

# Presentation of LIDILEM

1

LIDILEM  
Université Stendhal

*29 october 2014*

# Background

2

- **University Laboratory (Université Stendhal Grenoble 3) created by Louis Dabène in 1987**
- **Dedicated to linguistics and language teaching and learning (first and second language)**
- **<http://lidilem.u-grenoble3.fr>**

# Staff

3

- **Head : Marinette Matthey (sociolinguistics)**
- **About 50 professors and associate professors**
- **About 70 PhD students**

# Topics

4

- **3 teams/topics**
  - **Axe 1: Linguistic description, NLP, Corpora**
    - ✦ **2 programs :**
      1. **Syntax, semantics, pragmatics**
      2. **NLP**
  - **Axe 2: Sociolinguistics and language acquisition**
  - **Axe 3: Language teaching and learning, educational technology**

# Work related to phraseology

5

- Interest in phraseology for more than 10 years
- Several projects related to this topic:
  - ✦ **Emergence** (2003-2005) – Lexicon of scientific writing (O. Kraif, A. Tutin)
  - ✦ **Scientext** (ANR, 2007-2010) – Positioning in scientific writing (F. Grossmann, A. Tutin)
    - ✦ Large scientific corpora in French and English (treebanks)
  - ✦ **EMOLEX** (ANR Franco-allemand) – Lexicon of emotion in 5 european languages (I. Novakova, P. Blumenthal)
    - ✦ Large corpora (novels and newspapers in French, English, German, Spanish and Russian)
    - ✦ Some multilingual corpora (Olivier Kraif and Sascha Diwersy)
  - ✦ **Termith** (ANR Contint, Head ATILF, 2012-2015) - TERMinologie et Indexation de Textes en sciences Humaines. Work on scientific Phraseology (A. Tutin, O. Kraif)

# Work related to phraseology

6

## **1. Work about collocations, lexical associations, typology of multiword expressions**

- ✦ **Two fields of applications: lexicon of emotion and scientific writings**
- ✦ **Theoretical claim:**
  - Lexical associations are for most of them not irregular but governed by general rules
  - Phraseology is an essential property of natural languages. Few expressions are really « frozen »

## 2. Computational aspects and linguistic modelling:

- Extraction and modelling of collocations :
  - First, with simple FST techniques
  - Extraction and annotation of collocations in scientific writing.

### Concordance extraite par ConcQuest

Réduire

Fichier No1 - Emergence.fr.dat

composition , les pouvoirs et les responsabilités du comité , et donne un aperçu des	<b>éléments</b> qui devraient être <b>fournis</b>	dans le formulaire de protocole d' utilisation d' animaux .
e1-CCPA-Manuell-s53 [...] ) Les	<b>évaluations</b> sont <b>faites</b>	à partir des volumes 1 et 2 du Manuel et des déclarations de principes du CC
e1-CCPA-Manuell-s54 [...] Des	<b>évaluations</b> complètes sont <b>effectuées</b>	approximativement tous les trois ans .
e1-CCPA-Manuell-s62 [...] Pour quelques-unes de ces institutions , les	<b>évaluations</b> peuvent être <b>effectuées</b>	sur une base soit facultaire , soit départementale .
res législatives , les établissements ont le devoir de connaître la nature de toutes les	<b>expériences effectuées</b>	dans leurs laboratoires et de s' assurer de leur qualité .

Statistiques d'occurrences (9999 segments analysés.)

Requête (<base-colloc=/. /> [<> {0,6}]<type-verbe=/(produire|faire)/> | <type-verbe=/(produire|faire)/> [<> {0,6}]<base-colloc=/. /> ) : 40

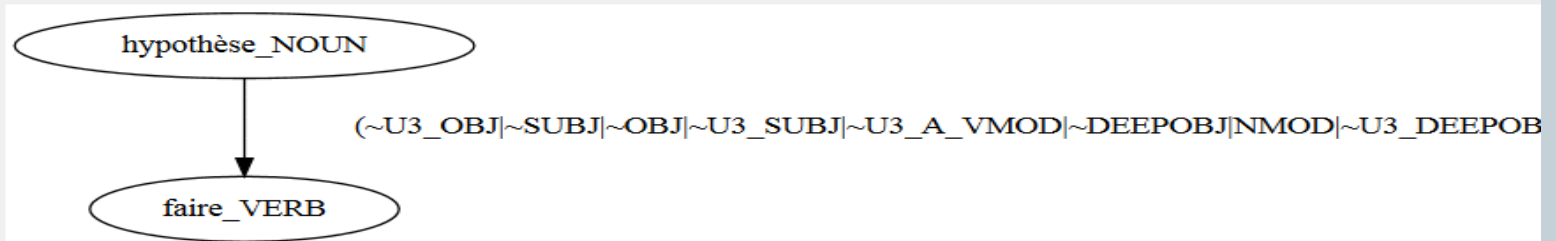
Répartition par textes

## 2. Extraction of collocations and other types of MWE from dependency annotated corpora (with Olivier Kraif and Sylvain Hatier)

- In the framework of the Emolex project
- In the framework of the Termith project

Collocatif *faire\_VERB* :

Forme canonique : faire hypothèse (201) ▲



`<l=hypothèse,c=NOUN,#1>&&<l=faire,c=VERB,#2>::((~U3_OBJ|~SUBJ|~OBJ|~U3_SUBJ|~U3_A_VMOD|~DEEPOBJ|NMOD|~U3_DEEPOBJ|U3_SUR_NMOD|~U3_SUR_VMOD),1,2)`

Show 10 entries

Search:

faire hypothèse

YMI... lien xml 10252 L'une fait de la régularité une hypothèse très forte... au point de supposer que les marqueurs d'un type de relatif



### 3. Lexicographic aspects

- Modelling of MWEs in corpora : dictionary cum corpus applications

Expressions indiquant le point de vue et l'attitude dans les écrits scientifiques

Accès par sens ▼

Pour exprimer...

Expressions :	Catégorie
... la certitude/l'évidence	
l'approximation	adverbe
... l'opinion positive	
... le point de vue	
... la quantité/l'intensité	
... l'opinion négative	
... la qualité	
en général (***)	
en moyenne (**)	
<b>au sens large (*)</b>	<b>Définition</b>
pour ainsi dire (*)	en ayant une définition extensive, large des phénomènes
	<b>Synonymes</b>
	grasso modo

Exemples

▼ 1/6 Contexte gauche: 10 mots Occurrence: Contexte droit: 10 mots

- lecteur , nous allons exploiter les différentes connaissances linguistiques ( **au sens large** du terme ) présentes dans les textes .
- Je vais vous parler de ma formation **au sens large** et de mes expériences dans l'enseignement et la
- est question de tenir compte de tous les textes ( pris ici **au sens large** puisque il s' agit de caricatures , mais nous verrons
- Il s' agirait davantage de la situation ou **au sens large** du contexte culturel et historique dans lequel s' inscrit.

- Teaching and learning activities about phraseology (Projet FULS)

# Interest in the AIM-WEST project

10

- 1. Linguistic study of MWEs and typology of MWEs**
- 2. Annotations of MWEs in corpora**