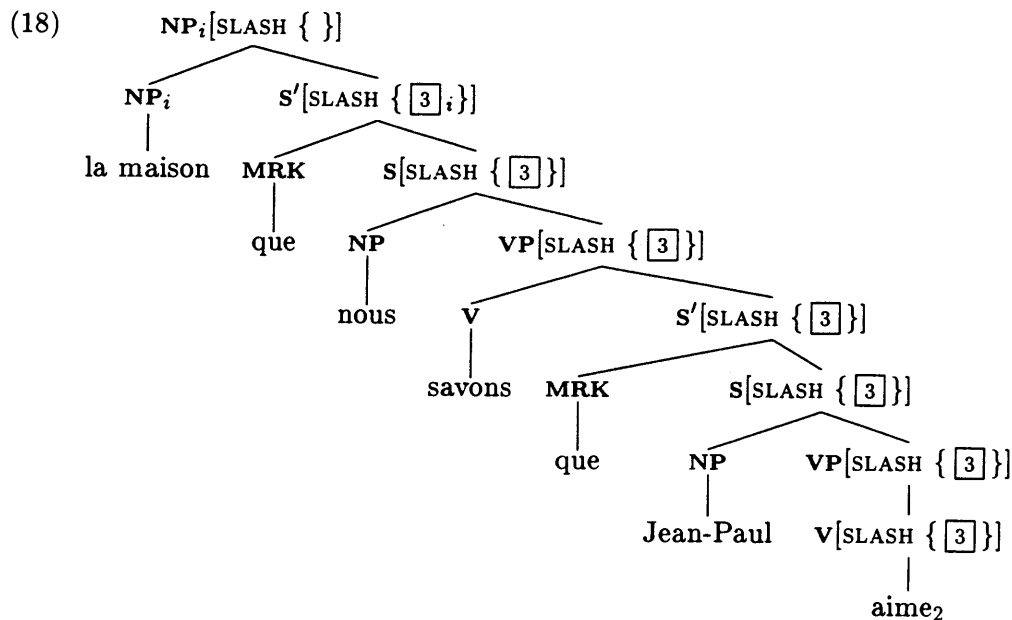
<p>(17) a. <i>aime</i>₁</p> <table style="border-left: 1px solid black; border-right: 1px solid black; border-bottom: 1px solid black; padding: 5px;"> <tr><td style="padding: 2px 5px;">HEAD</td><td style="padding: 2px 5px;"><i>verb</i>[fin]</td></tr> <tr><td style="padding: 2px 5px;">SUBJ</td><td style="padding: 2px 5px;">⟨$\boxed{1}$ NP[nom]_{3s}⟩</td></tr> <tr><td style="padding: 2px 5px;">COMPS</td><td style="padding: 2px 5px;">⟨$\boxed{2}$ NP[acc]⟩</td></tr> <tr><td style="padding: 2px 5px;">ARG-S</td><td style="padding: 2px 5px;">⟨$\boxed{1}$, $\boxed{2}$⟩</td></tr> <tr><td style="padding: 2px 5px;">NLOC SLASH</td><td style="padding: 2px 5px;">{ }</td></tr> </table>	HEAD	<i>verb</i> [fin]	SUBJ	⟨ $\boxed{1}$ NP[nom] _{3s} ⟩	COMPS	⟨ $\boxed{2}$ NP[acc]⟩	ARG-S	⟨ $\boxed{1}$, $\boxed{2}$ ⟩	NLOC SLASH	{ }	<p>b. <i>aime</i>₂</p> <table style="border-left: 1px solid black; border-right: 1px solid black; border-bottom: 1px solid black; padding: 5px;"> <tr><td style="padding: 2px 5px;">HEAD</td><td style="padding: 2px 5px;"><i>verb</i>[fin]</td></tr> <tr><td style="padding: 2px 5px;">SUBJ</td><td style="padding: 2px 5px;">⟨$\boxed{1}$ NP[nom]_{3s}⟩</td></tr> <tr><td style="padding: 2px 5px;">COMPS</td><td style="padding: 2px 5px;">⟨ ⟩</td></tr> <tr><td style="padding: 2px 5px;">ARG-S</td><td style="padding: 2px 5px;">⟨$\boxed{1}$, $\left[\begin{array}{l} \text{LOC} \quad \boxed{3} \\ \text{NLOC SLASH} \quad \{ \boxed{3} \} \end{array} \right] \rangle$</td></tr> <tr><td style="padding: 2px 5px;">NLOC SLASH</td><td style="padding: 2px 5px;">{ $\boxed{3}$ }</td></tr> </table>	HEAD	<i>verb</i> [fin]	SUBJ	⟨ $\boxed{1}$ NP[nom] _{3s} ⟩	COMPS	⟨ ⟩	ARG-S	⟨ $\boxed{1}$, $\left[\begin{array}{l} \text{LOC} \quad \boxed{3} \\ \text{NLOC SLASH} \quad \{ \boxed{3} \} \end{array} \right] \rangle$	NLOC SLASH	{ $\boxed{3}$ }
HEAD	<i>verb</i> [fin]																				
SUBJ	⟨ $\boxed{1}$ NP[nom] _{3s} ⟩																				
COMPS	⟨ $\boxed{2}$ NP[acc]⟩																				
ARG-S	⟨ $\boxed{1}$, $\boxed{2}$ ⟩																				
NLOC SLASH	{ }																				
HEAD	<i>verb</i> [fin]																				
SUBJ	⟨ $\boxed{1}$ NP[nom] _{3s} ⟩																				
COMPS	⟨ ⟩																				
ARG-S	⟨ $\boxed{1}$, $\left[\begin{array}{l} \text{LOC} \quad \boxed{3} \\ \text{NLOC SLASH} \quad \{ \boxed{3} \} \end{array} \right] \rangle$																				
NLOC SLASH	{ $\boxed{3}$ }																				

³See Pollard and Sag 1987, Flickinger 1987, and Flickinger and Nerbonne 1992.

⁴We follow the explicit proposals introduced by Pollard and Sag (1994), who adapt ideas developed originally by Gazdar (1981).

The systematic registration in argument structures of information about long-distance dependencies is the backbone of the HPSG analysis of parasitic gaps developed in Pollard and Sag (1994: chap. 9)⁵ and will also play a fundamental role in our unified treatment of extraction and pied piping in French, discussed below.

Slashed verbs like *aime*₂, by a principle of UG called the **NONLOCAL Feature Principle**, cause NLOC|SLASH specifications to be passed up the tree until an appropriate binding environment is found. Thus only *aime*₂, not *aime*₁, may terminate a long-distance (filler-gap) dependency, as illustrated in (18) and, through a cascade of identities, the filler at the top of the dependency will be linked to the relevant position of *aime*₂'s argument structure. Extraction is thus treated entirely in terms of constraint satisfaction, rather than as transformational 'movement'.



4 Clitics – The Basic Treatment

In this section, we account for the possibility of extraction out of NPs in French, leaving the discussion of the internal constraints on these structures to section 5. Our treatment has two components. First, building on the observation that extraction out of NP is possible only if the NP in question functions as subject or object of some verb, we cast our analysis in terms of a lexical rule that maps a verb normally selecting a complete term (a saturated subject or object) into a new verb form that selects for an incomplete term and a further *de*-phrase complement that

⁵The HPSG **Subject Condition** disallows slashed subjects unless there is another slashed element on the same ARG-S list, thus predicting contrasts like *Who did my talking to bother (*them)?*

is linked to the element missing from the incomplete term.⁶ The verb resulting from this lexical rule, since it now selects for a *de*-phrase complement associated with one of its dependents, may undergo further lexical processes, allowing the *de*-phrase in question to be extracted or cliticized. Second, in order to account for the common behavior of cliticization, relativization and question formation, we formulate our lexical rule for cliticization as an operation on the slashed verbs, i.e. those that have undergone the **CELR**. The **CELR** thus feeds the **Complement Cliticization Lexical Rule (CCLR)**.

None of these features of our analysis is gratuitous, the key to this lexicalized treatment being the analysis of clitics. Once it is accepted (following Miller (1992)) that clitics in French should be treated as verbal affixes, it follows that the entire analysis must be lexical in nature, for the various other regularities, e.g. the one expressed by the **CELR**, must feed cliticization. Moreover, the clitic *en* which corresponds to the complement of an incomplete NP is in no way different from the *en* that corresponds to a *de*-phrase complement that is the verb's semantic argument (e.g. *se souvenir de*). If the lexical rule composing the nominal complement into the complement structure of the higher verb simply feeds the **Complement Cliticization Lexical Rule**, then this fact is simply and elegantly accounted for.

Let us begin with the nature of cliticization. As demonstrated by Miller (1992), there are numerous arguments showing that clitics are lexically attached inflections, not cliticized pronouns. These involve the degree of selection with respect to the host (clitics are always attached to verbs not to other VP-initial elements), arbitrary gaps in the set of combinations (*P. le/*me lui a présenté; finis-tu/*-je?*), morphophonological idiosyncrasies (*P. *(y) va.; P. (*y) ira.*), rigid and idiosyncratic ordering (*P. me(dat) le(acc) donne.; P. le(acc) lui(dat) donne.*), clitics undergoing lexical phonological rules (obligatory liaison for nasal consonants), and the impossibility of clitic coordination (**P. [le et la] voit.* 'P. sees him and her.'). Here we adapt Miller's proposal to treat all clitics in terms of inflectional morphology.

Given the strict lexicalism of HPSG, this analysis of clitics entails that the verbs hosting clitics represent lexical items that are different from the corresponding verb forms that are not clitic hosts. **CCLR** introduces an element into the set value of the feature **CLTS** and different values for this feature will receive appropriately different morphological interpretation, giving rise to such lexical forms as *le-donne* ['it-gives'], *le-lui-donne* ['it-to-him-gives'], etc. An important aspect of our formulation of **CCLR** is that it unifies the grammar of extraction and cliticization by operating only on verb forms that have already undergone **CELR**, and hence have already acquired a slashed element in their argument structure. There is ample independent motivation for this aspect of our account, coming from the agreement of past participle complements of the auxiliary *avoir*, and floating of the quantifier *tous*. In standard written French, the past participle agrees in number and gender with a cliticized or *wh*-extracted direct object. Similarly, cliticization or extraction

⁶This basic idea, closely akin to Lambek's division categories is imported into HPSG in a number of interesting papers stemming from the work of Hinrichs & Nakazawa (1989, 1990). See also Moortgat (1984).

of the direct object allows for a 'floated' quantifier, occurring between the auxiliary and the participle:

(19) Past Participle Agreement:

Marie l'a *écrit/écrite (la lettre).
la lettre que Marie a *écrit/écrite.
Quelle lettre as-tu *écrit/écrite?
*Marie a écrite la lettre.

(20) Floating Tous, Toutes From Object Position:

Marie a vu tous les livres. ('Mary saw all the books')
*Marie a tous vu les livres. ('Mary saw all the books')
(?)Marie les-a vus tous. ('Mary saw them all')
Marie les-a tous vus. ('Mary saw them all')
(?)ces livres, que M. a vus tous. ('the books all of which M. saw')
ces livres, que M. a tous vus. ('the books all of which M. saw')

In our analysis, the only verbs with a slashed element in their argument structure will be slashed verbs like (17b) or cliticized forms that have undergone **CCLR**. Thus the grammar of past participle agreement and the floating of *tous* from objects both make reference to the slashed object in the verb's argument structure, in a fashion made precise by Miller & Sag (1993) and Abeillé et al (in preparation).⁷

CCLR, formulated as in (21), thus unifies the account of an important set of empirical phenomena.⁸

(21) Complement Clitic Lexical Rule (**CCLR**):

$$\left[\begin{array}{l} \text{HEAD } verb \left[\text{CLTS } S_1 \right] \\ \text{NLOC | SLASH } \left\{ \boxed{1} \right\} \cup S_2 \end{array} \right] \Rightarrow \left[\begin{array}{l} \text{HEAD } verb \left[\text{CLTS } S_1 \cup \left\{ \boxed{1} \right\} \right] \\ \text{NLOC | SLASH } S_2 \end{array} \right]$$

where $\boxed{1} \in \{NP[\text{acc}]_\alpha, NP[\hat{a}]_\alpha, NP[\hat{a}_2], NP[de_1], NP[de_2]\}$ and α ranges over permissible person, number and gender combinations.

CCLR takes as input any verb whose **SLASH** value (a set) contains an element of a certain kind and yields a verbal entry where that element has been removed from the **SLASH** value, and moved instead into the set value of the **CLTS** feature.⁹

⁷These facts could in principle be compatible with an analysis where **CCLR** feeds **CELR**, rather than the reverse, as we have assumed. Our particular formulation seems more natural, given that the cliticized forms correspond to only a proper subset of the possibilities for *wh*-extraction and that the slashed verbs are identical in form to uncliticized verbs.

⁸We treat the *à* and *de* in cliticizable complements as markers on the complement NP rather than as heads of PPs. NP[*à*₁] cliticizes as dative, cf. *je souris à Paul/je lui souris*, NP[*à*₂] cliticizes as *y*, cf. *je pense à ce problème/j'y pense*, NP[*de*₁] and NP[*de*₂] both cliticize as *en*, cf. *Je me souviens de ce problème/je m'en souviens*, and *Je reviens d'Angleterre/J'en reviens*, respectively.

⁹This procedural way of thinking of lexical rules is by no means necessary. For interesting alternatives, see Copestake 1992, Riehemann 1993, and Kathol 1994, inter alia.

EXTRACTION FROM THE FRENCH NP

This correctly accounts for the fact that cliticized verb forms, like the slashed verb forms that give rise to them, have reduced valence in French:¹⁰

- (22) a. Marie le voit (*Jean). ('Marie sees him')
 b. Marie le donne (*le livre) à Pierre. ('Marie gives it to Pierre')
 c. Marie le lui donne (*le livre)(*à Pierre). ('Marie gives it to him')
 d. Marie en vient (*de Paris). ('Marie is coming from there')

For example, *aime*₂, which is the output of the **CELR** in (16), is input to **CCLR**, which allows us to derive the schematic *aime*₃ in (23)b.

- (23) a. *aime*₂
- | | | |
|--------------|--|-----------|
| HEAD | <i>verb</i> | [CLTS { } |
| SUBJ | ⟨ [1] NP[nom] _{3s} ⟩ | |
| COMPS | ⟨ ⟩ | |
| ARG-S | ⟨ [1], [LOC [3]
NLOC SLASH { [3] }] ⟩ | |
| NLOC SLASH | { [3] } | |
- b. *aime*₃ ⇒ l-aime, nous-aime, ...
- | | | |
|--------------|--|----------------|
| HEAD | <i>verb</i> | [CLTS { [3] }] |
| SUBJ | ⟨ [1] NP[nom] _{3s} ⟩ | |
| COMPS | ⟨ ⟩ | |
| ARG-S | ⟨ [1], [LOC [3]
NLOC SLASH { [3] }] ⟩ | |
| NLOC SLASH | { } | |

Although space limitations prevent us from elaborating here, Abeillé and Godard (1993) [see also Abeillé et al (in preparation)] present an analysis of the French auxiliary system that assigns flat structures to sentences headed by the auxiliaries *avoir* and *être*, whose highly schematic lexical entries select (via the COMPS feature) for lexical participles followed by whatever complements those participles themselves select. The lexical entries for auxiliary verbs thus divide out the elements subcategorized by the participles and build flat structures via Schema 2, in accordance with the principles outlined in section 3 above. The lexical entries for auxiliaries may thus undergo **CELR** and **CCLR** unproblematically in the derivation of the verbs that head sentences like (24):

- (24) a. Marie l'a vu. ('Marie saw him')
 b. Marie l'a donné à Pierre. ('Marie gave it to Pierre')
 c. Marie le-lui-a donné. ('Marie gave it to him')
 d. Paul en-est venu. ('Paul came from there')

But, as noted by Abeillé & Godard, 1993, there is substantial evidence that adjectival and passive complements should be treated in terms of partially saturated phrases. Consider the VP *être fidèle à ses convictions* ['to be faithful to one's convictions']; as illustrated in (25), not only may the whole AP be cliticized as is expected, but also the adjective alone, or the NP[à] complement of the A:

¹⁰That is, there is no 'clitic doubling'. Our analysis also guarantees that there is no resumptive pronoun in standard French relative clauses (*Pierre à qui je crois que je (*lui) ai donné ce livre* ['Pierre to whom I think that I (to-him) have given this book']).

- (25) a. Fidèle à ses convictions, je pense qu'il l'est ('faithful to his convictions, I think he it-is')
- b. Fidèle, je pense qu'il l'est plus à ses amitiés qu'à ses convictions. ('faithful, I think he it-is more to his friendships than to his convictions')
- c. Je pense qu'il y est fidèle ('I think that he to-it-is faithful')

The full array of data follow if *être* (optionally) composes with its complement AP: both the adjective and the complement of the adjective are treated as complements of the verb *être*, and can thus be cliticized independently. Thus, the composition rule for nominals that we are about to discuss is one form of a more general structure in French, in which a matrix verb inherits its complements' complements. To allow for such structures, we propose the parameterized version of Schema 2 that is shown in (26):

(26) **Schema 2a** (French):

$X[\textit{nonfin}] \longrightarrow \textit{Head-Dtr} , \textit{Complement-Dtrs}$

The difference between this and the formulation of Schema 2 presented in section 3 is simply that the category on the left is not specified for a saturated COMPS list, and can thus be either saturated or unsaturated. We restrict this possibility to categories which are not finite, such as NP and AP, or infinitival VP.

Let us now turn to the composition rule, by which the complement of an NP may be turned into the complement of the matrix verb. As should be clear from the above examples (cf. (4)a,d and (6)a), there is no Subject Condition in French, i.e. no bar to extracting elements from within a subject NP. However, it is not possible to extract out of all NP arguments; the direct object NP (cf. (2), (5)a) contrasts with prepositional complements (PPs or NPs marked with *à* or *de*), as illustrated by the familiar data cited earlier:

- (27) a. *Pierre dont j'ai voté contre l'ami ('P. of-which I voted against the friend')
- b. *Pierre, dont j'ai parlé au frère ('P. of-which I talked to the brother')

Thus, the generalization is that the composition rule applies to verbs that have an unmarked argument NP (where 'unmarked' is a cover term for NPs that are not introduced by a marker like *à* or *de*), that is, either a subject (NP[*nom*]) or a direct object (NP[*acc*]). This is achieved in rule (28) by appealing to a variable feature F, which ranges over valence features, e.g. SUBJ and COMPS and specifying the composable argument as [*unm(arked)*]:

(28) **Nominal Composition Lexical Rule (NCLR):**

$$\left[\begin{array}{l} \text{HEAD} \quad \textit{verb} \\ \text{VAL} \mid \textit{F} \langle \dots, \textit{NP}[\textit{unm}], \dots \rangle \end{array} \right] \Longrightarrow \left[\begin{array}{l} \text{VAL} \mid \textit{F} \langle \dots, \textit{NP}[\textit{unm}], \boxed{1}, \dots \rangle \\ \text{COMPS} \langle \boxed{1} \rangle \end{array} \right]$$

EXTRACTION FROM THE FRENCH NP

531

Now there is evidence against positing two different lexical forms for *l'aime*, analogous to those motivated for *l'est* – the fact that the expected cliticization in (29) is unquestionably ungrammatical:

- (29) *Paul l'aime de ta pièce, mais pas de la mienne.
(‘Paul loves it of your play, but not of mine.’)

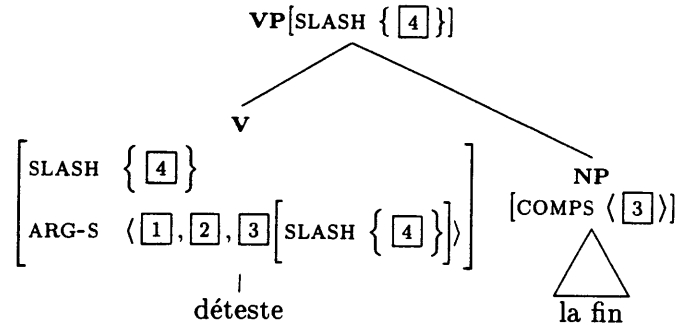
One possibility is to modify **NCLR** so that its output already selects for a slashed *de*-phrase, one that cannot simply be realized as an overt *de*-phrase, but which can be removed from the verbs **COMPS** list as it undergoes **CELR**. Either through this modification or some other, examples like (29) must be excluded.

Assuming some such modification has been made, consider how **NCLR** applies to the basic *déteste*₁ to give *déteste*₂, with an augmented **COMPS** list. *Déteste*₂ then undergoes **CELR** to give *déteste*₃, with a slashed complement, and *déteste*₃ may undergo **CCLR** to give *déteste*₄, which is morphologically realized as *en-déteste*:

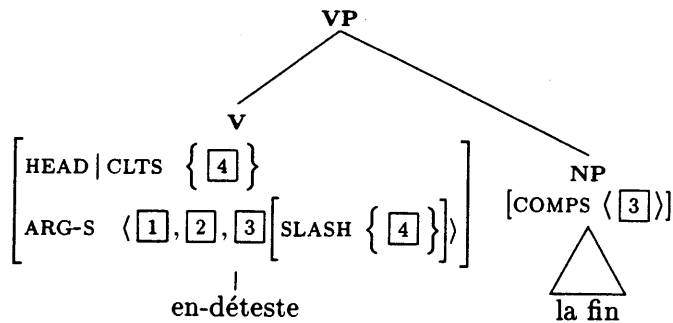
- (30) a. *déteste*₁ (‘hates’) b. *déteste*₂
- | | |
|---|--|
| $\left[\begin{array}{l} \text{HEAD } \textit{verb}[\textit{fin}] \\ \text{SUBJ } \langle [1] \text{NP}[\textit{nom}]_{3s} \rangle \\ \text{COMPS } \langle [2] \text{NP}[\textit{acc}] \rangle \\ \text{ARG-S } \langle [1], [2] \rangle \end{array} \right]$ | $\left[\begin{array}{l} \text{HEAD } \textit{verb}[\textit{fin}] \\ \text{SUBJ } \langle [1] \rangle \\ \text{COMPS } \langle [2] \text{NP}[\textit{acc}], [3] \text{NP}[\textit{de}] \rangle \\ \quad [\text{COMPS } \langle [3] \rangle] \\ \text{ARG-S } \langle [1], [2], [3] \left[\begin{array}{l} \text{LOC } [4] \\ \text{SLASH } \{ [4] \} \end{array} \right] \rangle \end{array} \right]$ |
| <p>c. <i>déteste</i>₃</p> $\left[\begin{array}{l} \text{HEAD } \textit{verb}[\textit{fin}] \\ \text{SUBJ } \langle [1] \rangle \\ \text{COMPS } \langle [2] \text{NP}[\textit{acc}] \rangle \\ \quad [\text{COMPS } \langle [3] \rangle] \\ \text{ARG-S } \langle [1], [2], [3] \left[\begin{array}{l} \text{LOC } [4] \\ \text{SLASH } \{ [4] \} \end{array} \right] \rangle \\ \text{NLOC SLASH } \{ [4] \} \end{array} \right]$ | <p>d. <i>déteste</i>₄ \Rightarrow <i>en-déteste</i></p> $\left[\begin{array}{l} \text{HEAD } \textit{verb}[\textit{fin}, \text{CLTS } \{ [4] \}] \\ \text{SUBJ } \langle [1] \rangle \\ \text{COMPS } \langle [2] \text{NP}[\textit{acc}] \rangle \\ \quad [\text{COMPS } \langle [3] \rangle] \\ \text{ARG-S } \langle [1], [2], [3] \left[\begin{array}{l} \text{LOC } [4] \\ \text{SLASH } \{ [4] \} \end{array} \right] \rangle \\ \text{NLOC SLASH } \{ \} \end{array} \right]$ |

*déteste*₃ is the matrix V in the relative clause *le livre dont il déteste la fin*, as illustrated in (31)a, and *déteste*₄ the V realized as *en-déteste* in the VP *en déteste la fin*, as illustrated in (31)b

(31) a.



b.



There is an important feature of our analysis that must be understood before proceeding to the next section. The two structures in (31) are the only kind of structures that contain unsaturated nominal phrases. Verbs that have not undergone **CCLR**, for example, select only NPs specified as [COMPS < >] and we will assume that the general islandhood of nonpredicative NPs is to be accounted for simply by preventing NPs from being slashed. Thus extraction from an NP or cliticization of an NP argument onto a matrix verb) must involve structures like those in (31), where the unsaturated element, i.e. the element of the NP's COMPS list is itself a slashed element. Such elements must be slashed because they arise only through **NCLR**, which provides an extra *de*-phrase on the verb's COMPS list that is identified with the element on the NP's COMPS list. As we have just seen, **NCLR** outputs must feed **CELR** and **CCLR**, both of which require that the element removed from the verb's COMPS list be slashed. Hence, in virtue of the interaction of our lexical rules and principles of UG (e.g. **VALP**), the element remaining on the NP's COMPS list in structures like (31) must always be slashed.

5 The Argument Structure of French NPs

Having shown how we represent the possibility of cliticization and extraction out of NP, we now turn to the constraints that account for the range of data given in section 2. These data follow from the interaction of general constraints on the argument structure hierarchy (ARG-S value) of nominals, including a general accessibility condition, valid both for extraction and pied-piping cases. The constraints in question all affect non-predicative nouns, whose argument structures are as sketched in (32)

(32) Non-Predicative Nouns

<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="padding: 2px 5px;">HEAD</td> <td style="padding: 2px 5px;"><i>noun</i>[-PRD]</td> </tr> <tr> <td style="padding: 2px 5px;">SPR</td> <td style="padding: 2px 5px;">L1</td> </tr> <tr> <td style="padding: 2px 5px;">COMPS</td> <td style="padding: 2px 5px;">L2</td> </tr> <tr> <td style="padding: 2px 5px;">ARG-S</td> <td style="padding: 2px 5px;">L1 $\cup_{\langle \rangle}$ L2</td> </tr> </table>	HEAD	<i>noun</i> [-PRD]	SPR	L1	COMPS	L2	ARG-S	L1 $\cup_{\langle \rangle}$ L2	The value of ARG-S must satisfy the various constraints discussed below
HEAD	<i>noun</i> [-PRD]								
SPR	L1								
COMPS	L2								
ARG-S	L1 $\cup_{\langle \rangle}$ L2								

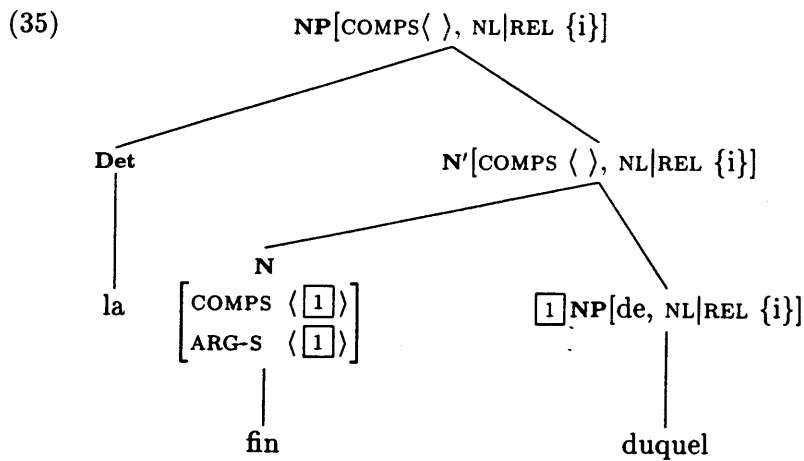
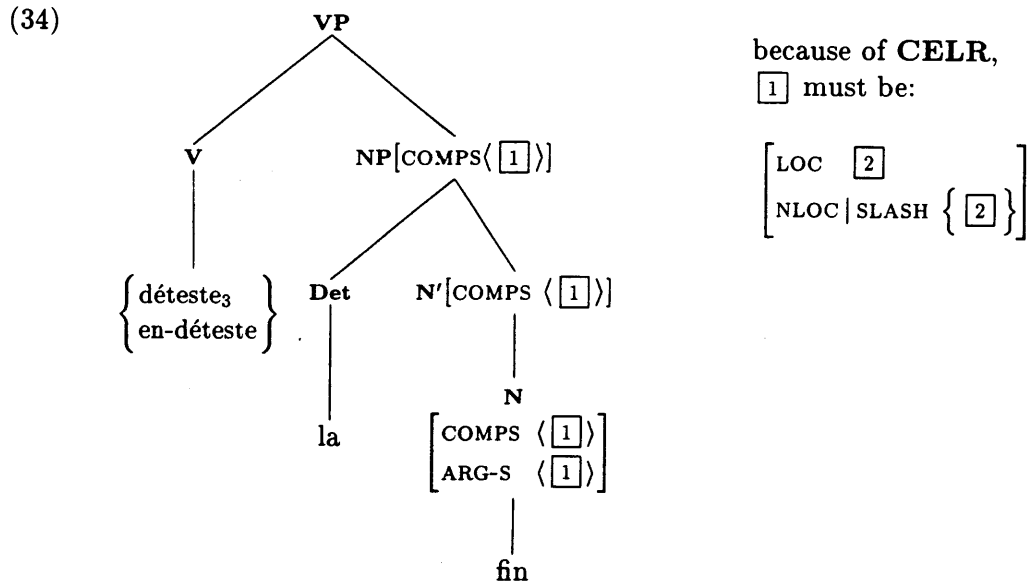
Here $\cup_{\langle \rangle}$ notes the shuffle operation or domain union (Reape in press) of the SPR and the COMPS values. This operation allows us to express the canonical relation between the values of a noun's valence features and its argument structure, namely that the valence lists add up to the ARG-S value, **modulo order**. That is, nominal argument structures differ from their verbal counterparts (see (14)) in not requiring that the ARG-S value be obtained strictly by appending one valence list to another.

The general constraint on accessibility itself can now be stated as in (33), which says that an argument of a noun can have a nonempty value for a NLOC feature F just in case that argument precedes all others within the noun's ARG-S list ('*neset*' denotes a non-empty set, and '<' means 'precedes' within a given list):

(33) Accessibility Condition (on Nominal ARG-S): $\left[\text{NLOC} \mid \text{F } neset \right] < \mathbf{x}$

The NLOC value includes the SLASH feature (for the missing extracted argument), the QUE feature (for interrogative *wh*-words and the pied-piped phrases that contain them) and the REL feature (for relative words and pied-piped phrases). Hence, the constraint in (33) has the effect of allowing the dependent of a nominal to be extracted from an NP, or pied-piped within an NP, only if the dependent is the first member of the head noun's ARG-S list. Thus the various **other** factors that constrain which elements can be initial within a nominal argument structure will constrain extraction and pied piping in a uniform manner.

To see this, we first present the simplest cases, where the noun has a unique argument. This argument is extracted in (34) and it is an NP-internal relative phrase that gives rise to pied-piping in (35):



There are various types of noun that take a single argument: relational Ns such as *frère* ['brother'], *élève (du professeur X)* ['student (of professor X)']; derived nominals corresponding to an adjective or a verb itself having one argument, e.g. *intelligence* ['intelligence'], *venue* ['coming']; Ns which are the head of partitive NPs, denoting a part of the whole denoted by the NP[de] complement, for example *la moitié (de ces gâteaux)* ['half (of these cakes)'], *trois élèves (de cette classe)* ['three students (of this class)'], *les aiguilles (de l'horloge)* ['the hands (of the clock)']. Although the **Accessibility Condition** in (33) applies to all these examples, none of them provides crucial support for (33). The interesting cases, of course, are those nouns that take several arguments.

These divide into two classes: nouns taking diverse kinds of arguments which may be independently extracted or realized as *wh*-NPs, and nouns taking various arguments, only one of which has this property.¹¹ Consider first nouns denoting material, fabricated objects. These occur with more than one relativizable complement. Thus a flat, non-representational N like *maison* may occur with both a possessor and an agent (*la maison de Le Corbusier de Mr X*); a representational N (*interprétation, démonstration*) may occur with an agent and a theme; and a representational N like *portrait* may combine with possessors, agents or themes, though at most two of these three argument types may be realized simultaneously (for independent reasons).

We propose that the data illustrated in (4)-(6) above should be explained by the **Accessibility Condition** in (33), in conjunction with two further constraints on argument structure of nouns. The first reads as in (36) (where [*poss*] applies both to the possessive determiner (e.g. *mon, ma*) and the NP[*de*] which is interpreted as a possessor:

(36) **Possessor Constraint** (on Nominal ARG-S): [*poss*] < X

Consider one instantiation of *portrait*, with a possessor and a theme arguments:

(37)
$$\left[\begin{array}{l} \text{COMPS } \langle \boxed{2} \text{NP}[\text{de}]_{th}, \boxed{1} \rangle \\ \text{ARG-S } \langle \boxed{1} \text{NP}[\text{de}, \text{poss}], \boxed{2} \rangle \end{array} \right]$$

It follows from the two constraints in (33) and (37) that only the possessor can be extracted or realized as a *wh*-word, as illustrated in (38):

- (38) a. Pierre a admiré le portrait [de Rosa la Rouge] [de Barnes].
(‘P. has admired the portrait of R.l.R. of Barnes.’)
- b. Pierre a admiré son portrait de Rosa la Rouge
(‘P. has admired his portrait of R.l.R.’)
- c. Pierre en a admiré le portrait de Rosa la Rouge.
(‘P. of-it has admired the portrait of R.l.R.’)
- d. *Rosa la Rouge dont Marie a admiré le portrait de Barnes
(‘R.l.R. of-which M. has admired the portrait of Barnes.’)
- e. *Rosa la Rouge au portrait de Barnes de laquelle je me suis intéressé
(‘R.l.R. in the portrait of Barnes of whom I got interested’)

The **Possessor Constraint** in (36) also accounts for the restrictions on the interpretation of the possessive determiner noted in Milner (1977, 1982). DET[*poss*] may not be interpreted as a theme if there is an overt possessor argument (similarly, the DET may not be interpreted as an agent if the two arguments are agent and possessor). Thus, while (38b) is good, (39) is unacceptable:

¹¹We will not address here the question of whether an NP is an argument of the basic noun or is introduced by a lexical rule.

(39) *Son portrait de Barnes

Such an NP contains two elements specified as [*poss*] and there is no way that both of them can be initial in the noun's ARG-S value, as (36) would require.¹²

Likewise, one cannot extract or pied-pipe from the theme in *son interprétation de la neuvième*, because the NP contains an NP[*poss*] element (*son*, interpreted here as Agent) which must be the highest argument:

- (40) a. *la neuvième, dont j'ai entendu son interprétation
(‘the Ninth, of-which I have heard its/his interpretation’)
- b. *J'en ai entendu son interprétation.
(‘I have heard its/his interpretation of it’)
- c. *la neuvième, à son interprétation de laquelle je me suis intéressé
(‘the Ninth, in its/his interpretation of which I'm interested’)

The second constraint on nominal argument structures, which interacts with the **Accessibility Condition** (33) is the following:

- (41) **Agent Constraint** (on Nominal ARG-S): $NP_{ag} < NP_{th}$

From the interaction of (33) and (41), it follows that the theme may not be extracted or realized as a *wh*-word if the agent is present, cf. the data in (4) and (42)-(43) :

- (42) a. La jeune fille dont le portrait est à la Fondation Barnes (‘the young woman of-which the portrait is at the B. F.’)
- b. Le portrait en serait à La Fondation Barnes (‘the portrait of-her would be at the B. F.’)
- c. ?Le portrait par Corot en serait à la Fondation Barnes (‘the portrait by C. of-her would be at the B. F.’)
- (43) a. *La jeune fille dont le portrait de Corot est à la Fondation Barnes (‘the young woman of-which the portrait of C. is at the B. F.’)
- b. *Le portrait de Corot en serait à la fondation Barnes (‘the portrait of C. of-her would be at the B. F.’)

Constraints (41) and (36) together also explain why the possessive DET may not be interpreted as a theme if there is an agent argument (Milner *idem*):

- (44) a. Son portrait de la jeune fille est à la fondation Barnes (‘his portrait of the young woman is at the B. F.’)

¹²By similar reasoning, the **Accessibility Condition** disallows both multiple extraction from and multiple pied-piping within the French NP.

EXTRACTION FROM THE FRENCH NP

- b. *Son portrait de Corot est à la fondation Barnes ('her portrait of C. is at the B. F.')

Note that this analysis implies that the agent and the possessor arguments with a flat or a representational N are optional not only in the syntax, but also in the ARG-S. It is only when the agent is not present that the theme may become the highest argument, cf. (42a,b), and only when the possessor is not present that the agent may be the highest argument – cf. (44). Our analysis also implies that the *par*-NP agent either does not belong to the ARG-S (it is a modifier, as in Milner 1986 and Zubizarreta 1987) or else it is a 'demoted' argument, as in Relational Grammar (also Pollard and Sag 1987).

Let us now turn to the more problematic cases: the Ns which have one extractable complement only, although they have several arguments. They are: (a) nouns derived from a V with a prepositional complement, cf. *participation* (*à*), *descente* (*à, vers*), *éloignement* (*de*) ['participation (in)', 'descent (to/towards)', 'estrangement (from)'], (b) nouns derived from an adjective with a prepositional complement, cf. *aptitude* (*à*) ['ability (for)'], and (c) psychological nouns, which take both an experiencer and a theme argument. With nouns of type (a) and (b), the argument that corresponds to the subject of the S is extractable, not the argument corresponding to the prepositional complement. This is the case whether the other argument is syntactically realized or not, as illustrated by examples we cited earlier ((1) and (8a,b)).

Similarly, with psychological Ns, the experiencer is extractable, but not the theme, even if the experiencer is not syntactically present, as shown in (45) and again by our earlier examples ((3) and (8c,d)):

- (45) a. La passion du jeu des aristocrates russes ('the passion of (= for) gambling of the Russian aristocrats')
 La peur des serpents de nos ancêtres ('the fear of snakes of our ancestors')
- b. Les aristocrates russes dont la passion du jeu est aujourd'hui incompréhensible ('the Russian aristocrats of-which the passion of gambling is impossible to understand nowadays')
 Nos ancêtres dont la peur des serpents a été bien étudiée ('our ancestors of-which the fear of snakes has been well documented')

This array of data follows from our analysis if the highest argument of nouns in class (a) and (b) above (those related to preposition-taking verbs and adjectives) is the argument that the verb or noun realizes as its subject. Additionally, the highest argument of the psychological noun must be the experiencer. Moreover, these elements must be present in the noun's argument structure, even when they are not syntactically realized. A more complete study of derived and psychological Ns is clearly called for here, but for the moment, our analysis includes the following four hypotheses: (1) a noun's ARG-S value sometimes may contain elements not selected by that noun's SPR or COMPS value. Such noncanonical ARG-Ss may thus contain an element which is not syntactically realized. (2) the lexical rules

which give the derived nominals from the basic A or V do not change the ARG-S (when the N has the same semantic sort as the V). (3) there are derived nominals based on the passive form of Vs of action (accomplishment, or achievement), and the highest argument of a passive corresponds to the second argument of an active V. (4) there are no derived nominals based on the passive form of psychological Vs. Hypotheses (1)-(3) would explain why the argument of *construction*, which corresponds to the direct object of the V, is extractable/relativizable, in contrast with the prepositional argument of *participation* in (1b), for instance: there is no passive form of Vs with prepositional complements in French.

- (46) a. L'Opéra dont la construction sera bientôt achevée ('the Opera of-which the construction will soon be finished')
- b. L'Opéra sera bientôt construit ('the Opera will soon be finished')
- c. La participation à la conférence est surtout américaine ('the participation to the conference is mainly american')
- d. *La conférence a été largement participée (à) ('the conference was widely participated to')

As for psychological nouns, what is important for our analysis is that, even if they are derived from a verb, they do not have a passive form, as can be seen from the well-known fact that no such noun takes a *par*-NP complement: **la passion du jeu par les aristocrates russes*, **la peur des serpents par nos ancêtres*.

Finally, let us return to the **Accessibility Condition** (33). Because it is stated in terms of NONLOCAL features, it makes a prediction beyond those already discussed. Because REL values are 'percolated' from daughter to mother in NPs and PPs throughout a pied-piping construction, (33) also entails that in complex pied-piping, only the highest argument of the highest argument of the pied-piped NP may be a *wh*-NP. This is exactly what we find, as shown by the contrast between (47) and (48):

- (47) a. Je me suis intéressé aux œuvres de Matisse/du père de P. Matisse ('I got interested in the works of M./of the father of P.M.')
- b. Matisse aux œuvres de qui je me suis intéressé ('M. in the works of whom I got interested')
- c. P. Matisse aux œuvres du père de qui je me suis intéressé ('P.M. in the works of the father of whom I got interested')
- (48) a. Il faut armer les parents contre [la peur [des examens des enfants]]. ('It is necessary to arm the parents against the fear of the exams of the children.')
- b. *les examens des enfants contre la peur desquels il faut armer les parents ('the exams of the children against the fear of which it is necessary to arm the parents')

- c. *les enfants contre la peur des examens desquels il faut armer les parents
(‘the children against the fear of the exams of whom it is necessary to
arm the parents’)

The most embedded NP in (47c) is a *wh*-word, that is, an NP which lexically bears the nonempty REL value. By (33), the NP[*de*] *de qui* then must be the highest argument of the noun *père* and the phrase *du père de qui* must be the highest argument of *œuvres*. All this follows simply because the HPSG treatment of pied-piping is based on percolation of REL specifications, governed by the **NONLOCAL Feature Principle**). This in turn causes the **Accessibility Condition** to ‘kick in’ at every nominal level in complex cases like (47c). On the other hand, although the most embedded phrase in (48) is the highest argument of the governing noun (as evidenced by the extractability in *les enfants dont les examens sont toujours une épreuve pour les parents* [‘the children of-which the exams are always a test for the parents’]), this NP itself is not the highest argument in the pied-piped NP. Thus, the relatives in (48) are correctly ruled out as violations of the **Accessibility Condition**.

6 Conclusion

We began this study from the strict lexicalist perspective of HPSG, whose rich lexical entries are organized by hierarchical inheritance and lexical rules. Our framework in turn leads us to analyze syntactic dependencies not in terms of movement operations, but rather in terms of constraints on feature structures, some universal – some language-particular. Our treatment of extraction, cliticization, and pied-piping all in terms of NONLOCAL features leads us to the fundamental constraint proposed in this paper: the **Accessibility Condition** on nominal argument structures. This condition, through its interaction with the other constructs of our HPSG treatment of French grammar, correctly predicts a wide range of subtle constraints on extraction from the French NP, unifying the restricted accessibility of extraction with that of pied-piping. Previous treatments that we are familiar with, all of which attempt to derive the constraints on extraction accessibility from general constraints on extraction, are unable to achieve a comparable result, chiefly because they have no way to express the linguistically significant generalization uniting extraction and pied piping. As is well known (see section 2 above), the constraints on pied piping are in general distinct from those on extraction. The relevant overlap in these constraints is precisely as described by the **Accessibility Condition** and the rather complex set of deductions that follow from its interaction with independently motivated aspects of our analysis.

Bibliography

- Abeillé, Anne, and Danièle Godard. 1993. *The Syntactic Structure of French Auxiliaries*. Unpublished manuscript – Paris: CNRS, U. Paris 7 and U. Paris 8.
- Abeillé, Anne, Danièle Godard, Philip Miller, and Ivan A. Sag. in progress. *The Major Syntactic Structures of French*.
- Ajdukiewicz, K. 1935. Die Syntaktische Konnexität. In *Studia Philosophica* 1: 1–27. [Translated as ‘Syntactic Connexion’ in S. McCall, ed., 1967, *Polish logic*. Oxford: Clarendon Press, pp. 207–231.]
- Copestake, Ann. 1992. *The Representation of Lexical Semantic Information*. Doctoral dissertation, University of Sussex [Published as Cognitive Science Research Paper CSRP 280].
- Flickinger, Daniel. 1987. *Lexical Rules in the Hierarchical Lexicon*. Doctoral dissertation, Stanford University.
- Flickinger, Daniel, and John Nerbonne. 1992. Inheritance and Complementation: A Case Study of *Easy* Adjectives and Related Nouns. In Walter Dalemans and Gerald Gazdar, eds., *Inheritance and Natural Language Processing*, special issue of *Computational Linguistics* 19.3: 269–309.
- Gazdar, Gerald. 1981. Unbounded Dependencies and Coordinate Structure. *Linguistic Inquiry* 12: 155–184.
- Giorgi A., and G. Longobardi. 1992. *The Syntax of Noun Phrases: Configurations, Parameters and Empty Categories*. Cambridge: CUP.
- Godard, Danièle. 1992. Extraction out of NP in French. *Natural Language and Linguistic Theory* 10: 233–277.
- Harris, Zellig S. 1946. From Morpheme to Utterance. *Language*: 161–183. [Reprinted in Martin Joos, ed., 1957, *Readings in Linguistics I*. Chicago: University of Chicago Press]
- Hinrichs, Erhard, and Tsuneko Nakazawa. 1989. Flipped Out: Aux in German. *CLS* 25. Chicago: Chicago Linguistic Society, 193–202.
- Hinrichs, Erhard, and Tsuneko Nakazawa. 1990. Subcategorization and VP Structure in German. In Hughes, Shaun, and Joe Salmons, eds., *Proceedings of the Third Symposium on Germanic Linguistics*. Amsterdam: Benjamins.
- Kathol, Andreas. 1994. On Agreement in HPSG. Unpublished paper, Ohio State University.
- Miller, Philip. 1992. *Clitics and Constituents in Phrase Structure Grammar*. New York: Garland. [Ph. D. dissertation, U. of Utrecht, 1991].

- Miller, Philip and Ivan A. Sag. 1993. French Clitic Movement Without Clitics or Movement. Unpublished manuscript: U. de Lille 3 and Stanford University. [presented at 1993 LSA Meeting – Los Angeles]
- Milner, Jean-Claude. 1978. *De la syntaxe à la Interprétation – Quantités, Insultes, Exclamations*. Paris: Le Seuil.
- Milner, Jean-Claude. 1982. *Ordres et Raisons de Langue*. Paris: Le Seuil.
- Milner, Jean-Claude. 1986. Introduction à un Traitement du Passif. *Collection URA 642, DRL, U. Paris 7*.
- Moortgat, Michael. 1984. A Fregean Restriction on Metarules. In Charles Jones and Peter Sells, eds., *Proceedings of the Fourteenth Annual Meeting of the North Eastern Linguistic Society*, 306–325.
- Pollard, Carl and Ivan A. Sag. 1987. *Information-Based Syntax and Semantics, Volume 1: Fundamentals*. CSLI Lecture Notes Series No. 13. Stanford: Center for the Study of Language and Information. [Distributed by University of Chicago Press]
- Pollard, Carl, and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press and Stanford: CSLI Publications.
- Pollock Jean-Yves. 1989. Opacity, Genitive Subjects, and Extraction from NP in English and French, *Probus* 1: 151–162.
- Reape, Mike. In press. Getting Things in Order. In Arthur Horck and Wietske Sijtsma, eds., *Discontinuous Constituency*. Berlin: Mouton de Gruyter.
- Riehemann, Susanne. 1993. *Word Formation in Lexical Type Hierarchies: A Case Study of bar-Adjectives in German*. Master's thesis, University of Tübingen.
- Ruwet, Nicolas. 1972. *Théorie Syntaxique et Syntaxe du Français*. Paris: Le Seuil.
- Sportiche, Dominique. 1989. Le Mouvement Syntaxique: Contraintes et Paramètres. *Langages* 95: 35–80.
- Stowell Timothy. 1989. Subjects, Specifiers and the X-bar Theory, in Mark Baltin and Anthony Kroch, eds., *Alternative Conceptions of Phrase Structure*. Chicago: University of Chicago Press, pp. 232–262.
- Zubizarreta, Maria-Luisa. 1987. *Levels of Representation in the Lexicon and in the Syntax*. Dordrecht: Foris Publications.

