

# Génération de terminologies et d'index par extraction automatique d'expressions polylexicales

Luka Nerima - Unité d'informatique  
Laboratoire d'analyse et de technologie du langage (LATL / CUI)

Collaboration : Eric Wehrli et Violeta Seretan (FTI)

# Expressions polylexicales: exemples

- ❖ analyse du discours
  - ❖ analyse de la parole
  - ❖ analyse en constituants
  - ❖ arbitraire du signe
- 
- ❖ curieux hasard
  - ❖ dernier rempart
  - ❖ deuxième avis
- 
- ❖ exercer profession
  - ❖ étouffer colère
  - ❖ déchaîner colère
  - ❖ éprouver sentiment
  - ❖ émettre hypothèse

# Expressions polylexicales: exemples

## ❖ N-prep-N

- ❖ analyse du discours
- ❖ analyse de la parole
- ❖ analyse en constituants
- ❖ arbitraire du signe

## ❖ Adj-N

- ❖ curieux hasard
- ❖ dernier rempart
- ❖ deuxième avis

## ❖ V-Obj

- ❖ exercer profession
- ❖ étouffer colère
- ❖ déchaîner colère
- ❖ éprouver sentiment
- ❖ émettre hypothèse

❖ ...

# Outil d'extraction

- ❖ Approche hybride: analyse syntaxique + statistiques
- ❖ Analyse par l'analyseur syntaxique FIPS (Eric Wehrli)
  - ❖ expressions dont les termes sont en relation syntaxique, p.e.: N-N, Adj-N, N-Adj, N-Prep-N, Suj-V, V-Obj, etc.
  - ❖ structure syntaxique « normalisée »
- ❖ Méthodes statistiques -> mesurer le degré d'association des termes de l'expression

# Exemple V-Obj: « éprouver des difficultés »

termes pas forcément contigus

- ❖ éprouvant de très sérieuses difficultés
- ❖ éprouvent toujours les plus grandes difficultés
- ❖ éprouver, comme pour d'autres entités plus grandes, ou moins européennes dans leurs caractéristiques, de grandes difficultés

transformations syntaxiques

- ❖ Les difficultés qu'elle aurait pu éprouver ...



# Application

- ❖ Projet de mise en ligne et d'édition numérique des manuscrits de Ferdinand de Saussure (-> Claire Forel)
- ❖ ... mais s'appliquerait aussi à Nietzsche, Newton, Balzac...

## Display Collocations

Collocations (289)

Source: C:\...\Corpus\premieresTranscriptions.odc

donnée de sens  
**rapport irréductible**  
 dire à ami  
 instance arracher  
 petit terrain  
 dieu combler  
 instance pouvoir  
 objet matériel  
 certain nombre  
 dieu irrité  
 terrain dans île  
 atteindre existence  
 contenir preuve  
 différence dans échelle  
 différence entre sens  
 éteindre ombre  
 grande puissance civilisateur  
 objet donner  
 objet premier  
 papa dévorer  
 papa tomber  
 peuple parler  
 plan important  
 richesse récompenser

irréductibles entre eux.  
 (1 signe/sa signific.) = 2 formes différence entre 2  
 formes - différence entre 3 sens. (1 signe / autre signe) Nh+ 0 (1 signific. / 1 autre signif.)

C'est là que nous appelons le QUATERNION FINAL, et en considérant les 4 termes dans leur rapports: le triple **rapport irréductible**. C'est peut-être à tort que nous renonçons à réduire ces trois rapports à un seul; mais il nous semble que cette tentative ??? dépasserait la compétence du linguiste.

l'idéourbasCapital  
 Ce n'est pas la même chose de parler du rapport de la forme et de l'idée, ou du rapport de l'idée et de la forme: parce que si l'on prend pour base la forme A on ... un certain nombre d'idées a, b, c; et que si l'on prend pour base l'idée a on... (+- ...) un certains nombres de formes A H Z. Il n'y a donc le résidu b, c d'une part et B, Z de l'autre ( rapport (abc)/A) et que si l'on prend pour base l'idée a on en ... 4

on voit donc que la parallèle dont nous faisons momentanément une unité positive et indépendante des formes, n'est pas positive pour la même raison qu'elle est pas indépendante des formes; mais ou n'est pas indépendante des formes pr <pour> le même raison qu'elle n'est pas positive;  
 Ce qu'est la Catégorie gamma? par rapport à la parallèle

Language: French

Crt: 1 of 2



Open source file

 Group by syntactic typeOrder:  score frequency alphabetical

Details

Type: Noun - Adjective

Score: 14.86

Rank: 2

Status: New collocation!

Options ...

Validate ...

Cancel

# Méthodologie

- ❖ Génération automatique (ou semi automatique) de terminologies et d'index sur corpus
  - ❖ -> annotation de documents
  - ❖ -> recherche de documents (au sens moteur de recherche)
- ❖ Génération incrémentale
  - ❖ pas de terminologie et / ou de catalogue à priori
  - ❖ facile à faire ou à refaire
  - ❖ découverte de nouvelles entrées terminologiques ?

# Démos

<http://latlapps.unige.ch>

## LatlApps

Analyse  
syntaxique

Traduction  
automatique  
(SIWIS)

Synthèse  
vocale

Visualisation  
de collocations

Extraction  
de collocations