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
de quoi fait-il prendre conscience à Marie ?

To be tried at http://alpage.inria.fr/parserdemo
FRMG depends on LEFFF syntactic lexicon (Sagot), and largely co-developed 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 Metagrammar 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
  - one verbal tree stands for many subcat frames, arg positions, realizations, ...

```
S | NP0  ↓ Ln  ↓ S  V  ↓ NP1  ↓ PP2
  ↓ VP

ordre libre
```

S
Coupling FRMG with a lexicon: Hypertags

FRMG

hypertag #286

[extracted -
  fun [fun0]
arg0  arg0
kind [kind0] subj |nosubj
pcas -
real [real0] - | CS | N2 | PP | S | cln | prel | pri
[extracted -
  fun [fun1]
arg1  arg1
kind [kind1] - | acomp | obj | prepacomp | prepobj
pcas [pcas1] + | - | apres | à | avec | de | par | ...
real [real1] - | CS | N | N2 | PP | S | V | adj | cla | ...
[extracted -
  fun [fun2]
arg2  arg2
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

LEF

hypertag «promettre»

[extracted -
  fun [fun0]
arg0  arg0
kind [kind0] subj |nosubj
pcas -
real [real0] - | CS | N2 | PP | S | cln | prel | pri
[extracted -
  fun [fun1]
arg1  arg1
kind [kind1] - | acomp | obj | prepacomp | prepobj
pcas [pcas1] + | - | apres | à | avec | de | par | ...
real [real1] - | CS | N | N2 | PP | S | V | adj | cla | ...
[extracted -
  fun [fun2]
arg2  arg2
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
<table>
<thead>
<tr>
<th><strong>FRMG</strong></th>
<th><strong>LEFFF</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>hypertag #286</strong></td>
<td><strong>hypertag &quot;promettre&quot;</strong></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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
  ⇒ 5,736 entries for 5,450 distinct ones (intensional level)

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

<table>
<thead>
<tr>
<th>kind</th>
<th>#tables</th>
<th>#entries</th>
<th>#distinct entries</th>
</tr>
</thead>
<tbody>
<tr>
<td>verbs</td>
<td>67</td>
<td>13,867</td>
<td>5,738</td>
</tr>
<tr>
<td>pred. nouns</td>
<td>78</td>
<td>12,696</td>
<td>8,531</td>
</tr>
</tbody>
</table>

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

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
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

<table>
<thead>
<tr>
<th>lexica</th>
<th>cat.</th>
<th>#entries</th>
<th>#verbs</th>
<th>ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEFFF</td>
<td>v</td>
<td>5,736</td>
<td>5,450</td>
<td>1.06</td>
</tr>
<tr>
<td>LGLEX</td>
<td>v</td>
<td>13,867</td>
<td>5,738</td>
<td>2.41</td>
</tr>
<tr>
<td></td>
<td>pnoun</td>
<td>12,696</td>
<td>8,531</td>
<td>1.48</td>
</tr>
<tr>
<td>DICOVALENCE</td>
<td>v</td>
<td>8,000</td>
<td>3,700</td>
<td>2.16</td>
</tr>
<tr>
<td>NEW LEFFF</td>
<td>v</td>
<td>12,613</td>
<td>7,933</td>
<td>1.58</td>
</tr>
</tbody>
</table>
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
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 disambiguation and conversion to Passage format

<table>
<thead>
<tr>
<th>Lexicon</th>
<th>Coverage</th>
<th>Groups</th>
<th>Relations</th>
<th>Time</th>
<th>Timeout</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>%</td>
<td>%</td>
<td>s</td>
<td>%</td>
</tr>
<tr>
<td>LEFFF</td>
<td>3,556</td>
<td>76.08</td>
<td>89.21</td>
<td>0.30</td>
<td>0.00</td>
</tr>
<tr>
<td>NEW LEFFF</td>
<td>3,495</td>
<td>74.81</td>
<td>88.65</td>
<td>0.43</td>
<td>0.03</td>
</tr>
<tr>
<td>LGLEX</td>
<td>3,437</td>
<td>73.60</td>
<td>87.97</td>
<td>0.84</td>
<td>0.03</td>
</tr>
<tr>
<td>DICOVALENCE</td>
<td>2,773</td>
<td>59.78</td>
<td>86.98</td>
<td>0.42</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Analysis per verbal relation

<table>
<thead>
<tr>
<th></th>
<th>lefff</th>
<th>NewLefff</th>
<th>Iglex</th>
<th>dv</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUJ-V</td>
<td>79.3</td>
<td>78.8</td>
<td>77.8</td>
<td>76.1</td>
</tr>
<tr>
<td>AUX-V</td>
<td>91.6</td>
<td>91.2</td>
<td>89.3</td>
<td>86.7</td>
</tr>
<tr>
<td>COD-V</td>
<td>72.5</td>
<td>72.2</td>
<td>66.5</td>
<td>65.5</td>
</tr>
<tr>
<td>CPL-V</td>
<td>62.4</td>
<td>62.6</td>
<td>59.5</td>
<td>61.6</td>
</tr>
<tr>
<td>ATB-S</td>
<td>66.5</td>
<td>59.5</td>
<td>45.8</td>
<td>8.7</td>
</tr>
</tbody>
</table>
**New**: evaluation on CONLL dependency version of FTB richer set of verbal dependencies, but still shallow level

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>de</td>
<td>de</td>
<td>P</td>
<td>P</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>quoi</td>
<td>quoi?</td>
<td>PRO</td>
<td>PROWH</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>fait</td>
<td>faire</td>
<td>V</td>
<td>V</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>-il</td>
<td>-il</td>
<td>CL</td>
<td>CLS</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>prendre</td>
<td>prendre</td>
<td>V</td>
<td>V</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>conscience</td>
<td>conscience</td>
<td>N</td>
<td>NC</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>à</td>
<td>à</td>
<td>P</td>
<td>P</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td>Marie</td>
<td>marie</td>
<td>N</td>
<td>NPP</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>?</td>
<td>?</td>
<td>PONCT</td>
<td>PONCT</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lexicon</th>
<th>Coverage (%)</th>
<th>LAS (%)</th>
<th>Time (s)</th>
<th>Timeout (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEFFF</td>
<td>88.29</td>
<td>82.44</td>
<td>0.62</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>NEW LEFFF</strong></td>
<td>86.96</td>
<td>81.57</td>
<td>0.80</td>
<td>0.02</td>
</tr>
<tr>
<td>LGLEX</td>
<td>84.80</td>
<td>78.94</td>
<td>1.62</td>
<td>0.08</td>
</tr>
<tr>
<td><strong>MST (stat parsing)</strong></td>
<td>-</td>
<td>88.2</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Analysis per CONLL verbal relation

Recall

Precision

Leff  Iglex  NewLeff

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

\[ \text{a form is suspect if occurring more often than expected in failed sentences, in co-occurrence with non-suspect forms} \]

\[ \Rightarrow \text{fix-point iterative algorithm, close to EM (Expectation-Maximization) and return the best sentences where a form is the main suspect} \]

\[ \Rightarrow \text{WEB-based interface to browse suspects, lexical info, and sentences} \]

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

\[ \text{a verb is suspect for lexicon L if occurring more often than expected in failed sentences that succeed for LEFFF, in co-occurrence with non-suspect verbs.} \]

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

<table>
<thead>
<tr>
<th></th>
<th>Coverage</th>
<th>Groups</th>
<th>Relations</th>
<th>Time</th>
<th>Timeout</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run1</td>
<td>64.60</td>
<td>83.99</td>
<td>57.02</td>
<td>3.82</td>
<td>0.31</td>
</tr>
<tr>
<td>Run6</td>
<td>73.60</td>
<td>87.97</td>
<td>63.03</td>
<td>0.84</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Evolutions:
- completing lexica and/or merging information (error mining, evaluation)
- better factorization of lexical entries in LGLEX, delaying use of more semantic entries at disambiguation time
- (alternatively) pre-parsing supporttagging 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