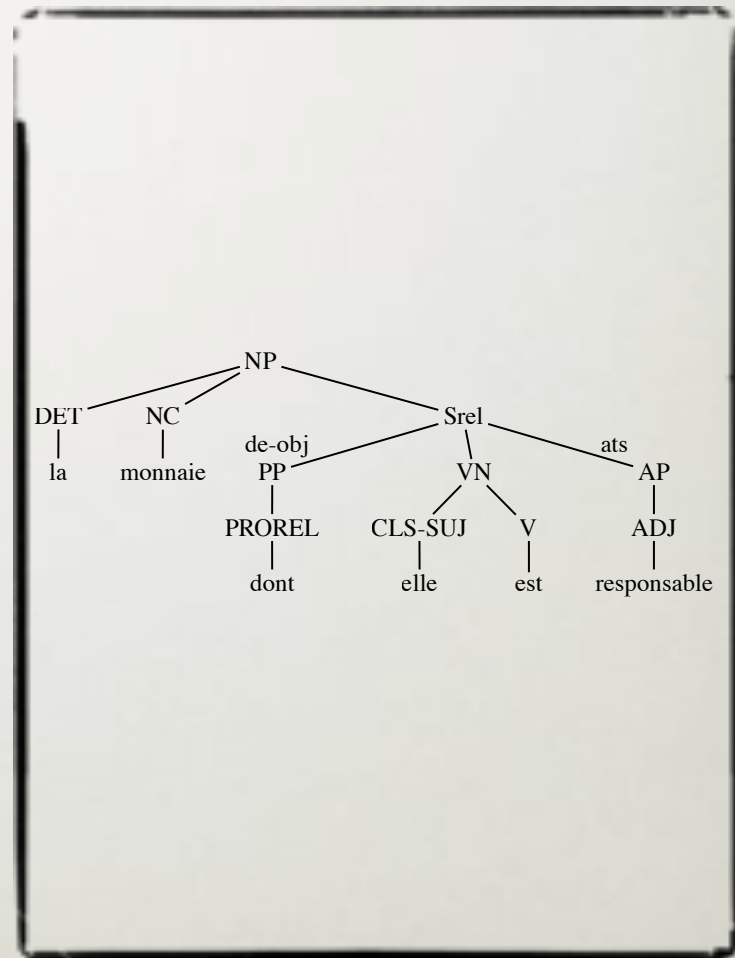


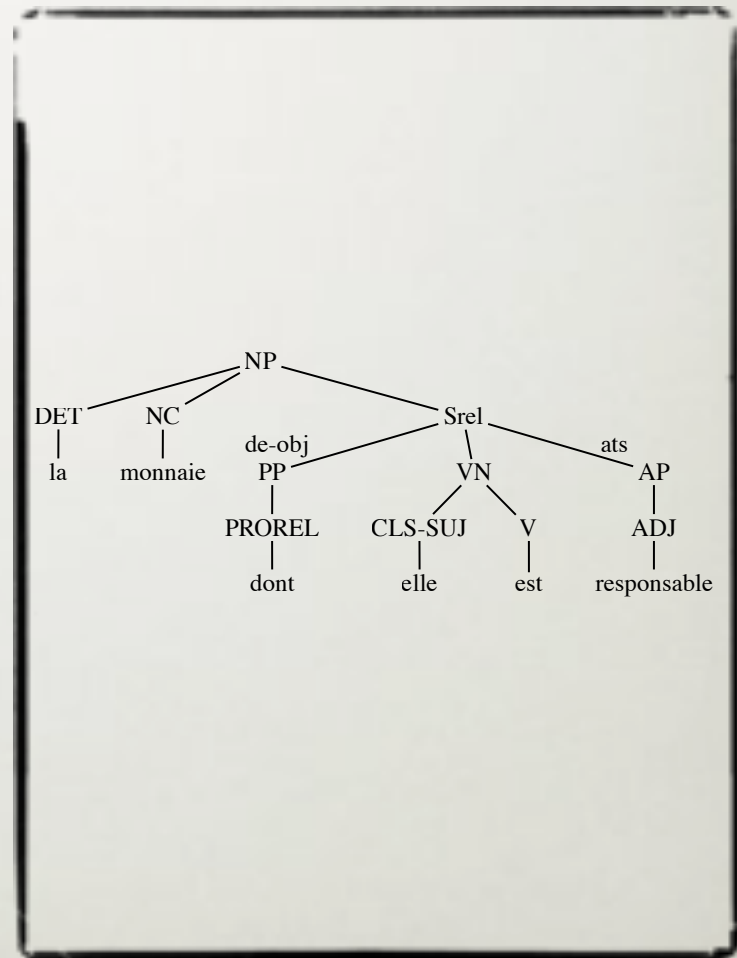
CATEGORIAL GRAMMAR
AND
FRENCH SYNTAX

RICHARD MOOT

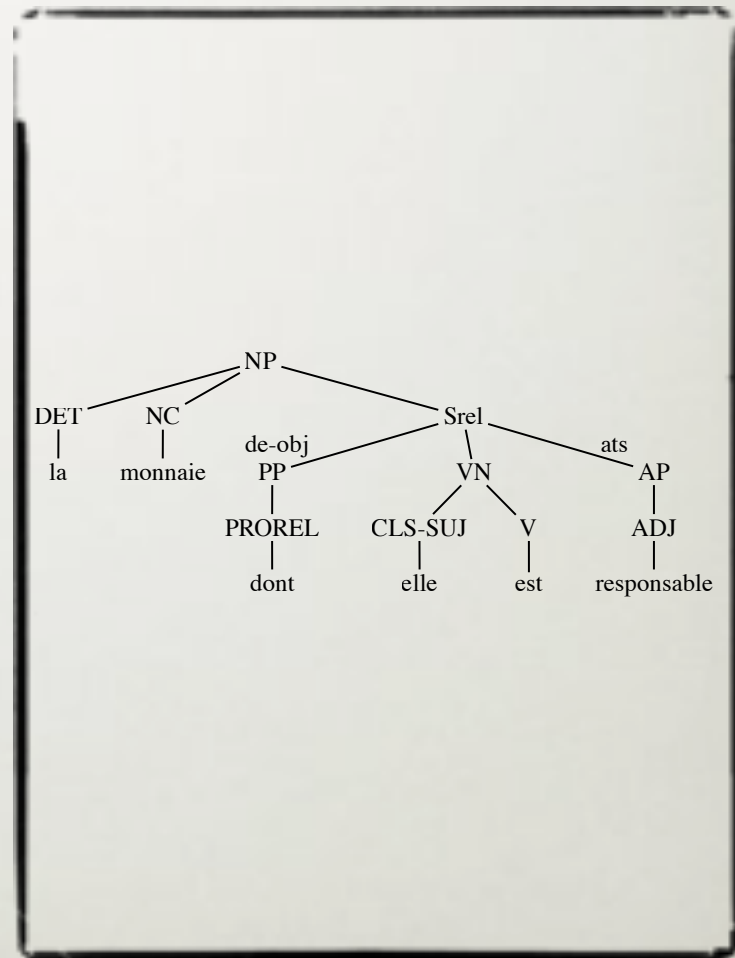
- In this first talk about the principles of categorial grammars for analyzing the french language, I will talk about the basic building blocks.



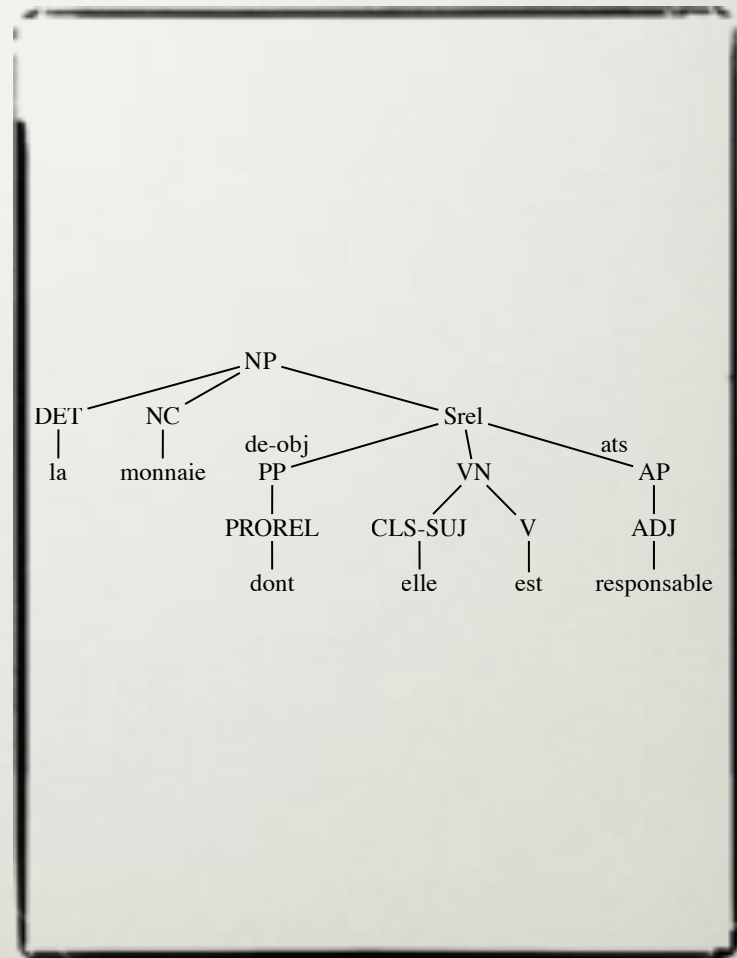
-
- The goal is to give an analysis which is as simple as possible.
 - We will use the annotated treebank of Paris VII as a guideline.



-
- I will start at the very beginning, with the basic building blocks (the atomic formulas)
 - I will (informally) introduce the concepts of categorial grammar as we progress.



-
- Even though we start simple, we will quickly see there are some tough choices to make.
 - I welcome any comments and suggestions for improvements!



ATOMIC FORMULAS

np	<i>Paris, Milan, M. John Major, cette époque</i>
n	<i>mois, secrétaire d'Etat, qualité</i>
pp	<i>à Milan, à son frère Michel, de contrefaçons</i>
s	<i>Il s'agissait de contrefaçons, Libération cherche à augmenter ses recettes</i>

"noun phrase"

"noun" or "common noun"

"prepositional phrase"

ATOMIC FORMULAS

- np
- n
- s
 - S_{ppart} , S_{ppres} , S_{inf} , S_{ainf} , S_{deinf}
- pp
 - $pp_{\grave{a}}$, pp_{de} , pp_{par} , ...

We will introduce a sort of "subtyping", some slightly more refined types to make useful selection restrictions on the arguments. These subtypes will be motivated by several examples on the slides which follow.

COMPLEX FORMULAS

- All atomic formulas are formulas
- if A and B are formulas, then
A/B and B \ A are formulas as well

$$\frac{A/B \quad B}{A} [/ E]$$

$$\frac{B \quad B \backslash A}{A} [\backslash E]$$

A/B looks for a formula B to its right to form an A

B \ A looks for a formula B to its left to form an A

Of course this is a logic and there are

COMPLEX FORMULAS

np	(np)\s/np	np
Jean		Marie

A transitive verb looks first for a noun phrase to its right ("Marie" in this case) to form an intransitive verb np\s

Then it combines with a noun phrase to its left ("Jean" in this case) to form a sentence.

COMPLEX FORMULAS

np	(np \ s) / np	np / n	n
Jean	aime	la	logique

The determiner "la" looks for a noun to its right to form a noun phrase.

Therefore the two expressions "la logique" and "Marie" can be assigned the same formula: np

COMPLEX FORMULAS

np	(np \s) / pp	pp / np	np / n	n
Il	travaillait	à	une	réforme

Some verbs select a prepositional argument (though sometimes the distinction between argument and modifier can be subtle, as we will see in the slides which follow)

COMPLEX FORMULAS

np	(np \ s) / pp	pp / np	np / n	n
Il	travaillait	à	une	table

Difference between arguments and modifiers

COMPLEX FORMULAS

np	(np \ s)	(s \ s) / np	np / n	n
Il	travaillait	à	une	table

Difference between arguments and modifiers

The difference between this analysis and the one on the previous slide is that here "à une table" is analyzed as a locative, the table is the place where the subject is working, whereas on the previous analysis, the table was the object being made by the subject.

COMPLEX FORMULAS

np	(np \ s) / pp	pp / np	np / n	n
Il	travaillait	à	une	réforme

COMPLEX FORMULAS

	np	(np \ s) / pp	pp / np	np / n	n
*	Il	travaillait	dans	une	réforme

?! This seems strange

pp's are generally difficult: they can be either arguments of the verb, adverbs (modifying the verb) or noun modifiers and the distinctions (especially between adverb and verb argument) are sometimes subtle

Fortunately, the functional annotation of the Paris VII corpus and resources such as Dicovalence help make the right decisions.

The verb "travailler" does not allow a locative pp as an argument. Though "travailler dans une école" is certainly possible, I prefer not to analyse this by having "travailler" select a locative pp argument, since this argument is not "inherent" in the verb. Other verbs, such as "rester" or "aller" do have a good reason to select a pp argument

COMPLEX FORMULAS

np	(np \ s) / pp _à	pp _à / np	np / n	n
Il	travaillait	à	une	réforme

?! This seems strange

It makes sense to have the verb in
which prepositions it select as argu

The new formula for the
verb requires a
preposition headed by "à"

The new formula for "à"
indicated that it is a
preposition headed by "à"

COMPLEX FORMULAS

np	(np \ s) / pp _à	pp _à / np	np / n	n
Il	travaillait	à	une	réforme

This multiplies the number of formulas for verbs, but is justified by

"travailler" combines with several prepositions: "à", "sur", "avec", "pour", "contre".

The difference between "travailler à" and "travailler sur" is a bit subtle: "travailler sur" suggest work on an item already in a more-or-less finished state, a work of refinement, whereas "travailler à" is more a work of creation.

COMPLEX FORMULAS

np	(np \ s) / pp _{sur}	pp _{sur} / np	np / n	n
Il	travaillait	sur	une	réforme

This multiplies the number formulas for many verbs, but is justified by semantic differences

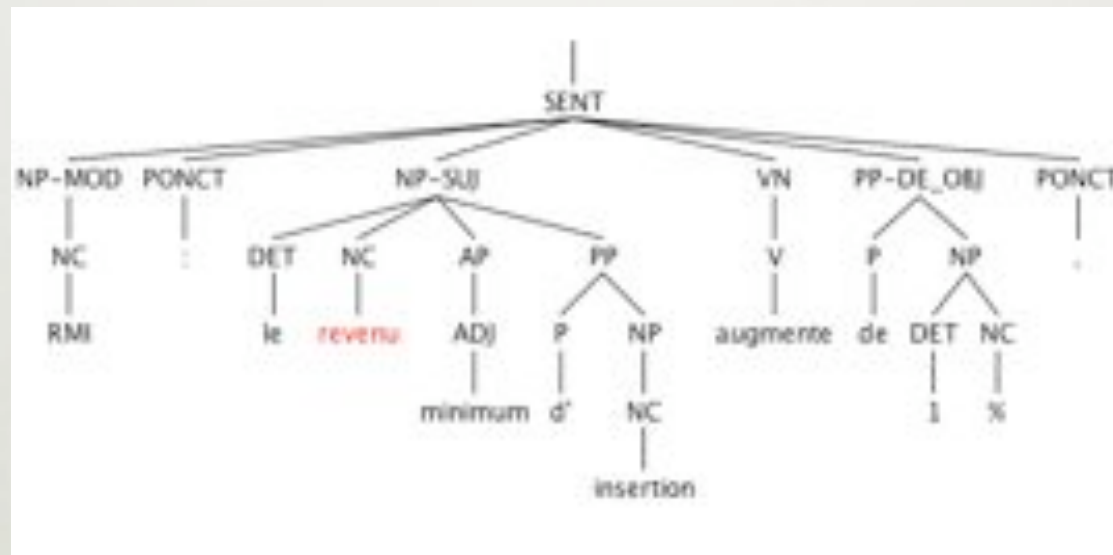
COMPLEX FORMULAS

np	(np \ s) / pp _{avec}	pp _{avec} / np	np / n	n
Il	travaillait	avec	le	gouvernement

This multiplies the number formulas for many verbs, but is justified by semantic differences

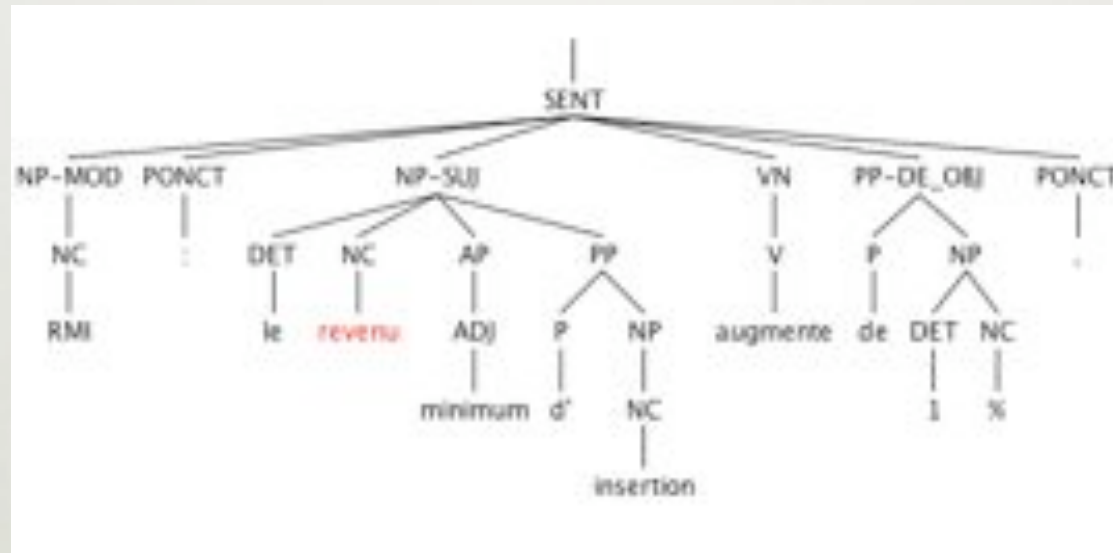
COMPLEX FORMULAS

Example from the corpus



COMPLEX FORMULAS

Example from the corpus



augmente: $(np \setminus s) / pp_{de}$

COMPLEX FORMULAS

np	(np \ s) / (np \ s)	(np \ s) / pp _a	pp _a / np	np
Jean	est	arrivé	à	Paris

Similarly, the verbs “*être*” and “*avoir*”
combine with past participle
simple tensed verb

Unfortunately, the Paris VII corpus analyses “est arrivé” as a single VN constituent (in general the verb cluster and clitics are treated in this way). This means that some rebracketing is necessary to get the analysis displayed here.

COMPLEX FORMULAS

np	(np \ s) / (np \ s)	(np \ s) / pp _a	pp _a / np	np
Jean	est	arrive	à	Paris

Similarly, the verbs “*être*” and “*avoir*” combine with past participles not with simple tensed verbs

COMPLEX FORMULAS

np	$(np \setminus s) / (np \setminus S_{ppart})$	$(np \setminus S_{ppart}) / pp_{\grave{a}}$	$pp_{\grave{a}} / np$	np
Jean	est	arrivé	à	Paris

We can code this selection restriction in the formula for “*est*”

A MORE COMPLEX EXAMPLE

- Le raffermissement a été provoqué par des ventes de marks au profit du yen , à de nouvelles appréhensions sur le sort de l' ex- URSS - hypothèque qui ne cessera de peser sur la devise allemande dans les prochains mois - et à l' évolution de la conjoncture en Allemagne .

other verbs such as "cesser" select an infinitive headed by "de" as an argument

COMPLEX FORMULAS

$(np \setminus s) / (np \setminus s_{deinf})$	$(np \setminus s_{deinf}) / (np \setminus s_{inf})$	$(np \setminus s_{inf}) / pp_{sur}$	pp_{sur} / np	np
cessera	de	peser	sur	...

ANOTHER COMPLEX EXAMPLE

- Dès le 9 août dernier , M. Pierre Jaans , directeur de l' Institut monétaire luxembourgeois (IML) , autorité bancaire centrale du Grand-Duché , avait menacé de demander à la Cour de Luxembourg de liquider la BCCI SA (Le Monde du 9 août 1991) .

"menacer" + deinf

"demander" + ppa + deinf

COMPLEX FORMULAS

$((np \setminus S_{inf}) / (np \setminus S_{deinf})) / pp_{\grave{a}}$	$pp_{\grave{a}}$	$(np \setminus S_{deinf}) / (np \setminus S_{inf})$	$(np \setminus S_{inf}) / np$	np
demander	à ...	de	liquider	...

More on these kinds of examples when I talk about semantics (subject control versus object control)
 "Jean persuade Marie de venir" "Jean aide Marie à venir" (Marie vient)
 "Jean aime venir" (Jean vient")

"Jean demande de louer un livre" (Jean loue un livre)
 "Jean demande Marie de louer un

ANOTHER COMPLEX EXAMPLE

- Apparemment , cette reprise tarde à venir , si tant est qu' elle vienne et , surtout , élément capital sur les marchés de change , l' écart de taux d' intérêt entre les Etats-Unis et l' Allemagne s' est encore creusé cette semaine , passant de 6 % à 6,50 % .



"tarder" + ainf

COMPLEX FORMULAS

np/n	n	(np \ s) / (np \ Sainf)	(np \ Sainf) / (np \ Sinf)	np \ Sinf
cette	reprise	tarde	à	venir

“Il me tarde de aller à Paris” is also possible, but again the semantics - as well as the syntax - are quite different (eg. impersonal “il” only)

PRESENT PARTICIPLE

- The present participle is most frequently used in the context “*en + (np \ S_{ppres})*” (eg. “*en voyageant*”, “*en franchissant les frontières*”)
- Therefore (s \ s) / (np \ S_{ppres}) is one of the possible formula assignments to “*en*”

OTHER INFINITIVE GROUPS

- infinitive groups headed by the preposition “*pour*” are not assigned the formula $(np \setminus s_{pourinf}) / (np \setminus s_{inf})$ but the formula $(s \setminus s) / (np \setminus s_{inf})$ and (more rarely) $pp_{pour} / (np \setminus s_{inf})$, eg. “*nécessaire pour contrôler l’inflation*”
- the case for “*sans*” is similar.

OTHER INFINITIVE GROUPS

- finally, the combination “*commencer / finir par + infinitive*” is analysed by assigning $pp_{par} / (np \setminus s_{inf})$ to the word “*par*”, eg. “*certain finissent par trouver des différences*”

OTHER INFINITIVE GROUPS

Good question:

why not use $pp_{\grave{a}} / (np \setminus s_{inf})$ and $pp_{de} / (np \setminus s_{inf})$?

OTHER INFINITIVE GROUPS

Good question:

why not use $pp_{\grave{a}} / (np \setminus s_{inf})$ and $pp_{de} / (np \setminus s_{inf})$?

- in favor: a verb like “*commencer*” is assigned $(np \setminus s) / pp_a$ which permits both “*commence à lire*” et “*commence à un livre*”
- there are quite a few verbs which follow this pattern

OTHER INFINITIVE GROUPS

Good question:

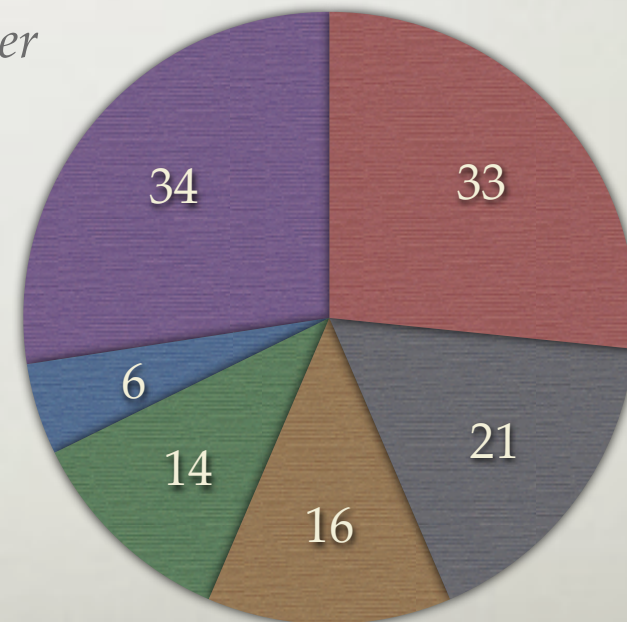
why not use $pp_{\grave{a}} / (np \setminus s_{inf})$ and $pp_{de} / (np \setminus s_{inf})$?

- problem: quite a few other verbs allow only an $\grave{a}inf / deinf$ argument, eg. “*les entreprises cherchent à diminuer leur coûts*” contre “**les entreprises cherchent à une réduction de leur coûts*”
- so it makes sense to distinguish $(np \setminus s_{ainf})$ from pp_a

THE EXTRACTED LEXICON

- Formula assignments to the present tense verb form “fait”
- 124 occurrences in the corpus with 19 different formulas assigned to it.

- $(np \setminus s)/np$
- $((np \setminus s)/pp_de)/np$
- $(np \setminus s)/(np/s_inf)$
- $((np \setminus s)/pp_a)/np$
- $((np \setminus s)/np)/(np \setminus s_inf)$
- *other*



ADJECTIVAL USES OF PARTICIPLES

La société Dassault systèmes , filiale de Dassault aviation , spécialisée dans la conception et la fabrication assistées par ordinateur (CFAO)

"assistées" is another example, which will have the formula $(n \setminus n) / pp_par$, which shows how these cases are very similar to passives

"spécialisée" is a past participle, but it is used as an adjective - that is to say a noun modifier - in the current sentence. This means that the correct formula to assign it would be $(n \setminus n) / pp_dans$

These cases are very frequent. It seems tempting to treat them with a (non-logical) lexical rule; this is the solution used for the CCGbank. It has the additional advantage that adverbs can keep their normal $((np \setminus s) \setminus (np \setminus s))$ formulas, combine with the past participle, then apply the non-logical rule to transform the result to $n \setminus n$

Currently, the extracted grammar gives explicit $(n \setminus n)$ categories to past participles used as adjectives (and $(n \setminus n) \setminus (n \setminus n)$ to adverbs modifying them). Lexical ambiguity is reduced (and supertagger performance is improved) by adopting the CCGbank strategy.

ADJECTIVAL USES OF PARTICIPLES

Les 60 % du capital restants demeureront tchèques

"restants" = present participle used as an adjective

"demeureront" is a copula verb and therefore has as one of its possibilities to select an adjective - like "tchèques" in the current example - to its right.

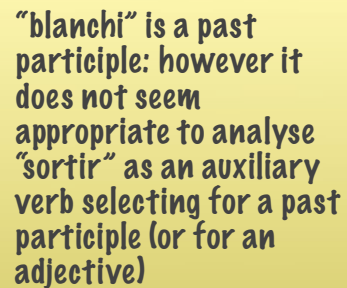
Les reporters demeureront à leur antenne

the second example shows how "demeureront" can also select a past participle, which is an indication that `n\n` is a good formula in this context as well

Question: does it make sense to analyse "être" the same way as "demeurer" that is to say replacing all `(np\s)/(np\s_ppart)` formulas for "être" by `(np\s)/(n\n)`?

ADVERBIAL USES OF PARTICIPLES

Au demeurant , M. Tapie se dit convaincu qu' il sortira blanchi de l' affaire Toshiba ...



"blanchi" is a past participle: however it does not seem appropriate to analyse "sortir" as an auxiliary verb selecting for a past participle (or for an adjective)

ADVERBIAL USES OF PARTICIPLES - “UNE FOIS”

Une fois les réformes lancées

une fois acquise la question de " l' équilibrage
structurel " et donc résolue celle du déficit financier
par des mesures nouvelles

BEYOND AB GRAMMARS

STRUCTURAL RULES

- Even though it is permitted to assign many formulas to a single word, sometime we can generalize the *pattern* of this set of formulas by using structural rules
- The most important (and the clearest) case, is the case of adverbs.

BEYOND AB GRAMMARS

STRUCTURAL RULES

... condamnent à l'unanimité le projet

- In the exemple above, “*condamner*” is a transitive verb taking “*le projet*” as its object argument.
- However, the adverb “*à l'unanimité*”, which modifies the verb is placed between the verb and its object.

BEYOND AB GRAMMARS

STRUCTURAL RULES

... condamnent à l'unanimité le projet

- What formula do we assign to “à *l'unanimité*”?
- Preferably, the same formula would work for “*condamnent le projet à l'unanimité*” as well

BEYOND AB GRAMMARS

STRUCTURAL RULES

... condamnent à l'unanimité le projet

- The multimodal solution is to say: “à l’unanimité” has type $s \setminus_1 s$ where the mode 1 indicates that the adverb (optionally) “move”
- this means “à” has form
- in addition we have $(s \setminus_1 s) / np \vdash (s \setminus s) / np$

The definition of movement is very liberal in the extracted grammar. It is basically: modify an “s” anywhere. This is like a logic where a commutative and a non-commutative implication live together. An interesting open question is: what are the right restrictions on movement?

BEYOND AB GRAMMARS

STRUCTURAL RULES

- In the resulting system a non-commutative and a commutative logic live together.
- The commutative implication $A \setminus_1 B$ is allows generalizations of the lexical formulas assigned to words.

BEYOND AB GRAMMARS

INTRODUCTION RULES

- Even though we have seen formulas such as $(np \setminus s) / (np \setminus s_{ppart})$ which take a complex formula as their argument, we have not (yet) done anything interesting with this property.

BEYOND AB GRAMMARS

INTRODUCTION RULES

... $[B]^i$

$[B]^i$...

$$\frac{A}{A/B} [/ I]^i$$

$$\frac{A}{B/A} [\backslash I]^i$$

$[B]$ is the rightmost free hypothesis in the proof

$[B]$ is the leftmost free hypothesis in the proof

“...” contains at least one other non-discharged formula

Each introduction rule discharges exactly one formula.

A unique integer i links the rule application with this formula

BEYOND AB GRAMMARS

INTRODUCTION RULES

n	$(n \setminus n) / (s / np)$	np	$(np \setminus s) / (np \setminus s_{ppart})$	$(np \setminus s_{ppart}) / np$
redaction	qu'	on	a	créée

BEYOND AB GRAMMARS

INTRODUCTION RULES

n	$(n \setminus n) / (s / np)$	np	$(np \setminus s) / (np \setminus s_{ppart})$	$(np \setminus s_{ppart}) / np$
redaction	qu'	on	a	créée

$$\frac{\text{créée} \quad (np \setminus s_{ppart}) / np \quad np}{np \setminus s_{ppart}} [/ E]$$

BEYOND AB GRAMMARS

INTRODUCTION RULES

n	$(n \setminus n) / (s / np)$	np	$(np \setminus s) / (np \setminus s_{ppart})$	$(np \setminus s_{ppart}) / np$
redaction	qu'	on	a	créée

$$\begin{array}{c}
 \text{créée} \\
 \frac{a \quad \frac{(np \setminus s_{ppart}) / np \quad np}{[E]}}{(np \setminus s) / (np \setminus s_{ppart}) \quad np \setminus s_{ppart}} [E] \\
 np \setminus s
 \end{array}$$

BEYOND AB GRAMMARS

INTRODUCTION RULES

n	$(n \setminus n) / (s / np)$	np	$(np \setminus s) / (np \setminus s_{ppart})$	$(np \setminus s_{ppart}) / np$
redaction	qu'	on	a	créée

$$\begin{array}{c}
 \text{créée} \\
 \frac{\text{a} \quad \frac{(np \setminus s_{ppart}) / np \quad [np]^1}{[/ E]}}{np \setminus s_{ppart} \quad [/ E]} \\
 \frac{\text{on} \quad \frac{(np \setminus s) / (np \setminus s_{ppart})}{np \setminus s} \quad [\setminus E]}{np} \\
 \frac{s}{s / np} \quad [/ I]^1
 \end{array}$$

BEYOND AB GRAMMARS

INTRODUCTION RULES

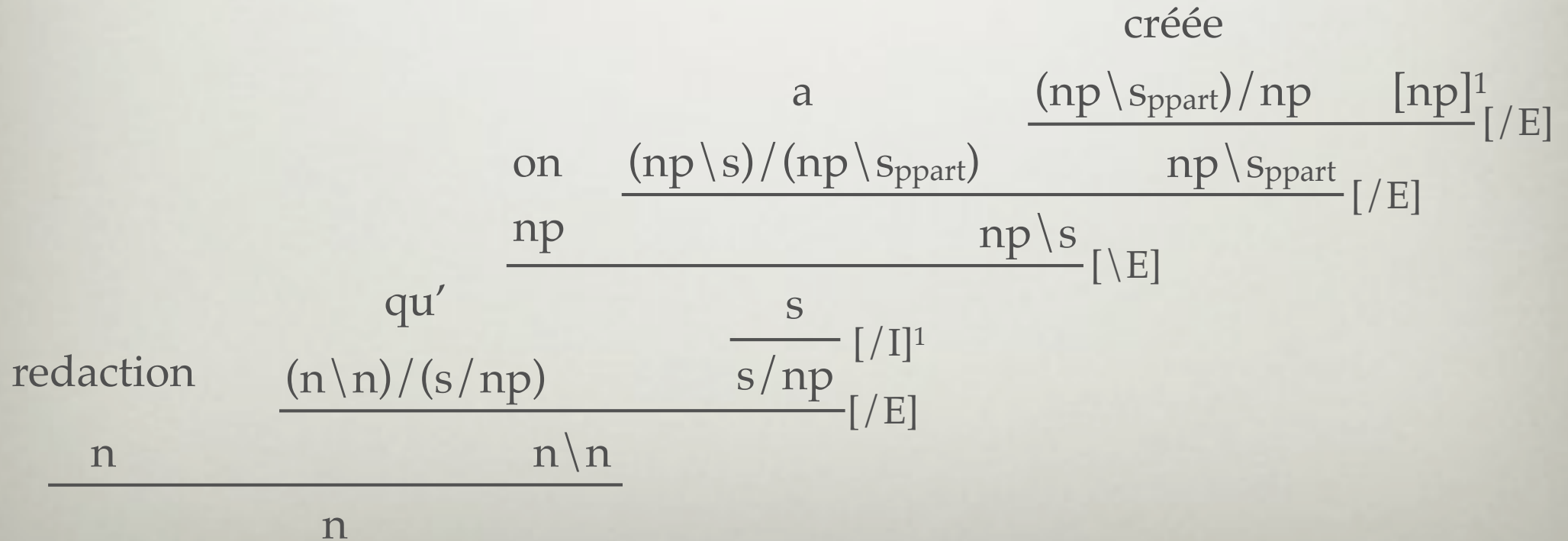
n	$(n \setminus n) / (s / np)$	np	$(np \setminus s) / (np \setminus s_{ppart})$	$(np \setminus s_{ppart}) / np$
redaction	qu'	on	a	créée

$$\begin{array}{c}
 \text{créée} \\
 \frac{\text{a} \quad \frac{(np \setminus s_{ppart}) / np \quad [np]^1}{[/ E]}}{np \setminus s_{ppart} \quad [/ E]} \\
 \frac{\text{on} \quad \frac{(np \setminus s) / (np \setminus s_{ppart})}{np \setminus s} \quad [\setminus E]}{np} \\
 \frac{\text{qu}' \quad \frac{(n \setminus n) / (s / np) \quad \frac{s}{s / np} \quad [/ I]^1}{[/ E]}}{n \setminus n}
 \end{array}$$

BEYOND AB GRAMMARS

INTRODUCTION RULES

n	$(n \setminus n) / (s / np)$	np	$(np \setminus s) / (np \setminus S_{ppart})$	$(np \setminus S_{ppart}) / np$
redaction	qu'	on	a	créée



BEYOND AB GRAMMARS

INTRODUCTION RULES

n	(n \ n) / (s / np)	np	(np \ s) / (np \ S _{ppart})	(np \ S _{ppart}) / np	s \ 1s
redaction	qu'	on	a	créée	ensemble

BEYOND AB GRAMMARS

INTRODUCTION RULES

n	$(n \setminus n) / (s / \diamond \square np)$	np	$(np \setminus s) / (np \setminus s_{ppart})$	$(np \setminus s_{ppart}) / np$	$s \setminus 1s$
redaction	que	on	a	créée	ensemble

BEYOND AB GRAMMARS

INTRODUCTION RULES

.... $[B]^i$

$$\frac{A}{A/B} [/ I]^i$$

$[B]$ is the rightmost free hypothesis in the proof

“...” contains at least one other non-discharged formula

Each introduction rule discharges exactly one formula.

A unique integer i links the rule application with this formula

$[B]^i$...

$$\frac{A}{B/A} [\backslash I]^i$$

$[B]$ is the leftmost free hypothesis in the proof

Lambek calculus version of the introduction rules.

Repeat of previous slide.

BEYOND AB GRAMMARS

INTRODUCTION RULES

.... $[B]^i$

$$\frac{A}{A / \diamond \square B} [/ I]^i$$

Note that this rule is just a "sugared" version of a set of derivation steps in a slightly more complicated logic. Details can be found in eg. Moortgat (1997)

$[B]$ is a free hypothesis occurring anywhere in the proof

there is at least one other non-discharged formula

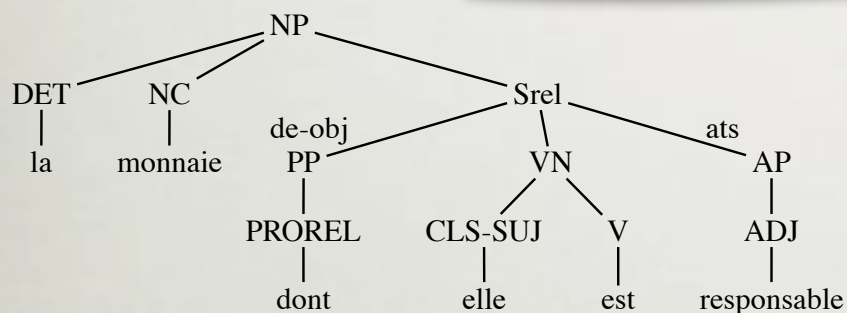
Each introduction rule discharges exactly one formula.

A unique integer i links the rule application with this formula

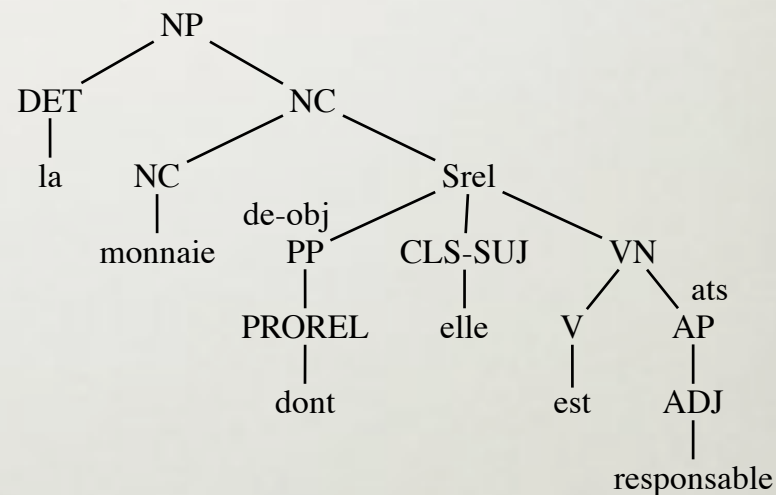
BEYOND AB GRAMMARS

A MORE COMPLEX EXAMPLE

Sentence as we find it in the corpus. "dont" is a relative pronoun like "que" but which selects a sentence missing de "de" preposition (instead of a sentence missing an np like "que")

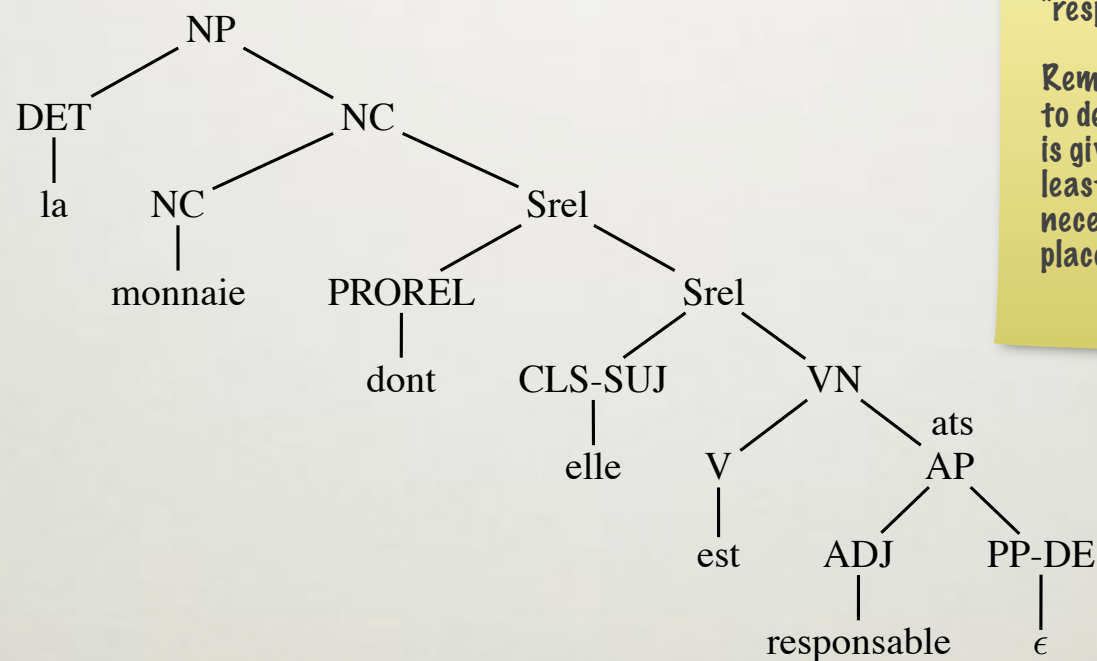


Note how "dont" is annotated as a "de-obj" argument, which is useful.



BEYOND AB GRAMMARS

A MORE COMPLEX EXAMPLE

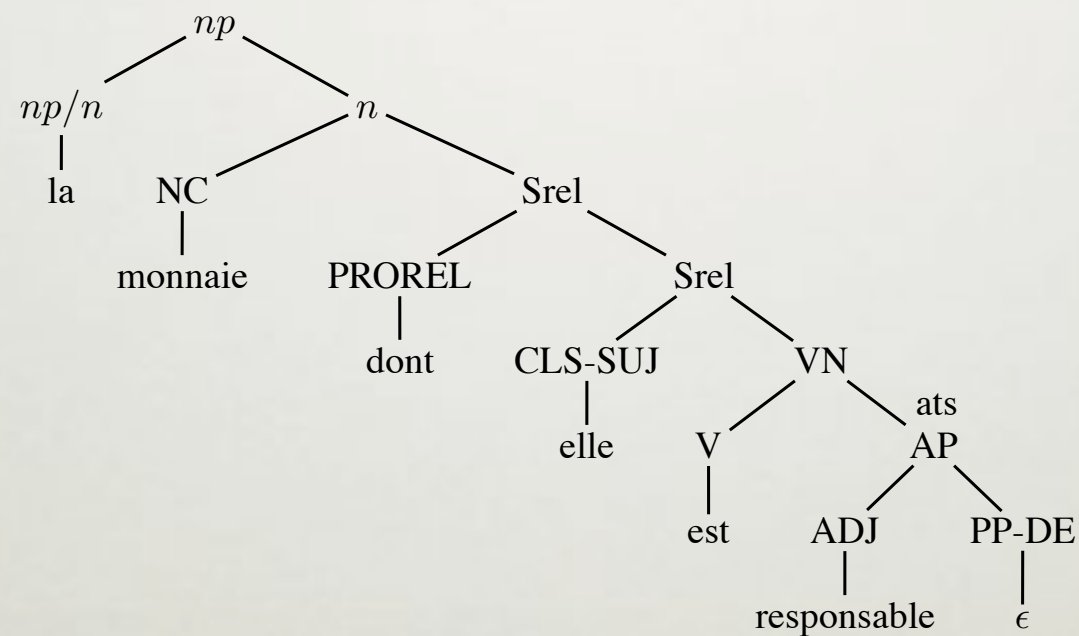


However, the “de” preposition belongs to “responsable” (some adjectives select for prepositions: “responsable de X” functions as an adjective just as “responsable”)

Remark however, that there is no way to derive this from the annotation as it is given. Manual intervention (or at least verification!) is unfortunately necessary to assure the correct placement of the hypothetical preposition.

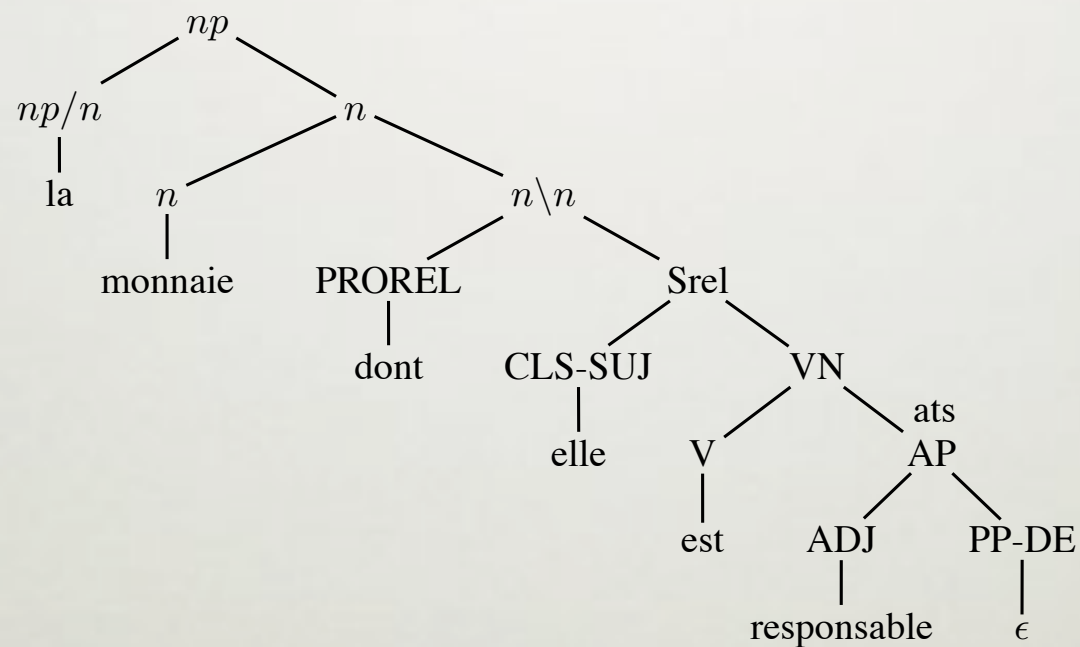
BEYOND AB GRAMMARS

A MORE COMPLEX EXAMPLE



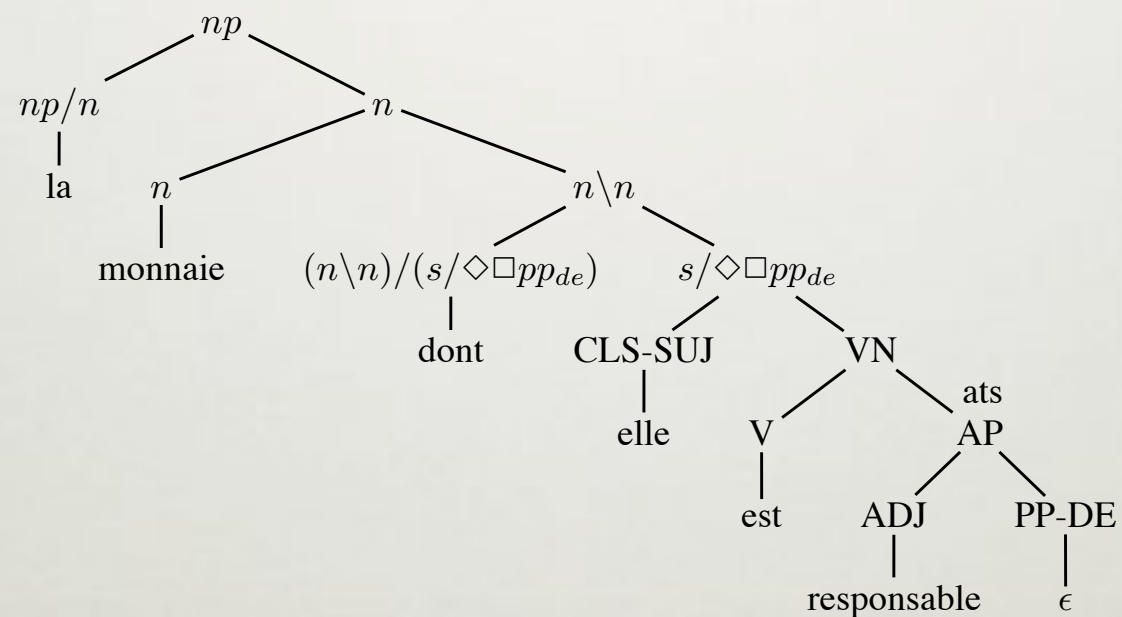
BEYOND AB GRAMMARS

A MORE COMPLEX EXAMPLE



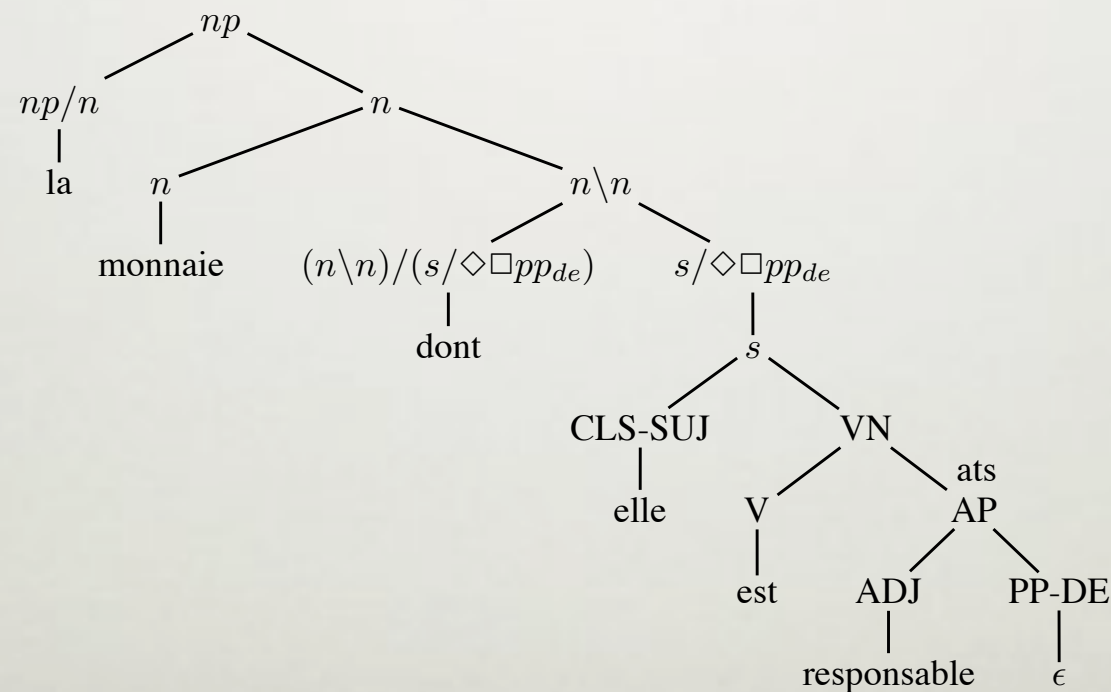
BEYOND AB GRAMMARS

A MORE COMPLEX EXAMPLE



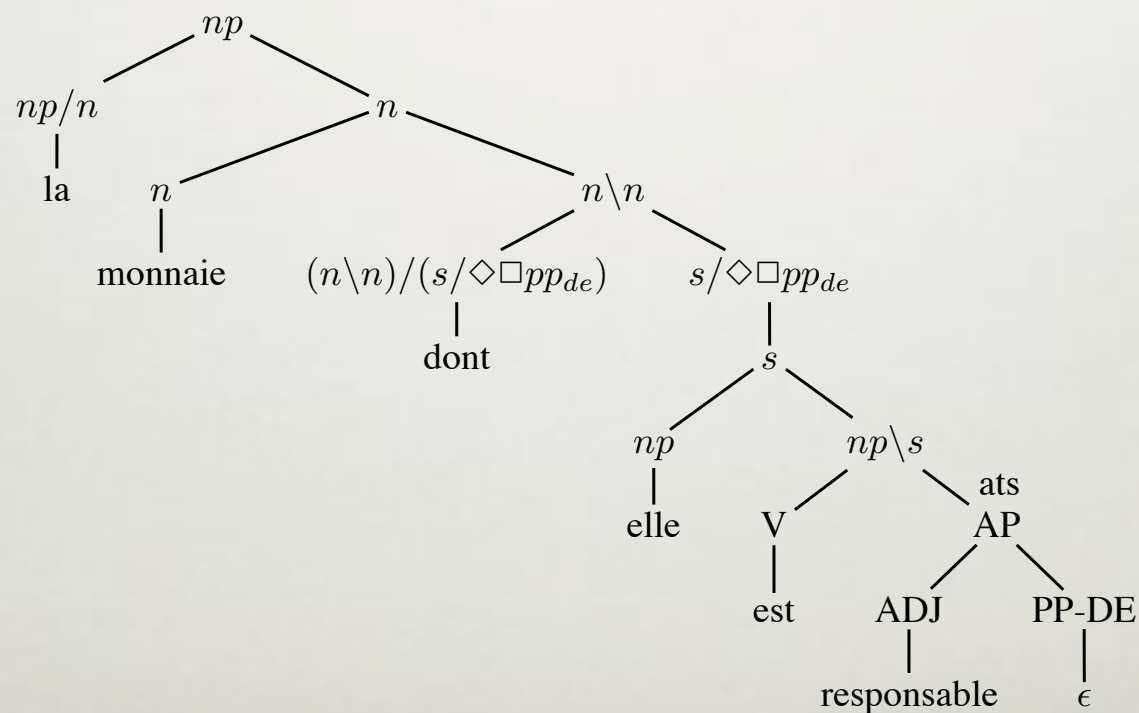
BEYOND AB GRAMMARS

A MORE COMPLEX EXAMPLE



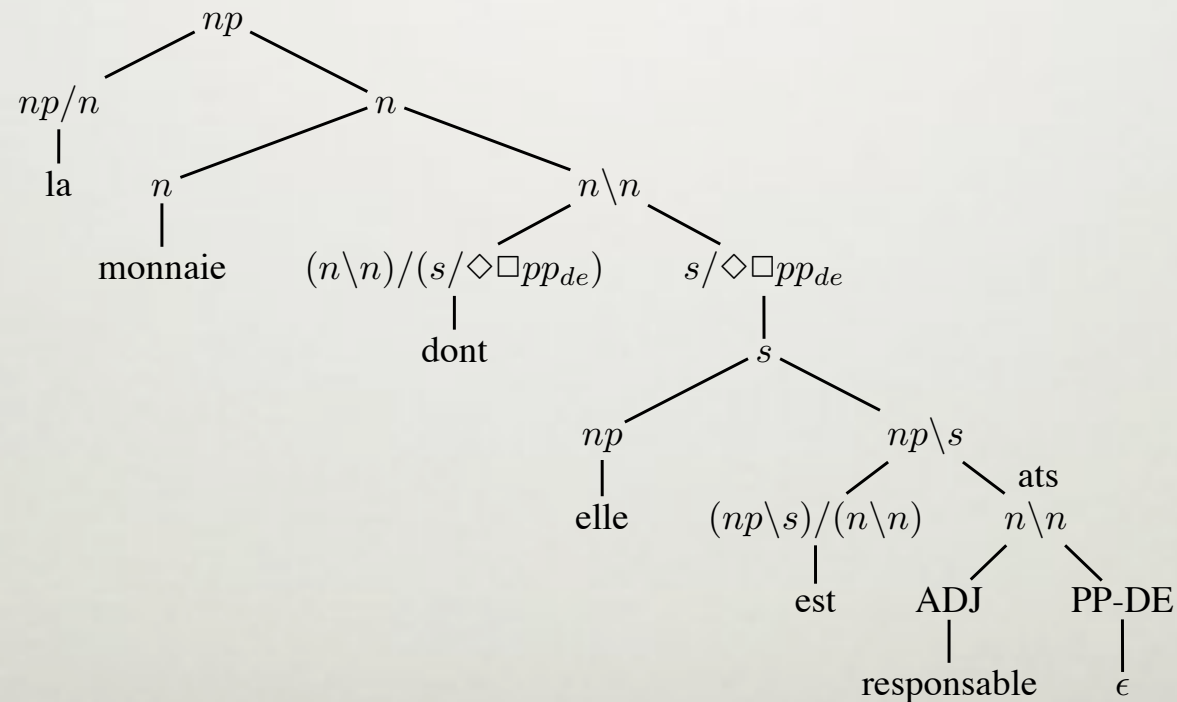
BEYOND AB GRAMMARS

A MORE COMPLEX EXAMPLE



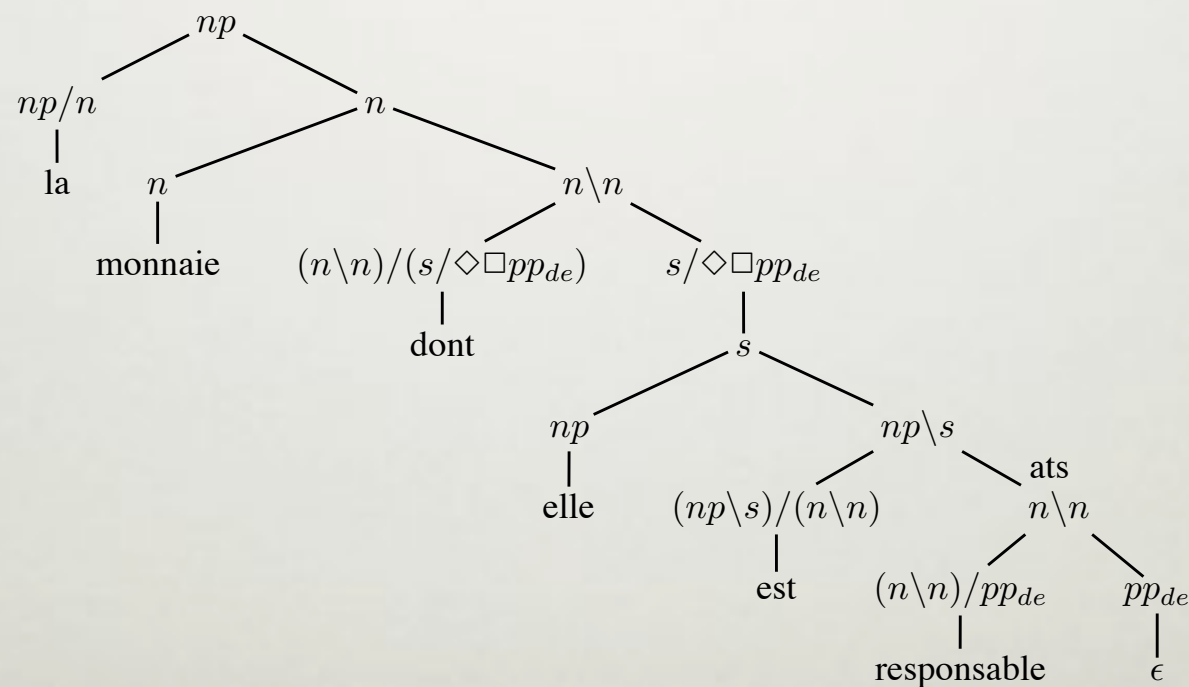
BEYOND AB GRAMMARS

A MORE COMPLEX EXAMPLE



BEYOND AB GRAMMARS

A MORE COMPLEX EXAMPLE



BEYOND AB GRAMMARS

INTRODUCTION RULES

np	$(np \setminus s) / (np \setminus s_{deinf})$	$(np \setminus s_{deinf}) / (np \setminus s_{inf})$	np	$np \setminus (np \setminus s_{inf})$
Il	convient	de	tout	reviser

BEYOND AB GRAMMARS

INTRODUCTION RULES

np	$(np \setminus s) / (np \setminus s_{deinf})$	$(np \setminus s_{deinf}) / (np \setminus s_{inf})$	$(np \setminus s_X) / ((np \setminus s_X) / np)$	$(np \setminus s_{inf}) / np$
Il	convient	de	tout	reviser

Advantage of this solution: the formula for "reviser" is a simple transitive verb form for an infinitive.

In addition we do not generate "Il convient de les exercices reviser" (the class of pronouns which can occur to the left of an embedded VP seems to be rather limited: "tout", "rien", "ce" and "beaucoup" are the only ones which I found in the extracted grammar)

The s_X here is a first-order quantifier which quantifies over the different types of sentences (even though I think only s_{inf} and s_{ppart} are possible here).

BEYOND AB GRAMMARS

CLITICS

une lettre qu' il lui a adressé

il a adressé une lettre à M. Eltsine

BEYOND AB GRAMMARS

CLITICS

une lettre qu' il lui a adressé

il a adressé une lettre à M. Eltsine

a: (np \ s) / (np \ s_{ppart})

adressé: ((np \ s_{ppart}) / pp_a) / np

BEYOND AB GRAMMARS

CLITICS

une lettre qu' il lui a adressé

il a adressé une lettre à M. Eltsine

que: $(n \setminus n) / s / \diamond \square np$

a: $(np \setminus s) / (np \setminus S_{ppart})$

adressé: $((np \setminus S_{ppart}) / pp_a) / np$

BEYOND AB GRAMMARS

CLITICS

une lettre qu' il lui a adressé

il a adressé une lettre à M. Eltsine

que: $(n \setminus n) / s / \diamond \square np$

lui: $((np \setminus s_x) \setminus (np \setminus s_x) / \square pp_a)$

a: $(np \setminus s) / (np \setminus s_{ppart})$

adressé: $((np \setminus s_{ppart}) / pp_a) / np$

CONCLUSIONS

- We have applied the basic principles of categorial grammar to the analysis of the french language, using an annotated corpus as a guide.