Introducing the French Lexical Network

Veronika Lux-Pogodalla

Toulouse, January 2012
FLN people

- scientific director: **Alain Polguère**

In its first three years, the FLN is

- supported by a grant from the Agence de Mobilisation Économique de Lorraine and Fonds Européen de Développement Régional, in the context of a R&D project called RELIEF
- developed in collaboration with the MVS company (esp. **N. Gader**)
FLN builds on previous research and lexicographical work:

- framework: Explanatory Combinatorial Lexicography [Mel’čuk et al.1995, Mel’čuk2006]
- practical experience: the Explanatory Combinatorial Dictionaries, the DiCo and Dicoube databases, the Lexique Actif du Français dictionary and the DicoPop web pages [Polguère2012].
1. Micro- and macrostructure of the FLN
   - Nodes in the FLN
   - Links in the FLN
   - Introduction to Lexical Functions

2. Methodology and tools for developing the FLN
   - Lexical coverage
   - Using Lexical Functions to expend the lexical coverage
   - Lexicographic editor

3. Conclusion
**Generic lexical model**

- FLN belongs to the -$Net$ family
  - network structure
  - NOT a dictionary
  - built in a lexicographic way
- FLN is equivalent to a repository of virtual dictionaries
  [Selva et al.2003]
  - does not focus on a limited set of lexical properties
  - includes lexicographic definitions, grammatical features, etc.
  - generic resource from which various dictionaries can be generated
Nodes in the FLN

- Lexical Units
  - lexemes such as
    - Fr. Coup I.1 [Il a reçu un coup sur la tête en tombant.] *He got a knock on his head when he fell.*
    - Fr. Coup I.2 [Le voleur lui a donné un coup sur la tête.]*. *The burglar stroke him a blow on his head.*
  - idioms Fr. ‘Coup de soleil’ *SUNBURN* (lit. ‘knock of sun’)

- Linguistic clichés
  (ex. Fr. Après vous !) *Go ahead!* (lit. ‘After you’)

Introducing the French Lexical Network

Toulouse, January 2012
Microstructure of nodes in the FLN

Six lexicographic zones:

1. GC: grammatical characteristics (part of speech, noun gender, specific inflectional behavior, etc.)
2. DF: definition
3. GP: headword’s government pattern, i.e. the description of its syntactic valency [Mel’čuk2004a, Mel’čuk2004b, Milićević2009]
4. LF: lexical function relations originating from the Lexical Unit
5. EX: for lexicographic exemples
6. PH: pointers to full phrasemes
Lexical Units are connected to other Lexical Units by Lexical Functions [Mel’čuk1996]

- lexical paradigmatic links (e.g. synonymy, antonymy, conversivity, actants names)
- syntagmatic links: collocations controlled by LUs (e.g. typical intensifier, support verb)

Wordforms in definitions will be connected to corresponding LU
Example: all lexical function links leaving from or leading to **CHANTAGE** in the FLN
Lexical functions

- A lexical function $F$ is a correspondence which associates a lexical item $L$, called the argument of $F$, with a set of lexical items $F(L)$ - the value of $F$.

- Lexical Functions provide a formalized language to model semantic relationship between lexical units (paradigmatic LFs) and relations of cooccurrence between lexical units (syntagmatic LFs).

- Database of Lexical Functions used in RLF:
  - around 250 simple standard Lexical Functions
  - around 300 complex standard Lexical Functions
  - around 30 semi-standard or non-standard Lexical Functions
  - organised in around 100 families
Simple Lexical Functions: some examples

- **Syn**
  VÉLO → BICYCLETTE

- **Magn**
  PLUIE → BATTANTE

- **S₁**
  CHIEN → MAÎTRE

- **Real₁**
  VOITURE → CONDUIRE
Lexical Functions: support verb and verb of realisation

<table>
<thead>
<tr>
<th>Support verbs</th>
<th>Verbs of realisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Oper}_i$</td>
<td>$\text{Real}_i$</td>
</tr>
<tr>
<td>$I$</td>
<td>$I$</td>
</tr>
<tr>
<td>$\downarrow$</td>
<td>$\downarrow$</td>
</tr>
<tr>
<td>$A_i$</td>
<td>$A_i$</td>
</tr>
<tr>
<td>$\downarrow$</td>
<td>$\downarrow$</td>
</tr>
<tr>
<td>$L$</td>
<td>$L$</td>
</tr>
<tr>
<td>$\downarrow$</td>
<td>$\downarrow$</td>
</tr>
<tr>
<td>$\text{Func}_i$</td>
<td>$\text{Fact}_i$</td>
</tr>
<tr>
<td>$I$</td>
<td>$I$</td>
</tr>
<tr>
<td>$\downarrow$</td>
<td>$\downarrow$</td>
</tr>
<tr>
<td>$L$</td>
<td>$L$</td>
</tr>
<tr>
<td>$\downarrow$</td>
<td>$\downarrow$</td>
</tr>
<tr>
<td>$A_i$</td>
<td>$A_i$</td>
</tr>
<tr>
<td>$\downarrow$</td>
<td>$\downarrow$</td>
</tr>
<tr>
<td>$\text{Labor}_{ij}$</td>
<td>$\text{Labreal}_{ij}$</td>
</tr>
<tr>
<td>$I$</td>
<td>$I$</td>
</tr>
<tr>
<td>$\downarrow$</td>
<td>$\downarrow$</td>
</tr>
<tr>
<td>$A_i$</td>
<td>$A_i$</td>
</tr>
<tr>
<td>$\downarrow$</td>
<td>$\downarrow$</td>
</tr>
<tr>
<td>$A_j$</td>
<td>$A_j$</td>
</tr>
<tr>
<td>$\downarrow$</td>
<td>$\downarrow$</td>
</tr>
<tr>
<td>$L$</td>
<td>$L$</td>
</tr>
<tr>
<td>Support verbs</td>
<td>Verbs of realisation</td>
</tr>
<tr>
<td>---------------------------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>( \text{Oper}_1(\text{PROMESSE}) = \text{FAIRE} [\text{ART} \sim] )</td>
<td>( \text{Real}_1(\text{PROMESSE} ) = \text{TENIR} [\text{ART} \sim] )</td>
</tr>
<tr>
<td>( \text{Func}_1(\text{AIDE}) = \text{VENIR} [\text{de N}] )</td>
<td>( \text{Fact}_2(\text{MÉMOIRE}) = \text{GARDER} [\text{N}] )</td>
</tr>
<tr>
<td>( \text{Labor}_{12}(\text{INTERROGATOIRE}) = \text{SOUMETTRE} [\text{N à ART} \sim] )</td>
<td>( \text{Labreal}_{12}(\text{PIÈGE}) = \text{PRENDRE} [\text{N à ART} \sim] )</td>
</tr>
</tbody>
</table>
Operators to combine Lexical Functions

- concatenation: $F_1 F_2$
  
  PreparReal$_1$
  VÉLO $\rightarrow$ ENFOURCHER
  
  $S_1 CausOper_1$
  CRIME $\rightarrow$ MOBILE

- $\&$ operator: $F_1 \& F_2$
  
  $S_{\text{Loc}} \& S_2$
  HABITER $\rightarrow$ MAISON

- $\vee$ operator: $F_1 \vee F_2$
  
  Able$_1 \vee A_1$
  JOUER $\rightarrow$ JOUEUR$_{\text{ADJ}}$

- $+$ operator: $F_1 + F_2$
  
  AntiBon + AntiMagn$_{\text{taille}}$
  CHIEN $\rightarrow$ ROQUET
1. Micro- and macrostructure of the FLN

2. Methodology and tools for developing the FLN
   - Lexical coverage
   - Using Lexical Functions to expend the lexical coverage
   - Lexicographic editor

3. Conclusion
Lexical coverage [Polguère and Sikora ITRE]

- Priming wordlist: 3,739 vocables
  - "basic French" e.g. Échelle Dubois-Buyse
  - "Éduscol" vocabulary list
  - Robert Benjamin nomenclature
  - Vocabulary wordlist compiled at the Université de Montréal for the Quebec ministry of education
- Target after three years: at least 10,000 vocables
- On the long run: basic contemporary French
Expending the lexical coverage

- Induction process to make the wordlist gradually grow
- Based on Lexical Functions
  - first step: subset of paradigmatic lexical functions (close semantic derivation)
  - second step: subset of other common standard lexical functions
  - ...
Lexical Functions used in the first induction step (1)

- **Syn**: exact synonyms of L;
  
  VÉLO → BICYCLETTE

- **Anti**: exact antonyms of L;
  
  LÉGAL → ILLÉGAL

- **Syn_{sex}**: intersecting synonym of L that denotes the same individual/animal as L but of the opposite sex;
  
  ACTEUR → ACTRICE, CHIEN → CHIENNE

- **Fem**: lexical unit denoting a female creature corresponding to the male creature denoted by L;
  
  POULE → COQ

- **Masc**: lexical unit denoting a male creature corresponding to the female creature denoted by L;
  
  COQ → POULE
Lexical Functions used in the first induction step (2)

- **V₀**: verbal conversion of L; \textit{miaulement} → \textit{miauler}
- **S₀**: nominal conversion of L; \textit{soupirer} → \textit{soupir}
- **\textbf{Adj₀}**: adjectival conversion of L; \textit{abdomen} → \textit{abdominal}
- **\textbf{Adv₀}**: adverbial conversion of L; \textit{lent} → \textit{lentement}
- **Sᵢ**: nouns meaning ‘i\textsuperscript{th} actant of L’;
  \textit{conduire} → \textit{conducteur} \[= S₁\]
- **Aᵢ**: adjectives meaning ‘that is the i\textsuperscript{th} actant of L’;
  \textit{respecter} → \textit{respecteux} \[= A₁\]
- **\textbf{Ableᵢ}**: adjectives meaning ‘that has the ability to be the i\textsuperscript{th} actant of L’;
  \textit{admirer} → \textit{admirable} \[= \textbf{Able}_2\]
- **strict Mult**: collective nouns that do include in their definition the meaning of L.
  \textit{feuille} → \textit{feuillage}
Some FLs used in the second induction step

- **$\text{Syn}_\subset$:** less specific synonyms of $L$;
  
  \[\text{VÉNÉRATION} \rightarrow \text{RESPECT}\]

- **$\text{Syn}_\supset$:** more specific synonyms of $L$;
  
  \[\text{RESPECT} \rightarrow \text{VÉNÉRATION}\]

- **$\text{Contr}$**: contrastive;
  
  \[\text{CHAT} \rightarrow \text{CHIEN}\]

- **$\text{Gener}$**: generic term;
  
  \[\text{ARMOIRE} \rightarrow \text{MEUBLE}\]

- **$S_{\text{instr}}, S_{\text{loc}}, S_{\text{res}}, S_{\text{med}}$**;
  
  \[\begin{align*}
  \text{PEINDRE} & \rightarrow \text{PINCEAU}, \\
  \text{LUTTER} & \rightarrow \text{ARÈNE}, \\
  \text{COPIER} & \rightarrow \text{COPIE}, \\
  \text{PEINDRE} & \rightarrow \text{PEINTURE}
  \end{align*}\]
Some FLs used in the second induction step

- **Magn** and **IncepPredPlus**;
  AMOUR → FOUG
  NOMBRE → AUGMENTER

- **AntiMagn** and **IncepPredMinus**;
  MAJORITÉ → COURTE,
  VENT → MOLLIR

- **Bon** and **AntiBon**;
  CHOIX → BON
  CHOIX → MAUVAIS

- **Ver** and **AntiVer**;
  COMPLIMENT → MÉRITÉ
  COMPLIMENT → IMMÉRITÉ

- **Loc**; PERSONNEL → AU SEIN DE

- **Oper**, **Real**, **Func**, **Fact**, **Labor**, **Labreal**;

- **Son**. CHIEN → ABOYER
Some statistics about FLN

Vocables, i.e. entries \( [= V] \): 11051
Lexical units, i.e. senses \( [= LU] \): 14876
Polysemy rate \( [= LU/V] \): 1.346
Lexical function links \( LU_1 \rightarrow LU_2 \) \( [= LFL] \): 19643
Connectivity rate \( [= LFL/LU] \): 1.32
Growth of the FLN

[Graph showing the growth of the FLN from 01/04/2012 to 15/01/2013 with data points for number of vocables (V), number of lexical units (LU), and polysemy rate (LU/V).]
Growth of the FLN
Dedicated lexicographic editor Dicet

- is a customization of the *Dixit* publishing tool
- provides support for user rights management and includes workflow functionalities
- shares features with tools such as TLex [de Schryver and de Pauw 2007]
- has knowledge about information that has to be provided in each zone
- supervises the entering of information via special lexicographic tools
- supports *building* rather than *writing* articles
- uses formal information provided to compute a textual presentation (*article-view*) of the headword’s description
Correspondance between lexical function lexicographic tools and the article-view

Syn. menace, extorsion

V₀ faire chanter [N- $2]

$1, qui fait $-
S₁ auteur [de ART ~], //maître-chanteur, //maître-chanteuse

$2, qui subit $-
S₂ cible [de ART ~], victime [de ART ~]

Ce par quoi $1 fait $-
Smed moyen [de ~]

[ $1] faire $-
Oper₁ faire [ART ~], exercer [ART ~], se livrer [à ART ~], recourir [à ART ~], user [de ~], pratiquer [ART ~]

[ $2] subir $-
Oper₂ subir [ART ~]

[ $2] réagir comme $1 le souhaite vis-à-vis de $-
Real₁ céder [à ART ~], obérer [à ART ~], obtenir [à ART ~]

[ $2] réagir en opposition à ce que $1 souhaite vis-à-vis de $-
AntiReal₂ refuser [ART ~]
1. Micro- and macrostructure of the FLN

2. Methodology and tools for developing the FLN

3. Conclusion
Perspectives

1. semi-automate the building of FLN
2. further develop other Lexical Networks
3. demonstrate the usefulness of RLF in applications


Modèle lexicographique de croissance du vocabulaire fondé sur un processus aléatoire, mais systématique.
