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

#### $\mathsf{Elsa}\ \mathsf{Tolone}^1$

1. Computational Linguistics Group

IGM, Université Paris-Est (France) February 24, 2010

### Context

- Lexicon-Grammar tables are a large-coverage lexical resource developed manually for years
- They contain syntactic and semantico-syntactic information
- Such information is arguably very useful for parsing
- But Lexicon-Grammar tables are not 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 parser

# Objectives

- Three major objectives
  - 1. convert Lexicon-Grammar tables to an NLP format,
  - 2. plug the resulting lexicon, named  $Iglex_{Lefff}$ , 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 et al. 2006] used by FRMG
  - $\rightarrow$  this allows for a comparison between  $\mathsf{FRMG}_{\mathsf{Lefff}}$  and  $\mathsf{FRMG}_{\mathit{lglex}}$

- 2. Lexicon-Grammar verb tables for French
  - 3. Converting *lglex* into a NLP lexicon

Conclusions and perspectives

#### 1. Parsing

#### 2. Lexicon-Grammar verb tables for French

- 2.1. Lexicon-Grammar tables
- 2.2. Table of classes
- 2.3. *Iglex*

#### 3. Converting *lglex* into a NLP lexicon

- 3.1. Converting *lglex* into a NLP lexicon
- 3.2. Integration in the FRMG parser

2. Lexicon-Grammar verb tables for French 3. Converting *lglex* into a NLP lexicon Conclusions and perspectives

# 1. Parsing

2. Lexicon-Grammar verb tables for French 3. Converting /g/ex into a NLP lexicon Conclusions and perspectives

## Parsing

#### Building the syntactic structure of a sentence in order to give them a meaning

 $\rightarrow$  explicit the dependency relationships between different words (e.g., between subject and object) Difficulty : **Complexity and size of the language** 

#### Applications :

- Reading comprehension
- Information extraction
- Translation

2. Lexicon-Grammar verb tables for French 3. Converting /g/ex into a NLP lexicon Conclusions and perspectives

#### Parsers

- symbolic = grammar + lexicon that specifies the syntactic behavior of each word of the language
  - $\rightarrow$  developed manually
- probabilistic = model acquired from a manually annotated corpus
- Symbolic parser used : FRMG [Thomasset & de La Clergerie 2005]

2. Lexicon-Grammar verb tables for French 3. Converting /g/ex into a NLP lexicon Conclusions and perspectives

### Example of dependencies



Paul s'adresse à Max (Paul talks to Max)

2. Lexicon-Grammar verb tables for French

3. Converting *lglex* into a NLP lexicon Conclusions and perspectives 2.1. Lexicon-Grammar tables

- 2.2. Table of classes
- **2.3.** Iglex

### 2. Lexicon-Grammar verb tables for French

2. Lexicon-Grammar verb tables for French

3. Converting *lglex* into a NLP lexicon Conclusions and perspectives 2.1. Lexicon-Grammar tables 2.2. Table of classes

2.3. Iglex

## Lexicon-Grammar tables

- Specifications :
  - Developed manually for over 30 years by the Computational Linguistics Group of IGM (Université Paris-Est)
  - Describe the syntactic features of words
  - Rich lexical ressources
  - Not directly usable in a symbolic parser
- [Gross 1975]
  - Study the syntax in a basic sentence (or subcategorization frame)
     e.g.: N<sub>0</sub> V N<sub>1</sub>
  - Use in French of verbs, adverbs, predicative nouns and adjectives and frozen expressions
    - $\rightarrow$  they share some features
  - The different meanings are distinguished (e.g.: cuisiner (to cook something/to cook someone))

2. Lexicon-Grammar verb tables for French

3. Converting *lglex* into a NLP lexicon Conclusions and perspectives

# 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

2.1. Lexicon-Grammar tables

2.2. Table of classes

2.3. Iglex

- 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.: N<sub>0</sub> := Nhum → names of people
- ► These features define **61 verb classes** for 13,400 simple verbs

2.1. Lexicon-Grammar tables

2.2. Table of classes

2. Lexicon-Grammar verb tables for French

3. Converting *lglex* into a NLP lexicon Conclusions and perspectives

## Table V33

INU =: INNUM	N0 =: N-hum	N0 =: Nnr	Рру	Ppv =: se figé	Ppv =: en figé	Ppv =: les figé	Nég	<ent></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></opt>
	-	-	<e></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></e>	-	-	-	-	renoncer	-	-	+	+	-	-	+	-	-	Max renonce à son héritage

Defining feature:  $N_0 \vee a N_1$ 

2. Lexicon-Grammar verb tables for French

3. Converting *lglex* into a NLP lexicon Conclusions and perspectives

### Outcome

- Inventory :
  - ▶ 61 classes of simple verbs
  - > 32 classes of adverbs (adverbs in -ment and frozen adverb)
  - 59 classes of predicative nouns (nouns with argument(s) that are studied with their light verb)
     e.g.: Luc monte une attaque contre le fort

2.1. Lexicon-Grammar tables

2.2. Table of classes

2.3. Iglex

(Luc is lauching an attack against the fort)

 65 classes of frozen expressions
 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)

#### Avantages :

- Rich description
- Large-coverage
- Solid linguistic base

2.1. Lexicon-Grammar tables

2.2. Table of classes

2.3. Iglex

2. Lexicon-Grammar verb tables for French

3. Converting *lglex* into a NLP lexicon Conclusions and perspectives

## Problems

► Different names for the same feature → Harmonization of the column headings e.g.: [extrap] and *il* V N<sub>0</sub> W

- Features not defined clearly
  - $\rightarrow$  Material on features
- Implicit defining features (literature)
  - $\rightarrow$  Constant + or for the whole of the table
- All features are not encoded in each table
  - $\rightarrow$  Symbol +, or  $\mathit{o}$  for the whole of the table

Work in progress for verbs, adverbs, predicative nouns, and frozen expressions

Parsing
 Lexicon-Grammar verb tables for French
 Converting /g/ex into a NLP lexicon

Conclusions and perspectives

### Table of classes

Defining features are not represented in the tables

- $\rightarrow$  to be dealt with in a table of classes for simple verbs:
  - one row for each class
  - one column for each feature (overall, after normalization among tables)
  - at the intersection of a row and a column,
    - o = the feature is explicitly dealt with in the corresponding table

2.1. Lexicon-Grammar tables

2.2. Table of classes

2.3. Iglex

► + (resp. -) = the corresponding feature is valid (resp. not valid) for all entries in the corresponding class

The table of simple verb classes has just been completed [Constant & Tolone 2008]

2. Lexicon-Grammar verb tables for French

3. Converting *lglex* into a NLP lexicon Conclusions and perspectives 2.1. Lexicon-Grammar tables

2.2. Table of classes

**2.3.** Iglex

#### The table of simple verb classes

table	N0 =: Nhum	N0 =: N-hum	N0 =: Nnr	N0 =: V1-inf W	<ent></ent>	Ppv =: se figé	N0 V	N0 V N1	zone 1	N0 V à N1	N1 =: Nhum	N1 =: N-hum	N0 V Prep N1 V0-inf W	N0 V N1 V0-inf W	N0 V V0-inf W
V_2	+	-	-	-	0	0	-	-	-	-	-	+	0	0	+
V_4	-	-	+	+	0	-	0	+	-	-	о	0	-	-	-
V_31R	0	0	-	-	0	0	+	-	-	-	-	-	-	-	-
V_31H	+	-	-	-	0	0	+	-	-	-	-	-	-	-	-
V_33	0	0	0	-	0	0	0	-	-	+	0	0	-	-	-
V_32H	0	-	0	-	0	0	-	+	-	-	+	-	-	-	-

2.1. Lexicon-Grammar tables

2.3. Iglex

2. Lexicon-Grammar verb tables for French

3. Converting *lglex* into a NLP lexicon Conclusions and perspectives



The table of simple verb classes permits the extraction of a **syntactic lexicon** of simple verbs from Lexicon-Grammar tables [Constant & Tolone 2008]:

- text or XML format
- named Iglex
- generated from the original Excel verb tables by the LGExtract tool

 $\mathit{lglex}$  is the starting point of the conversion process towards the Alexina format

2. Lexicon-Grammar verb tables for French

3. Converting *lglex* into a NLP lexicon Conclusions and perspectives 2.1. Lexicon-Grammar tables

2.2. Table of classes

2.3. Iglex

### Iglex: an example

ID=V\_35L\_242 lexical-info=[locs=(loc=[id="1",list=()],loc=[id="2",list=()]),cat="verb",verb=[lemma="ruisseler"], aux-list=(),prepositions=()] args=( const=[dist=(comp=[cat="NP",source="true",introd-prep=(),origine=(orig="Loc N1 =: de N1 source"), introd-loc=(prep="de")]),pos="1"], const=[dist=(comp=[cat="NP",introd-prep=(),origine=(orig="Loc N2 =: vers N2 destination", orig="Loc N2 =: dans N2 destination"),introd-loc=(prep="vers",prep="dans"),destination="true"]),pos="2"], const=[pos="0",dist=(comp=[cat="NP",introd-prep=(),nothum="true",origine=(orig="N0 =: N-hum"), introd-loc=()])]) all-constructions=[absolute=(construction="o::N0 V Loc N1 source Loc N2 destination",construction="o::N0 V", construction="0::N0 être V-ant",construction="true::N0 V Loc N1"), relative=(construction="Ppv =: y",construction="Ppv =: en",construction="[extrap]")] example=[example="L'eau ruisselle de la gouttière sur les passants"]

#### Remains to be done: interpret a number of columns

3.1. Converting *lglex* into a NLP lexicon

3.2. Integration in the FRMG parser

### Le Lefff and the conversion

- ► The Lefff (Lexique des Formes Fléchies du Français) is a morphological and syntactic lexicon for French [Sagot *et al.* 2006]
  → used by the FRMG parser: This allows a comparison between FRMG<sub>Lefff</sub> et FRMG<sub>lglex</sub>
- ► The conversion of Lexicon-Grammar tables into the Alexina framework is not straightforward → made for the simple verbs [Tolone & Sagot 2009]

3.1. Converting Iglex into a NLP lexicon

3.2. Integration in the FRMG parser

### The previous example after conversion

#### ruisseler<sup>35L</sup><sub>242</sub> v-er:std 100;Lemma;v; <**Suj**:cln|sn,**Dloc**:(de-sn|en),**Loc**:(vers-sn|dans-sn|y)>; cat=v; %active

 L'eau ruissèle du haut de la montagne vers le bas de la vallée (Water streams from the top of the mountain towards the bottom of the valley)

3.1. Converting Iglex into a NLP lexicon

3.2. Integration in the FRMG parser

## Integration in the FRMG parser

- We replaced the Lefff with a modified version of the Lefff in which verb entries are replaced by lglex<sub>Lefff</sub>
- The result is a variant of FRMG, named FRMG<sub>lglex</sub> unlike the standard variant denoted by FRMG<sub>Lefff</sub>.
- FRMG<sub>lglex</sub> gives better results than FRMG<sub>Lefff</sub> for some chunks or some relations with the different corpus, and vice-versa [Tolone & Sagot 2009]

 $\rightarrow$  in the future:  $\mbox{coupling both parser variants}$  to improve the results

- The lexical ambiguity is higher in FRMG<sub>Iglex</sub> because the number of entries is higher. This implies:
  - a higher ambiguity for the parser
  - and therefore a higher error rate in the disambiguation step

#### **Conclusions and perspectives**

#### Long-term

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

- continue to improve the tables
- end the table of classes for each category
- 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)