

1. Les tables du Lexique-Grammaire
  2. Le lexique *LGLex*
  3. Conversion de *LGLex* en un lexique au format *Alexina*
  4. Intégration dans l'analyseur syntaxique *FRMG*
  5. Évaluation et discussion
- Conclusions et perspectives

# Analyse syntaxique à l'aide des tables du Lexique-Grammaire du français

Elsa Tolone

Soutenance de thèse, LIGM, Université Paris-Est, France  
31 mars 2011

1. Les tables du Lexique-Grammaire
  2. Le lexique *LGLex*
  3. Conversion de *LGLex* en un lexique au format *Alexina*
  4. Intégration dans l'analyseur syntaxique *FRMG*
  5. Évaluation et discussion
- Conclusions et perspectives

## Analyse syntaxique

- ▶ **Construire la structure grammaticale d'une phrase**  
pour lui donner du sens  
→ expliciter les relations de dépendance entre les mots  
(entre sujet et verbe, verbe et objet, etc.)  
Difficulté : **Complexité et grandeur de la langue**
- ▶ Applications :
  - ▶ Compréhension de texte
  - ▶ Extraction d'information
  - ▶ Traduction

## Analyseurs syntaxiques

- ▶ Analyseurs :
  - ▶ **symboliques** = grammaire + lexique syntaxique qui spécifie le comportement grammatical de chaque mot de la langue
  - ▶ **probabilistes** = modèle acquis à partir d'un corpus annoté manuellement
- ▶ Analyseur syntaxique symbolique retenu :  
FRMG [Thomasset & de La Clergerie 2005]  
→ formalisme lexical utilisé par FRMG : *Alexina*, format du lexique *Lefff* [Sagot 2010]

1. Les tables du Lexique-Grammaire

2. Le lexique *LGLex*

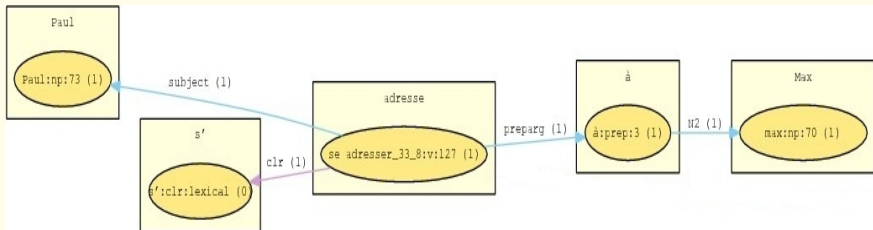
3. Conversion de *LGLex* en un lexique au format *Alexina*

4. Intégration dans l'analyseur syntaxique *FRMG*

5. Évaluation et discussion

Conclusions et perspectives

## Exemple de dépendances avec *FRMG*



*Paul s'adresse à Max*

## Tables du Lexique-Grammaire

- ▶ Les **tables du Lexique-Grammaire** du français = ressource lexicale très riche
  - informations **syntaxiques** et syntaxico-sémantiques
- ▶ Mais **pas directement exploitables** telles quelles dans un analyseur syntaxique
  - ▶ informations importantes non représentées car implicites pour une table donnée
  - ▶ format de représentation des informations lexicales non formalisé
  - ▶ il faut interfacer ces informations avec un analyseur syntaxique symbolique réel

## Objectifs

Triple objectif :

- ▶ **convertir** les tables du Lexique-Grammaire en un format TAL  
→ permettre la diffusion dans des formats adaptés pour les futures utilisations
- ▶ **coupler** le lexique syntaxique obtenu, nommé  $LGLex_{Leff}$ , avec un analyseur syntaxique  
→ avoir un format directement utilisable par un analyseur
- ▶ **évaluer** l'analyseur obtenu en comparant  $FRMG_{Leff}$  et  $FRMG_{LGLex}$   
→ savoir si le Lexique-Grammaire peut être utilisé dans un analyseur syntaxique

1. Les tables du Lexique-Grammaire
  2. Le lexique *LGLex*
  3. Conversion de *LGLex* en un lexique au format Alexina
  4. Intégration dans l'analyseur syntaxique FRMG
  5. Évaluation et discussion
- Conclusions et perspectives

1. Les tables du Lexique-Grammaire

2. Le lexique *LGLex*

3. Conversion de *LGLex* en un lexique au format Alexina

4. Intégration dans l'analyseur syntaxique FRMG

5. Évaluation et discussion

## 1. Les tables du Lexique-Grammaire

### 2. Le lexique *LGLex*

### 3. Conversion de *LGLex* en un lexique au format *Alexina*

### 4. Intégration dans l'analyseur syntaxique *FRMG*

### 5. Évaluation et discussion

Conclusions et perspectives

### 1.1. Tables du Lexique-Grammaire

### 1.2. Tables des classes

### 1.3. Constructions de base

# 1 Les tables du Lexique-Grammaire



## Tables du Lexique-Grammaire

Développées **manuellement** depuis plus de 40 ans par le LADL [Gross 1975], puis l'Équipe d'Informatique Linguistique du LIGM (Université Paris-Est)

- ▶ Étude de la syntaxe d'une phrase élémentaire (ou **cadre de sous-catégorisation**)  
ex : N0 V à N1
- ▶ Catégories concernées : verbes, adverbes, noms et adjectifs prédicatifs et expressions figées pour le français  
→ certaines **propriétés en commun** = **classes**
- ▶ Les différents emplois sont distingués (ex : *cuisiner*)

## Principe

- ▶ Chaque classe est codée dans une **table** :
  - ▶ lignes = entrées lexicales (mots)
  - ▶ colonnes = propriétés syntaxico-sémantiques
- ▶ **Propriétés** :
  - ▶ constructions
  - ▶ autres propriétés (distributionnelles, morphologiques, transformationnelles, sémantiques, etc.)  
ex : N0 = : Nhum → nom humain
- ▶ Chaque propriété **testée pour chaque entrée**  
→ codage binaire :
  - ▶ + : propriété acceptée
  - ▶ - : propriété non acceptée
- ▶ **Propriétés définitoires** pour chaque classe  
dont une **construction de base**

1. Les tables du Lexique-Grammaire

2. Le lexique LGLex

3. Conversion de LGLex en un lexique au format Alexina

4. Intégration dans l'analyseur syntaxique FRMG

5. Évaluation et discussion

Conclusions et perspectives

1.1. Tables du Lexique-Grammaire

1.2. Tables des classes

1.3. Constructions de base

# Extrait de la table des verbes de la classe 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

Construction de base : N0 V à N1

## Bilan

- ▶ Inventaire [Tolone 2009] :
  - ▶ 67 classes de **verbes simples**
    - ▶ 13 867 entrées pour 5 738 lemmes distincts
  - ▶ 78 classes de **noms prédicatifs simples et composés** (noms avec argument(s) qui sont étudiés avec leur verbes support)
    - ▶ 12 696 entrées pour 8 531 lemmes distincts  
ex : *Luc monte une attaque contre le fort*
  - ▶ 69 classes d'**expressions figées, principalement verbales et adjectivales**
    - ▶ 39 628 entrées pour 38 658 lemmes distincts  
ex : *arriver à la cheville de qq'un, être à la retraite*
  - ▶ 32 classes d'**adverbes simples et (semi-)figés** (adverbes en *-ment* et locutions adverbiales)
    - ▶ 10 488 entrées pour 9 326 lemmes distincts  
ex : [*changer*] du jour au lendemain

## Historique des versions

Lexique riche à large couverture disponible sur

<http://infolingu.univ-mlv.fr/>

- ▶ version 1 = format d'origine
- ▶ version 2 = 1ères tables en ligne (60% seulement)
- ▶ version 3 = depuis ma thèse, dans les différents formats de lexique :
  - ▶ tables → lisibilité pour les linguistes
  - ▶ *LGLex* → format d'échange
  - ▶ *LGLex<sub>Leff</sub>* → format d'intégration dans un analyseur

## Problèmes

- ▶ **Noms différents** pour une même propriété  
→ Harmonisation des intitulés de colonnes  
ex : [extrap] et il V N0 W
- ▶ Propriétés **pas définies clairement**  
→ Documentation des propriétés
- ▶ **Propriétés définitoires implicites** (littérature)  
→ Constante + ou – pour l'ensemble de la table
- ▶ Entrées pas à **jour** dans les tables  
→ Ajout des entrées manquantes
- ▶ Entrées **non codées** (~)  
→ Codage d'entrée

## Tables des classes

Propriétés définitoires → à décrire dans des **tables des classes**  
(1 par catégorie) :

- ▶ colonnes = toutes les propriétés syntaxiques répertoriées pour la catégorie concernée (après harmonisation)
- ▶ lignes = classes définies pour cette catégorie
- ▶ intersection ligne/colonne :
  - ▶  $o$  = propriété explicitement codée dans la table concernée
  - ▶  $+$  (resp.  $-$ ) = la propriété correspondante est vérifiée (resp. non vérifiée) par tous les éléments de la classe
  - ▶  $?$  = non encore renseigné

Table des classes créée pour chaque catégorie [Tolone et al. 2010]

# 1. Les tables du Lexique-Grammaire

## 2. Le lexique *LGLex*

### 3. Conversion de *LGLex* en un lexique au format *Alexina*

### 4. Intégration dans l'analyseur syntaxique *FRMG*

### 5. Évaluation et discussion

### Conclusions et perspectives

## 1.1. Tables du Lexique-Grammaire

### 1.2. Tables des classes

### 1.3. Constructions de base

# Extrait de la tables des classes des verbes

table	N0 =; Nhum	N0 =; N-hum	N0 =; Nnr	N0 =; V1-inf W	<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	+	-	-	0	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	-	+	-	-	+	-	-	-	-



## 1. Les tables du Lexique-Grammaire

### 2. Le lexique *LGLex*

### 3. Conversion de *LGLex* en un lexique au format *Alexina*

### 4. Intégration dans l'analyseur syntaxique *FRMG*

### 5. Évaluation et discussion

### Conclusions et perspectives

### 1.1. Tables du Lexique-Grammaire

### 1.2. Tables des classes

### 1.3. Constructions de base

# Verbes simples : les complétives et les infinitives

## ► Modifications de la construction de base [Tolone 2010]

Tables	Construction de base d'origine	Construction de base actuelle	Colonnes codées dans la table
6, 7, 8, 9, 10, 14, 15	N0 V Qu P	N0 V N1	N1 = : Qu Pind N1 = : Qu Psubj N1 = : si P ou si P N1 = : ce(ci+la) N1 = : Nhum N1 = : N-hum N1 = : le fait Qu P

► *Max croit que Luc est malade*

► *Max craint qu'Ida ne parte*

[Gross 1975]

## 1. Les tables du Lexique-Grammaire

### 2. Le lexique *LGLex*

### 3. Conversion de *LGLex* en un lexique au format *Alexina*

### 4. Intégration dans l'analyseur syntaxique *FRMG*

### 5. Évaluation et discussion

### Conclusions et perspectives

### 1.1. Tables du Lexique-Grammaire

### 1.2. Tables des classes

### 1.3. Constructions de base

# Verbes de BGL (2)

## ▶ Éclatement en plusieurs classes [Tolone et al. 2010]

Table	Construction de base d'origine	Construction de base actuelle
35L	N0 V Loc N1	N0 V Loc N1 source Loc N2 destination

▶ *Paul a bondi du tabouret sur la table*

▶ *Paul a bondi du tabouret*

▶ *Paul a bondi sur la table*

Table	Construction de base actuelle
35LS	N0 V Loc N1 source

▶ *Le train a déraillé de la voie*

Table	Construction de base actuelle
35LD	N0 V Loc N1 destination

▶ *Le bateau s'enfonce dans les flots*

[Boons et al. 1976a]

## Expressions figées (M. Gross)

- ▶ Reconstitution des constructions de base à partir des colonnes lexicales après avoir été harmonisées [Tolone et al. 2010]

Table	Construction de base
EPCPN	N0 être Prép1 Det1 C1 Prép2 N2

<ENT>être	<ENT>Prép1	<ENT>Det1	<ENT>C1	<ENT>Prép2
<i>être</i>	<i>en</i>	<E>	<i>accord</i>	<i>avec</i>
<i>être</i>	<i>de</i>	<i>la même</i>	<i>force</i>	<i>que</i>
<i>être</i>	<i>à</i>	<i>la</i>	<i>frontière</i>	<i>entre</i>

**Table:** Extrait de la table EPCPN des expressions figées (colonnes lexicales)

## Adverbes en *-ment*

- Clarification des constructions de base [Tolone et al. 2010]

Table	Construction de base	Exemple
ADVM...	N0 V Adv W	<i>Ce livre est en vente <b>exclusivement</b> sur ce site</i> <i>*<b>Exclusivement</b>, ce livre est en vente sur ce site</i>
ADVMP ADVMS ADVMTF	Adv, N0 V W	<i>Ce livre est en vente <b>régulièrement</b> sur ce site</i> <i><b>Régulièrement</b>, ce livre est en vente sur ce site</i> <i>*<b>Régulièrement</b>, ce livre n'est pas en vente sur ce site</i>
ADVMP	Adv, N0 ne V pas W	<i>Ce concert est <b>musicalement</b> une réussite</i> <i><b>Musicalement</b>, ce concert est une réussite</i> <i><b>Musicalement</b>, ce concert n'est pas une réussite</i>

[Molinier & Levrier 2000]

## Adverbes (semi-)figés

- ▶ Reconstitution de la structure de l'adverbe à partir des colonnes lexicales après avoir été harmonisées [Tolone et al. 2010]

Table	Structure de l'adverbe	Exemple
PCPC	Prép1 Det1 C1 Prép2 Det2 C2	[changer] du jour au lendemain

- ▶ Définition des constructions de base (approximation)

Table	Construction de base
P...	N0 V Adv W Adv, N0 V W Adv, N0 ne V pas W

[Gross 1986]

1. Les tables du Lexique-Grammaire

2. Le lexique *LGLex*

3. Conversion de *LGLex* en un lexique au format Alexina

4. Intégration dans l'analyseur syntaxique FRMG

5. Évaluation et discussion

Conclusions et perspectives

## 2. Le lexique *LGLex*

## *LGLex*

Table des classes → possibilité d'extraire un **lexique syntaxique** pour chaque catégorie à partir des tables [Constant & Tolone 2010] :

- ▶ nommé lexique *LGLex*
- ▶ conversion depuis les tables Excel ou CSV grâce à l'outil *LGExtract*
- ▶ format d'échange, reposant sur les mêmes concepts linguistiques que ceux qui sont manipulés dans les tables
- ▶ format textuel ou XML
- ▶ version 3.3 contenant *LGLex* pour chaque catégorie disponible sur <http://infolingu.univ-mlv.fr/>

## Format du lexique *LGLex*

- ▶ **ID=catégorie\_numTable\_numEntrée**
- ▶ **lexical-info**=[...] → lemme et ses informations lexicales (auxiliaires, verbes supports, déterminants, prépositions)
- ▶ **args**=(...) → arguments et leurs distributions accompagnées d'autres informations (traits sémantiques, mode de la complétive, argument contrôlé par l'infinitive, prépositions)
- ▶ **all-constructions**=[absolute=(...), relative=(...)] → liste des constructions acceptées (soit nommées de façon complète avec tous les éléments dans l'ordre, soit transformations à partir de la construction de base, mais aussi réductions de GN, structures des adverbes figés, etc.)
- ▶ **example**=[...] → exemple illustrant l'entrée



## Extrait du lexique *LGLex* (verbe)

ID=V\_36SL\_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")],  
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"]

- ▶ entrée *clouer*V\_36SL\_28
- ▶ Reste à faire : **interpréter** les constructions

1. Les tables du Lexique-Grammaire

2. Le lexique *LGLex*

3. Conversion de *LGLex* en un lexique au format Alexina

4. Intégration dans l'analyseur syntaxique FRMG

5. Évaluation et discussion

Conclusions et perspectives

3.1. L'analyseur syntaxique FRMG

3.2. Le *Lefff* et Alexina

3.3. Conversion de *LGLex*

3.5. Lexique obtenu : *LGLex<sub>Lefff</sub>*

## 3. Conversion de *LGLex* en un lexique au format Alexina

1. Les tables du Lexique-Grammaire

2. Le lexique *LGLex*

3. Conversion de *LGLex* en un lexique au format Alexina

4. Intégration dans l'analyseur syntaxique FRMG

5. Évaluation et discussion

Conclusions et perspectives

3.1. L'analyseur syntaxique FRMG

3.2. Le *Lefff* et *Alexina*

3.3. Conversion de *LGLex*

3.5. Lexique obtenu : *LGLexLefff*

## L'analyseur syntaxique FRMG

FRMG est un analyseur **TAG** du français [Thomasset & de La Clergerie 2005] :

- ▶ issu de la compilation d'une **méta-grammaire**
- ▶ très **compacte** grâce à la factorisation des arbres
- ▶ exploitant les fonctionnalités de **DyALog**  
(environnement de programmation en logique)

FRMG s'intègre dans une chaîne de traitement :

- ▶ en amont
  - ▶ **SXPipe** : segmentation, token, corrections, entités nommées
  - ▶ **Lefff** : lexique morphosyntaxique et syntaxique du français
- ▶ en aval, module de **désambiguïsation** (basé sur des heuristiques)

## Le *Lefff*

- ▶ Le *Lefff* (Lexique des Formes Fléchies du Français) est un lexique morphologique et syntaxique pour le français [Sagot 2010]
  - ▶ à large couverture (536 375 entrées correspondant à 110 477 lemmes distincts couvrant toutes les catégories)
  - ▶ librement distribué (licence LGPL-LR)
- ▶ Il repose sur l'architecture Alexina de modélisation et d'acquisition de lexiques morphologiques et syntaxiques → validé ensuite très largement manuellement, notamment pour les verbes

1. Les tables du Lexique-Grammaire

2. Le lexique *LGLex*

3. Conversion de *LGLex* en un lexique au format Alexina

4. Intégration dans l'analyseur syntaxique FRMG

5. Évaluation et discussion

Conclusions et perspectives

3.1. L'analyseur syntaxique FRMG

3.2. Le *Lefff* et Alexina

3.3. Conversion de *LGLex*

3.5. Lexique obtenu : *LGLex<sub>Lefff</sub>*

# Alexina

## Architecture à deux niveaux

- ▶ Le lexique **intensionnel**
  - ▶ associe à chaque entrée (emploi d'un lemme) un cadre de sous-catégorisation canonique
  - ▶ liste les redistributions possibles à partir de ce cadre
- ▶ Le processus de **compilation** du lexique intensionnel en lexique **extensionnel** construit différentes entrées pour chaque forme fléchie du lemme et chaque redistribution possible

## Alexina sur un exemple

- ▶ Exemple d'entrée intentionnelle :

*clarifier* 1 *v-er:std*

*100;Lemma;v;*

<**Suj:***cln|scompl|sinf|sn*,**Obj:***(cl|scompl|sn)*>;

*%ppp\_employé\_comme\_adj,%actif,%se\_moyen\_impersonnel,*

*%passif\_impersonnel,%passif*

- ▶ **Fonctions syntaxiques** (cf. Dicovalence) : Suj, Obj, Objà, Objde, Loc, Dloc, Att, Obl/Obl2
- ▶ **Réalisations** : directes (*sn, sa, sinf, scompl, qcompl*) ; clitiques (*cln, cl, cld, y, en*) ; prépositionnelles (*prep+directe*, p.ex. *par-sn, à-sinf, de-scompl*)
- ▶ **Redistributions** : *%actif, %passif, %se\_neutre, %actif\_impersonnel*, etc.

1. Les tables du Lexique-Grammaire

2. Le lexique *LGLex*

3. Conversion de *LGLex* en un lexique au format Alexina

4. Intégration dans l'analyseur syntaxique FRMG

5. Évaluation et discussion

Conclusions et perspectives

3.1. L'analyseur syntaxique FRMG

3.2. Le *Lefff* et Alexina

3.3. Conversion de *LGLex*

3.5. Lexique obtenu : *LGLex<sub>Lefff</sub>*

## Le processus de conversion

- ▶ Conversion des tables du Lexique-Grammaire vers le formalisme Alexina **pas directe** [Tolone & Sagot 2011]
  - ▶ **definition formelle** ou **interprétation dynamique** de tous les noms de propriétés  
Exemple :
    - ▶  $N0 V N1 \rightarrow$  construction de base
    - ▶  $N0 V \rightarrow$  effacement du N1
    - ▶  $N0 V N1 \text{ à } N2 \rightarrow$  allongement de la construction de base
    - ▶  $N0 V Qu P \rightarrow$  réalisation du N1
    - ▶ [passif par]  $\rightarrow$  redistribution passive
  - ▶ informations additionnelles importantes complétées grâce à des heuristiques ou à partir d'autres ressources lexicales
    - ▶ le nom de chaque fonction syntaxique, les phénomènes d'attribution, les informations morphologiques, etc.

1. Les tables du Lexique Grammaire

2. Le lexique *LGLex*

3. Conversion de *LGLex* en un lexique au format Alexina

4. Intégration dans l'analyseur syntaxique FRMG

5. Évaluation et discussion

Conclusions et perspectives

3.1. L'analyseur syntaxique FRMG

3.2. Le *Leff* et Alexina

3.3. Conversion de *LGLex*

3.5. Lexique obtenu : *LGLex<sub>Leff</sub>*

## L'exemple précédent du lexique *LGLex* (verbe)

ID=V\_36SL\_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"]
```

► entrée *clouer*V\_36SL\_28



## Résultat de la conversion sur l'exemple précédent

*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 a cloué ces planches*
- ▶ *Cette planche a été clouée au mur par Max*
- ▶ *Cette planche est clouée au mur par Max*
- ▶ etc.

## Autre extrait du lexique *LGLex* (nom)

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)
- ▶ (Ce geste+Luc) a la considération de Max

[G. Gross 1989]

1. Les tables du Lexique-Grammaire

2. Le lexique *LGLex*

3. Conversion de *LGLex* en un lexique au format Alexina

4. Intégration dans l'analyseur syntaxique FRMG

5. Évaluation et discussion

Conclusions et perspectives

3.1. L'analyseur syntaxique FRMG

3.2. Le *Lefff* et *Alexina*

3.3. Conversion de *LGLex*

3.5. Lexique obtenu : *LGLex<sub>Lefff</sub>*

## Résultat de la 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
```

Version 3.3 contenant les verbes et les noms disponible sur  
<http://infolingu.univ-mlv.fr/>

► Reste à convertir : **expressions figées** et **adverbes**

1. Les tables du Lexique-Grammaire

2. Le lexique  $LGLex$

3. Conversion de  $LGLex$  en un lexique au format Alexina

4. Intégration dans l'analyseur syntaxique FRMG

5. Évaluation et discussion

Conclusions et perspectives

3.1. L'analyseur syntaxique FRMG

3.2. Le *Lefff* et Alexina

3.3. Conversion de  $LGLex$

3.5. Lexique obtenu :  $LGLex_{Lefff}$

## Lexique obtenu : $LGLex_{Lefff}$

- ▶ **verbes** :  $LGLex_{Lefff}$  contient 16 955 entrées pour 5 723 lemmes verbaux différents (2,96 entrées/lemme)
  - ▶ À titre de comparaison, le *Lefff* contient seulement 7 072 entrées verbales pour 6 818 lemmes verbaux distincts (1,04 entrées/lemme)
- ▶ **noms prédicatifs** :  $LGLex_{Lefff}$  contient 17 517 entrées pour 8 363 lemmes différents (2,09 entrées/lemme)
  - ▶ Le *Lefff* ne contient que 218 entrées de noms prédicatifs (1 entrée/lemme)

1. Les tables du Lexique-Grammaire

2. Le lexique *LGLex*

3. Conversion de *LGLex* en un lexique au format *Alexina*

**4. Intégration dans l'analyseur syntaxique FRMG**

5. Évaluation et discussion

Conclusions et perspectives

## 4. Intégration dans l'analyseur syntaxique FRMG

## Intégration dans l'analyseur syntaxique FRMG

Il faut donc :

- ▶ remplacer les entrées verbales du *Lefff* par *LGLex<sub>Lefff</sub>*
- ▶ ajouter les entrées nominales de *LGLex<sub>Lefff</sub>*
- ▶ conserver les autres entrées du *Lefff*
- ▶ compléter le résultat par diverses entrées verbales venant du *Lefff*, qui ne font pas partie du lexique *LGLex*
  - ▶ entrées pour les auxiliaires et semi-auxiliaires
  - ▶ certains verbes à montée
  - ▶ les verbes impersonnels
- ▶ construire la base de données lexicales correspondantes
- ▶ spécifier à FRMG d'utiliser cette dernière

1. Les tables du Lexique-Grammaire

2. Le lexique *LGLex*

3. Conversion de *LGLex* en un lexique au format Alexina

4. Intégration dans l'analyseur syntaxique FRMG

**5. Évaluation et discussion**

Conclusions et perspectives

5.1. Protocole

5.2. Résultats

5.3. Discussion

## 5. Évaluation et discussion

## Protocole utilisé

- ▶ Nous avons évalué  $FRMG_{Leff}$  et  $FRMG_{LGLex}$  en analysant la partie annotée manuellement de la 1ère campagne Passage de 2007 [Hamon *et al.* 2008]
  - ▶ 4 306 phrases du corpus EASy annoté en 2004 + 500 nouvelles phrases : styles variés (journalistique, médical, oral, questions, littéraire, etc.)
- ▶ métriques utilisées : celles définies à l'occasion de la campagne EASy d'évaluation des analyseurs syntaxiques [Paroubek *et al.* 2006]
  - ▶ évaluation en *chunks* et en *relations* ( $\sim$  dépendances entre mots pleins)



## Précautions

- ▶ FRMG utilise les entrées converties à partir des tables, et non pas les entrées telles qu'elles sont dans les tables  
→ erreurs dûes à la conversion
- ▶ la majorité des noms prédicatifs n'ont pas pu être évalués car FRMG ne prend que ceux n'ayant pas de déterminant
- ▶ Passage ne permet pas d'évaluer toutes les informations contenues dans les tables (informations sémantiques par ex.)
- ▶ le *Lefff* a été développé en parallèle aux campagnes EASy et Passage, contrairement aux tables
- ▶ *LGLex<sub>Lefff</sub>* ne contient pas toutes les entrées verbales nécessaires, puisque tous les verbes ne sont pas codés

## Résultats

Résultats comparatifs Passage de  $FRMG_{Leff}$  et  $FRMG_{LGLex}$   
(pourcentages de f-mesure) :

Sous-corpus	Chunks		Relations	
	$FRMG_{Leff}$	$FRMG_{LGLex}$	$FRMG_{Leff}$	$FRMG_{LGLex}$
general_lemonde	<b>88.22%</b>	84.60%	<b>62.73%</b>	59.01%
litteraire_2	<b>88.91%</b>	88.46%	<b>65.28%</b>	62.43%
mail_9	<b>82.60%</b>	81.90%	<b>58.55%</b>	56.00%
medical_3	85.04%	<b>85.89%</b>	64.79%	<b>65.26%</b>
oral_delic_4	78.80%	<b>81.79%</b>	<b>51.67%</b>	51.14%
questions_amaryllis	<b>91.30%</b>	90.73%	<b>66.56%</b>	64.77%
<i>total</i>	<b>87.05%</b>	<i>85.53%</i>	<b>63.10%</b>	<i>60.25%</i>

Temps d'analyse plus élevés avec  $FRMG_{LGLex}$  qu'avec  $FRMG_{Leff}$  :  
temps médian par phrase de 0,62 s contre 0,26 s

- ceci provient du nombre d'entrées par lemme 3 plus élevé dans *LGLex* que dans le *Leff*

- ▶ FRMG<sub>LGLex</sub> donne de meilleurs résultats que FRMG<sub>Leff</sub> pour certaines relations
  - ▶ relations « classiques » MOD\_A (modifieur-adjectif) et MOD\_R (modifieur-adverbe)
  - ▶ relations « difficiles » MOD\_P (modifieur-préposition) et APP (apposition)
- ▶ la relation ATB\_SO (attribut du sujet ou de l'objet) est celle pour laquelle la différence en rappel est la plus importante (34,0% contre 58,4%) ;
  - ▶ les tables du Lexique-Grammaire encodent très peu d'informations sur les phénomènes d'attribution, mais cela peut aussi être dû aux erreurs de la référence Passage

1. Les tables du Lexique-Grammaire

2. Le lexique *LGLex*

3. Conversion de *LGLex* en un lexique au format *Alexina*

4. Intégration dans l'analyseur syntaxique *FRMG*

5. Évaluation et discussion

**Conclusions et perspectives**

# Conclusions et perspectives

## Version 3.3 des tables du Lexique-Grammaire

L'ensemble des tables disponible sous licence LGPL-LR sur <http://infolingu.univ-mlv.fr/>

- ▶ Verbes simples :
  - ▶ 67 tables + la table des classes (551 propriétés)
  - ▶ index de toutes les entrées
  - ▶ documentation des propriétés
  - ▶ formules définitoires de chaque table
  - ▶ arbre classification
- ▶ Noms prédicatifs simples et composés :
  - ▶ 78 tables + la table des classes (496 propriétés)
- ▶ Expressions figées :
  - ▶ 69 tables + la table des classes (276 propriétés)
- ▶ Adverbes simples et (semi-)figées :
  - ▶ 32 tables + la table des classes (159 propriétés)

## Poursuite de la méthode

Optimiser l'utilisation des données lexicales du Lexique-Grammaire pour l'analyse syntaxique

- ▶ coder les **entrées manquantes ou non codées** dans les tables
- ▶ coder la **table des classes** pour chaque catégorie
- ▶ **améliorer/corriger le processus de conversion**
- ▶ appliquer cette technique aux tables du Lexique-Grammaire pour les **autres catégories**
- ▶ généraliser cette méthode pour les **autres langues** pour lesquelles des tables du Lexique-Grammaire à large-couverture sont disponibles (par exemple, le grec)

## Amélioration de FRMG<sub>LGLex</sub> et de *LGLex*

- ▶ Amélioration de FRMG<sub>LGLex</sub> :
  - ▶ **coupler les 2 variantes de l'analyseur** pour garder un maximum d'analyses complètes, bien meilleures que les analyses partielles
  - ▶ **détecter des erreurs**, notamment au moyen de techniques automatiques, telles que la fouille d'erreurs dans les sorties d'analyseurs syntaxiques [Sagot & de La Clergerie 2008])
  - ▶ intégrer les **noms prédicatifs** avec déterminants
- ▶ Amélioration du lexique obtenu :
  - ▶ grâce à un corpus annoté, déduire les **poids** permettant de repérer les emplois rares
  - ▶ **fusionner** avec d'autres ressources lexicales
  - ▶ utilisation du lexique dans d'autres systèmes de TAL

## Références

- ▶ [Boons et al. 1976a] Boons J.-P., Guillet A. et Leclère C. *La structure des phrases simples en français : Classes de constructions transitives*. Rapport interne, LADL, CNRS, Paris 7. 1976.
- ▶ [Constant & Tolone 2010] Matthieu Constant et 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, pages 79-93. Aracne. 2010.
- ▶ [Gross 1989] Gaston Gross. *Les constructions converses du français*. Droz. Genève, Suisse.
- ▶ [Gross 1986] Maurice Gross. *Grammaire transformationnelle du français : Syntaxe de l'adverbe*, volume 3. Paris : ASSTRIL. 1986.
- ▶ [Gross 1975] Maurice Gross. *Méthodes en syntaxe : Régime des constructions complétives*. Hermann. Paris, France. 1975.
- ▶ [Hamon et al. 2008] Hamon O., Mostefa D., Ayache C., Paroubek P., Vilnat A. et La Clergerie E. Passage : from French Parser Evaluation to Large Sized Treebank. Actes de LREC'08. Maroc.



## Références (2)

- ▶ **[Molinier & Levrier 2000]** Molinier C. et Levrier F. *Grammaire des adverbes : description des formes en -ment*. Genève : Droz. 2000.
- ▶ **[Paroubek et al. 2006]** Patrick Paroubek, Isabelle Robba, Anne Vilnat et Christelle Ayache. Data, Annotations and Measures in EASy : the Evaluation Campaign for Parsers of French. Actes de LREC'06. Gênes, Italie. 2006.
- ▶ **[Sagot 2010]** Benoît Sagot. The Lefff, a freely available and large-coverage morphological and syntactic lexicon for French. Actes de LREC'10, 8 pages. La Valette, Malte. 2010.
- ▶ **[Sagot et de La Clergerie 2008]** Benoît Sagot et Éric de La Clergerie. Fouille d'erreurs sur les sorties d'analyseurs syntaxiques. *Traitement Automatique des Langues (T.A.L.)*, vol.49, num.1, pages 41-60. Hermès. Paris, France. 2008.

## Références (3)

- ▶ [Thomasset & de La Clergerie 2005] François Thomasset et Éric de La Clergerie. Comment obtenir plus des méta-grammaires. Actes de TALN'05. Dourdan, France. 2005.
- ▶ [Tolone 2010] Elsa Tolone. Constructions définitives des tables du Lexique-Grammaire : Le cas des complétives et des infinitives. Actes de MajecSTIC'10 (session poster), 4 pages. Bordeaux, France. 2010.
- ▶ [Tolone et al. 2010] Elsa Tolone, Stavroula Voyatzi et Christian Leclère. Constructions définitives des tables du Lexique-Grammaire. Actes de LGC'10, pages 321-331. Belgrade, Serbie. 2010.
- ▶ [Tolone 2009] Elsa Tolone. Les tables du Lexique-Grammaire au format TAL. Actes de MajecSTIC'09, 8 pages. Avignon, France. 2009.
- ▶ [Tolone & Sagot 2011] Elsa Tolone et Benoît Sagot. Using Lexicon-Grammar tables for French verbs in a large-coverage parser. LNAI. Springer Verlag. 2011. À paraître.

## Liens

- ▶ Manuscrit :  
<http://www-igm.univ-mlv.fr/~tolone/phd.pdf>
- ▶ Ressources : <http://infolingu.univ-mlv.fr/>
  - > Ressources Linguistiques
  - > Lexique-Grammaire
  - > Téléchargement