Using Lexicon-Grammar tables for French in a large-coverage parser

Elsa Tolone[1]

1. LIGM, Université Paris-Est (France)

Seminar of PLN group – FaMAF,Universidad Nacional de Córdoba, Argentina
August 13, 2010

Context

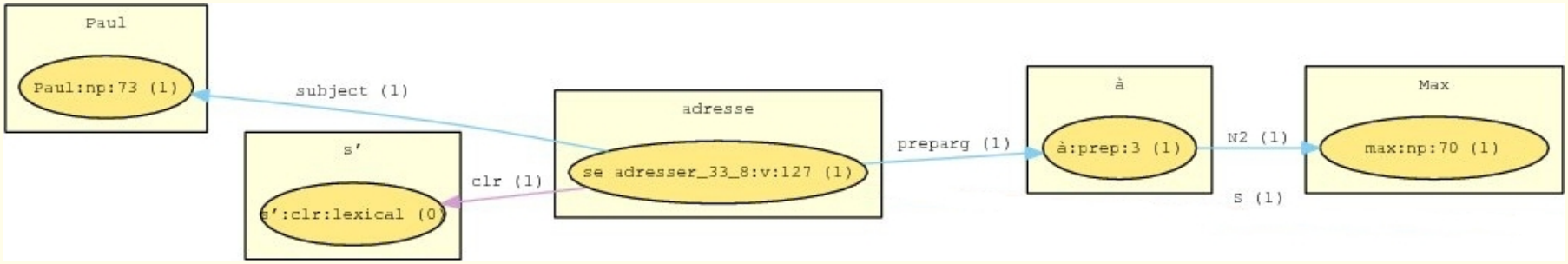
- **Lexicon-Grammar tables** for French are a large-coverage lexical resource
- They contain **syntactic** and semantico-syntactic information
- Such information is arguably very **useful for parsing**
- But Lexicon-Grammar tables are **not directly usable** as such in a parser
  - features that are shared by all entries in a given table are not explicitly given
  - lexical features are not properly formalized
  - these data need to be integrated in a real-life symbolic parser

Objectives

- Three major objectives
  1. **convert** Lexicon-Grammar tables to an NLP format,
  2. **plug** the resulting lexicon, named $LGLex_{\mathrm{Lefff}}$, with a parser
  3. **evaluate** the resulting parser
- NLP tools used:
  - parser: FRMG [Thomasset & de La Clergerie 2005]
  - lexical formalism: Alexina, formalism used by the Lefff lexicon [Sagot 2010] used by FRMG

$\rightarrow$ this allows for a comparison between $\mathrm{FRMG}_{\mathrm{Lefff}}$ and FRMG_{LGLex}

Example of dependencies



Paul s'adresse à Max (*Paul talks to Max*)
→ entry *s'adresser V_33_8*

1. Lexicon-Grammar tables for French

Lexicon-Grammar tables

Developed **manually** for over 40 years by the LADL [Gross 1975], and the Computational Linguistics Group of LIGM (Université Paris-Est)

- Study the syntax in a basic sentence (or **subcategorization frame**)
  e.g.: N0 V N1
- Use in French of verbs, adverbs, predicative nouns and adjectives and frozen expressions
  → they **share some features**
- The different meanings are distinguished
  (e.g.: *cuisiner* (*to cook something/to cook someone*))

Principle

- Each class is described in a **table**:
  - one row for each (lemma-level) entry
  - one column for each feature that is relevant for the class
  - at the intersection of a row and a column, $+$ (resp. $-$) $=$ the corresponding feature is valid (resp. not valid) for the corresponding entry
- A class is defined by a set of **"defining features"**
- For a given table, the defining features often include:
  - a basic defining feature, often a subcategorization frame,
  - often additional features (distributional, morphological, transformational, semantic, etc.)
    e.g.: N0 =: Nhum $\rightarrow$ names of people

Table V_33

N0 =: Nhum	N0 =: N-hum	N0 =: Nnr	Ppv	Ppv =: se figé	Ppv =: en figé	Ppv =: les figé	Nég	<ENT>	N0 V	N0 être V-ant	N1 =: Nhum	N1 =: N-hum	N1 =: le fait Qu P	Ppv =: lui	Ppv =: y	N0hum V W sur ce point	[extrap]	<OPT>
+	-	-	<E>	-	-	-	-	renaître	+	+	-	+	-	-	+	-	-	Max renaît au bonheur de vivre
+	-	-	se	+	-	-	-	rendre	+	-	+	+	+	-	+	+	+	Max s'est rendu à mon opinion
+	-	-	se	+	-	-	-	rendre	+	-	+	-	-	-	-	-	-	Le caporal s'est rendu à l'ennemi
+	-	-	<E>	-	-	-	-	renoncer	-	-	+	+	-	-	+	-	-	Max renonce à son héritage

Defining feature: N0 V à N1

Inventory

- Inventory [Tolone 2009]:
  - 64 classes of **simple verbs**
    - 13 862 entries for 5 739 distinct lemmas
  - 32 classes of **simple and frozen adverbs** (adverbs in *-ment* and frozen adverbs)
    - 10 487 entries for 9 273 distinct lemmas
      e.g.: *[changer] du jour au lendemain* (*[to change] overnight*)
  - 78 classes of **simple and frozen predicative nouns** (nouns with argument(s)that are studied with their light verb)
    - 12 696 entries for 8 530 distinct lemmas
      e.g.: *Luc monte une attaque contre le fort*
      (*Luc is lauching an attack against the fort*)
  - 69 classes of **verbal frozen expressions**
    - 39 627 entries for 38 626 distinct lemmas
      e.g.: *Tu n'arrives pas à la cheville de Marie*
      (*You can't hold a candle to Mary*,
      literaly *You don't arrive at the ankle of Mary*)

Additional entries

- simple predicative nouns:
  - 8 classes with also **simple predicative adjectives**
    - 1 408 entries for 1 287 distinct lemmas
      e.g.: *Max a une certaine aptitude à supporter la douleur*
      → *Max est apte à supporter la douleur*
      (*Max has a certain ability to withstand pain*
      → *Max is able to withstand pain*)
- adverbs :
  - 1 classe with also **compound predicative adjectives**
    - 304 entries for 300 distinct lemmas
      e.g.: *être doux comme un ange*
      (*to be gentle as an angel*)

Problems

- **Different names** for the same feature
  → Harmonization of the column headings
- Features **not defined clearly**
  → Documentation of features
- **Implicit defining features** (literature)
  → Constant + or − for the whole of the table
- All features are **not encoded** in each table
  → Symbol +, − or *o* for the whole of the table
- **Outdated** entries in the tables
  → Added missing entries
- **Uncoded** entries
  → Coding entries

LGLex

The improvement of the tables permits the extraction of a **syntactic lexicon** for each categories from Lexicon-Grammar tables [Constant & Tolone 2010]:

- text or XML format
- named *LGLex*
- generated from the original Excel verb tables by the *LGExtract* tool
- the version 3.2 contained *LGLex* of tables which are distributed for each category on `http://infolingu.univ-mlv.fr/`

LGLex is the starting point of the conversion process towards the Alexina format

Format of *LGLex* lexicon

- **ID=category_numTable_numEntry**
- **lexical-info**=[...] → lemma and lexical information (auxiliaries, support verbs, determiners, prepositions)
- **args**=(...) → arguments and their nature with other information (semantic features, mood of the complementizer phrase, argument controled by the infinitive, prepositions)
- **all-constructions**=[absolute=(...), relative=(...)] → list of accepted constructions (either entirely specifying the accepted construction with all its constituents in the order, or transformations from the main construction, but also GN reductions, structure of the multiword adverbs, etc.)
- **example**=[...] → an illustrative example of the entry

LGLex: an example (verb)

ID=V_36L_28
*lexical-info=[cat="verb",verb=[***lemma="clouer"***],prepositions=(***preposition=[id="2",list=(prep="avec")]***)*
locatifs=(locatif=[id="2",list=()],aux-list=())]
args=(
*const=[***pos="0"***,dist=(comp=[***cat="NP",hum="true"***,introd-prep=(),introd-loc=(),*
origin=(orig="N0 =: Nhum")],
comp=[cat="NP",introd-prep=(),introd-loc=(),nothum="true",
origin=(orig="N0 =: N-hum")]))],
*const=[***pos="1"***,dist=(comp=[***cat="NP",nothum="true"***,introd-prep=(),introd-loc=(),*
origin=(orig="N1 =: N-hum")]))])
*const=[***pos="2"***,dist=(comp=[***cat="NP",destination="true"***,introd-prep=(),***introd-loc=(prep="à",prep="sur")***,*
origin=(orig="Loc N2 =: à N2 destination",orig="Loc N2 =: sur N2 destination")]))],
*all-constructions=[absolute=(construction="true::***N0 V N1 Loc N2***",construction="o::***N0 V N1***",*
*construction="true::***N0 V N1 Prép N2***", construction="true::N0 V N1 et N2"*
construction="o::N0 V N1 de N3 attache",construction="o::N0 V N1hum Loc N2abs",
construction="o::N3 attache V N1",construction="o::N0 V N1 + 2"),
*relative=(construction="***[passif par]***")]*
example=[example="Max a cloué cette planche(avec+contre+sur)celle-là"]

- entry *clouer*V_36L_28 (*to nail*)
- Remains to be done: **interpret** a number of columns

2. The Lefff and the Alexina format

Le Lefff

- The Lefff (Lexique des Formes Fléchies du Français) is a morphological and syntactic lexicon for French [Sagot 2010]
  - large coverage (536 375 entries corresponding to 110 477 distinct lemmas covering all categories)
  - freely available (LGPL-LR license)
- It relies on the **Alexina** framework for the modeling and acquisition of morphological and syntactic lexicons

Alexina

Two-level architecture

- The **intensional** lexicon
  - associates with each entry (meaning of a lemma) a canonical subcategorization frame
  - lists all possible redistributions (restructurations) from this frame
- The **compilation** process of the intensional lexicon into the **extensional** lexicon generates different entries for each inflected form and each possible redistribution

Alexina on an example

- Example of an intensional entry:

  clarifier$_1$ *v-er:std*
  Lemma;v;
  *<***Suj***:cln|scompl|sinf|sn,***Obj***:(cla|scompl|sn)>;*
  %active,%se_moyen_impersonal,
  %passive_impersonal,%passive

- **Syntactic functions** (cf. Dicovalence): Suj, Obj, Objà, Objde, Loc, Dloc, Att, Obl/Obl2
- **Realizations**: direct (sn, sa, sinf, scompl, qcompl); clitic (cln, cla, cld, y, en); prepositional (prep+direct, e.g., par-sn, à-sinf, de-scompl)
- **Redistributions**: %active, %passive, %se_neutre, %impersonnal_active, etc.

3. Converting *LGLex* into an Alexina lexicon

The conversion process

- The conversion of Lexicon-Grammar tables into the Alexina framework is **not straightforward**
  - It requires a **formal definition** or a **dynamic interpretation** of all feature names
- We won't enter into the details of this conversion process

LGLex: the previous example (verb)

ID=V_36L_28
*lexical-info=[cat="verb",verb=[***lemma="clouer"***],prepositions=(***preposition=[id="2",list=(prep="avec")]***)*
locatifs=(locatif=[id="2",list=()],aux-list=())]
args=(
*const=[***pos="0"***,dist=(comp=[***cat="NP",hum="true"***,introd-prep=(),introd-loc=(),*
origin=(orig="N0 =: Nhum")],
comp=[cat="NP",introd-prep=(),introd-loc=(),nothum="true",
origin=(orig="N0 =: N-hum")])],
*const=[***pos="1"***,dist=(comp=[***cat="NP",nothum="true"***,introd-prep=(),introd-loc=(),*
origin=(orig="N1 =: N-hum")])])
*const=[***pos="2"***,dist=(comp=[***cat="NP",destination="true"***,introd-prep=(),***introd-loc=(prep="à",prep="sur")***,*
origin=(orig="Loc N2 =: à N2 destination",orig="Loc N2 =: sur N2 destination")])],
*all-constructions=[absolute=(construction="true::***N0 V N1 Loc N2***",construction="o::***N0 V N1***",*
*construction="true::***N0 V N1 Prép N2***", construction="true::N0 V N1 et N2"*
construction="o::N0 V N1 de N3 attache",construction="o::N0 V N1hum Loc N2abs",
construction="o::N3 attache V N1",construction="o::N0 V N1 + 2"),
*relative=(construction="***[passif par]***")]*
example=[example="Max a cloué cette planche(avec+contre+sur)celle-là"]

- entry *clouer*V_36L_28 (*to nail*)

The previous example after conversion

clouer V_36SL_28 *v-er:std*
100;Lemma;v;
*<***Suj***:cln|sn,***Obj***:sn,***Loc***:(à-sn|sur-sn|avec-sn)>;*
*cat=v;***@SujNhum***;***@ObjN-hum***;*
%actif,%passif,%ppp_employé_comme_adj

- *Max a cloué ces planches au mur*
  (*Max has nailed this boards on the wall*)
  - *Max a cloué ces planches*
    (*Max has nailed this boards*)

The previous example after conversion (2)

clouer V_36SL_28 *v-er:std*
100;Lemma;v;
*<**Suj**:cln|sn,**Obj**:sn,**Loc**:(à-sn|sur-sn|avec-sn)>;*
cat=v;@SujNhum;@ObjN-hum;
%actif,%passif,%ppp_employé_comme_adj

Entry in the **intensional** lexicon :

- entry identifier : categorie_numTable_numEntry
- morphological class, which defines the patterns that build its inflected forms, using inflection classes from the Lefff
- category (or part-of-speech)
- initial sub-categorization frame
- additional information represented by macros
- the list of possible redistributions

The previous example after conversion (3)

clouer V_36SL_28 *v-er:std*
100;Lemma;v;
*<**Suj**:cln|sn,**Obj**:sn,**Loc**:(à-sn|sur-sn|avec-sn)>;*
cat=v;@SujNhum;@ObjN-hum;
*%actif,**%passif**,%ppp_employé_comme_adj*

One entry in the **extensional** lexicon:

clouées *100 v*
pred=clouer V_36SL_28
*<**Suj**:sn,**Loc**:(à-sn|sur-sn|avec-sn),**Obl**:(par-sn)>;*
@passive**,**@pers**,cat=v,**@Kfp

- *Ces planches ont été clouées au mur par Max*
  (This boards have been nailed on the wall by Max)

Another extract of *LGLex* lexicon (noun)

ID=N_aa_18
*lexical-info=[cat="noun",noun=[notperm=[***complete="considération"***],noun1="considération"]],*
Vsup=[cat="verb",list=(value="ressentir",value="éprouver",value="avoir")],
Vsup2=[cat="verb",list=(value="avoir",value="être l'objet de",value="faire l'objet de")],
*detN=[list-det-modif=(det-modif=[***det="du+de l'+de la"***],modif="false"],*
*det-modif=[det="un+une",modif="true"]),prepositions=(***preposition=[id="1",list=(prep="pour")]***)]*
args=(
const=[pos="0",dist=(comp=[cat="NP",hum="true"])],
*const=[***pos="1"***,dist=(comp=[***cat="NP",nothum="true"***],comp=[***cat="NP",hum="true"***])])*
*all-constructions=[absolute=(construction="true::***N0 avoir Det N Prép N1***"),*
*construction="true::***N1 avoir Det N de N0***",*
reductionsGN=(construction="::le N de N0 Prép N1"),relative=(),verbales=()]
example=[example="Max a de la considération pour (ce geste+Luc)"]

- *Max a de la* **considération** *pour (ce geste+Luc)*
  (*Max has the account for (this gesture+Luc)*,
  literaly *Max has consideration for (this gesture+Luc)*)
- *(Ce geste+Luc) a de la considération pour Max*
  (*(This gesture+Luc) has the account for Max*)

Result of the conversion

considération N_aa_18	*nc-2f* *100;Lemma;cf;* *<***Suj***:cln|sn,***Obl***:pour-sn>;* *cat=nc;@SujNhum;@OblN-hum;@OblNhum;* *lightverb=avoir|ressentir|éprouver;* *%default*
considération N_aa_18	*nc-2f* *100;Lemma;cf;* *<***Suj***:cln|sn,***Objde***:de-sn>;* *cat=nc;@SujNhum;* *lightverb=avoir;* *%default*

The version 3.2 contained verbs and predicative nouns of tables which are distributed on `http://infolingu.univ-mlv.fr/`

- Remains to be convert: **frozen expressions** and **adverbs**

The resulting lexicon: $LGLex_{\mathrm{Lefff}}$

- **verbes** : $LGLex_{\mathrm{Lefff}}$ contains 16 955 entries for 5 723 unique verb lemmas (2,96 entries per lemma)
  - to be compared with the last published version of the Lefff: 7 072 verb entries for 6 818 unique verb lemmas (1,04 entries per lemma)

  At the extensional level, the Lefff contains 361 268 entries, whereas $LGLex_{\mathrm{Lefff}}$ contains 862 546
- **predicatives nouns** : $LGLex_{\mathrm{Lefff}}$ contains 17 517 entries for 8 363 distinct lemmas (2,09 entries per lemma)
  - The Lefff contains only 218 entries of predicative nouns (1 entry per lemma)

  At the extensional level, the *LGLex* contains 32 389 entries

4. Integration in the FRMG parser

The FRMG parser

FRMG is a parser **TAG** of French:

- from the compilation of a **meta-grammar**
- very **compact** thanks to the factorization of trees (1 986 300 defactorized trees $\rightarrow \sim$ 75 630 trees)
- exploiting the functionalities of **DyALog** $\rightarrow$ logic programming environment (language, compiler, virtual machine)

FRMG fits into a processing chain:

- upstream, with **SXPipe** and the lexicon **Lefff**
  - **SXPipe** : segmentation, token, corrections, named entities $\rightarrow$ returns a **DAG** (graph of words)
  - **Lefff**: morphological and syntactic lexicon for French $\rightarrow$ connection lexicon/grammar: anchoring with **hypertag**
- downstream, with a module of **disambiguation**

Integration in the FRMG parser

- We replaced the Lefff with a modified version of the Lefff in which verb entries are replaced by $LGLex_{\mathrm{Lefff}}$
- additional Lefff entries must be added for
  - (semi-)auxiliaries
  - several raising verbs
  - impersonal verb constructions
  - light verbs

The result is a **variant of FRMG**, named FRMG_{LGLex} unlike the standard variant denoted by $\mathrm{FRMG}_{\mathrm{Lefff}}$

5. Evaluation and discussion

Protocol used

- We evaluated $\text{FRMG}_{\text{Lefff}}$ and FRMG_{LGLex} by parsing the manually annotated part of the EASy corpus [Paroubek *et al.* 2005]
  - 4 306 sentences of various genres (journalistic, medical, oral, questions, literacy, etc.)
- evaluation metrics: those of the first EASy parsers' evaluation campaign that took place in December 2005 [Paroubek *et al.* 2006]
  - evaluation in **chunks** and **relations** ($\sim$ dependencies between lexical words)

Preliminary remarks

FRMG$_{LGLex}$'s results must be analyzed with the following facts in mind:

- FRMG$_{LGLex}$'s verb entries are the result of a conversion process from the original tables
  $\rightarrow$ this conversion process certainly introduces errors
- EASy not allow to evaluate all the information contained in tables (e.g. semantic features)
- the Lefff was developed in parallel with the EASy campaigns (unlike Lexicon-Grammar tables)

Results

Comparative results of $\text{FRMG}_{\text{Lefff}}$ and FRMG_{LGLex} (in terms of f-measure):

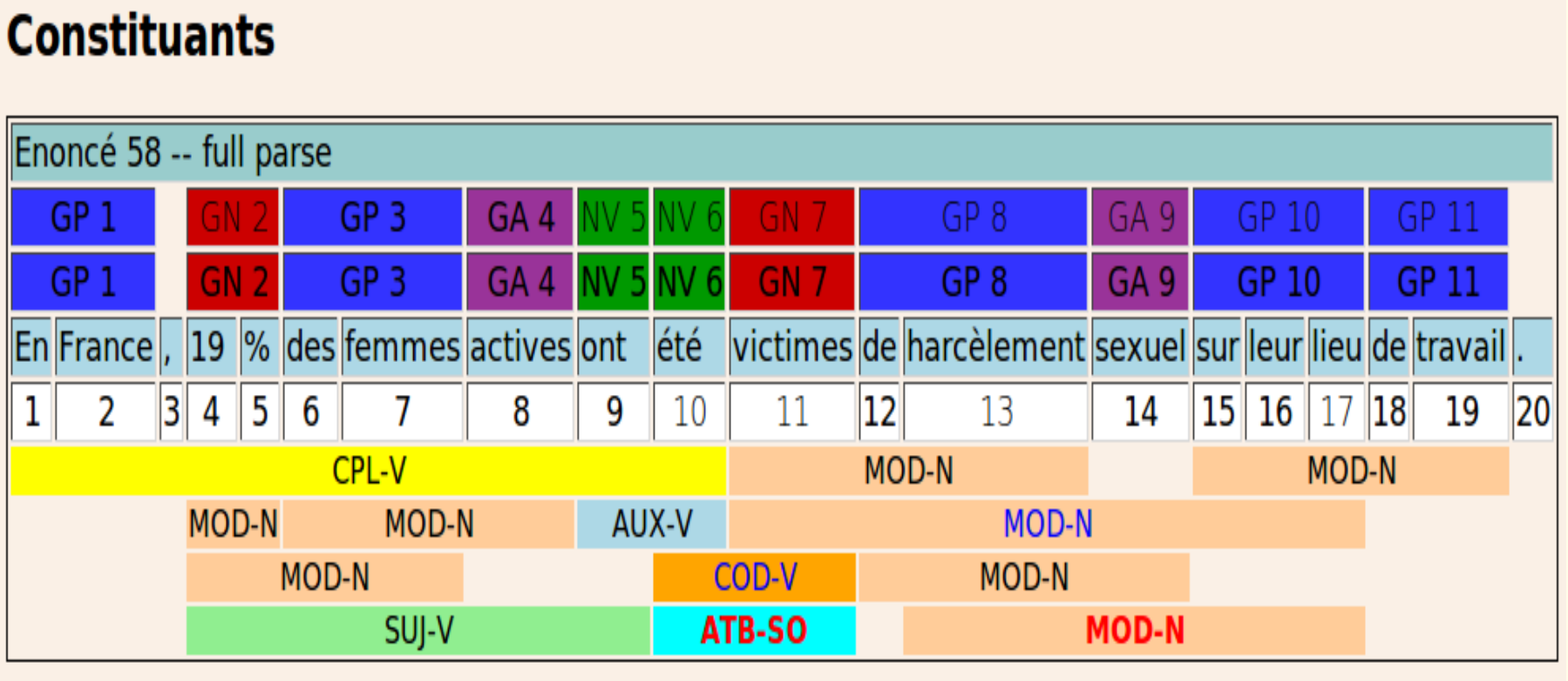
	Chunks		Relations	
Sub-corpus	$\text{FRMG}_{\text{Lefff}}$	FRMG_{LGLex}	$\text{FRMG}_{\text{Lefff}}$	FRMG_{LGLex}
`general_lemonde`	**88.22%**	84.60%	**62.73%**	59.01%
`litteraire_2`	**88.91%**	88.46%	**65.28%**	62.43%
`mail_9`	**82.60%**	81.90%	**58.55%**	56.00%
`medical_3`	85.04%	**85.89%**	64.79%	**65.26%**
`oral_delic_4`	78.80%	**81.79%**	**51.67%**	51.14%
`questions_amaryllis`	**91.30%**	90.73%	**66.56%**	64.77%
total	**87.05%**	*85.53%*	**63.10%**	*60.25%*

Parsing times higher with FRMG_{LGLex} than with $\text{FRMG}_{\text{Lefff}}$: the median parsing time per sentence is 0,62s vs. 0,26s

- this comes from the higher average number of entries per verb lemma (approx. 3) in *LGLex* than in the Lefff
  $\rightarrow$ more ambiguity

- FRMG_{LGLex} gives better results than $\text{FRMG}_{\text{Lefff}}$ for some relations
  - "standard" relations MOD-A (adjective modifier) and MOD-R (adverb modifier)
  - "tough" relations MOD-P (preposition modifier) and APP (apposition)
- the ATB-SO relation (subject or object attribute) is the relation with the highest difference in terms of recall (34,0% vs. 58,4%)
  - this is because Lexicon-Grammar tables encode very little information about attribution phenomena, but it can be due to errors of reference

Example: annotation of components



Example: annotation of relations

Relations

1. Sujet - Verbe

sujet	verbe
G2 F5 67%	G5 F9

2. Auxiliaire - Verbe

auxiliaire	verbe
G5 F9	G6 F10

3. COD - Verbe

COD	verbe
G7	G6

4. Complément - Verbe

complément	verbe
G1 F2 67%	G6 F10

7. Attribut - Sujet / Objet

attribut	verbe	sujet / objet
F11	F10	sujet

8. Modifieur - Nom

modifieur	nom	à propager
G3 F7 67%	G2 F5 67%	
F4	F5	
G4 F8	G3 F7 67%	
G8 F13 67%	G7 F11	
G9 F14	G8 F13 67%	
G11 F19 67%	G10 F17 50%	
G10	G7	
F17	F13	

Conclusions and perspectives

Short-term

- Many sentences receive a full parse from FRMG_{LGLex} but not from $\text{FRMG}_{\text{Lefff}}$, and vice-versa
  - $\rightarrow$ **coupling both parser variants** could prove useful, since full parses have a higher f-measure than partial parses
- Lefff and $LGLex_{\text{Lefff}}$ are **complementary** in many aspects
- $\rightarrow$ use automatic techniques to improve each resource thanks to the other (e.g., via statistical analysis of parsing results [Sagot & de La Clergerie 2008])

Long-term

Optimize the use of lexical data in Lexicon-Grammar for parsing

- coding the **missing and uncoding entries** in the tables
- **improve/correct the conversion process**
- generalize the technique to Lexicon-Grammar tables for **other categories**
- generalize the technique to **other languages** for which large-coverage Lexicon-Grammar tables are available (e.g., Greek)

Links

- `http://www-igm.univ-mlv.fr/~tolone/`
- Le*ffe* (Léxico de las Formas Flexionadas del Español) [Molinero et al. 2009]
  - Miguel Ángel Molinero Álvarez, Benoît Sagot and Lionel Nicolas. A morphological and syntactic wide-coverage lexicon for Spanish: The Leffe. Proceedings of RANLP'09, 6 pp. Borovets, Bulgaria. 2009.
  - *Victoria* project (La Coruña (Galicia), Spain / Paris and Nice, France) → Simplify the acquisition, expansion and correction:
    - morphological rules
    - coverage lexicon
    - grammars

References

- [Constant & Tolone 2010] Matthieu Constant and Elsa Tolone. A generic tool to generate a lexicon for NLP from Lexicon- Grammar tables. Lingue d'Europa e del Mediterraneo, Grammatica comparata, vol.1, pp.79-93. Aracne. 2010.
- [Gross 1975] Maurice Gross. Méthodes en syntaxe : Régime des constructions complétives. Hermann. Paris, France. 1975.
- [Paroubek et al. 2005] Patrick Paroubek, Louis-Gabriel Pouillot, Isabelle Robba and Anne Vilnat. Premier bilan de la participation du LORIA à la campagne d'évaluation EASy. Proceedings of the EASy workshop of TALN'05. Dourdan, France. 2005.
- [Paroubek et al. 2006] Patrick Paroubek, Isabelle Robba, Anne Vilnat and Christelle Ayache. Data, Annotations and Measures in EASy: the Evaluation Campaign for Parsers of French. Proceedings of LREC'06. Genoa, Italy. 2006.

References (2)

- [Sagot et de La Clergerie 2008] Benoît Sagot and Éric de La Clergerie. Fouille d'erreurs sur les sorties d'analyseurs syntaxiques. Traitement Automatique des Langues (T.A.L.), vol.49, num.1, pp. 41-60. Hermès. Paris, France. 2008.
- [Sagot 2010] Benoît Sagot. The Lefff, a freely available and large-coverage morphological and syntactic lexicon for French. Proceedings of LREC'10, 8 pp. Valletta, Malta. 2010.
- [Thomasset & de La Clergerie 2005] François Thomasset and Éric de La Clergerie. Comment obtenir plus des méta-grammaires.Proceedings of TALN'05. Dourdan, France. 2005.
- [Tolone 2009] Elsa Tolone. Les tables du Lexique-Grammaire au format TAL. Proceedings of MajecSTIC'09. Avignon, France. 2009.
- [Tolone & Sagot 2009] Elsa Tolone et Benoît Sagot. Using Lexicon-Grammar tables for French verbs in a large-coverage parser. Proceedings of LTC'09, pp. 200-204. Poznań, Poland 2009.

Example of DAG

- ##DAG BEGIN
  1 {<F id="E1F1">une</F>} "une" 2
  2 {<F id="E1F2">pomme</F>} "pomme" 3
  2 {<F id="E1F2">pomme</F> <F id="E1F3">de</F>
  <F id="E1F4">terre</F>} "pomme_de_terre" 5
  3 {<F id="E1F3">de</F>} "de" 4
  4 {<F id="E1F4">terre</F>} "terre" 5
  5 {<F id="E1F5">pour</F>} "pour" 6
  6 {<F id="E1F6">20</F>} "_NUMBER" 7
  7 {<F id="E1F7">francs</F>} "francs" 8
  8 {<F id="E1F8">le</F>} "le" 9
  9 {<F id="E1F9">2</F> <F id="E1F10">avril</F>
  <F id="E1F11">2006</F>} "_DATE_arto" 10
  ##DAG END
- Each line is a transition:
  starting_state {comment} "word" arrival_state