

# De GLÀFF à GLAWI et retour... d'expérience sur 6 ans d'exploitation du Wiktionnaire

*Thématiques actuelles de la recherche en TAL, 9/11/2015*

**Franck Sajous**

CLLE-ERSS, CNRS & Université de Toulouse 2



# Sommaire

- 1 Wiktionary(s) et Wiktionnaire
  - Wiktionary et TAL : aperçu et contexte historique
  - Macro- et micro-structure, format d'encodage
- 2 Ressources issues du Wiktionnaire
  - GLÀFF
  - GLAWI

## Objectifs

- Présentation des principales caractéristiques du Wiktionnaire  
→ potentiels, notamment pour la linguistique et le TAL
- Illustration par la description de réalisations fondées sur le Wiktionnaire

## Générique (par ordre d'apparition)

Bruno Gaume, Laurent Prévot, Emmanuel Navarro, Yannick Chudy,  
Nabil Hathout et Basilio Calderone

# Wiktionary/Wiktionnaire

## « Le compagnon lexical de Wikipédia »

- projet de lexicographie collaborative
- 171 éditions de langue
- *dumps* mis à disposition librement
- Wiktionary =
  - ensemble du projet  
<http://www.wiktionary.org>
  - version de langue anglaise  
<http://en.wiktionary.org>



## Wiktionnaire

- édition de langue française de Wiktionary  
<http://fr.wiktionary.org>

# Une grande richesse lexicale

## Contenu des articles

- étymologies
- définitions : gloses et exemples
- relations lexico-sémantiques
- traductions
- relations morphologiques
- paradigmes flexionnels
- transcriptions phonémiques
- variantes orthographiques

<http://fr.wiktionary.org/wiki/mousse>

## mousse

↳ Voir aussi : **moussé**, **Moussé**

### Sommaire

- [-] Français
  - [×] Étymologie
  - [-] Nom commun 1
    - [×] Synonymes
    - [×] Dérivés
    - [×] Traductions
  - [+] Nom commun 2
  - [+] Nom commun 3
  - [+] Adjectif
  - [×] Forme de verbe
  - [×] Prononciation
  - [×] Voir aussi
  - [×] Références

Voir : [Navarro et al., 2009], [Sajous et al., 2010], [Meyer, 2013], [Sajous et al., 2013b], [Hathout et al., 2014a], [Sajous et al., 2014]

# Premières utilisations

## Wiktionary : un terrain de jeu pour le TAL

- une ressource libre
- avec une nomenclature exceptionnelle
- que l'on peut télécharger !

## Calculs de proximité sémantique

- [Zesch et al., 2008] : calculs vectoriels
- [Navarro et al., 2009] : marches aléatoires

# *"We desperately need linguistic resources!"* [Sekine, 2010]

## Contexte : lexiques existants pour le français

- **Flexionnels** : Leffh [Clément et al., 2004] (gratuit&libre), Morphalou [Romary et al., 2004] (gratuit, non libre)
- **Phonologiques** : BDLex [Pérennou and de Calmès, 1987], ILPho [Boula De Mareuil et al., 2000]  
→ payants, non libres
- **Pour la psycholinguistique** : Lexique [New, 2006]  
→ gratuit, libre, couverture limitée
- **À large couverture** : WOLF [Sagot and Fišer, 2008]  
construit entièrement automatiquement  
→ quantité vs. qualité [Eckard et al., 2012]  
Dans la dernière version, seulement 10 % des synsets contiennent une définition en français

# Pourquoi ne pas utiliser Wiktionary. . .

... pour construire des ressources lexicales ?

## WiktionaryX [Sajous et al., 2010]

```
<entry pos="N" id="en_N_computer#0">
  <defs>
    <toplevel-def>
      <gloss domain="Computing"> A programmable device that performs
        mathematical calculations and operations [...]/</gloss>
    </toplevel-def>
    <toplevel-def>
      <gloss register="Dated">A person employed to perform computations.</gloss>
      <example> Only a few years ago Mr. Powers, an American computer,
        disproved a hypothesis about prime numbers[...]</example>
    </toplevel-def>
  </defs>
  <syn>
    <item target="en_N_processor">processor</item>
    <item target="en_N_machine">machine</item>
  </syn>
  <hypo>
    <item target="en_N_desktop">desktop</item>
    <item target="en_N_laptop">laptop</item>
  </hypo>
  <trans>
    <item lang="ar" target="ar_N_حاسوب">حاسوب</item>
    <item lang="ar" target="ar_N_كمبيوتر">كمبيوتر</item>
    <item lang="bg" target="bg_N_компютър">компютър</item>
    <item lang="hr" target="hr_N_računalo">računalo</item>
  </trans>
</entry>
```

## GLÀFF [Sajous et al., 2013b, Hathout et al., 2014b]

```
affluent | Ncms | affluent | a.fly.ã | a.fly.A~
affluents | Afpmp | affluent | a.fly.ã | a.fly.A~
affluents | Ncmp | affluent | a.fly.ã | a.fly.A~
affluent | Vmip3p- | affluer | a.fly | a.fly
```

# Experts vs. foules : pour une approche agnostique

## La « sagesse des foules »

- légitimité ? Voir [Penta, 2011, Sajous et al., 2014]
- “*Wisdom of crowds vs. wisdom of linguists*”  
[Zesch and Gurevych, 2010]

**Wiktionnaire au moins complémentaires des ressources traditionnelles** : créations lexicales, domaines de spécialité, variations diatopiques, *subcultures*, formes argotiques, formes désuètes, etc.

- « *Faut-il brûler les dictionnaires ?* » [Dal and Namer, 2012]  
Formes attestées (fréquentes) absentes des dictionnaires : *américanité*, *bien-penseance*, *contre(-)productif*, *doublonner*, *européanité*, *fiabiliser*, *gravage*, *livrage*, *mesurette*  
→ toutes dans le Wiktionnaire



## Experts vs. foules : pas d'opposition binaire

Imports	Articles	%
0	242499	83.42
1	48162	16.57
2	46	0.02

Sources	Articles	%
DAF8	27945	57.91
Littré	20278	42.02
Larousse XIXe	24	0.05

→ Wiktionnaire : experts **ET** foules

Références	Articles	%
0	260362	89.56
1	27818	9.57
2	2268	0.78
3	208	0.07

Sources	Articles	%
Littré	6497	19.56
DAF8	6311	19.00
TLFi	6256	18.84
Rivarol	4358	13.12
Meyer	3523	10.61
FranceTerme	2922	8.80
Mistral	650	1.96
ODS5	394	1.19
GDT	200	0.60
DAF9	195	0.59
Bescherelle	116	0.35
Gaffiot	105	0.32

# Macro-structure

http://fr.wiktionary.org/wiki/affluent

entrée = url = forme orthographique



- Page d'accueil
- Index alphabétique
- Portails thématiques
- Page au hasard
- Page au hasard par langue
- Poser une question

Contribuer

Aide

Boîte à outils

Autres langues

- العربية
- Cymraeg
- Deutsch
- EAAnꞵꞵꞵ
- English
- Eesti
- Euskara
- فارسی
- Suomi
- Gallego
- Magyar
- Bahasa Indonesia
- Ido
- Italiano
- Malagasy
- македонски
- Nederlands
- Polski
- پښتو
- Русский
- Svenska
- தமிழ்
- Тыа
- Türkçe
- Tiếng Việt

Article Discussion

Lire Modifier Afficher l'historique

## affluent

**Français** [modifier]

### Étymologie

(1374)Du latin **affluens**(+ **affluent**+) → voir **affluer**.

### Adjectif

**affluent**

- (Géographique)Qui se **jette dans** un **autre** en **parlant** d'un **cours d'eau**.  
Le Rhin et les rivières **affluentes**.
- (Médecine)Qui **affluent**, qui se **portent** en **abondance vers quelque partie du corps**(se dit des **humeurs**).  
La sérosité, la salive **affluente**.

	Singulier	Pluriel
Masculin	<b>affluent</b> <i>/a.fly.ɑ̃/</i>	<b>affluents</b> <i>/a.fly.ɑ̃/</i>
Féminin	<b>affluente</b> <i>/a.fly.ɑ̃t/</i>	<b>affluentes</b> <i>/a.fly.ɑ̃t/</i>

### Nom commun

**affluent** */a.fly.ɑ̃/* masculin

- (Géographique) **Cours d'eau** qui se **jette dans** un **autre**.  
Voici le Madon, **affluent** de la Moselle, dont les eaux limoneuses invitent peu à la baignade, malgré la grande chaleur.  
(Gustave Fraipont: Les Vosges , 1923)

Singulier	Pluriel
<b>affluent</b>	<b>affluents</b>
	<i>/a.fly.ɑ̃/</i>

### Traductions

anglais : **affluent**<sup>[en]</sup>  
catalan : **afluent**<sup>[ca]</sup>  
espagnol : **afuente**<sup>[es]</sup>  
espéranto : **branĉrivero**<sup>[eo]</sup>  
ido : **enfluinto**<sup>[io]</sup>  
néerlandais : **zijrivier**<sup>[nl]</sup>  
occitan : **afluent**<sup>[oc]</sup>

### Forme de verbe

**affluent** */a.fly/*

- Troisième personne du pluriel de l'indicatif présent de **affluer**.
- Troisième personne du pluriel du subjonctif présent de **affluer**.

Conjugaison du verbe <b>affluer</b>		
Indicatif	Présent	ils/elles <b>affluent</b>
Subjonctif	Présent	qu'ils/elles <b>affluent</b>

# Macro-structure

<http://fr.wiktionary.org/wiki/affluente>

affluente

**Français** [modifier]



**Forme d'adjectif**

**affluente** féminin /a.fly.ôt/

1 Féminin singulier de [affluent](#) .

# Macro-structure

http://fr.wiktionary.org/wiki/affluente

## affluente

**Français** [modifier]



Wiktionnaire  
Le dictionnaire libre

Page d'accueil  
 Index Alphanumérique  
 Portails Thématiques  
 Page au hasard  
 Page au hasard par langue  
 Poser une question

Contribuer  
 Aide  
 Babes à outils

Autres langues  
 العربية  
 Cymraeg  
 Deutsch  
 Ελληνικά  
 English  
 Español  
 Français  
 Galego  
 Magyar  
 Bahasa Indonesia  
 Ido  
 Italiano  
 Malayalam  
 বাংলা  
 Nederlands  
 Polski  
 Pycckий  
 Svenska  
 தமிழ்  
 Türkçe  
 中文

Article Discussion
Lire Modifier Afficher l'historique

### affluent

**Français** [modifier]

**Etymologie**

(1374)Du latin **affluens**+ **affluens** (« → voir **affluer**».

**Adjectif**

**affluent**

- (Géographie)Qui se **jette dans** un **autre** **parlant** un **course** eau.  
Le Rhin et les rivières **affluentes**.
- (Médecine)Qui **affluent**, qui se **portent** **abondance** vers **quelque** **partie** du **corps** se dit des **humeurs**.  
La sérosité, la salive **affluente**.

**Nom commun**

**affluent** (a, f) / masculin

- (Géographie) **Cours** d'**eau** qui se **jette dans** un **autre**.  
Voici le Madon, **affluent** de la Moselle, dont les eaux limoneuses invitent peu à la baignade, malgré la grande chaleur.  
(modifier | réviser | voir | historique)

**Traductions**

anglais : affluent<sup>[…]</sup>  
 catalan : afuent<sup>[…]</sup>  
 espagnol : afuente<sup>[…]</sup>  
 espéranto : branĉrivero<sup>[…]</sup>  
 ido : enfluanto<sup>[…]</sup>  
 maltais : zifruer<sup>[…]</sup>  
 occitan : afuent<sup>[…]</sup>

**Forme de verbe**

**affluent** (a, f) /

- Troisième personne du pluriel de l'indicatif présent de **affluer**.
- Troisième personne du pluriel du subjonctif présent de **affluer**.

Créer un compte Connexion

**Forme d'adjectif**

**affluente** féminin /a, f) / ô /

1 Féminin singulier de **affluent** .

# Macro-structure

<http://fr.wiktionary.org/wiki/affluer>

## affluer

**Français** [modifier]



### Verbe

**affluer** /a.fly.e/ intransitif 1<sup>er</sup> groupe (conjugaison)

- 1 Couler vers, en parlant des eaux qui se rendent et se réunissent dans un même lit.

Plusieurs ruisseaux et plusieurs rivières **affluent** dans la Seine, dans le Rhône.

- 2 (Figuré) Abonder, arriver en abondance.

Quand tout dort encore, les larges avenues de Carpentras voient **affluer** des amoncellements de cerises et de fraises. — (Ludovic Naudeau, La France se regarde. -1931)

# Macro-structure

http://fr.wiktionary.org/wiki/Annexe:Conjugaison\_en\_français/affluer



Wiktionnaire  
Le dictionnaire libre

- Page d'accueil
- Index alphabétique
- Portails thématiques
- Page au hasard
- Page au hasard par langue
- Poser une question

Contribuer

- Journal des contributeurs
- La Wikidémie
- Communauté
- Discuter sur IRC
- Modifications récentes
- Faire un don

Aide

Boîte à outils

Annexe [Discussion](#) Lire [Modifier](#) [Afficher l'historique](#)

## Annexe:Conjugaison en français/affluer

Annexe:Conjugaison en français

Conjugaison de **affluer**, verbe du 1<sup>er</sup> groupe, conjugué avec l'auxiliaire avoir

### Modes impersonnels

Mode	Présent	Passé
<b>Infinitif</b>	affluer /a.fluɛ/	avoir afflué /a.vwaʁ_a.fluɛ/
<b>Gérondif</b>	en affluant /ɔ̃.n_a.fluɑ̃/en ayant	afflué /ɔ̃.n_ɛ.jò.t_a.fluɛ/
<b>Participle</b>	affluant /a.fluɑ̃/	afflué /a.fluɛ/

### Indicatif

	Présent	Passé composé
j'afflue	/ʒ_a.flɥ/	j'ai afflué /ʒ_e_a.fluɛ/
tu affluas	/ty_a.flɥ/	tu as afflué /ty_a.z_a.fluɛ/
il/elle/on afflue	/il(ɛ/lʷ)ʒ_a.flɥ/	il/elle/on a afflué /il(ɛ/lʷ.n)_a.t_a.fluɛ/
nous affluons	/nu.z_a.flɥ/	nous avons afflué /nu.z_a.vɔ.z_a.fluɛ/
vous affluez	/vu.z_a.fluɛ/	vous avez afflué /vu.z_a.ve.z_a.fluɛ/
ils/elles affluent	/il(ɛ)l.z_a.flɥ/	ils/elles ont afflué /il(ɛ)l.z_ɔ.t_a.fluɛ/
	Imparfait	Plus-que-parfait
j'affluais	/ʒ_a.fluɛ/	j'avais afflué /ʒ_a.ve.z_a.fluɛ/
tu affluais	/ty_a.fluɛ/	tu avais afflué /ty_a.ve.z_a.fluɛ/
il/elle/on affluait	/il(ɛ/lʷ)ʒ_a.fluɛ/	il/elle/on avait afflué /il(ɛ/lʷ.n)_a.ve.t_a.fluɛ/
nous affluions	/nu.z_a.flɥ/	nous avions afflué /nu.z_a.vɔ.z_a.fluɛ/
vous affluiez	/vu.z_a.fluɛ/	vous aviez afflué /vu.z_a.vje.z_a.fluɛ/
ils/elles affluaient	/il(ɛ)l.z_a.fluɛ/	ils/elles avaient afflué /il(ɛ)l.z_a.ve.t_a.fluɛ/
	Passé simple	Passé antérieur
j'affluai	/ʒ_a.fluɛ/	j'eus afflué /ʒ_y.z_a.fluɛ/
tu affluas	/ty_a.fluɛ/	tu eus afflué /ty_y.z_a.fluɛ/
il/elle/on afflua	/il(ɛ/lʷ)ʒ_a.fluɛ/	il/elle/on eut afflué /il(ɛ/lʷ.n)_y.t_a.fluɛ/
nous affluâmes	/nu.z_a.fluam/	nous eûmes afflué /nu.z_ym.z_a.fluɛ/
vous affluâtes	/vu.z_a.fluɑt/	vous eûtes afflué /vu.z_yt.z_a.fluɛ/
ils/elles affluèrent	/il(ɛ)l.z_a.fluɛr/	ils/elles eurent afflué /il(ɛ)l.z_yw.t_a.fluɛ/
	Futur simple	Futur antérieur
j'affluerais	/ʒ_a.flɥ.ɛr/	j'aurai afflué /ʒ_o.ve_a.fluɛ/
tu afflueras	/ty_a.flɥ.ɛr/	tu auras afflué /ty_o.ɛ.a.z_a.fluɛ/
il/elle/on affluera	/il(ɛ/lʷ)ʒ_a.flɥ.ɛr/	il/elle/on aura afflué /il(ɛ/lʷ.n)_y.t_o.ɛ.a_a.fluɛ/
nous affluerons	/nu.z_a.flɥ.ɛr/	nous aurons afflué /nu.z_o.ɛ.ɔ.z_a.fluɛ/
vous affluerez	/vu.z_a.flɥ.ɛr/	vous aurez afflué /vu.z_o.ve.z_a.fluɛ/

# Macro-structure

[http://fr.wiktionary.org/wiki/Annexe:Conjugaison\\_en\\_français/affluer](http://fr.wiktionary.org/wiki/Annexe:Conjugaison_en_français/affluer)

The image shows a screenshot of the Wiktionary page for 'affluer'. The page is divided into several sections, with arrows pointing from the main page to detailed views of specific sections:

- Annexe:Conjugaison en français**: The main page title, with an arrow pointing to the detailed view of the conjugation table.
- Modes impersonnels**: A section containing 'Mode', 'Infinitif', 'Gérondif', and 'Participle'. An arrow points to the detailed view of the 'Forme de verbe' section.
- Indicatif**: A section containing 'Présent' and 'Imparfait'. An arrow points to the detailed view of the 'Forme de verbe' section.
- Passé simple**: A section containing 'Présent' and 'Imparfait'. An arrow points to the detailed view of the 'Forme de verbe' section.
- Futur simple**: A section containing 'Présent' and 'Imparfait'. An arrow points to the detailed view of the 'Forme de verbe' section.

The detailed views show the conjugation table for 'affluer' and the 'Forme de verbe' section, which includes the following information:

**Forme de verbe**

affluer *(a, fly)*

- Première personne du singulier de l'indicatif présent du verbe **affluer**
- Troisième personne du singulier de l'indicatif présent du verbe **affluer**
- Première personne du singulier du subjonctif présent du verbe **affluer**
- Troisième personne du singulier du subjonctif présent du verbe **affluer**
- Deuxième personne du singulier de l'impératif du verbe **affluer**

Conjugaison du verbe affluer		
Indicatif	Présent	j'afflue il/elle/on afflue
	Subjonctif	Présent que j'afflue qu'il/elle/on afflue
Impératif	Présent	(2 <sup>e</sup> personne du singulier) afflue

# Micro-structure : un format d'encodage malheureux

## Pas de structuration en bases de données

- milieu des années 1960s : première vision du dictionnaire sous forme de base de données par L. Urdang [Rundell and Kilgarriff, 2011]
- 2003 : lancement de Wiktionary
  - Format d'encodage : wikicode
  - Pas de syntaxe formelle définie !
  - MediaWiki Parser : *"a mess of regular expressions"* (E. Jones, auteur de wikipedia2text)



Wiktionnaire  
Le dictionnaire libre

[Page d'accueil](#)  
[Index alphabétique](#)  
[Portails thématiques](#)  
[Page au hasard](#)  
[Page au hasard par langue](#)  
[Poser une question](#)

► [Contribuer](#)

► [Aide](#)

► [Boîte à outils](#)

Autres langues

العربية  
 Cymraeg  
 Deutsch  
 Ελληνικά  
 English  
 Eesti  
 Euskara  
 فارسی  
 Suomi  
 Galego  
 Magyar  
 Bahasa Indonesia  
 Ido  
 Italiano  
 Malagasy  
 മലയാളം  
 Nederlands  
 Polski  
 پښتو  
 Русский

Article [Discussion](#)

Lire [Modifier](#) [Afficher l'historique](#)

## affluent

**Français** [[modifier](#)]



### Étymologie

(1374)Du latin **affluens** (« **affluent** ») → voir **affluer**.



### Adjectif

#### affluent

- (Géographie) Qui se **jette dans** un**autre** en **parlant** d'un **cours** d'eau.  
Le Rhin et les rivières **affluentes**.
- (Médecine)Qui **affluent**, qui se **portent** en **abondance vers quelque partie** du **corps** (se dit des **humeurs**).  
La sérosité, la salive**affluente**.

	Singulier	Pluriel
Masculin	<b>affluent</b> <i>/a.fly.ɑ̃/</i>	<b>affluents</b> <i>/a.fly.ɑ̃/</i>
Féminin	<b>affluente</b> <i>/a.fly.ɑ̃t/</i>	<b>affluentes</b> <i>/a.fly.ɑ̃t/</i>



### Nom commun

#### affluent /a.fly.ɑ̃/ masculin

- (Géographie) **Cours d'eau**qui se **jette dans** un **autre**.  
Voici le Madon, **affluent** de la Moselle, dont les eaux limoneuses invitent peu à la baignade, malgré la grande chaleur.  
(Gustave Fraipont ; Les Vosges , 1923)

	Singulier	Pluriel
	<b>affluent</b>	<b>affluents</b>
		<i>/a.fly.ɑ̃/</i>

### Traductions

anglais : <sup>(en)</sup> **affluent** <sup>(en)</sup>  
 catalan : <sup>(ca)</sup> **afluent** <sup>(ca)</sup>  
 espagnol : <sup>(es)</sup> **afluente** <sup>(es)</sup>  
 espéranto : <sup>(eo)</sup> **branĉrivero** <sup>(eo)</sup>  
 ido : <sup>(io)</sup> **enfluento** <sup>(io)</sup>  
 néerlandais : <sup>(nl)</sup> **zijrivier** <sup>(nl)</sup>  
 occitan : <sup>(oc)</sup> **affluent** <sup>(oc)</sup>



### Forme de verbe

#### affluent /a.fly/

Conjugaison du verbe **affluer**



Wiktionnaire  
Le dictionnaire libre

[Page d'accueil](#)  
[Index alphabétique](#)  
[Portails thématiques](#)  
[Page au hasard](#)  
[Page au hasard par langue](#)  
[Poser une question](#)

► [Contribuer](#)

► [Aide](#)

► [Boîte à outils](#)

Autres langues

العربية  
 Cymraeg  
 Deutsch  
 Ελληνικά  
 English  
 Eesti  
 Euskara  
 فارسی  
 Suomi  
 Galego  
 Magyar  
 Bahasa Indonesia  
 Ido  
 Italiano  
 Malagasy  
 മലയാളം  
 Nederlands  
 Polski  
 پښتو  
 Русский

Article [Discussion](#)

Lire [Modifier](#) [Afficher l'historique](#)

## affluent

**Français** [[modifier](#)]

### Étymologie

(1374)Du latin **affluens** (« **affluent** ») → voir **affluer**.

### Adjectif

**affluent**

- (Géographie) Qui se **jette dans** un**autre** en **parlant** d'un **cours** d'eau.  
Le Rhin et les rivières **affluentes**.
- (Médecine)Qui **affluent**, qui se **portent** en **abondance vers quelque partie** du **corps** (se dit des **humeurs**).  
La sérosité, la salive**affluente**.

	Singulier	Pluriel
Masculin	<b>affluent</b> <i>/a.fly.ɑ̃/</i>	<b>affluents</b> <i>/a.fly.ɑ̃/</i>
Féminin	<b>affluente</b> <i>/a.fly.ɑ̃t/</i>	<b>affluentes</b> <i>/a.fly.ɑ̃t/</i>

### Nom commun

**affluent** */a.fly.ɑ̃/* masculin

- (Géographie) **Cours d'eau**qui se **jette dans** un **autre**.  
Voici le Madon, **affluent** de la Moselle, dont les eaux limoneuses invitent peu à la baignade, malgré la grande chaleur.  
(Gustave Fraipont ; Les Vosges , 1923)

### Traductions

anglais : **affluent**<sup>(en)</sup>  
 catalan : **afluent**<sup>(ca)</sup>  
 espagnol : **afluente**<sup>(es)</sup>  
 espéranto : **branĉrivero**<sup>(eo)</sup>  
 ido : **enfluanto**<sup>(io)</sup>  
 néerlandais : **zijrivier**<sup>(nl)</sup>  
 occitan : **afluent**<sup>(oc)</sup>

### Forme de verbe

**affluent** */a.fly/*

Conjugaison du verbe **affluer**



## Wikicode

## Modification de « affluent »

Gratistique    Fichier    ▶ Avancé    ▶ Caractères spéciaux    ▶ Aide

```
== {{langue|fr}} ==
{{-étym-}}
: {{date|1374}} Du {{étyl|la|fr|mot=affluens|sens=[[affluent]]}}
{{cf|affluer|lang=fr}}.

{{-adj-|fr}}
{{fr-accord-cons|a.fly.ô|t}}
'''affluent'''

# {{géographie|fr}} Qui se [[jeter|jette]] [[dans]] un [[autre]] en
[[parlant]] d'un [[cours]] d'eau.
#* "La sérosité, la salive '''affluente'''."

{{-nom-|fr}}
{{fr-rég|a.fly.ô}}
'''affluent''' {{pron|a.fly.ô|fr}} {{m}}

# {{géographie|fr}} [[cours d'eau|Cours d'eau]] qui se [[jeter|jette]]
[[dans]] un [[autre]].

{{-flex-verb-|fr}}
{{fr-verbe-flexion|affluer|ind.p.3p=oui|sub.p.3p=oui}}
'''affluent''' {{pron|a.fly|fr}}

# "Troisième personne du pluriel de l'indicatif présent de" [[affluer]].
# "Troisième personne du pluriel du subjonctif présent de" [[affluer]].

{{-pron-}}
```

## Wikicode

## Modification de « affluent »

Graſtaltique    Fichier    ▶ Avancé    ▶ Caractères spéciaux    ▶ Aide

```
== {{langue|fr}} ==
{{-étym-}}
: {{date|1374}} Du {{étyl|la|fr|mot=affluens|sens=[[affluent]]}}
{{cf|affluer|lang=fr}}.
```

```
{{-adj-|fr}}
```

```
{{fr-accord-cons|a.fly.ã|t}}
```

```
'''affluent'''
```



**{{fr-accord-cons|a.fly.ã|t}}**

```
# {{géographie|fr}} Qui se [[jeter|jette]] [[dans]] un [[autre]] en
[[parlant]] d'un [[cours]] d'eau.
```

```
#* "La sérosité, la salive '''affluente'''."
```

```
{{-nom-|fr}}
```

```
{{fr-rég|a.fly.ã}}
```

```
'''affluent''' {{pron|a.fly.ã|fr}} {{m}}
```

```
# {{géographie|fr}} [[cours d'eau|Cours d'eau]] qui se [[jeter|jette]]
[[dans]] un [[autre]].
```

```
{{-flex-verb-|fr}}
```

```
{{fr-verbe-flexion|affluer|ind.p.3p=oui|sub.p.3p=oui}}
```

```
'''affluent''' {{pron|a.fly|fr}}
```

```
# "Troisième personne du pluriel de l'indicatif présent de" [[affluer]].
```

```
# "Troisième personne du pluriel du subjonctif présent de" [[affluer]].
```

```
{{-pron-}}
```

## Wikicode

## Modification de « affluent »

Graſtaltique    Fichier    ▶ Avancé    ▶ Caractères spéciaux    ▶ Aide

```

== {{langue|fr}} ==
{{-étym-}}
: {{date|1374}} Du {{étyl|la|fr|mot=affluens|sens=[[affluent]]}}
{{cf|affluer|lang=fr}}.

```

```

{{-adj-|fr}}

```

```

{{fr-accord-cons|a.fly.ã|t}}

```

```

''affluent''

```

```

# {{géographie|fr}} Qui se [[jeter|jette]] [[dans]] un [[autre]] en
[[parlant]] d'un [[cours]] d'eau.
#* ''La sérosité, la salive ''affluente''.''

```

```

{{-nom-|fr}}

```

```

{{fr-rég|a.fly.ã}}

```

```

''affluent'' {{pron|a.fly.ã|fr}} {{m}}

```

```

# {{géographie|fr}} [[cours d'eau|Cours d'eau]] qui se [[jeter|jette]]
[[dans]] un [[autre]].

```

```

{{-flex-verb-|fr}}

```

```

{{fr-verbe-flexion|affluer|ind.p.3p=oui|sub.p.3p=oui}}

```

```

''affluent'' {{pron|a.fly|fr}}

```

```

# ''Troisième personne du pluriel de l'indicatif présent de'' [[affluer]].

```

```

# ''Troisième personne du pluriel du subjonctif présent de'' [[affluer]].

```

```

{{-pron-}}

```

**{{fr-accord-cons|a.fly.ã|t}}**

	Singulier	Pluriel
Masculin	<b>affluent</b> /a.fly.ã/	<b>affluents</b> /a.fly.ã/
Féminin	<b>affluente</b> /a.fly.õt/	<b>affluentes</b> /a.fly.õt/

## Wikicode

## Modification de « affluent »

Gratulation Fichier ▶ Avancé ▶ Caractères spéciaux ▶ Aide

```

== {{langue|fr}} ==
{{-étym-}}
: {{date|1374}} Du {{étyl|la|fr|mot=affluens|sens=[[affluent]]}}
{{cf|affluer|lang=fr}}.

{{-adj-|fr}}
{{fr-accord-cons|a.fly.ã|t}}
'''affluent'''

# {{géographie|fr}} Qui se [[jeter|jette]] [[dans]] un [[autre]] en
[[parlant]] d'un [[cours]] d'eau.
#* "La sérosité, la salive '''affluente'''."

{{-nom-|fr}}
{{fr-rég|a.fly.ã}}
'''affluent''' {{pron|a.fly.ã|fr}} {{m}}

# {{géographie|fr}} [[cours d'eau|Cours d'eau]] qui se [[jeter|jette]]
[[dans]] un [[autre]].

{{-flex-verb-|fr}}
{{fr-verb-flexion|affluer|ind.p.3p=oui|sub.p.3p=oui}}
'''affluent''' {{pron|a.fly|fr}}

# "Troisième personne du pluriel de l'indicatif présent de" [[affluer]].
# "Troisième personne du pluriel du subjonctif présent de" [[affluer]].

{{-pron-}}

```

**{{fr-accord-cons|a.fly.ã|t}}**

	Singulier	Pluriel
Masculin	<b>affluent</b> /a.fly.ã/	<b>affluents</b> /a.fly.ã/
Féminin	<b>affluente</b> /a.fly.õt/	<b>affluentes</b> /a.fly.õt/

affluent|Afpms|affluent|a.fly.ã  
 affluents|Afpmp|affluent|a.fly.ã  
 affluente|Afpfs|affluent|a.fly.õt  
 affluentes|Afpfp|affluent|a.fly.õt

## Wikicode : redondance

## affluent

 Adjectif

## affluent

- (Géographie)* Qui se **jette dans un autre** en parlant d'un cours d'eau.
- (Médecine)* Qui **affluent**, qui se **portent en abondance vers quelque partie du corps**.

	Singulier	Pluriel
Masculin	<b>affluent</b> <i>/a.fly.ɑ̃/</i>	affluents <i>/a.fly.ɑ̃/</i>
Féminin	affluente <i>/a.fly.ɑ̃t/</i>	affluentes <i>/a.fly.ɑ̃t/</i>



## Nom commun

affluent */a.fly.ɑ̃/* masculin

- (Géographie)* Cours d'eau qui se **jette dans un autre**.

	Singulier	Pluriel
	<b>affluent</b>	affluents
	<i>/a.fly.ɑ̃/</i>	<i>/a.fly.ɑ̃/</i>



## Forme de verbe

affluent */a.fly/*

- Troisième personne du pluriel de l'indicatif présent de **affluer**.
- Troisième personne du pluriel du subjonctif présent de **affluer**.

Conjugaison du verbe <i>affluer</i>		
INDICATIF	Présent	ils/elles affluent
SUBJONCTIF	Présent	qu'ils/elles affluent

```

{{-adj-|fr}}
{{fr-accord-cons|a.fly.ɑ̃|t}}
''affluent''
# {{géographie|fr}} Qui se [[jeter|jette]] [[dans]] un [[autre]] en [[parlant]] d'un [[cours]] d'eau.

{{-nom-|fr}}
{{fr-rég|a.fly.ɑ̃}}
''affluent'' {{pron|a.fly.ɑ̃|fr}} {{m}}
# {{géographie|fr}} [[cours d'eau|Cours d'eau]] qui se [[jeter|jette]] [[dans]] un [[autre]].

{{-flex-verb-|fr}}
{{fr-verbe-flexion|affluer|ind.p.3p=oui|sub.p.3p=oui|}}
''affluent'' {{pron|a.fly|fr}}
# ''Troisième personne du pluriel de l'indicatif présent de'' [[affluer]].
# ''Troisième personne du pluriel du subjonctif présent de'' [[affluer]].

```

## Wikicode : redondance

## affluent

 Adjectif

## affluent

1. (*Géographie*) Qui se **jette dans un autre** en parlant d'un cours d'eau.
2. (*Médecine*) Qui **affluent**, qui se **portent en abondance vers quelque partie du corps**.

	Singulier	Pluriel
Masculin	<b>affluent</b> <i>/a.fly.ɑ̃/</i>	affluents <i>/a.fly.ɑ̃/</i>
Féminin	affluente <i>/a.fly.ɑ̃t/</i>	affluentes <i>/a.fly.ɑ̃t/</i>



## Nom commun

affluent */a.fly.ɑ̃/* masculin

1. (*Géographie*) Cours d'eau qui se **jette dans un autre**.

Singulier	Pluriel
<b>affluent</b>	affluents
	<i>/a.fly.ɑ̃/</i>



## Forme de verbe

affluent */a.fly/*

1. *Troisième personne du pluriel de l'indicatif présent de affluer.*
2. *Troisième personne du pluriel du subjonctif présent de affluer.*

Conjugaison du verbe *affluer*

INDICATIF	Présent	ils/elles affluent
SUBJONCTIF	Présent	qu'ils/elles affluent

```

{{-adj-|fr}}
{{fr-accord-cons|a.fly.ɑ̃|t}}
''affluent''
# {{géographie|fr}} Qui se [[jeter|jette]] [[dans]] un [[autre]] en [[parlant]] d'un [[cours]] d'eau.

{{-nom-|fr}}
{{fr-rég|a.fly.ɑ̃}}
''affluent'' {{pron|a.fly.ɑ̃|fr}} {{m}}
# {{géographie|fr}} [[cours d'eau|Cours d'eau]] qui se [[jeter|jette]] [[dans]] un [[autre]].

{{-flex-verb-|fr}}
{{fr-verbe-flexion|affluer|ind.p.3p=oui|sub.p.3p=oui|}}
''affluent'' {{pron|a.fly|fr}}
# ''Troisième personne du pluriel de l'indicatif présent de'' [[affluer]].
# ''Troisième personne du pluriel du subjonctif présent de'' [[affluer]].

```



## Wikicode : redondance

## affluent

 Adjectif

## affluent

- (Géographie) Qui se jette dans un autre en parlant d'un cours d'eau.
- (Médecine) Qui affluent, qui se portent en abondance vers quelque partie du corps.

	Singulier	Pluriel
Masculin	affluent <i>/a.fly.õ/</i>	affluents <i>/a.fly.õ/</i>
Féminin	affluente <i>/a.fly.õt/</i>	affluentes <i>/a.fly.õt/</i>



## Nom commun

affluent */a.fly.õ/* masculin

- (Géographie) Cours d'eau qui se jette dans un autre.

	Singulier	Pluriel
	affluent	affluents
	<i>/a.fly.õ/</i>	<i>/a.fly.õ/</i>



## Forme de verbe

affluent */a.fly/*

- Troisième personne du pluriel de l'indicatif présent de affluer.
- Troisième personne du pluriel du subjonctif présent de affluer.

## Conjugaison du verbe affluer

INDICATIF	Présent	ils/elles affluent
SUBJONCTIF	Présent	qu'ils/elles affluent

```

{{-adj-|fr}}
{{fr-accord-cons|a.fly.õ|t}}
''affluent''
# {{géographie|fr}} Qui se [[jeter|jette]] [[dans]] un [[autre]] en [[parlant]] d'un [[cours]] d'eau.

{{-nom-|fr}}
{{fr-rég|a.fly.õ}}
''affluent'' {{pron|a.fly.õ|fr}} {{m}}
# {{géographie|fr}} [[cours d'eau|Cours d'eau]] qui se [[jeter|jette]] [[dans]] un [[autre]].

{{-flex-verb-|fr}}
{{fr-verbe-flexion|affluer|ind.p.3p=oui|sub.p.3p=oui|}}
''affluent'' {{pron|a.fly|fr}}
# ''Troisième personne du pluriel de l'indicatif présent de'' [[affluer]]
# ''Troisième personne du pluriel du subjonctif présent de'' [[affluer]].

```

# Conséquence du choix d'encodage

## Conséquences directes

- pas de syntaxe définie formellement → pas de vérification automatique possible lors de l'édition
- pas de références croisées → information dupliquée potentiellement inconsistante

## Effets secondaires

- sur la participation :
  - potentielle restriction des contributeurs
- sur l'extraction d'information :
  - wikicode volatile
  - wikicode non stable d'une langue à l'autre

## Wikicode volatile...

Date	Code wiki
2004-03-22	===Nom===
2004-04-01	===[[Nom]]===
2004-04-01	===[[Nom commun]]===
2004-09-18	=== [[Nom commun]] ===
2004-10-04	=== Nom commun ===
2004-11-02	{{-nom-}}
2006-09-29	{{-nom- fr}}
...	
2015	=== {{S nom fr}} ===

Date	Code wiki
2004-03-22	====Traductions====
2004-04-01	====[[Traduction]]s====
2004-10-04	==== Traductions ====
2004-11-01	{{-trans-}}
2005-04-13	{{-trad-}}
2009-03-12	{{-trad-trier}}
...	
2015	==== {{S traductions}} ====

Date	Code wiki
mars 2004	* albanien : [[fjalor]] * [[bréton]] : [[geriadur]]
octobre 2004	* {{sq}} : [[fjalor]] * {{br}} : [[geriadur]]
2015	* {{T sq}} : {{trad+ sq fjalor}} {{m}} * {{T br}} : {{trad+ br geriadur}} {{m}}

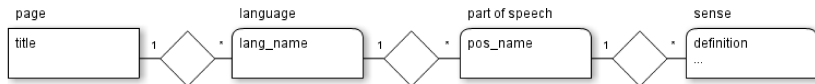
# Leveraging the Crowdsourcing of Lexical Resources

## for Bootstrapping a Linguistic Data Cloud [Hellmann et al., 2013]

**Abstract.** We present a declarative approach implemented in a comprehensive open-source framework based on *DBpedia* to extract lexical-semantic resources – an ontology about language use – from *Wiktionary*. The data currently includes language, part of speech, senses, definitions, synonyms, translations and taxonomies (hyponyms, hyperonyms, synonyms, antonyms) for each lexical word. Main focus is on flexibility to the loose schema and configurability towards differing language-editions of *Wiktionary*. This is achieved by a declarative mediator/wrapper approach. The goal is to allow the addition of languages just by configuration without the need of programming, thus enabling the swift and

name	active	available	RDF	#triples	ld languages
JWKT	✓	dumps	✗	-	✗ en, de
wikokit	✓	source + dumps	✓	n/a	✗ en, ru
texai	✗	dumps	✓	~ 2.7 million	✗ en
lemon scraper	✓	dumps	✓	~16k per lang	✗ 6
blexisma	✗	source	✗	-	✗ en
WISIGOTH	✗	dumps	✗	-	✗ en, fr
lexvo.org	✓	dumps	✓	~353k	✓ en

# Leveraging the Crowdsourcing of Lexical Resources for Bootstrapping a Linguistic Data Cloud [Hellmann et al., 2013]



```
===Synonyms===  
* [[building]]  
* [[company]]
```

Since the goal is to emit a link to each resource per line, we can write the ET in the following style, using the popular scraping paradigms such as regular expressions:

```
===Synonyms===  
(* [[\${target}]]  
)+
```

# GLÀFF, un Gros Lexique À tout Faire du Français

[Sajous et al., 2013a, Hathout et al., 2014b]

## Un lexique flexionnel

- à large couverture ;
- sous licence libre ;
- qui contient des informations :
  - morphosyntaxiques ;
  - phonémiques ;
  - fréquentielles.

affluent|Ncms|affluent|a.fly.ã|a.fly.A~|22|0.76|38|1.31|232|1.05|444|2.02|...  
affluents|Afpmp|affluent|a.fly.ã|a.fly.A~|2|0.06|15|0.51|5|0.02|183|0.83|...  
affluents|Ncmp|affluent|a.fly.ã|a.fly.A~|16|0.551|38|1.31|212|0.96|444|2.02|...  
affluent|Vmip3p-|affluer|a.fly|a.fly|9|0.318|187|6.48|369|1.67|1207|5.49|...

## Disponible

<http://redac.univ-tlse2.fr/lexiques/glaff.html>

## Tailles des nomenclatures

	Formes fléchies catégorisées			Lemmes catégorisés		
	Simple	Non simple	Total	Simple	Non simple	Total
<b>Lexique</b>	147 912	4 696	152 608	46 649	3 770	50 419
<b>BDLex</b>	431 992	4 360	436 352	47 314	1 792	49 106
<b>Lefff</b>	466 668	3 829	470 497	54 214	2 303	56 517
<b>Morphalou</b>	524 179	49	524 228	65 170	7	65 177
<b>GLÀFF</b>	1 401 578	24 270	1 425 848	172 616	13 466	186 082

Taille des nomenclatures

	Lexique	BDLex	Lefff	Morphalou	GLÀFF
<b>Lexique</b>		26,03	25,20	22,46	8,95
<b>BDLex</b>	76,02		79,87	70,40	28,75
<b>Lefff</b>	79,50	86,28		72,32	30,04
<b>Morphalou</b>	79,58	85,43	81,24		32,03
<b>GLÀFF</b>	<b>84,83</b>	<b>93,26</b>	<b>90,23</b>	<b>85,66</b>	

Couverture inter-lexiques

## % Couverture Lexiques/Corpus (« taille utile »)

Seuil : fréquence $\geq$		1	10	100	1000
<b>LM10</b> 1991-2000 (200M mots)	Nb formes	300 606	77 936	29 388	7 838
	Lexique	29,6	76,3	93,8	98,6
	BDLex	37,8	80,9	95,5	98,7
	Lefff	39,6	83,2	96,0	► <b>98,9</b>
	Morphalou	39,1	80,3	93,3	97,5
	GLÀFF	► <b>45,2</b>	► <b>86,2</b>	► <b>96,5</b>	98,7
<b>Wikipédia2008</b> (260M mots)	Nb formes	953 920	136 531	35 621	7 956
	Lexique	9,1	43,0	78,6	95,7
	BDLex	12,3	48,0	79,4	95,3
	Lefff	12,9	49,7	80,6	95,7
	Morphalou	13,1	48,9	78,7	94,2
	GLÀFF	► <b>16,4</b>	► <b>55,5</b>	► <b>83,2</b>	► <b>96,1</b>
<b>FrWaC</b> (1600M mots)	Nb formes	1 624 620	255 718	74 745	22 100
	Lexique	5,8	30,8	66,0	89,5
	BDLex	9,4	37,5	69,6	90,0
	Lefff	9,9	39,2	71,6	91,2
	Morphalou	10,1	38,7	69,4	88,5
	GLÀFF	► <b>13,1</b>	► <b>45,4</b>	► <b>76,4</b>	► <b>92,8</b>



## Couverture : vocabulaires spécifiques

### Exemples de formes spécifiques à GLÀFF et attestées dans LM10

*employabilité* (271), *attractivité* (245), *diabolisation* (254),  
*hébergeur* (171), *homophobie* (162), *brevetabilité* (157),  
*anticorruption* (110), *transversalité* (77), *institutionnellement* (73),  
*indétrônable* (24), *fatwa* (5)

Rappel : année la plus récente de LM10 = 2000

# Transcriptions phonémiques

Ressources		Transcriptions phonémiques		Syllabation Identiques
		Identiques	Comparables	
BDLex	Lexique	58,3%	96,9%	98,9%
GLÀFF	Lexique	79,5%	97,8%	98,5%
GLÀFF	BDLex	61,7%	96,9%	98,3%

Transcriptions *comparables* :

annulation des oppositions entre voyelles moyennes [o]/[ɔ], [e]/[ɛ] et [œ]/[ə]/[ø]

## GLAFFOLI (GLÀFF OnLine Interface)

<http://redac.univ-tlse2.fr/glaffoli/>[\[ Search \]](#) [\[ Preferences \]](#) [\[ About/Help \]](#)

Form	CONTAINS	ill	IPA	REGEX	^[^~]+\$
Lemma	CONTAINS				aeəgəabɔʒeεɛʀɥvʒd
POS	STARTS WITH	Afp	SAMPA	CONTAINS	
<input type="checkbox"/> attested forms only					
					<input type="button" value="Reset"/> <input type="button" value="Search"/>

Showing results 1-100 of 235 &gt;

[\[ Export as CSV \]](#)

Form	POS	Lemma	IPA	SAMPA	Frantext 20 <sup>e</sup>		LM10		FrWaC	
					Form ↓ ↑	Lemma ↓ ↑	Form ↓ ↑	Lemma ↓ ↑	Form ↓ ↑	Lemma ↓ ↑
<b>achilletalonesques</b>	Afpfp	achilletalonesque	a.ʃil.ta.lɔ̃.nɛsk	a.Sil.ta.lO~.nEsk	0 0	0 0	0 0	0 0	0 0	0 0
<b>capillaires</b>	Afpfp	capillaire	ka.pi.lɛʀ	ka.pi.IER	12 0.415	20 0.693	87 0.395	144 0.655	1123 0.895	3019 2.407
<b>capillotractées</b>	Afpfp	capillotracté	ka.pi.lɔ̃.tʀak.te	ka.pi.lO.tRak.te	0 0	0 0	0 0	0 0	2 0.001	11 0.008
<b>baillaire</b>	Afpfs	baillaire	ba.si.lɛʀ	ba.si.IER	1 0.034	2 0.069	2 0.009	2 0.009	36 0.028	44 0.035
<b>ancillaire</b>	Afpfs	ancillaire	ɑ̃.si.lɛʀ	A~.si.IER	10 0.346	25 0.866	10 0.045	25 0.113	66 0.052	128 0.102

## GLAWI : GLÀFF et WiktionaryX

## WiktionaryX

```

<entry pos="N" id="en_N_computer#0">
  <defs>
    <toplevel-def>
      <gloss domain="Computing"> A programmable device that performs
        mathematical calculations and operations [...]</gloss>
    </toplevel-def>
    <toplevel-def>
      <gloss register="Dated">A person employed to perform computations.</gloss>
      <example> Only a few years ago Mr. Powers, an American computer,
        disproved a hypothesis about prime numbers[...] </example>
    </toplevel-def>
  </defs>
  <syn>
    <item target="en_N_processor">processor</item>
    <item target="en_N_machine">machine</item>
  </syn>
  <hypo>
    <item target="en_N_desktop">desktop</item>
    <item target="en_N_laptop">laptop</item>
  </hypo>
  <trans>
    <item lang="ar" target="ar_N_حاسوب">حاسوب</item>
    <item lang="ar" target="ar_N_كمبيوتر">كمبيوتر</item>
    <item lang="bg" target="bg_N_компютър">компютър</item>
    <item lang="hr" target="hr_N_računalo">računalo</item>
  </trans>
</entry>

```

## GLÀFF

affluent | Ncms | affluent | a.fly.ã | a.fly. A  
 affluents | Afpmp | affluent | a.fly.ã | a.fly. A  
 affluents | Ncmp | affluent | a.fly.ã | a.fly. A  
 affluent | Vmip3p- | affluer | a.fly | a.fly

## GLAWI : GLÀFF et WiktionaryX

## WiktionaryX

```

<entry pos="N" id="en_N_computer#0">
  <defs>
    <toplevel-def>
      <gloss domain="Computing"> A programmable device that performs
        mathematical calculations and operations [...]</gloss>
    </toplevel-def>
    <toplevel-def>
      <gloss regis
    </toplevel-def>
    <example>
      disproved
    </toplevel-def>
  </defs>
  <syn>
    <item target="en
    <item target="en
  </syn>
  <hypo>
    <item target="en_t
    <item target="en_N_laptop">laptop</item>
  </hypo>
  <trans>
    <item lang="ar" target="ar_N_حاسوب">حاسوب</item>
    <item lang="ar" target="ar_N_كمبيوتر">كمبيوتر</item>
    <item lang="bg" target="bg_N_компютър">компютър</item>
    <item lang="hr" target="hr_N_računalo">računalo</item>
  </trans>
</entry>

```

## Lemmes uniquement

définitions, relations sémantiques et traductions

→ pas de formes fléchies, pas de prononciations

## GLÀFF

affluent | Ncms | affluent | a.fly.ã | a.fly.A

affluents | Afpmp | afflu

affluents | Ncmp | afflu

affluent | Vmip3p- | affl

## Lemmes + formes fléchies

prononciations

## GLAWI : GLÀFF et WiktionaryX

*“GLÀFF et WiktionaryX”... mais pas seulement*

- une ressource unique unifiant GLÀFF et WiktionaryX
- + relations morphologiques, étymologies...
- analyse plus fine du wikicode pour extraire autant d'information que possible
- en allant plus loin dans la processus de standardisation

Point de départ unique pour concevoir d'autres lexiques sur mesure.

POS	Nb lemmes	Nb formes fléchies
NC	179 340	272 170
NP	57 371	8 019
Adjectifs	56 296	93 295
Verbes	36 928	1 251 809
Adverbes	5 552	
Total	335 487	1 625 293

# Description : structure générale des articles

```
<article>
<title>mousse</title>
<pageId>7930</pageId>
<meta>
  <category>Lexique en français de la navigation</category>
  <category>Noms multigenres en français</category>
  <reference>TLFi</reference>
</meta>
<text>
  <pronunciations>
    <pron area="France">mus</pron>
  </pronunciations>
  <pos type="nom" lemma="1" locution="0" homoNb="1" gender="f" number="s">
    <paradigm>
      <wiki>{{fr-rég|mus}}</wiki>
      <inflection form="mousse" gracePOS="Ncfs" pron="mus"/>
      <inflection form="mousses" gracePOS="Ncfp" pron="mus"/>
    </paradigm>
    ...
  </pos>
  <pos type="nom" lemma="1" locution="0" homoNb="2" gender="m" number="s">
    ...
  </pos>
  <pos type="nom" lemma="1" locution="0" homoNb="3" gender="m" number="s">
    ...
  </pos>
  <pos type="adjectif" lemma="1" locution="0" gender="epicene" number="s">
    ...
  </pos>
  <pos type="verbe" lemma="0" locution="0">
    <inflectionInfos>
      <inflected gracePOS="Vmip1s-" lemma="mousser"/>
      <inflected gracePOS="Vmip3s-" lemma="mousser"/>
      <inflected gracePOS="Vmsp1s-" lemma="mousser"/>
      <inflected gracePOS="Vmsp3s-" lemma="mousser"/>
      <inflected gracePOS="Vmmp2s-" lemma="mousser"/>
    </inflectionInfos>
  </pos>
</text>
</article>
```

## Un format *ad hoc*

- Structure colle à celle du Wiktionnaire
- Conçue pour représenter toute l'information qu'il contient
- Standards existants (RDF/Lemon, TEI P5 chapitre 9, LMF) pas adaptés
- Possibilité d'extraction (partielle) et conversion

## Description : métadonnées

- *mousse*

```
<meta>  
  <category>Lexique en français de la navigation</category>  
  <category>Noms multigenres en français</category>  
  <reference>TLFi</reference>  
</meta>
```

- *nénuphar*

```
<meta>  
  <category>Plantes en français</category>  
  <category>Fleurs en français</category>  
  <import>DAF8</import>  
  <spellingVariation norm="nénufar"/>  
</meta>
```



Description : sections partie du discours : *mousse* Ncfs#1

```
<pos type="nom" lemma=1 loc=0 homoNb=1 gender="f" number="s">  
  <definitions> ... </definitions>  
  <subsection type="translations">  
    <trans lang="de">Moos</trans>  
    <trans lang="en">foam</trans>  
    <trans lang="en">moss</trans>  
    <trans lang="es">espuma</trans>  
  </subsection>  
  <subsection type="lexSemRel">  
    <item type="synonym">écume</item>  
    <item type="synonym">bière</item>  
  </subsection>  
  <subsection type="morphoRel">  
    <item type="derivative">moussant</item>  
    <item type="derivative">mousser</item>  
    <item type="derivative">mousseux</item>  
  </subsection>  
</pos>
```

## Description : traductions

POS	Nb sections		Nb traductions
noms	71 133	40 %	383 612
adjectifs	30 797	12 %	60 360
verbes	11 484	31 %	70 615
adverbes	3 014	54 %	14 478
Total	102 428	37 %	529 065

● *piste cyclable*

```
<translations>
  <trans lang="de">Radweg</trans>
  <trans lang="en">bicycle path</trans>
  <trans lang="it">pista ciclabile</trans>
  <trans lang="it">ciclopista</trans>
  <trans lang="nl">fietspad</trans>
  <trans lang="nb">sykkelvei</trans>
  <trans lang="nn">sykkelveg</trans>
  <trans lang="pt">ciclovia</trans>
  <trans lang="sv">cykelväg</trans>
</translations>
```

## Description : relations sémantiques

POS	Sections	Relations		
noms	31 332	17 %	synonym	46 605
			near-synonym	2 454
			antonym	4 625
			hyperonym	20 093
			hyponym	21 472
			holonym	1 115
			meronym	2 566
			adjectifs	5 613
			near-synonym	833
			antonym	3 858
			hyperonym	483
			hyponym	1 062
			holonym	23
			meronym	34

POS	Sections		Relations	
verbes	5 157	14 %	synonym	9 322
			near-synonym	643
			antonym	1 675
			hyperonym	86
			hyponym	162
			troponym	125
adverbes	1 491	27 %	synonym	3 041
			near-synonym	196
			antonym	494

● *communisme*

```

<subsection type="semRel">
  <item type="synonym">collectivisme</item>
  <item type="synonym">marxisme</item>
  <item type="antonym">capitalisme</item>
  <item type="hyperonym">idéologie</item>
  <item type="hyponym">anarcho-communisme</item>
  <item type="hyponym">bolchévisme</item>
  <item type="hyponym">trotskisme</item>
</subsection>

```

# Relations morphologiques

POS	Nb sections	Relations	
nouns	16 948	compound	1 118
		derivative	50 506
		related	22 874
adjectives	4 939	compound	309
		derivative	9 481
		related	6 767
verbs	5 443	compound	109
		derivative	10 684
		related	5 170
adverbs	899	derivative	488
		related	1 284

- *nom*

```

<subsection type="morpho">
  <item type="compound">nom commun</item>
  <item type="compound">nom collectif</item>
  <item type="compound">prête-nom</item>
  <item type="derivative">nommer</item>
  <item type="derivative">nommage</item>
  <item type="derivative">nomination</item>
  <item type="related">anonyme</item>
</subsection>

```

# Étymologies

85 % des pages consacrées à un lemme contiennent une section étymologique.

- *monoxyde*

```
<etym>
<wiki>{{date|1759}} du {{étyl|grc|μονόξυλος|monoxylos}}
  {{cf|mono-|-xyle|lang=fr}}.</wiki>
<xml> <date>1759</date> du grec ancien
  <foreignWord lang="grec ancien" translit="monoxylos">μονόξυλος</foreignWord>
  <cf value="mono-|-xyle" lang="fr"/>.</xml>
<txt>du grec ancien μονόξυλος voir mono- et -xyle.</txt>
</etym>
```

- *pillaver*

```
<etym>
<wiki>du {{étyl|rom|fr|mot=piyav|sens=boire}}.</wiki>
<xml>du romani
  <foreignWord lang="romani" sense="boire">piyav</foreignWord>
</xml>
<txt>du romani piyav.</txt>
</etym>
```

# Transcriptions phonémiques

94 % des entrées ont 1 ou plusieurs transcriptions  
variations diatopiques éventuelles

- *moins*

```
<pronunciations>  
  <pron area="Marseille">mwẽs</pron>  
  <pron area="Paris">mwẽ</pron>  
</pronunciations>
```

- *ananas*

```
<pronunciations>  
  <pron area="France">a.na.na</pron>  
  <pron area="France">a.na.nas</pron>  
  <pron area="Canada">a.na.nɑ</pron>  
  <pron area="Canada">a.na.nɔ</pron>  
</pronunciations>
```

## Description : définitions

mousse, Ncfs, homographe n° 1, sens n° 5

```

<pos type="nom" lemma=1 loc=0 homoNb=1 gender="f" number="s">
  <definitions>
    <definition>
      <gloss>
        <labels>
          <label type="sem" value="métonymie"/>
          <label type="attitudinal" value="familier"/>
        </labels>
        <wiki>{{méton|fr}} {{familier|fr}} [[bière|Bière]].</wiki>
        <txt>Bière.</txt>
        <xml><innerLink ref="bière">Bière</innerLink>.</xml>
        <conll>1 Bière bière NC nc n=s 0 root</conll>
      </gloss>
      <example>
        <wiki>''Une bonne ''mousse'' sans faux-col est un oxymore.''</wiki>
        <xml><i>Une bonne <b>mousse</b> sans faux-col est un oxymore.</i></xml>
        <txt>Une bonne mousse sans faux-col est un oxymore.</txt>
        <conll> 1 Une      une      DET  DET  g=f|n=s      3      det
              2 bonne   bon     ADJ  adj  g=f|n=s      3      mod
              3 mousse  mousse NC   nc   n=s          6      suj
              4 sans    sans    P     P     -            3      dep
              5 faux-col _       NC   -     -            4      prep
              6 est     être    V     v     n=s|p=3|t=pst 0      root
              7 un     un      DET  DET  g=m|n=s      8      det
              8 oxymore _       NC   -     -            6      ats</conll>
      </example>
    </definition>
  </definitions>
</pos>

```

# Du Wiktionnaire à GLAWI : processus d'extraction

≠ seulement délimiter le contenu par des balises

## Les informations du Wiktionnaire peuvent

- être lacunaires
- être dupliquées (plusieurs pages/endroits différents d'une même page)
- apparaître sous plusieurs formes :
  - patrons prédéfinis
  - alias
  - texte saisi « en dur » par les contributeurs

## Contribution de GLAWI

- inférer l'information manquante
- vérifier la cohérence des informations dupliquées et éparpillées
- normaliser les informations polymorphes



# Marques lexicographiques

## Intérêt de traiter correctement ces marques

- les supprimer des définitions
- les encoder formellement  
(pour permettre des recherches ciblant une marque donnée)

## Détection : nécessité d'un inventaire exhaustif

- Wiktionary anglais : un seul patron (label)  
`{{label|dated}}`, `{{label|transitive}}`, `{{label|oenology}}`
- Wiktionnaire : `{{désuet}}`, `{{transitif}}`, `{{oenologie}}`  
`{{oenologie}}`, `{{œnologie}}`, `{{oenol}}`, `{{œnol}}`, (oenologie)  
5 alias dans le Wiktionnaire → une seule valeur dans GLAWI
- Plus de 6 000 marques (patrons et alias) inventoriées

# Marques lexicographiques

<b>Diafrequential</b>	<b>6 166</b>	<b>Diatopic</b>	<b>8 726</b>
rare	4 215	Québec	1 717
extrêmement rare	1 016	France	1 138
très rare	301	Canada	971
plus courant	190	Suisse	962
courant	186	Belgique	637
plus rare	176	Lorraine	299
moins courant	62	Occitanie	246
peu usité	20	Normandie	134
		Provence	123
		Acadie	122
<b>Diachronic</b>	<b>24 450</b>	Louisiane	90
vieilli	9 431	Réunion	89
désuet	6 043	Afrique	64
avant 1835	1 654	Congo-Kinshasa	47
néologisme	820	Ardennes	46
archaïque	661	Languedoc-Roussillon	44
1986	73	Bretagne	40
1990	72		
766 autres années	5 841	362 autres régions	1 957

# Marques lexicographiques

Loanwords	1 493	Domains	155 532	Attitudinal	17 340
anglicisme	1 446	localités	49 060	familier	8333
indo-européen commun	22	géographie	11 935	argot	2166
hispanisme	11	botanique	6 461	populaire	1870
germanisme	7	zoologie	5 460	péjoratif	1587
gaulois	4	médecine	5 258	vulgaire	770
catalan	3	chimie	3 358	littéraire	513
		histoire	2 804	ironique	450
		marine	2 644	plaisanterie	345
<b>Semantics</b>	<b>23 860</b>	religion	2 559	injurieux	258
figuré	10 859	linguistique	2 177	exagération	254
par extension	6 666	agriculture	2 071	soutenu	247
en particulier	2 574	anatomie	2 005	poétique	183
analogie	1 213	informatique	1 718	verlan	128
métonymie	886	droit	1 698	enfantin	81
ellipse	793	physique	1 579	euphémisme	66
spécialement	704	militaire	1 572	très familier	65
métaphore	75	musique	1 570	informel	13
hyperbole	30	minéralogie	1 531	dérision	13
apocope	24	biologie	1 515	mélioratif	1
généralement	19	...			
litote	10	littérature	299		
figure	7	367 autres domaines	45 946		

# Marques lexicographiques

Loanwords	1 493	Domains	155 532	Attitudinal	17 340
anglicisme	1 446	localités	49 060	familier	8333
indo-européen commun	22	géographie	11 935	argot	2166
hispanisme	11	botanique	6 461	populaire	1870
germanisme	7	zoologie	5 460	péjoratif	1587
gaulois	4	médecine	5 258	vulgaire	770
catalan	3	chimie	3 358	<b>littéraire</b>	513
		histoire	2 804	ironique	450
		marine	2 644	plaisanterie	345
<b>Semantics</b>	<b>23 860</b>	religion	2 559	injurieux	258
figuré	10 859	linguistique	2 177	exagération	254
par extension	6 666	agriculture	2 071	soutenu	247
en particulier	2 574	anatomie	2 005	poétique	183
analogie	1 213	informatique	1 718	verlan	128
métonymie	886	droit	1 698	enfantin	81
ellipse	793	physique	1 579	euphémisme	66
spécialement	704	militaire	1 572	très familier	65
métaphore	75	musique	1 570	informel	13
hyperbole	30	minéralogie	1 531	dérision	13
apocope	24	biologie	1 515	mélioratif	1
généralement	19	...			
litote	10	<b>littérature</b>	299		
figure	7	367 autres domaines	45 946		

# Marques lexicographiques

Loanwords	1 493	Domains	155 532	Attitudinal	17 340
anglicisme	1 446	localités	49 060	familier	8333
indo-européen commun	22	géographie	11 935	argot	2166
hispanisme	11	botanique	6 461	populaire	1870
germanisme	7	zoologie	5 460	péjoratif	1587
gaulois	4	médecine	5 258	vulgaire	770
catalan	3	chimie	3 358	<b>littéraire</b>	513
		histoire	2 804	ironique	450
		marine	2 644	plaisanterie	345

## Le Petit Robert

**LITTÉR.** 1° terme des études littéraires (HIST. LITTÉR. : terme d'histoire littéraire); 2° littéraire : désigne un mot qui n'est pas d'usage familier, qui s'emploie surtout dans la langue écrite élégante. Ce mot a généralement des synonymes d'emploi plus courant

métaphore	75	musique	1 570	informel	13
hyperbole	30	minéralogie	1 531	dérision	13
apocope	24	biologie	1 515	mélioratif	1
généralement	19	...			
litote	10	<b>littérature</b>	299		
figure	7	367 autres domaines	45 946		

## Wiktionary

- littérature : *autofiction, dickensien, parallélisme, paysagisme...*
- littéraire : *heur, inobédient, jaculatoire, pusillanime...*

# Marques lexicographiques

## Normaliser n'est pas corriger

- questions :
  - *populaire* est-elle une marque (encore) pertinente ?
  - y a-t-il une différence entre *désuet* et *vieilli* ?
  - entre *littéraire* et *soutenu*, *familier* et *informel* ?

→ questions posées par des méta-lexicographes [Baider et al., 2011], non tranchées par les wiktionnaristes
- Wiktionnaire : pas de « classe » de marques dans le code, mais regroupement en catégories dans les pages d'aide.  
Même catégories (« registre de langue ») pour les emprunts, les marques diachroniques et diafréquentielles ; catégories différentes pour *familier/informel*.
- GLAWI : classement par catégories cohérentes.  
Couples {valeur → catégorie} manuellement établis  
catégories utilisées : *attitudinal*, *diachronic*, *diafrequential*, *diatopic*, *domain*, *grammar*, *loan*, *semantics*, *other*

# Paradigmes flexionnels

**affluent**

**Adjectif**

affluent

- (Géographie) Qui se jette dans un autre en parlant d'un cours d'eau.
- (Médecine) Qui **affluent**, qui se portent en abondance vers quelque partie du corps.

	Singulier	Pluriel
Masculin	affluent <span>/a.fy.ɑ̃/</span>	affluents <span>/a.fy.ɑ̃/</span>
Féminin	affluente <span>/a.fy.ɑ̃t/</span>	affluentes <span>/a.fy.ɑ̃t/</span>

**Nom commun**

affluent /a.fy.ɑ̃/ masculin

- (Géographie) Cours d'eau qui se jette dans un autre.

	Singulier	Pluriel
	affluent	affluents
	<span>/a.fy.ɑ̃/</span>	

**Forme de verbe**

affluent /a.fy.ɑ̃/

- Troisième personne du pluriel de l'indicatif présent de **affluer**.
- Troisième personne du pluriel du subjonctif présent de **affluer**.

Conjugaison du verbe affluer

INDICATIF	Présent	ils/elle
SUBJONCTIF	Présent	qu'ils/elle

Modes impersonnels

Mode	Présent	Passé
Infinitif	affluer <span>/a.fy.ɑ̃/</span>	avoir afflué <span>/a.va.ʁ.a.fy.ɑ̃/</span>
Gérondif	en affluant <span>/ɑ̃.a.fy.ɑ̃/</span>	en ayant afflué <span>/ɑ̃.a.ʁ.a.fy.ɑ̃/</span>
Participe	affluant <span>/a.fy.ɑ̃/</span>	afflué <span>/a.fy.ɑ̃/</span>

Indicatif

	Présent	Passé composé
j'afflue	<span>/ʁ.a.fy.ɑ̃/</span>	j'ai afflué <span>/ʁ.a.ʁ.a.fy.ɑ̃/</span>
tu afflues	<span>/tʁ.a.fy.ɑ̃/</span>	tu as afflué <span>/tʁ.a.ʁ.a.fy.ɑ̃/</span>
il/elle/on afflue	<span>/il.ʁ.a.fy.ɑ̃/</span>	il/elle/on a afflué <span>/il.ʁ.a.ʁ.a.fy.ɑ̃/</span>
nous affluons	<span>/nu.ʁ.a.fy.ɑ̃/</span>	nous avons afflué <span>/nu.ʁ.a.ʁ.a.fy.ɑ̃/</span>
vous affluez	<span>/vu.ʁ.a.fy.ɑ̃/</span>	vous avez afflué <span>/vu.ʁ.a.ʁ.a.fy.ɑ̃/</span>
ils/elles affluent	<span>/il.ʁ.a.fy.ɑ̃/</span>	ils/elles ont afflué <span>/il.ʁ.a.ʁ.a.fy.ɑ̃/</span>
	Imparfait	Plus-que-parfait
j'affluais	<span>/ʁ.a.fy.ɑ̃/</span>	j'avais afflué <span>/ʁ.a.vɛ.ʁ.a.fy.ɑ̃/</span>
tu affluais	<span>/tʁ.a.fy.ɑ̃/</span>	tu avais afflué <span>/tʁ.a.vɛ.ʁ.a.fy.ɑ̃/</span>
il/elle/on affluait	<span>/il.ʁ.a.fy.ɑ̃/</span>	il/elle/on avait afflué <span>/il.ʁ.a.vɛ.ʁ.a.fy.ɑ̃/</span>
nous affluions	<span>/nu.ʁ.a.fy.ɑ̃/</span>	nous avions afflué <span>/nu.ʁ.a.vɛ.ʁ.a.fy.ɑ̃/</span>
vous affluiez	<span>/vu.ʁ.a.fy.ɑ̃/</span>	vous aviez afflué <span>/vu.ʁ.a.vɛ.ʁ.a.fy.ɑ̃/</span>
ils/elles affluaient	<span>/il.ʁ.a.fy.ɑ̃/</span>	ils/elles avaient afflué <span>/il.ʁ.a.vɛ.ʁ.a.fy.ɑ̃/</span>

**Verbe**

affluer /a.fy.ɑ̃/ intransitif 1<sup>er</sup> groupe (conjugaison)

- Couler vers, en parlant des eaux qui se rendent et se réunissent dans un même lit.
  - Plusieurs ruisseaux et plusieurs rivières **affluent** dans la Seine, dans le Rhône.
- (Par analogie) Couler vers, en parlant des humeur du corps
  - Il faut empêcher le sang d'**affluer** vers telle partie.
- (Figuré) Abonder, arriver en abondance.

**affluente**

**Forme d'adjectif**




affluente féminin /a.fy.ɑ̃t/

- Féminin singulier de **affluent**.

Information potentiellement lacunaire, redondante (cohérente ou non), éparpillée. . .

# Paradigmes flexionnels : exemples de traitements

Inférer l'information manquante :

<h2>arrivages</h2> <p><b>Français</b> [ modifier   modifier le wikicode ]</p> <p> <b>Forme de nom commun</b></p> <p>arrivages <span><span><span></span><sup>a</sup></span><span><span></span><sup>li</sup></span><span><span></span><sup>vaʒ</sup></span><span><span></span><sup>\</sup></span></span></p> <p>1. Pluriel de <a href="#">arrivage</a>.</p>	<h2>arrivage</h2> <p><b>Français</b> [ modifier   modifier le wikicode ]</p> <p> <b>Étymologie</b> [ modifier   modifier le wikicode ]</p> <p>→ voir <i>arriver</i> et <i>-age</i>.</p> <p> <b>Nom commun</b> [ modifier   modifier le wikicode ]</p> <p>arrivage <span><span><span></span><sup>a</sup></span><span><span></span><sup>li</sup></span><span><span></span><sup>vaʒ</sup></span><span><span></span><sup>\</sup></span></span> <i>masculin</i></p>
--	--

*lemma* (arrow from [arrivage](#) to [arrivages](#))

*gender* (arrow from [arrivage](#) to [arrivages](#))

Détecter les incohérences :

## clavardeuses

**Français** [ modifier | modifier le wikicode ]

 **Forme de nom commun**

clavardeuses *masculin*

1. Féminin pluriel de [clavardeur](#).



# Paradigmes flexionnels : limites du traitement automatique

## Paradigmes verbaux : cas indécidables

- paradigmes incomplets : flexions manquantes dans Wiktionnaire ou défection réelle ?
- plusieurs formes pour une même flexion : incohérence ou surabondance ?
  - *payer*, Vmip1s- : paie, paye
  - *contredire*, Vmmp2p- : contredisez, contredites
  - *végéter*, Vmif1s- : végéterai, végèterai

# Paradigmes flexionnels : limites du traitement automatique

## Paradigmes verbaux : cas indécidables

- paradigmes incomplets : flexions manquantes dans Wiktionnaire ou défection réelle ?
- plusieurs formes pour une même flexion : incohérence ou surabondance ?
  - *payer*, Vmip1s- : **paie**, **paye**  
→ *surabondance effective*
  - *contredire*, Vmmp2p- : contredisez, contredites
  
  - *végéter*, Vmif1s- : végéterai, végèterai

# Paradigmes flexionnels : limites du traitement automatique

## Paradigmes verbaux : cas indécidables

- paradigmes incomplets : flexions manquantes dans Wiktionnaire ou défection réelle ?
- plusieurs formes pour une même flexion : incohérence ou surabondance ?
  - *payer*, Vmip1s- : paie, paye  
→ *surabondance effective*
  - *contredire*, Vmmp2p- : contredisez, contredites  
→ *forme erronée trouvée dans le tableau de conjugaison*
  - *végéter*, Vmif1s- : végéterai, végèterai

# Paradigmes flexionnels : limites du traitement automatique

## Paradigmes verbaux : cas indécidables

- paradigmes incomplets : flexions manquantes dans Wiktionnaire ou défection réelle ?
- plusieurs formes pour une même flexion : incohérence ou surabondance ?
  - payer, Vmip1s- : paie, paye  
→ *surabondance effective*
  - contredire, Vmmp2p- : contredisez, contredites  
→ *forme erronée trouvée dans le tableau de conjugaison*
  - végéter, Vmif1s- : végéterai, végèterai  
→ *réforme de l'orthographe de 1976*

# Paradigmes flexionnels : limites du traitement automatique

## Paradigmes verbaux : cas indécidables

- paradigmes incomplets : flexions manquantes dans Wiktionnaire ou défection réelle ?
- plusieurs formes pour une même flexion : incohérence ou surabondance ?
  - *payer*, Vmip1s- : **paie**, **paye**  
→ *surabondance effective*
  - *contredire*, Vmmp2p- : **contredisez**, **contredites**  
→ *forme erronée trouvée dans le tableau de conjugaison*
  - *végéter*, Vmif1s- : **végéterai**, **vègèterai**  
→ *réforme de l'orthographe de 1976*

## Traitement possible ?

- utilisation d'inventaires des verbes défectifs et surabondants
- *payer* → *entrepayer*, *prépayer*, *surpayer*
  - *pleuvoir* → *pleuvasser*, *pleuvigner*, *pleuviner*, *pleuvioter*, *pleuvocher*, *pleuvoter*, *repleuviner*, *repleuvioter*, *repleuvoir*, *repleuvoter*

# Autour de GLÀFF et de GLAWI

## Autres réalisations

- PsychoGLÀFF [Calderone et al., 2014]  
[redac.univ-tlse2.fr/lexicons/psychoglaaff\\_en.html](http://redac.univ-tlse2.fr/lexicons/psychoglaaff_en.html)
- Extraction automatique de relations morphologiques [Hathout et al., 2014a]  
(170 000 de relations extraites, exactitude de 94 à 99%)
- Enrichissement de la base NHUMA [Flaux et al., 2014]  
Noms d'humains créateurs (*symphoniste, sculpteur, romancier*). Exploitation des définitions de GLAWI → +15%
- Autres lexiques spécifiques extraits à la demande : emprunts, équivalents masculin/féminin, mots désuets, néologismes, gentilés, lexiques de domaines spécialisés, etc.
- Normalisation de graphies : étiquetage, RI, etc.

## Normalisation de graphies

Forme	Autre forme/forme std	Type
nénuphar maîtriser	nénufar maïtriser	variante orthographique correcte (avant/après réforme de 1990)
quinquenat	quinquennat	faute courante
enkikiner	enquiquiner	forme consciemment « dégradée » (texto/forum)
c'qui	ce qui	oralisation
coeur	cœur	ligature
VOIP	VoIP	casse des sigles
abreusement	abreuvement	concurrence suffixale
écart type copier-coller	écart-type copier/coller	composé
multi-joueur	multijoueur	dérivation
climato-sceptique graticiel	climatosceptique grat uiciel	mot valise
débit de boisson	débit de boissons	composé pluriel
coulibiac halal mozzarella chai clubbeur	koulibiak hallal mozzarella tchaï clubber	emprunt
erratums	errata	pluriel de singulatif
& al.	et al.	symbole/littéral
N'Djaména	Ndjamenana	NP

# Conclusion

## GLAWI : un dictionnaire électronique du français en XML

- une couverture sans précédent, des mises à jour régulières
- un point de départ pour des observations linguistiques et la création de nouveaux lexiques (rendues possibles par le format exploitable de GLAWI)
- une licence libre!!!

## Ressources lexicales : évolutions du contexte

- Une tendance à l'alignement des ressources
  - uniquement possible entre ressources... libres!
- Vers une généralisation de la libération des ressources lexicales?
  - Flexique [Bonami et al., 2014]
  - Morphalou 3 (Morphalou 2 + Lefff + Dela + ...)



# Conclusion

## GLAWI : un dictionnaire électronique du français en XML

- une couverture sans précédent, des mises à jour régulières
- un point de départ pour des observations linguistiques et la création de nouveaux lexiques (rendues possibles par le format exploitable de GLAWI)
- une licence libre !!!

## Ressources lexicales : évolutions du contexte

- Une tendance à l'alignement des ressources
  - uniquement possible entre ressources... libres !
- Vers une généralisation de la libération des ressources lexicales ?
  - Flexique [Bonami et al., 2014]
  - Morphalou 3 (Morphalou 2 + Lefff + Dela + ...)

GLAWI = seule ressource libre contenant des définitions en français

# Perspectives

## À venir prochainement

- autres langues (versions anglaise, espagnole et italienne de GLAWI en préparation)
- mise à jour de GLÀFF à partir de GLAWI  
→ GLÀFF, le retour... (du titre)

# Perspectives

## À venir prochainement

- autres langues (versions anglaise, espagnole et italienne de GLAWI en préparation)
- mise à jour de GLÀFF à partir de GLAWI  
→ GLÀFF, le retour... (du titre)

# Perspectives

## À venir prochainement

- autres langues (versions anglaise, espagnole et italienne de GLAWI en préparation)
- mise à jour de GLÀFF à partir de GLAWI  
→ GLÀFF, le retour... (du titre)

## Projets

- mesure de l'impact de l'intégration de GLÀFF dans un étiqueteur morphosyntaxique sur les performances de celui-ci
- développement d'une interface graphique pour GLAWI

# Perspectives

## À venir prochainement

- autres langues (versions anglaise, espagnole et italienne de GLAWI en préparation)
- mise à jour de GLÀFF à partir de GLAWI  
→ GLÀFF, le retour... (du titre)

## Projets

- mesure de l'impact de l'intégration de GLÀFF dans un étiqueteur morphosyntaxique sur les performances de celui-ci
- développement d'une interface graphique pour GLAWI

# Perspectives

## À venir prochainement

- autres langues (versions anglaise, espagnole et italienne de GLAWI en préparation)
- mise à jour de GLÀFF à partir de GLAWI  
→ GLÀFF, le retour... (du titre)



## Projets

- mesure de l'impact de l'intégration de GLÀFF dans un étiqueteur morphosyntaxique sur les performances de celui-ci
- développement d'une interface graphique pour GLAWI

# Perspectives

## À venir prochainement

- autres langues (versions anglaise, espagnole et italienne de GLAWI en préparation)
- mise à jour de GLÀFF à partir de GLAWI  
→ GLÀFF, le retour... (du titre)



## Projets

- mesure de l'impact de l'intégration de GLÀFF dans un étiqueteur morphosyntaxique sur les performances de celui-ci
- développement d'une interface graphique pour GLAWI



# Perspectives

## À venir prochainement

- autres langues (versions anglaise, espagnole et italienne de GLAWI en préparation)
- mise à jour de GLÀFF à partir de GLAWI  
→ GLÀFF, le retour... (du titre)






## Projets

- mesure de l'impact de l'intégration de GLÀFF dans un étiqueteur morphosyntaxique sur les performances de celui-ci
- développement d'une interface graphique pour GLAWI







## Références |

-  Baider, F., Lamprou, E., and Monville-Burston, M. (2011). *La marque en lexicographie : états présents, voies d'avenir*. Collection "La lexicothèque.". Lambert-Lucas.
-  Bonami, O., Caron, G., and Plancq, C. (2014). Construction d'un lexique flexionnel phonétisé libre du français. In *Actes du 4<sup>e</sup> Congrès Mondial de Linguistique Française (CMLF 2014)*, pages 2583–2596, Berlin.
-  Boula De Mareuil, P., Yvon, F., D'Alessandro, C., Aubergé, V., Vaissière, J., and Amelot, A. (2000). A French Phonetic Lexicon with variants for Speech and Language Processing. In *Proc. of the 2nd Intl Conference on Language Resources and Evaluation (LREC)*, pages 273–276.



## Références II

-  Calderone, B., Hathout, N., and Sajous, F. (2014).  
From GLÀFF to PsychoGLÀFF : a large psycholinguistics-oriented French lexical resource.  
*In Proceedings of the 16th EURALEX International Conference (EURALEX 2014)*, Bolzano, Italy.
-  Clément, L., Lang, B., and Sagot, B. (2004).  
Morphology based automatic acquisition of large-coverage lexica.  
*In Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)*, pages 1841–1844, Lisboa, Portugal.



## Références III

-  Dal, G. and Namer, F. (2012).  
Faut-il brûler les dictionnaires ? Ou comment les ressources numériques ont révolutionné les recherches en morphologie.  
*In Actes du 3eme Congrès Mondial de Linguistique Française, Lyon, volume 1, pages 1261–1276.*
-  Eckard, E., Barque, L., Nasr, A., and Sagot, B. (2012).  
Dictionary-Ontology Cross-Enrichment. Using TLFi and WOLF to enrich one another.  
*In COLING Workshop on Cognitive Aspects of the Lexicon, pages 81–93, Mumbai.*

## Références IV

-  Flaux, N., Lagae, V., and Stosic, D. (2014).  
Romancier, symphoniste, sculpteur : les noms d'humains  
créateurs d'objets idéaux.  
*In Actes du 4eme Congrès Mondial de Linguistique Française,  
Berlin.*
-  Hathout, N., Sajous, F., and Calderone, B. (2014a).  
Acquisition and enrichment of morphological and  
morphosemantic knowledge from the French Wiktionary.  
*In Proceedings of the COLING Workshop on Lexical and  
Grammatical Resources for Language Processing, pages 65–74,  
Dublin, Ireland. Association for Computational Linguistics and  
Dublin City University.*

## Références V

-  Hathout, N., Sajous, F., and Calderone, B. (2014b).  
GLÀFF, a Large Versatile French Lexicon.  
In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland.
-  Hellmann, S., Brekle, J., and Auer, S. (2013).  
Leveraging the crowdsourcing of lexical resources for bootstrapping a linguistic data cloud.  
In Takeda, H., Qu, Y., Mizoguchi, R., and Kitamura, Y., editors, *Semantic Technology*, volume 7774 of *Lecture Notes in Computer Science*, pages 191–206. Springer Berlin Heidelberg.

## Références VI

-  Meyer, C. M. (2013).  
*Wiktionary : The Metalexicographic and the Natural Language Processing Perspective.*  
PhD thesis, Technische Universität Darmstadt.
-  Navarro, E., Sajous, F., Gaume, B., Prévot, L., Hsieh, S., Kuo, I., Magistry, P., and Huang, C.-R. (2009).  
Wiktionary and NLP : Improving synonymy networks.  
In *Proceedings of the 2009 ACL-IJCNLP Workshop on The People's Web Meets NLP : Collaboratively Constructed Semantic Resources*, pages 19–27, Suntec, Singapore.  
Association for Computational Linguistics.

## Références VII



New, B. (2006).

Lexique 3 : Une nouvelle base de données lexicales.

In *Verbum ex machina. Actes de la 13ème conférence sur le Traitement automatique des langues naturelles*,  
Louvain-la-Neuve.




Penta, D. J. (2011).

The wiki-fication of the dictionary : defining lexicography in the digital age.

In *Proceedings of the MiT7 Conference "unstable platforms : the promise and peril of transition"*, Cambridge, Massachusetts.

## Références VIII



-  Pérennou, G. and de Calmès, M. (1987).  
BDLEX lexical data and knowledge base of spoken and written French.  
In *Proceedings of the European Conference on Speech Technology, ECST 1987*, pages 1393–1396, Edinburgh, Scotland, UK.
-  Romary, L., Salmon-Alt, S., and Francopoulo, G. (2004).  
Standards going concrete : from LMF to Morphalou.  
In Zock, M. and Saint-Dizier, P., editors, *COLING 2004 Enhancing and using electronic dictionaries*, pages 22–28, Geneva. COLING.






## Références IX

-  Rundell, M. and Kilgarriff, A. (2011).  
Automating the creation of dictionaries : Where will it all end ?  
In Meunier, F., De Cock, S., Gilquin, G., and Paquot, M.,  
editors, *A Taste for Corpora. In honour of Sylviane Granger*,  
pages 257–282. John Benjamins.
-  Sagot, B. and Fišer, D. (2008).  
Construction d'un wordnet libre du français à partir de  
ressources multilingues.  
In *Actes de TALN'2008*, Avignon.
-  Sajous, F., Hathout, N., and Calderone, B. (2013a).  
GLÀFF, un Gros Lexique À tout Faire du Français.  
In *Actes de la 20e conférence sur le Traitement Automatique  
des Langues Naturelles (TALN'2013)*, pages 285–298, Les  
Sables d'Olonne, France.

## Références X

-  Sajous, F., Hathout, N., and Calderone, B. (2014).  
Ne jetons pas le Wiktionnaire avec l'oripeau du web ! Études et réalisations fondées sur le dictionnaire collaboratif.  
*In Actes du 4e Congrès Mondial de Linguistique Française (CMLF 2014)*, pages 663–680, Berlin.
-  Sajous, F., Navarro, E., Gaume, B., Prévot, L., and Chudy, Y. (2013b).  
Semi-automatic Enrichment of Crowdsourced Synonymy Networks : the WISIGOTH System Applied to Wiktionary.  
*Language Resources and Evaluation, special issue on Collaboratively Constructed Language Resources*, pages 1–34.

## Références XI

-  Sajous, F., Navarro, E., Gaume, B., Prévot, L., and Chudy, Y. (2010).  
Semi-automatic Endogenous Enrichment of Collaboratively Constructed Lexical Resources : Piggybacking onto Wiktionary.  
In Loftsson, H., Rögnvaldsson, E., and Helgadóttir, S., editors, *Advances in Natural Language Processing*, volume 6233 of *LNCS*, pages 332–344. Springer Berlin / Heidelberg.
-  Sekine, S. (2010).  
We Desperately Need Linguistic Resources! –Based on the Users' Point of View.
-  Zesch, T. and Gurevych, I. (2010).  
Wisdom of Crowds versus Wisdom of Linguists - Measuring the Semantic Relatedness of Words.  
*Journal of Natural Language Engineering.*, 16(01) :25–59.

## Références XII

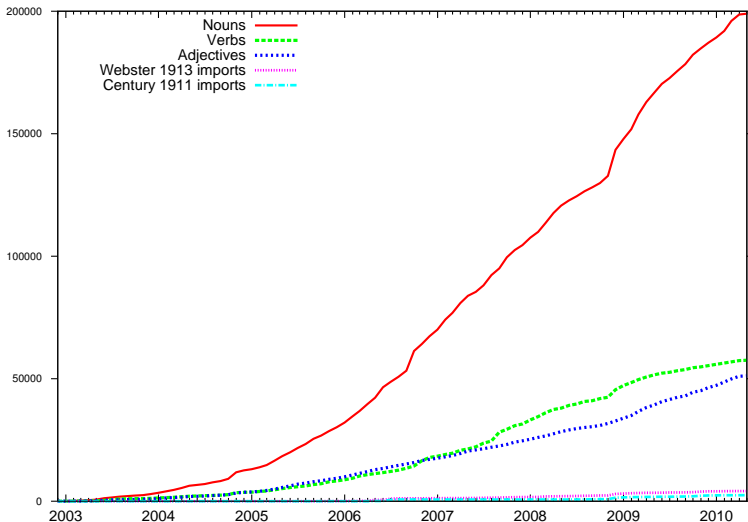


Zesch, T., Müller, C., and Gurevych, I. (2008).

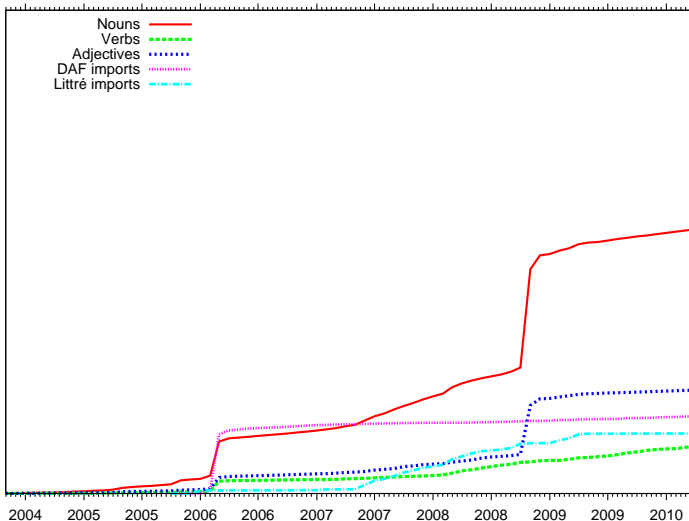
Extracting Lexical Semantic Knowledge from Wikipedia and Wiktionary.

In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech.

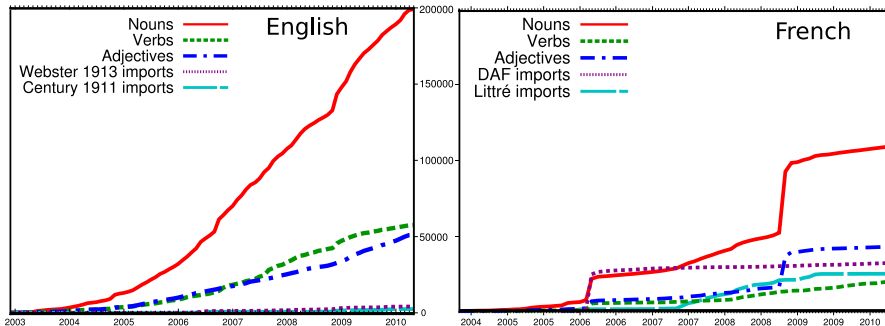
# Imports in English Wiktionary (2003-2010)



# Imports in Wiktionnaire (2003-2010)



# Imports in French and English Wiktionary



# Phonemic transcriptions

*plus* :

```
<text>
  <pos type="adverbe" lemma=1 loc=0 homoNb=1 gracePOS="Rgp">
    <pronunciations>
      <pron>ply</pron>
      <pron>plys</pron>
    </pronunciations>
  </pos>
  <pos type="adverbe" lemma="1" locution="0" homoNb="2" gracePOS="Rgp">
    <pronunciations>
      <pron>ply</pron>
    </pronunciations>
  </pos>
</text>
```

*moins* :

```
<text>
  <pronunciations>
    <pron region="Marseille, Haut Languedoc">mwēs</pron>
    <pron region="Paris, France">mwē</pron>
  </pronunciations>
</text>
```



# Transcriptions phonémiques

10 différences de transcription les plus fréquentes

BDLex / Lexique			GLÀFF/ Lexique			GLÀFF/ BDLex		
≠	%	cumul	≠	%	cumul	≠	%	cumul
ɛ/e	48,2	48,2	ɔ/o	60,0	60,0	e/ɛ	66,5	66,5
ɔ/o	32,2	80,4	ə	14,2	74,2	ɔ/o	10,6	77,1
o/ɔ	11,0	91,4	e/ɛ	6,9	81,1	ə	5,9	83,0
y/ɥ	1,8	93,2	ɛ/e	5,0	86,1	o/ɔ	4,4	87,4
ə/ø	1,4	94,6	ɑ/a	4,9	91,0	ɑ/a	3,8	91,2
ə/œ	1,4	96,0	s/z	1,3	92,3	ɥ/y	1,6	92,8
u/w	0,8	96,9	ə/ø	0,9	93,2	œ/ə	1,1	93,9
b/p	0,7	97,6	œ/ø	0,5	93,7	ø/ə	0,9	94,8
s/z	0,5	98,1	i	0,4	94,1	i	0,8	95,6
j	0,3	98,4	o/ɔ	0,4	94,5	w/u	0,8	96,4

90% des divergences : voyelles moyennes et schwa

## Exemples de transcriptions divergentes

≠	Forme	Transcriptions			
		BDLex	Lexique	GLÀFF	DPF
ɛ/e	été	/ɛ.te/	/e.te/	/e.te/	/ete/
s/z	stalinisme	/sta.li.nism/	/sta.li.nizm/	/sta.li.nism/	/stalinism/, /stalinizm/
b/p	obturer	/ɔb.ty.ʁe/	/ɔp.ty.ʁe/	/ɔp.ty.ʁe/	/ɔptyʁe/, /ɔbtyʁe/
o/ɔ	pomme	/pom/	/pɔm/	/pɔm/	/pɔm/
ə/ø/œ	heureux	/ə.ʁø/	/ø.ʁø/	/œ.ʁø/	/øʁø, œʁø/
y/ɥ	gradué	/gʁa.dy.e/	/gʁa.dɥe/	/gʁa.dɥe/	/gradɥe/, /gradɥe/, /gradye/
u/w	jouer inouï	/ʒu.e/ /i.nu.i/	/ʒwe/ /i.nwi/	/ʒwe/ /i.nwi/	/ʒwe/, /ʒue/ /inwi/, /inui/
a/ɑ	pâte	/pat/	/pat/	/pat/	/pat/, /pat/
+i, -i, +j, -j	riiez	/ʁi.i.je/	/ʁi.je/	/ʁij.je/	-
+ə, -ə	contenu	/kɔ̃.tə.ny/	/kɔ̃.tə.ny/	/kɔ̃t.ny/	/kɔ̃t(ə)ny/

## Exemples de transcriptions divergentes

≠	Forme	Transcriptions			
		BDLex	Lexique	GLÀFF	DPF
ɛ/e	été	/ɛ.te/	/e.te/	/e.te/	/ete/
s/z	stalinisme	/sta.li.nism/	/sta.li.nizm/	/sta.li.nism/	/stalinism/, /stalinizm/
b/p	obturer	/ɔb.ty.ʁe/	/ɔp.ty.ʁe/	/ɔp.ty.ʁe/	/ɔptyre/, /ɔbtyre/
o/ɔ	pomme	/pom/	/pɔm/	/pɔm/	/pɔm/
ə/ø/œ	heureux	/ə.ʁø/	/ø.ʁø/	/œ.ʁø/	/øʁø, œʁø/
y/ɥ	gradué	/gʁa.dy.e/	/gʁa.dɥe/	/gʁa.dɥe/	/gradɥe/, /gradɥe/, /gradye/
u/w	jouer inouï	/ʒu.e/ /i.nu.i/	/ʒwe/ /i.nwi/	/ʒwe/ /i.nwi/	/ʒwe/, /ʒue/ /inwi/, /inui/
a/ɑ	pâte	/pat/	/pat/	/pat/	/pat/, /pat/
+i, -i, +j, -j	riiez	/ʁi.i.je/	/ʁi.je/	/ʁij.je/	-
+ə, -ə	contenu	/kɔ̃.tə.ny/	/kɔ̃.tə.ny/	/kɔ̃t.ny/	/kɔ̃t(ə)ny/

## Exemples de transcriptions divergentes

≠	Forme	Transcriptions			
		BDLex	Lexique	GLÄFF	DPF
ε/e	été	/ɛ.te/	/e.te/	/e.te/	/ete/
s/z	stalinisme	/sta.li.nism/	/sta.li.nizm/	/sta.li.nism/	/stalinism/, /stalinizm/
b/p	obturer	/ɔb.ty.ʁe/	/ɔp.ty.ʁe/	/ɔp.ty.ʁe/	/ɔptyʁe/, /ɔbtyʁe/
o/ɔ	pomme	/pom/	/pɔm/	/pɔm/	/pɔm/
ə/ø/œ	heureux	/ə.ʁø/	/ø.ʁø/	/œ.ʁø/	/øʁø, œʁø/
y/ɥ	gradué	/gʁa.dy.e/	/gʁa.dɥe/	/gʁa.dɥe/	/gradɥe/, /gradɥe/, /gradye/
u/w	jouer inouï	/ʒu.e/ /i.nu.i/	/ʒwe/ /i.nwi/	/ʒwe/ /i.nwi/	/ʒwe/, /ʒue/ /inwi/, /inui/
a/ɑ	pâte	/pat/	/pat/	/pat/	/pat/, /pat/
+i, -i, +j, -j	riiez	/ʁi.i.je/	/ʁi.je/	/ʁij.je/	-
+ə, -ə	contenu	/kɔ̃.tə.ny/	/kɔ̃.tə.ny/	/kɔ̃t.ny/	/kɔ̃t(ə)ny/

## Divergences moins prédictibles : les emprunts

- *shaker* : /ʃɛi.kæʁ/, /ʃɛj.kœʁ/, /ʃɛ.kœʁ/
- *chili* : /ʃi.li/, /tʃi.li/
- *ginseng* : /ʒin.sɑ̃g/, /ʒin.sɑ̃g/, /ʒin.sɛŋ/