

Introducing the French Lexical Network

Veronika Lux-Pogodalla



Toulouse, January 2012

Introduction

FLN people

- scientific director: **Alain Polguère**
- lexicographers: **D. Beuseroy, J.-L. Benoit, C. Delaite, A. Ferté, X. Gouvert, S. Haton, É. Jactel, V. Lux-Pogodalla, C. Ménétrier, S. Pescarini, D. Sikora**

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 Dicouebe 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')

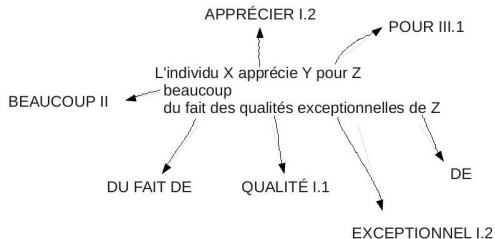
Microstructure of nodes in the FLN

Six lexicographic zones:

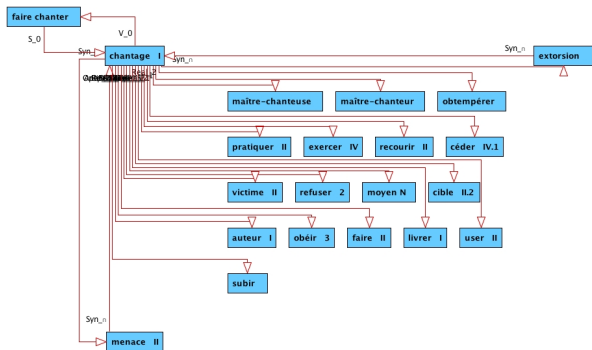
- 1 GC: grammatical characteristics (part of speech, noun gender, specific inflectional behavior, etc.)
- 2 DF: definition
- 3 GP: headword's government pattern, i.e. the description of its syntactic valency [Mel'čuk2004a, Mel'čuk2004b, Milićević2009]
- 4 LF: lexical function relations originating from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

Links in the FLN

- 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 I 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

MvsDicet

Fonctions lexicales - RLF (ATILF) - velupo - velupo (ln_fr/mysq/relief.atilf.fr:3306)

Identifiant : 236 Formule : **AntiReal₁**

Statut : 0 : Validée Type de lien : syntagmatic

Type de standard : standard Famille : Real₁

Structure

Composante(s) standard(s) A :

| Nom | Indice | Exposant |
|------|--------|----------|
| Real | 1 | |
| Anti | | |

Opérateur de combinaison :

Composante(s) standard(s) B :

| Nom | Indice | Exposant |
|-----|--------|----------|
|-----|--------|----------|

Composante non standard A :

Composante non standard B :

Glose(s) de vulgarisation associée(s) à la fonction lexicale

| Id association | Id glose | Formule | Statut |
|----------------|----------|--|-------------|
| 1576 | 1391 | [Individu \$1] relever ~ | 0 : Validée |
| 1577 | 1392 | [\$1] arrêter le fonctionnement de ~ | 0 : Validée |
| 1578 | 1393 | [\$1] échouer dans ~ | 0 : Validée |
| 1579 | 1394 | [\$1] lutter contre ~ | 0 : Validée |
| 1580 | 1395 | [\$1] ne pas se laisser influencer par ~ | 0 : Validée |

Notes

Dans le DiCo, a été encodé comme {NonPerm1Fact0} dans les articles de ANGOISSE, AVERSION, DÉCEPTION, DÉGOÛT, DÉPIT, EFFROI, HONTE#1, RÉPUGNANCE#1, RÉPUGNANCE#2, RÉPULSION#1. C'est une erreur.

Création par apol le 2012-02-26 00:00:00 Modification par eljac le 2012-12-11 15:30:45

Fonctions lexicales - RLF (ATILF) - velupo - velupo (ln_fr/mysq/relief.atilf.fr:3306)

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

| Support verbs | Verbs of realisation |
|---|---|
| <p>Oper_i</p> <pre> graph TD Oper_i -- I --> A_i Oper_i -- II --> L </pre> | <p>Real_i</p> <pre> graph TD Real_i -- I --> A_i Real_i -- II --> L </pre> |
| <p>Func_i</p> <pre> graph TD Func_i -- I --> L Func_i -- II --> A_i </pre> | <p>Fact_i</p> <pre> graph TD Fact_i -- I --> L Fact_i -- II --> A_i </pre> |
| <p>Labor_{ij}</p> <pre> graph TD Labor_ij -- I --> A_i Labor_ij -- II --> A_j Labor_ij -- III --> L </pre> | <p>Labreal_{ij}</p> <pre> graph TD Labreal_ij -- I --> A_i Labreal_ij -- II --> A_j Labreal_ij -- III --> L </pre> |

| Support verbs | Verbs of realisation |
|---|--|
| $\text{Oper}_1(\text{PROMESSE}) = \text{FAIRE}[\text{ART} \sim]$ | $\text{Real}_1(\text{PROMESSE}) = \text{TENIR}[\text{ART} \sim]$ |
| $\text{Func}_1(\text{AIDE}) = \text{VENIR} [\text{de N}]$ | $\text{Fact}_2(\text{MÉMOIRE}) = \text{GARDER}[\text{N}]$ |
| $\text{Labor}_{12}(\text{INTERROGATOIRE}) = \text{SOUMETTRE} [\text{N à ART} \sim]$ | $\text{Labreal}_{12}(\text{PIÈGE}) = \text{PRENDRE} [\text{N à ART} \sim]$ |

Operators to combine Lexical Functions

- concatenation: $F_1 F_2$
PreparReal₁
 VÉLO → ENFOURCHER
S₁CausOper₁
 CRIME → MOBILE
- & operator: $F_1 \& F_2$
S_{Loc} & S₂
 HABITER → MAISON
- v operator: $F_1 \vee F_2$
Able₁ ∨ A₁
 JOUER → JOUEUR_{ADJ}
- + operator: = $F_1 + F_2$
AntiBon + AntiMagn_{taille}
 CHIEN → ROQUET

- 1 Micro- and macrostructure of the FLN
- 2 Methodology and tools for developing the FLN
 - Lexical coverage
 - Using Lexical Functions to expand the lexical coverage
 - Lexicographic editor
- 3 Conclusion

Lexical coverage [Polguère and Sikora|TRE]

- 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

Expanding 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; MIAULEMENT_N → MIAULER_V
- **S₀**: nominal conversion of L; SOUPIRER → SOUPIR
- **Adj₀**: adjectival conversion of L; ABDOMEN → ABDOMINAL
- **Adv₀**: adverbial conversion of L; LENT_{Adj} → LENTEMENT
- **S_i**: nouns meaning 'ith actant of L';
CONDUIRE → CONDUCTEUR [= S₁]
- **A_i**: adjectives meaning 'that is the ith actant of L';
RESPECTER → RESPECTEUR [= A₁]
- **Able_i**: adjectives meaning 'that has the ability to be the ith actant of L';
ADMIRER → ADMIRABLE [= Able₂]
- strict **Mult**: collective nouns that do include in their definition the meaning of L.
FEUILLE → FEUILLAGE

Some FLs used in the second induction step

- **Syn_C**: less specific synonyms of L;
VÉNÉRATION → RESPECT
- **Syn_D**: more specific synonyms of L;
RESPECT → VÉNÉRATION
- **Contr**: contrastive;
CHAT → CHIEN
- **Gener**: generic term;
ARMOIRE → MEUBLE
- **S_{instr}, S_{loc}, S_{res}, S_{med}**;
PEINDRE → PINCEAU,
LUTTER → ARÈNE,
COPIER → COPIE,
PEINDRE → PEINTURE

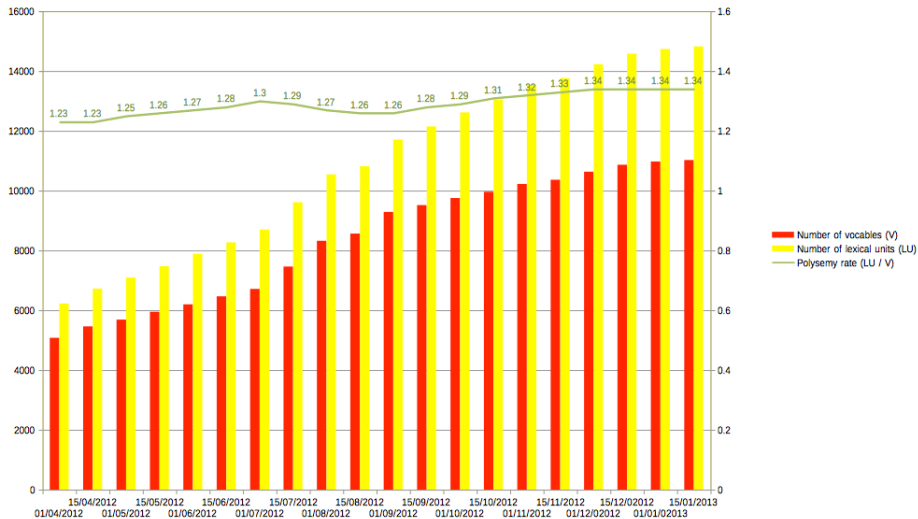
Some FLs used in the second induction step

- **Magn** and **IncepPredPlus**;
AMOUR → FOU
NOMBRE → AUGMENTER
- **AntiMagn** and **IncepPredMinus**;
MAJORITÉ → COURTE,
VENT → MOLLIR
- **Bon** and **AntiBon**;
CHOIX → BON
CHOIX → MAUVAIS
- **Ver** and **AntiVer**;
COMPLIMENT → MÉRITE
COMPLIMENT → IMMÉRITE
- **Loc_{in}**; PERSONNEL → AU SEIN DE
- **Oper_i**, **Real_i**, **Func_i**, **Fact_i**, **Labor_{ij}**, **Labreal_{ij}**;
- **Son.** CHIEN → ABOYER

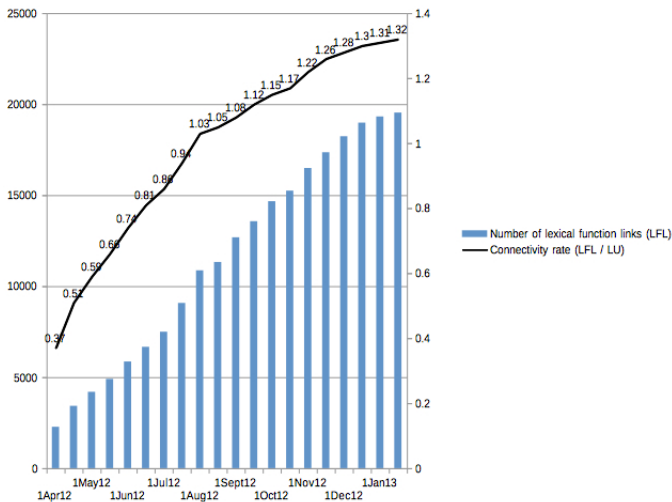
Some statistics about FLN

| | |
|---|---------|
| 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



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 Pauw2007]
- 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

chantage

- chantage I
- chantage II

Vocable (26711)

Name: chantage

Prefix: Subscript:

Status: 1 : In validation Activity: 1 : Assigned

Comments: @delinquance

Creation by mvs on June 16, 2011 5:08:49 PM
Modification by nomen on October 11, 2012 1

Sense (26709)

Number: 1

Status: 2 : In treatment

Syn₁ menace II; extorsion

V₀ faire chanter [N=S2]

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

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

Ce par quoi S₁ fait ~
S_{med} moyen_n [de -]

{S1} faire ~
Oper₁ faire II.1 [ART -], exercer IV [ART -], se livrer I [à ART -], recourir II [à ART -], user II [de -], pratiquer II [ART -]

{S2} subir ~
Oper₂ subir [ART -]

{S2} réagir comme S₁ le souhaite vis-à-vis de ~
Real₂ céder IV.1 [à ART -], obéir 3 [à ART -], obtempérer [à ART -]

{S2} réagir en opposition à ce que S₁ souhaite vis-à-vis de ~
AntiReal₂ refuser 2 [ART -]

[GC] [DF] [NB] [GP] [LF] [EX] [PH]

- 2-II
- 3-III
- Syn
- Conv-ij
- Anti
- Non
- Contr
- Epit
- Gener
- Figur





- Syn₁
 - Syn_n
- V₀
 - V₀
- S₁
 - S₁
- S₂
 - S₂
- S_{med}
 - S_{med}






Comments

- 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

-  de Schryver, G.-M. and de Pauw, G. (2007).
Dictionary Writing System (DWS) + Corpus Query Package (CQP):
The Case of TshwaneLex.
Lexikos, 17:226–246.
-  Mel'čuk, I. (1996).
Lexical Functions: A Tool for the Description of Lexical Relations in
the Lexicon.
In Wanner, L., editor, *Lexical Functions in Lexicography and Natural
Language Processing*, volume 31 of *Language Companion Series*,
pages 37–102. Benjamins, Amsterdam/Philadelphia.
-  Mel'čuk, I. (2004a).
Actants in semantics and syntax I: actants in semantics.
Linguistics, 42(1):1–66.
-  Mel'čuk, I. (2004b).
Actants in semantics and syntax II: actants in syntax.
Linguistics, 42(2):247–291.

-  Mel'čuk, I. (2006).
Explanatory Combinatorial Dictionary.
In Sica, G., editor, *Open Problems in Linguistics and Lexicography*,
pages 225–355. Polimetrica, Monza.
-  Mel'čuk, I., Clas, A., and Polguère, A. (1995).
Introduction à la lexicologie explicative et combinatoire.
Duculot, Paris/Louvain-la-Neuve.
-  Milićević, J. (2009).
Schéma de régime : le pont entre le lexique et la grammaire.
Langages, 176:94–116.
-  Polguère, A. (2012).
Like a Lexicographer Weaving Her Lexical Network.
In *Proceedings of the 3rd workshop on Cognitive Aspects of the
Lexicon (CogALex-III) - COLING 2012*, pages 1–4, Mumbai.
-  Polguère, A. and Sikora, D. (À PARAÎTRE).

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

In Masseron, C., Garcia-Debanc, C., and Ronveaux, C., editors, *Enseigner le lexique. Pratiques sociales, objets à enseigner et pratiques d'enseignement*, Collection de l'AiRDF 5. Presses de l'Université Laval, Laval (Québec).



Selva, T., Verlinde, S., and Binon, J. (2003).

Vers une deuxième génération de dictionnaires électroniques.

Traitement Automatique des Langues (TAL), 44(2):177–197.