Evaluating and improving syntactic lexica by plugging them within a parser

Elsa Tolone, Benoît Sagot, Éric de La Clergerie

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

LREC’12
Istanbul
May 23-25 2012
FRMG is a deep wide-coverage TAG parser for French, co-developed with LEFFF lexicon.

de quoi fait-il prendre conscience à Marie ?
What does he make Mary become aware of ?

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 meta-grammar

- A large-coverage Meta-grammar 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: 323 trees (and only 38 verbal trees)
  - One tree \(\equiv\) many ordinary TAG trees
  - One verbal tree stands for many subcat frames, arg positions, realizations, ...

\[
\begin{align*}
S &\quad \downarrow NP0 \quad \downarrow S \\
  &\quad \downarrow cln \\
VP &\quad \downarrow V \\
  &\quad \downarrow NP1 \quad \downarrow PP2
\end{align*}
\]

Free order
Coupling FRMG with a lexicon: Hypertags

FRMG

hypertag #286

arg0

kind subj |nosubj

pcas -

real |CS | N2 | PP | S | cln | prel | pri

extracted -

fun fun0

arg1

kind - | acomp | obj | prepacomp | prepobj

pcas + | - | apres | à | avec | de | par | ...

real |CS | N | N2 | PP | S | V | adj | cla | ...

extracted -

fun fun1

arg2

kind - | prepacomp | prepobj | prepcomp

pcas | prepvcomp | scomp | vcomp | wh-

comp

pcas - | + | apres | à | ...

real |CS | N | N2 | PP | S | ...

cat v

diathesis active

refl refl

ctrsubj ctr

imp imp
Coupling FRMG with a lexicon: Hypertags

**FRMG**

hypertag #286

- extracted
- fun
- kind
- pcas
- real

arg0

- kind
- subj |nosubj
- pcas
- real
- CS | N | N2 | PP | S | cln | prel | pri

arg1

- kind
- acomp | obj | prepacomp | prepobj
- pcas
- real
- CS | N | N2 | PP | S | V | adj | cla | ...

arg2

- kind
- prepacomp | prepobj | prepscomp | prepvcomp
- pcas
- real
- CS | N | N2 | PP | S | ...

cat
- v
- diathesis
- active

refl
- refl
- ctrsubj
- imp

**LEFFF**

hypertag «promettre»

- extracted
- fun
- kind
- pcas
- real

arg0

- kind
- subj | -
- pcas
- ...

arg1

- kind
- obj | scomp
- pcas
- ...

arg2

- kind
- prepobj | -
- pcas
- ...

refl
- -

ctrsubj
- suj

imp
- -
Coupling FRMG with a lexicon: Hypertags

**FRMG**

hypertag #286

arg0

fun

kind

pcas

real

extracted -

|

fun

kind

pcas

real

extracted -

|

fun

kind

pcas

real

extracted -

| cat

diathesis

refl

ctrsubj

imp

**LEFFF**

hypertag «promettre»

arg0

kind

cas

real

extracted -

|

fun

diathesis

refl

ctrsubj

imp

...
**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

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 = [ctrlsubj = suj];
Resulting from the conversion of LADL tables (Gross)

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>#lemma</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
  3,738 verbs for 8,313 entries
Evolution of LEFFF towards a more semantic lexicon

- finer-grained: one meaning per entry

- automatic merge 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

### Verbs from each lexicon:

<table>
<thead>
<tr>
<th>Lexica</th>
<th>#Entries</th>
<th>#Lemmas</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEFFF</td>
<td>7,108</td>
<td>6,827</td>
<td>1.04</td>
</tr>
<tr>
<td>LGLEX</td>
<td>13,867</td>
<td>5,738</td>
<td>2.41</td>
</tr>
<tr>
<td>DICOVALENCE</td>
<td>8,313</td>
<td>3,738</td>
<td>2.22</td>
</tr>
<tr>
<td>NEW LEFFF</td>
<td>12,613</td>
<td>7,933</td>
<td>1.58</td>
</tr>
</tbody>
</table>

### Other categories imported from LEFFF, shared by all lexica

<table>
<thead>
<tr>
<th>Category</th>
<th>#Int. Entries</th>
<th>#Lemmas</th>
<th>#Ext. Entries</th>
</tr>
</thead>
<tbody>
<tr>
<td>nouns</td>
<td>41,816</td>
<td>41,592</td>
<td>86,675</td>
</tr>
<tr>
<td>adjectives</td>
<td>10,556</td>
<td>10,517</td>
<td>34,359</td>
</tr>
<tr>
<td>adverbs</td>
<td>4,111</td>
<td>3,676</td>
<td>4,155</td>
</tr>
<tr>
<td>prepositions</td>
<td>260</td>
<td>259</td>
<td>728</td>
</tr>
<tr>
<td>proper nouns</td>
<td>52,499</td>
<td>52,202</td>
<td>52,571</td>
</tr>
<tr>
<td>other</td>
<td>1,007</td>
<td>854</td>
<td>1,589</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>Cover (%)</th>
<th>Chunks (%)</th>
<th>Rel (%)</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEFFF</td>
<td>83.45</td>
<td>89.03</td>
<td>66.76</td>
<td>0.35</td>
</tr>
<tr>
<td>NEW LEFFF</td>
<td>82.19</td>
<td>88.74</td>
<td>66.09</td>
<td>0.55</td>
</tr>
<tr>
<td>LGLEX</td>
<td>80.61</td>
<td>87.89</td>
<td>63.19</td>
<td>1.10</td>
</tr>
<tr>
<td>DICOVALENCE</td>
<td>71.44</td>
<td>88.08</td>
<td>64.49</td>
<td>0.38</td>
</tr>
<tr>
<td>Old DICOVALENCE</td>
<td>65.69</td>
<td>87.06</td>
<td>62.72</td>
<td>0.42</td>
</tr>
</tbody>
</table>
Analysis per verbal relation (F-measure)
Experiments on French TreeBank

Evaluation on CONLL dependency version of FTB (journalistic style)
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

On FTB test part (1200 sentences)

<table>
<thead>
<tr>
<th>Lexicon</th>
<th>Cover. (%)</th>
<th>LAS (%)</th>
<th>Time (s)</th>
<th>New LAS (%)</th>
<th>δ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEFFF</td>
<td>89.53</td>
<td>82.21</td>
<td>0.61</td>
<td>85.82</td>
<td>4.4</td>
</tr>
<tr>
<td>NEW LEFFF</td>
<td>88.76</td>
<td>81.36</td>
<td>0.94</td>
<td>83.83</td>
<td>3.0</td>
</tr>
<tr>
<td>LGLEX</td>
<td>86.73</td>
<td>78.75</td>
<td>1.95</td>
<td>81.82</td>
<td>3.9</td>
</tr>
<tr>
<td>DICOVALENCE</td>
<td>75.28</td>
<td>79.38</td>
<td>0.69</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MST</td>
<td>-</td>
<td>88.20</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Analysis per CONLL verbal relation

Recall

Precision

--- Lefff  --- LGLex  --- NewLefff  --- Dicovalence

--- Lefff  --- LGLex  --- NewLefff  --- Dicovalence
Error mining

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

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

\(\sim\) 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

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

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 expected LADL table
  - réaffirmer (to reaffirm 28), réélire (to reelect, 10), mixer (to mix, 7), zapper (to omit, 4)...
  - *la mémoire ... qui zappe les détails ...*
  - *the memory ... which omits the details ...*

- existing entries, but missing coding information
  - susciter (to spark off, 41; 36DT & 38R), recruter (to recruit, 14; 38R), ...

- mandatory args (for LGLEX), but missing ones in the sentences
  - kidnapper (to kidnap, 12; 36DT N0 V N1 Prep N2): *Les deux Italiens ont été kidnappés le 18 décembre*

- misc. situations

For NEW LEFFF: most significant error is *estimer* (to consider): missing clausal argument *(to consider that S)*
Conclusion

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

Future:

- specific training per lexica
- 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
- assign probabilities to entries and/or frames
- enrich FRMG with some new features, to take into account richer lexical information

Thank you