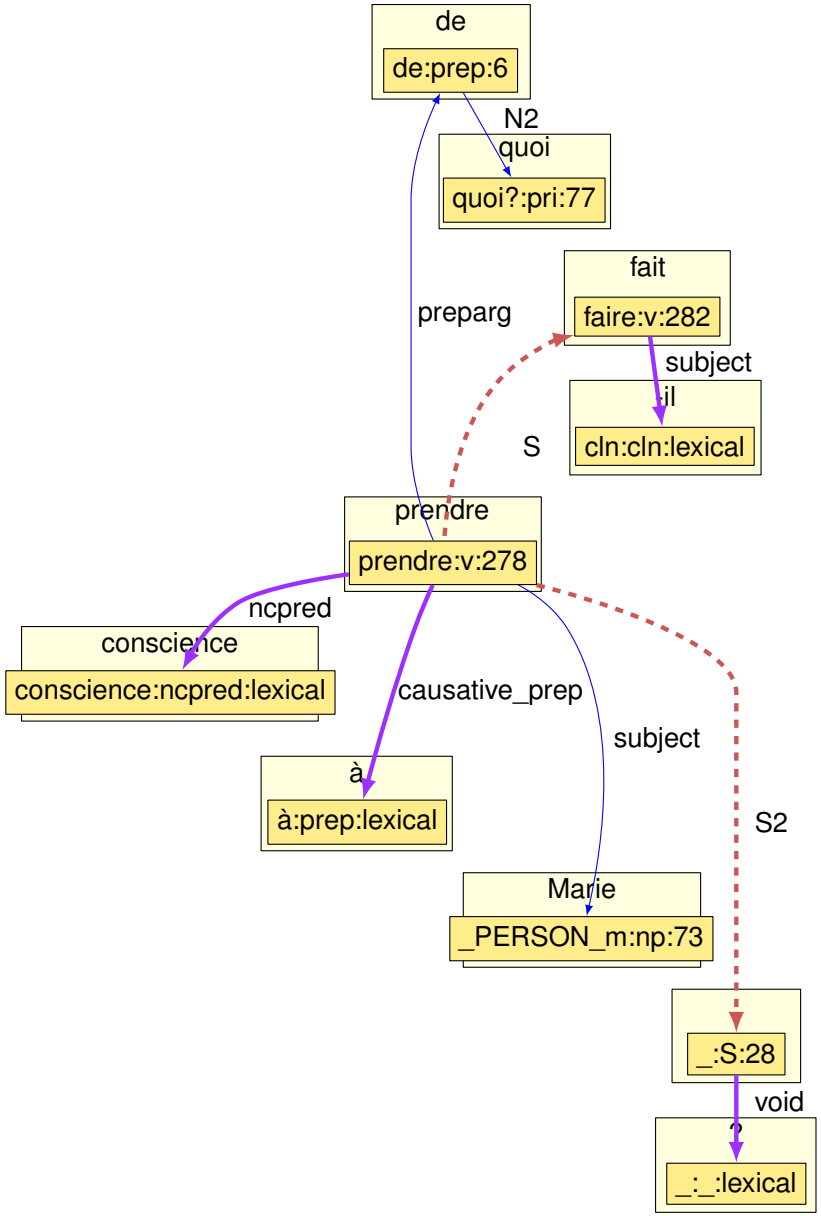
Evaluating syntactic lexica through their integration in the FRMG parser

Elsa Tolone, Éric de La Clergerie, Benoît Sagot

LIGM, Université Paris-Est, France & FaMAF, Universidad Nacional de Córdoba, Argentine
INRIA Paris-Rocquencourt
http://alpage.inria.fr

Colloque Lexique et grammaire 2011
Nicosie, Chypre, 5-8 Octobre 2011

Deep parsing with FRMG (and LEFFF)



de quoi fait-il prendre conscience à Marie ?

To be tried at `http://alpage.inria.fr/parserdemo`

Evaluating lexica ?

FRMG depends on LEFFF syntactic lexicon (Sagot),
and largely co-developped with it, but

- can we use FRMG with other lexica ?
- is it easy to plug a new lexicon ?
- can we then run fine-grained evaluations of syntactic lexica ?
- can we use feedback information to improve a lexicon ?

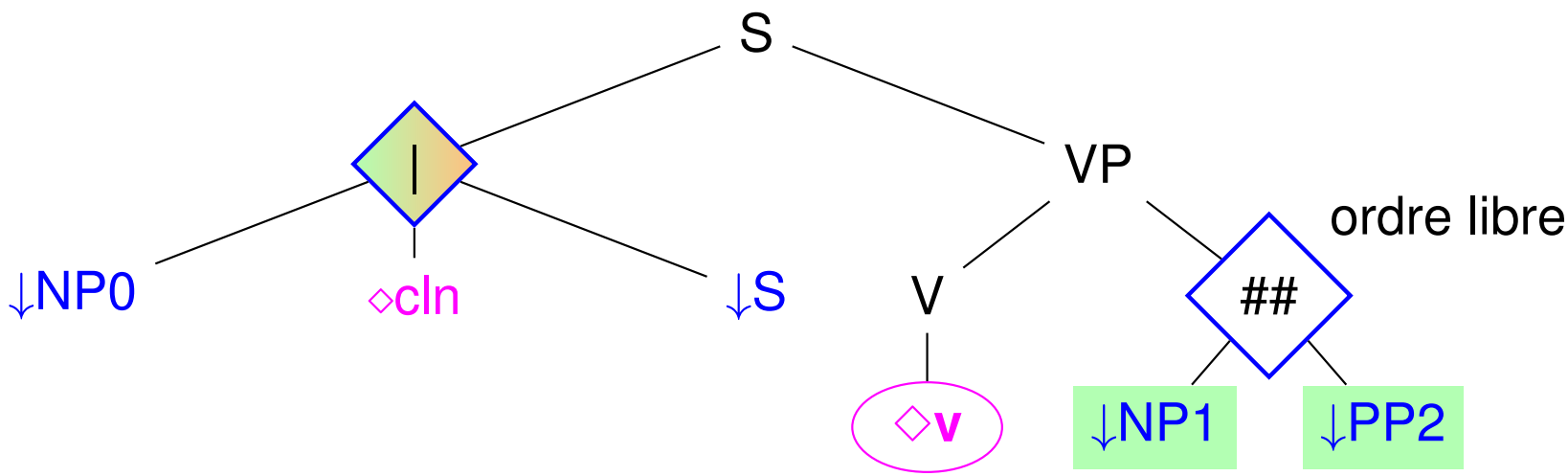
We tried to answer these questions, with 3 new lexica:
LGLEX, DICOVALENCE, and NEW LEFFF

FRMG: a French metagrammar

- A large-coverage Metagramar for French
  abstract descriptive layer, constraint-based, modularity, inheritance

- generation of a TAG/TIG grammar
  extended domain of locality, capture of subcategorization frames

- with **factorized** trees
  - current version: 290 trees (and only 32 verbal trees)
  - one tree $\equiv$ many ordinary TAG trees
  - $\Rightarrow$ one verbal tree stands for many subcat frames, arg positions, realizations, ...



Coupling FRMG with a lexicon: Hypertags

FRMG

hypertag #286

arg0	*arg0*	extracted	-
		fun	*fun0*
		kind	*kind0* **subj** |**nosubj**
		pcas	-
		real	*real0* - | **CS** | **N2** | **PP** | **S** | **cln** | **prel** | **pri**
arg1	*arg1*	extracted	-
		fun	*fun1*
		kind	*kind1* - | **acomp** | **obj** | **prepacomp** | **prepobj**
		pcas	*pcas1* + | - | **apres** | **à** | **avec** | **de** | **par** | ...
		real	*real1* - | **CS** | **N** | **N2** | **PP** | **S** | **V** | **adj** | **cla** | ...
arg2	*arg2*	extracted	-
		fun	*fun2*
		kind	*kind2* - | **prepacomp** | **prepobj** | **prepscomp** | **prepvcomp** | **scomp** | **vcomp** | **wh-comp**
		pcas	*pcas2* - | + | **apres** | **à** | ...
		real	*real2* - | **CS** | **N** | **N2** | **PP** | **S** | ...
cat		**v**	
diathesis		**active**	
refl		*refl*	
ctrsubj		*ctr*	
imp		*imp*	

Coupling FRMG with a lexicon: Hypertags

FRMG

hypertag #286

```
[
  arg0  arg0 [ extracted  -
               fun        fun0
               kind       kind0   subj |nosubj
               pcas       -
               real       real0   - | CS | N2 | PP | S | cln | prel | pri ]

  arg1  arg1 [ extracted  -
               fun        fun1
               kind       kind1   -  | acomp |  obj  | prepacomp | prepobj
               pcas       pcas1   + |  -  | apres | à | avec | de | par | ...
               real       real1   - | CS | N | N2 | PP | S | V | adj | cla | ... ]

  arg2  arg2 [ extracted  -
               fun        fun2
               kind       kind2   -  | prepacomp | prepobj | prepscomp
                                  | prepvcomp | scomp | vcomp | wh-comp
               pcas       pcas2   - | + | apres | à | ...
               real       real2   - | CS | N | N2 | PP | S | ... ]

  cat        v
  diathesis  active
  refl       refl
  ctrsubj    ctr
  imp        imp
]
```

LEFFF

hypertag «promettre»

```
[
  arg0   [ fun   suj
           kind  subj | -
           pcas  - ]

  arg1   [ fun   obj
           kind  obj | scomp
           pcas  - ]

  arg2   [ fun   objà
           kind  prepobj | -
           pcas  à | - ]

  refl   -
  ctrsubj suj
  imp    -
]
```

Coupling FRMG with a lexicon: Hypertags

FRMG

hypertag #286

arg0 [arg0]
- extracted -
- fun [fun0] **suj**
- kind [kind0] **subj** |**nosubj**
- pcas **-**
- real [real0] **-** | **CS** | **N2** | **PP** | **S** | **cln** | **prel** | **pri**

arg1 [arg1]
- extracted -
- fun [fun1] **obj**
- kind [kind1] **-** | acomp | **obj** | prepacomp | prepobj
- pcas [pcas1] + | **-** | apres | à | avec | de | par | ...
- real [real1] **-** | **CS** | **N** | **N2** | **PP** | **S** | **V** | **adj** | **cla** | ...

arg2 [arg2]
- extracted -
- fun [fun2] **objà**
- kind [kind2] **-** | prepacomp | **prepobj** | prepscomp | prepvcomp | scomp | vcomp | whcomp
- pcas [pcas2] **-** | + | apres | **à** | ...
- real [real2] **-** | **CS** | **N** | **N2** | **PP** | **S** | ...

cat **v**

diathesis **active**

refl [refl] **-**

ctrsubj [ctr] **suj**

imp [imp] **-**

LEFFF

hypertag «promettre»

arg0
- fun **suj**
- kind **subj** | **-**
- pcas **-**

arg1
- fun **obj**
- kind **obj** | **scomp**
- pcas **-**

arg2
- fun **objà**
- kind **prepobj** | **-**
- pcas **à** | **-**

refl **-**

ctrsubj **suj**

imp **-**

Alexina and Lefff (Sagot)

ALEXINA is a lexical formalism

- with an *intensional* level for lemma
- and the generation of an *extensional* level for forms
- the descriptions use a set of primitive features, and macros

LEFFF is a wide-coverage morphosyntactic and syntactic lexicon for French, covering all categories

LEFFF is partially factorized:
one entry may cover several meanings and several subcat frames
$\Rightarrow$ 5,736 entries for 5,450 distinct ones (intensional level)

Freely available at `http://gforge.inria.fr/projects/alexina/`

Lefff example: promettre

- Intensional level

```
promettre   v55  100;Lemma;v;
<Suj:cln|scompl|sinf|sn,
 Obj:(cla|de-sinf|scompl|sn),
 Objà:(cld|à-sn)>;
@CtrlSujObj,cat=v;
%actif,%passif,%ppp_employé_comme_adj,%passif_impersonnel,%
    passif
```

- Extensional level

```
promet   100   v
[pred="promettre_____1<Suj:cln|scompl|sinf|sn,
␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣Obj:(cla|de-sinf|scompl|sn),
␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣Objà:(cld|à-sn)>",
 @CtrlSujObj,@pers,cat=v,@P3s]
```

- A macro

```
@CtrlSujObl = [ctrsubj = suj];
```

LgLex

Resulting from the conversion of LADL tables

1. first, into LGlex format (Constant and Tolone)
2. then, into alexina format (Sagot and Tolone)

- Wide-coverage lexicon

kind	#tables	#entries	#distinct entries
verbs	67	13,867	5,738
pred. nouns	78	12,696	8,531

- Fine-grained: many entries for some verbs
  53 entries for *tenir* (LEFFF: 6 entries)

- Some features can't be represented in ALEXINA and/or exploited in FRMG
  for instance: added determiner for predicative nouns in FRMG
  missing: semantic restrictions

Freely available at `http://infolingu.univ-mlv.fr`

Dicovalence

Developed by Mertens and van den Eynde

- based on pronominal approach (Benveniste)
- fine grained: one entry for one meaning
- small coverage
  3,700 verbs for 8,000 entries

New Lefff

Evolution of LEFFF towards a more semantic lexicon

- finer-grained: one meaning per entry
- automatic fusion of LEFFF and DICOVALENCE,
  plus manual validation of 505 verbs (986 entries):
  - the 100 most frequent lemmas
  - the *dubious lemmas*: more output entries than the sum of corresponding LEFFF and DICOVALENCE entries
- still a wide-coverage lexicon
  7,933 verbs, for 12,613 entries

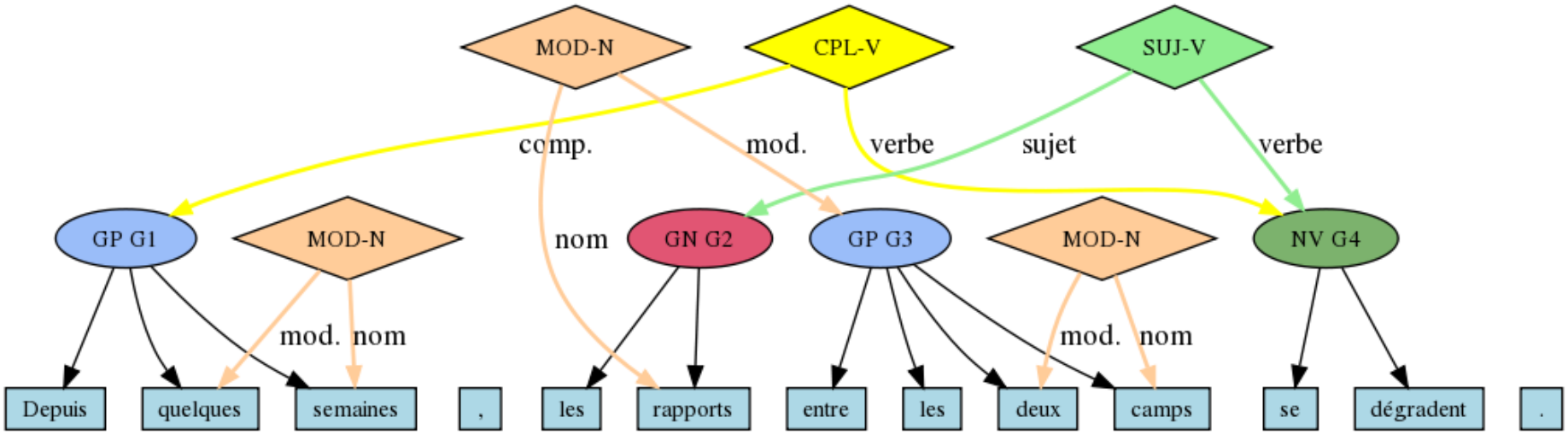
The lexica at a glance

lexica	cat.	#entries	#verbs	ratio
LEFFF	v	5,736	5,450	1.06
LGLEX	v	13,867	5,738	2.41
	pnoun	12,696	8,531	1.48
DICOVALENCE	v	8,000	3,700	2.16
NEW LEFFF	v	12,613	7,933	1.58

EASy/Passage evaluation campaign

French parsing evaluation campaigns organized within EASy and Passage actions.
We use the EASy reference corpus as benchmark

- around 4K sentences, manually annotated (but with errors !)
- various styles: journalistic, literacy, medical, mail, oral, questions
- constituency and dependency based format (shallow level)
  - 6 kinds of *chunks*: GN, NV, GA, GR, GP, PV
  - 14 kinds of *relations*: SUJ-V, AUX-V, COD-V, ATB-SO , CPL-V, MOD-V, MOD-N, MOD-A, MOD-R, MOD-P, COMP, COORD, APPOS, JUXT



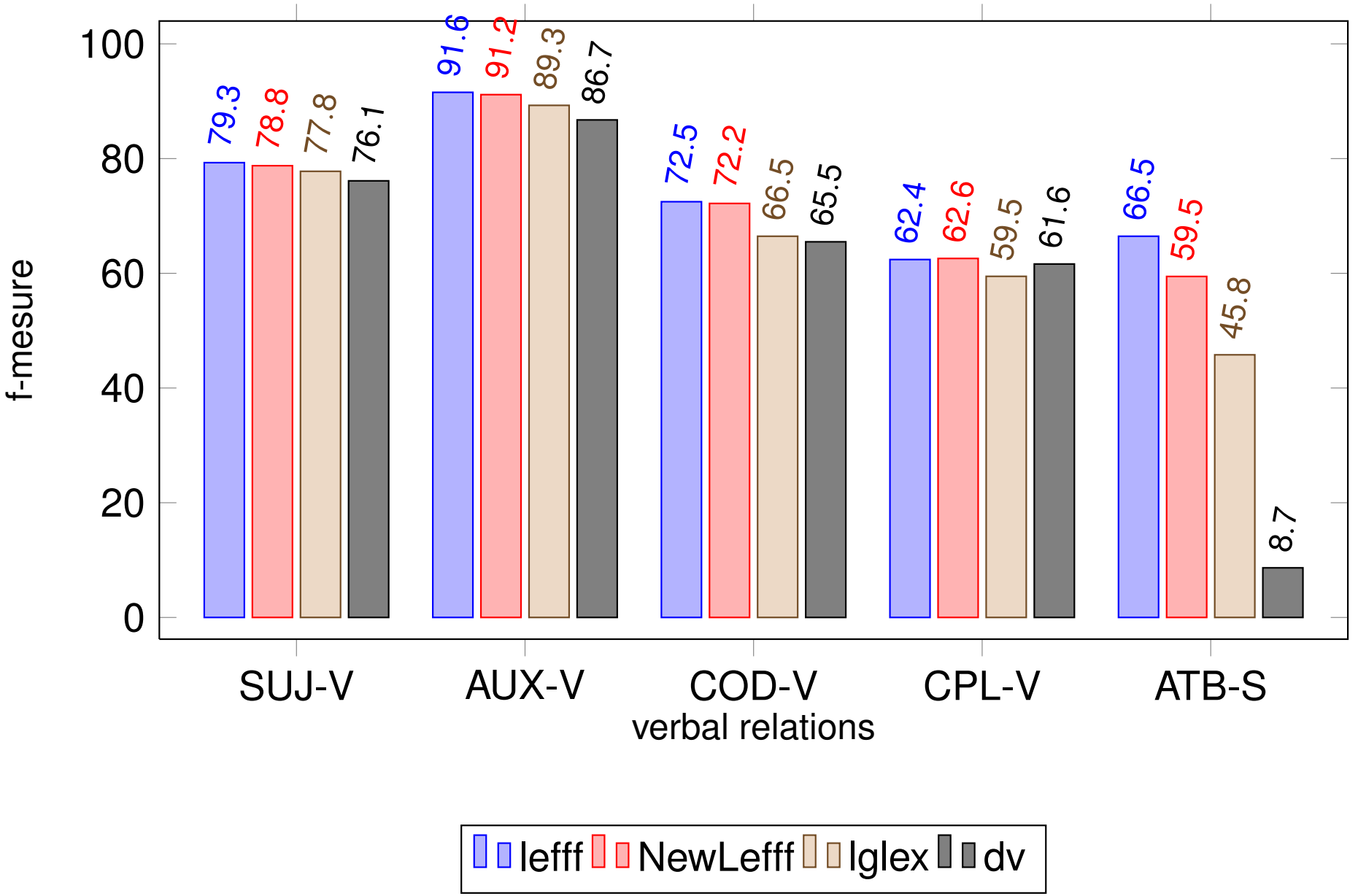
Overall results

Setting:

- each sentence segmented with SxPipe, no prior tagging, use of a lexicon
- FRMG returns either
  - full parses (possibly by relaxing some agreement constraints)
  - sequences of partial parses, covering the sentence
  - nothing in case of timeout (100s)
- whenever possible, FRMG returns a shared dependency forest of all possibilities
- then heuristic-based disambiguisation and conversion to Passage format

Lexicon	Coverage		Groups	Relations	Time	Timeout
	#	%	%	%	s	%
Lefff	3,556	76.08	89.21	66.36	0.30	0.00
New Lefff	3,495	74.81	88.65	65.41	0.43	0.03
LgLex	3,437	73.60	87.97	63.03	0.84	0.03
Dicovalence	2,773	59.78	86.98	61.91	0.42	0.00

Analysis per verbal relation



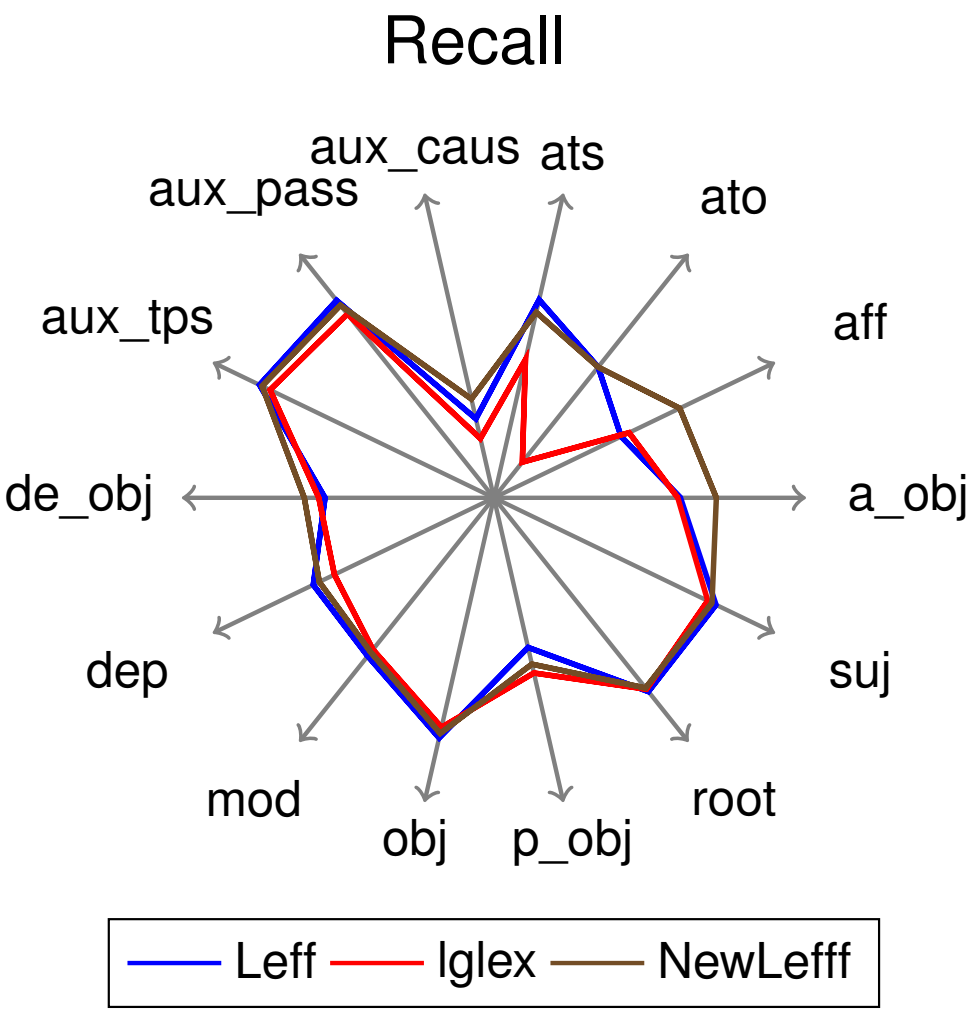
Bonus: experiments on French TreeBank

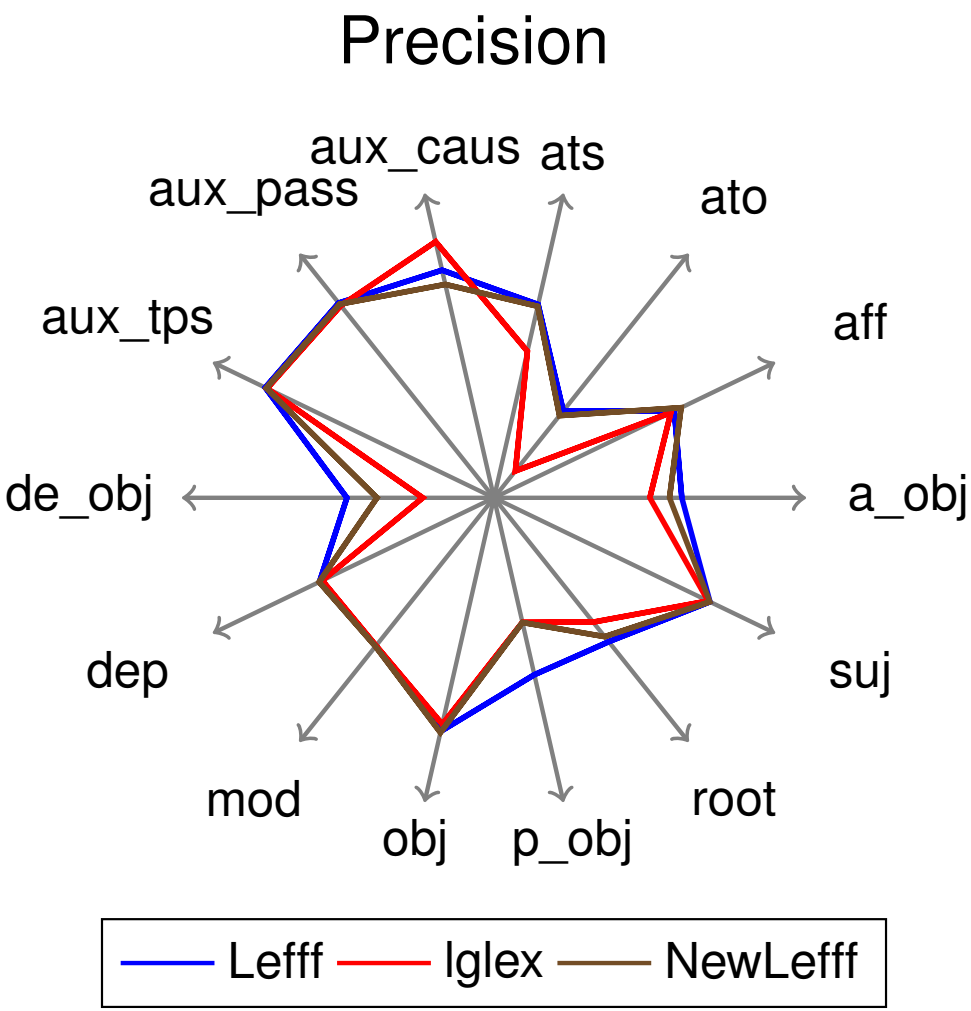
New: evaluation on CONLL dependency version of FTB
richer set of verbal dependencies, but still shallow level

```
1   de          de          P       P       5       de_obj
2   quoi        quoi?       PRO     PROWH   1       obj
3   fait        faire       V       V       5       aux_caus
4   -il         -il         CL      CLS     5       suj
5   prendre     prendre     V       V       0       root
6   conscience  conscience   N       NC      5       obj
7   à           à           P       P       5       mod
8   Marie       marie       N       NPP     7       obj
9   ?           ?           PONCT   PONCT   5       ponct
```

Lexicon	Coverage (%)	LAS (%)	Time (s)	Timeout (%)
LEFFF	88.29	82.44	0.62	0.02
NEW LEFFF	86.96	81.57	0.80	0.02
LGLEX	84.80	78.94	1.62	0.08
MST (stat parsing)	-	88.2	-	-

Analysis per CONLL verbal relation





Error mining

Original motivation: find lexical entries that are incorrect or incomplete through full parse failures:

> *a form is suspect if occurring more often than expected in failed sentences, in co-occurrence with non-suspect forms*

$\rightsquigarrow$ fix-point iterative algorithm, close to EM (Expectation-Maximization) and return the best sentences where a form is the main suspect
$\Rightarrow$ WEB-based interface to browse suspects, lexical info, and sentences

May be used for any lexica, but can also be adapted for contrasting lexica

> *a verb is suspect for lexicon L if occurring more often than expected in failed sentences that succeed for* LEFFF*, in co-occurrence with non-suspect verbs.*

Tried on a 100Ksent. toy corpus (wikipedia, wikisource, europarl, AFP news) but could be tried on CPC (100Mwords) or even bigger (700Mwords)

Some suspects (for LGLEX)

A first typology of errors on the first 15th suspects for LGLEX:

- missing entries in the right table
  - réaffirmer (28s), réélire (10), se réimplanter (5), mixer (7)
    Mixé par Jimi Hazel , assisté de Bruce Calder , enregistré chez Jimi à l' « Electric Lady Studios » à New York
- existing entries, but missing codes
  - susciter (41; 36DT & 38R), recruter (14; 38R), délocaliser (9; 38L), zapper (4; 35L: N0 V Loc N1 source Loc N2 dest)
  - *Elle a également " déploré " la mémoire de " plus en plus sélective " de la jeune femme , " qui zappe les détails qui font désordre "*
- mandatory args (for LGLEX), but missing in the sentences
  kidnapper (12; 36DT N0 V N1 Prep N2) : *Les deux Italiens ont été kidnappés le 18 décembre*
- misc. situations

Conclusion

- Relatively easy to plug new lexica into FRMG
- Rather good results with the tried lexica, even if lower than with LEFFF (better than FRMG+LEFFF in 2007 Passage campaign)
- Room for needed and normal co-adaptation FRMG-lexicon

	Coverage	Groups	Relations	Time	Timeout
Run1	64.60	83.99	57.02	3.82	0.31
Run6	73.60	87.97	63.03	0.84	0.03

Evolutions:

- completing lexica and/or merging information (error mining, evaluation)
- better factorization of lexical entries in LGLEX, delaying use of more semantic entries at disambiguisation time
- (alternatively) pre-parsing *suppertagging* or *hypertagging* phases to prune search space
- assign probabilities to entries and/or frames
- enrich FRMG with some new features, to take into account richer lexical information