GLÀFF, un Gros Lexique À tout Faire du Français

UE TAL

7/10/2013

Franck Sajous, Nabil Hathout et Basilio Calderone
CLLE-ERSS, CNRS & Université de Toulouse 2

introduction|Ncfs|introduction|ɛ̃.tʁɔ.dyk.sjɔ̃

GLÀFF est un lexique :

- gros ;
- construit à partir du Wiktionnaire ;
- sous licence libre ;
- qui contient des informations morphosyntaxiques et phonémiques.

talentueuse|Afpfs|talentueux|ta.lɑ̃.tɥøz
talentueusement|Rgp|talentueusement|ta.lɑ̃.tɥøz.mɑ̃
talentueuses|Afpfp|talentueux|ta.lɑ̃.ty.øz
talent|Ncms|talent|ta.lɑ̃
talent|Vmip3p-|taler|tal
talent|Vmsp3p-|taler|tal
taleraient|Vmcp3p-|taler|ta.lə.ʁɛ

Plan

Ressources connexes (pour le français)

Lexiques morphosyntaxiques (gratuits)

- ABU (1999) : 300 000 entrées
  `http://abu.cnam.fr`
- Lefff [Clément et al., 2004] : 500 000 entrées
- Morphalou [Romary et al., 2004] : 525 000 entrées
  licence contraignante

avec transcriptions phonémiques

- BDLex [Pérennou and de Calmès, 1987] : 440 000 entrées
  (ELRA-S0004/Academic-Research : 2000 €)
- Brulex [Content et al., 1990] : 36 000 entrées, gratuit
- ILPho [Boula De Mareuil et al., 2000] : 319 000 entrées
  (ELRA-S0163/Academic-Research : 100 €)
- Lexique [New, 2006] : 135 000 entrées, gratuit

Ressources connexes

Une situation encore insatisfaisante

Pas de lexique libre, à large couverture, contenant des transcriptions phonémiques.
Obstacle pour :

- les outils (e.g. large couverture : étiqueteurs, transcriptions : phonétiseurs)
- la recherche
  - morphologie flexionnelle et dérivationnelle (Hathout 2009 ; Boyé 2011)
  - variation phonologique, phonotaxe et structures syllabiques (Goldsmith et Xanthos 2009 ; Maïonchi-Pino et al. 2013)
- la diffusion de travaux exploitant ces ressources ($\rightarrow$ œuvres dérivées)

Autre problème : ressources existantes non mises à jour

Wiktionary et sa branche francophone, le Wiktionnaire

Bande-annonce

- dictionnaire multilingue libre, accessible en ligne
- projet satellite («*compagnon lexical*») de Wikipédia
- 2 millions d'articles pour le français
- alimenté/mis à jour constamment... par «*les foules*»
- contient (notamment) :
  - des définitions et des exemples
  - des relations sémantiques
  - des traductions
  - des transcriptions phonémiques
- téléchargeable : `http://dumps.wikimedia.org`

Le Wiktionnaire pour le TAL

«2 millions d'articles pour le français»

Nomenclature impressionnante, mais sont comptabilisés :

- locutions
- variantes orthographiques
- mots d'autres langues

Restent 1,4 millions d'entrées (monolexicales)

Qualité ?

- Polémique : [Giles, 2005] vs. [Encyclopaedia Britannica, 2006]
- [Zesch and Gurevych, 2010] : ressources fondées sur la *« sagesse des foules »* sérieusement compétitives

Le Wiktionnaire pour le TAL

Qualité : notre point de vue

- Wiktionary ≠ Wikpédia
  «neutralité de point de vue» moins essentielle...
  mais contributeurs actifs «particuliers»...
- Qualité en tant que base de calcul de similarité
  ≠ qualité intrinsèque du dictionnaire
- Notre but : lexique pour le TAL ≠ dictionnaire
- Et par ailleurs...
  - Wiktionary a un large couverture
  - est distribué sous licence libre
  - contient des defs, relations sémantiques, traductions, transcriptions phonémiques
  - est mis à jour constamment (→ potentiellement seul à rendre compte des formes néologiques)

TAL, Wiktionnaire & Wiktionary

Travaux antérieurs

- [Zesch et al., 2008] : calcul de similarité sémantique
- [Navarro et al., 2009] : potentiel comme source de données primaire (lexique électronique FR+EN)
  → WiktionaryX [Sajous et al., 2010]
  `http://redac.univ-tlse2.fr/lexiques/wiktionaryx.html`
- [Krizhanovsky, 2010] : extracteur pour le Wiktionary russe
- [Sajous et al., 2011] : tentative de guidage des foules
  → description détaillée du Wiktionnaire + WiktionraryEN
- [Anton Pérez et al., 2011] : intégration du Wiktionary portuguais dans Onto.PT [Gonçalo Oliveira and Gomes, 2010]
- [Sérasset, 2012] : DBnary, extraction d'un réseau multilingue
- [Meyer and Gurevych, 2012] : OntoWiktionary
  [Gurevych et al., 2012] : UBY

Articles : entrée = url = forme orthographique

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

Créer un compte Connexion

Article Discussion Lire Modifier Afficher l'historique

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
Svenska

affluent

Français [modifier]

Étymologie

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

Adjectif

affluent

1 (Géographie)Qui se jette dans un autre en parlant d'un cours d'eau.
Le Rhin et les rivières **affluentes**.

2 (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	**affluent** /a.fly.ɑ̃/	affluents /a.fly.ɑ̃/
Féminin	affluente /a.fly.ɑ̃t/	affluentes /a.fly.ɑ̃t/

Nom commun

affluent /a.fly.ɑ̃/ masculin

1 (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
affluent	affluents
/a.fly.ɑ̃/	

Traductions

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

Forme de verbe

affluent /a.fly/

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

Conjugaison du verbe affluer

Articles : entrée = url = forme orthographique

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

affluente

Français [modifier]

Forme d'adjectif

affluente féminin /a.fly.ɑ̃t/

1 Féminin singulier de affluent.

Catégories : français | Formes d'adjectifs en français

Tables de conjugaison

http://fr.wiktionary.org/wiki/Annexe:Conjugaison_en_français/affluer

Annexe Discussion Lire Modifier Afficher l'historique

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:Conjugaison en français/affluer

Annexe:Conjugaison en français

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

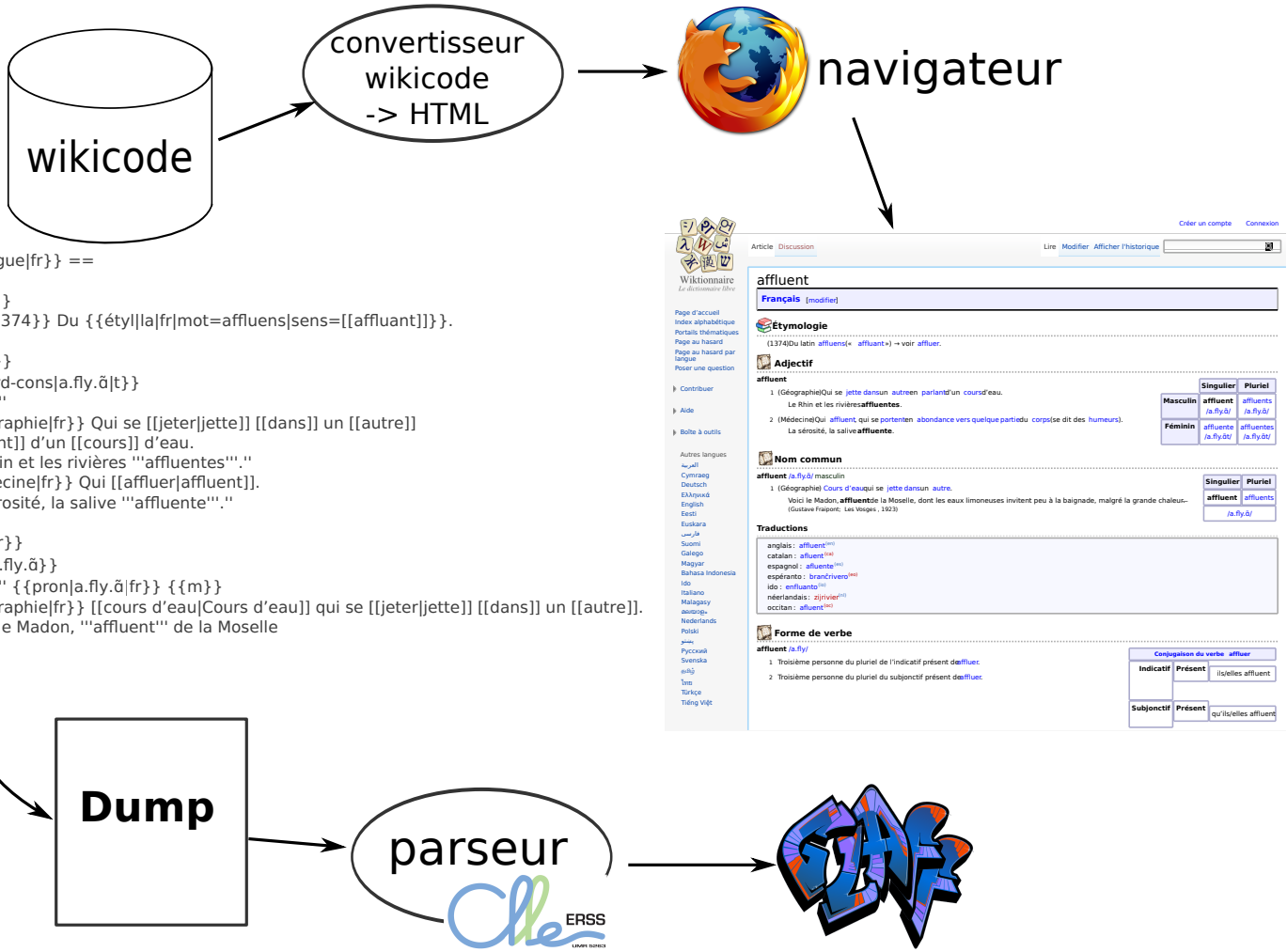
Modes impersonnels

Mode	Présent			Passé		
Infinitif		affluer	/a.flɥe/	avoir	afflué	/a.vwaʁ_a.flɥe/
Gérondif	en	affluant	/ɑ̃.n_a.flɥɑ̃/	en ayant	affluée	/ɑ̃.n_ɛ.jɑ̃.t_a.flɥe/
Participe		affluant	/a.flɥɑ̃/		affluée	/a.flɥe/

Indicatif

Présent			Passé composé			
j'/je/j'	affluе	/ʒ_a.fly/	j'ai	affluée	/ʒ_e a.flɥe/
tu	affluеs	/ty a.fly/	tu as	affluée	/ty a.z_a.flɥe/
il/elle/on	affluе	/[il/ɛl/ɔ̃]a.fly/	il/elle/on a	affluée	/[i.l/ɛ.l/ɔ̃.n]_a.t_a.flɥe/
nous	affluons	/nu.z_a.flɥɔ̃/	nous avons	affluée	/nu.z_a.vɔ̃.z_a.flɥe/
vous	affluez	/vu.z_a.flɥe/	vous avez	affluée	/vu.z_a.ve.z_a.flɥe/
ils/elles	affluent	/[il/ɛl].z_a.fly/	ils/elles ont	affluées	/[i/ɛ]l.z_ɔ̃.t_a.flɥe/
Imparfait			**Plus-que-parfait**			
j'/je/j'	affluais	/ʒ_a.flɥɛ/	j'avais	affluée	/ʒ_a.vɛ.z_a.flɥe/
tu	affluais	/ty a.flɥɛ/	tu avais	affluée	/ty a.vɛ.z_a.flɥe/
il/elle/on	affluait	/[il/ɛl/ɔ̃]a.flɥɛ/	il/elle/on avait	affluée	/[i.l/ɛ.l/ɔ̃.n]_a.vɛ.t_a.flɥe/
nous	affluions	/nu.z_a.fly.jɔ̃/	nous avions	affluée	/nu.z_a.vjɔ̃.z_a.flɥe/
vous	affluiez	/vu.z_a.fly.je/	vous aviez	affluée	/vu.z_a.vje.z_a.flɥe/
ils/elles	affluaient	/[il/ɛl].z_a.flɥɛ/	ils/elles avaient	affluées	/[i/ɛ]l.z_a.vɛ.t_a.flɥe/
Passé simple			**Passé antérieur**			

Moteur MediaWiki



Moteur MediaWiki

navigateur

«XML dumps»

microstructure pas encodée en XML, mais en wikicode !

```
<page>
    <title>affluent</title>
    <text>
    == {{langue|fr}} ==
    {{-adj-|fr}}
    {{fr-accord-cons|a.fly.|t}}
    '''affluent'''
    # {{géographie|fr}} Qui se [[jeter|jette]] [...]
    </text>

</page>
```

affluent|Vmip3p-|affluer|a.fly

Encodage : wikitexte

MediaWiki

"The parser serves as the de facto standard for the MediaWiki syntax, as no formal syntax has been defined. Due to this lack of a formal definition, it has been difficult [...] to port the parsing to another language."

`http://en.wikipedia.org/wiki/Mediawiki`

- pas de vérification automatique lors d'une modification
- syntaxe évolue
- syntaxe différente d'une édition de langue à l'autre

Extraire, trier, inférer et vérifier l'information

→ nécessité d'un parseur qui prenne en compte les erreurs de syntaxe, les incohérences, les informations lacunaires et la redondance

Encodage : wikitexte

affluent

Français [modifier]

Étymologie

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

Adjectif

affluent

1 (Géographie) Qui se jette dans un autre en parlant d'un cours d'eau.

Le Rhin et les rivières **affluentes**

2 (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	**affluent** /a.fly.ɑ̃/	affluents /a.fly.ɑ̃/
Féminin	affluente /a.fly.ɑ̃t/	affluentes /a.fly.ɑ̃t/

Nom commun

affluent /a.fly.ɑ̃/ masculin

1 (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
affluent	affluents
/a.fly.ɑ̃/	

Traductions

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

Encodage : wikitexte

affluent

Français [modifier]

Étymologie

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

Adjectif

affluent

1 (Géographie) Qui se jette dans un autre en parlant d'un cours d'eau.
Le Rhin et les rivières **affluentes**

2 (Médecine) Qui afflue, qui se porte en abondance vers quelque partie du corps (se dit des humeurs).
La sérosité, la salive **affluente**.

	Singulier	Pluriel
Masculin	**affluent** /a.fly.ɑ̃/	affluents /a.fly.ɑ̃/
Féminin	affluente /a.fly.ɑ̃t/	affluentes /a.fly.ɑ̃t/

Nom commun

affluent /a.fly.ɑ̃/ masculin

1 (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
affluent	affluents
/a.fly.ɑ̃/	

Traductions

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

Encodage : wikitexte

Modification de « affluent »

Graphique Fichier Avancé Caractères spéciaux Aide

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

{{-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.
#* ''La sérosité, la salive '''affluente'''.''

{{-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]].

{{-pron-}}
```

Encodage : wikitexte

Modification de « affluent »

Graphalique Fichier Avancé Caractères spéciaux Aide

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

{{-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.
#* ''La sérosité, la salive '''affluente'''.''

{{-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]].

{{-pron-}}
```

{{fr-accord-cons|a.fly.ɑ̃|t}}

Encodage : wikitexte

Modification de « affluent »

Graphtalique Fichier Avancé Caractères spéciaux Aide

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

{{-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.
#* ''La sérosité, la salive '''affluente'''.''

{{-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]].

{{-pron-}}
```

{{fr-accord-cons|a.fly.ɑ̃|t}}

	Singulier	Pluriel
Masculin	**affluent** /a.fly.ɑ̃/	affluents /a.fly.ɑ̃/
Féminin	affluente /a.fly.ɑ̃t/	affluentes /a.fly.ɑ̃t/

Encodage : wikitexte

Modification de « affluent »

Graphalique Fichier Avancé Caractères spéciaux Aide

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

{{-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.
#* ''La sérosité, la salive '''affluente'''.''

{{-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]].

{{-pron-}}
```

{{fr-accord-cons|a.fly.ɑ̃|t}}

	Singulier	Pluriel
Masculin	**affluent** /a.fly.ɑ̃/	affluents /a.fly.ɑ̃/
Féminin	affluente /a.fly.ɑ̃t/	affluentes /a.fly.ɑ̃t/

affluent|Afpms|affluent|a.fly.ɑ̃
affluents|Afpmp|affluent|a.fly.ɑ̃
affluente|Afpfs|affluent|a.fly.ɑ̃t
affluentes|Afpfp|affluent|a.fly.ɑ̃t

Information lacunaire, redondance et cohérence

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

```
{{-flex-adj-|fr}}
'''logicielles''' {{pron||fr}} {{fplur}}
```

→ `logicielles|Afpfp|logiciel` (prononciation ?)

Information lacunaire, redondance et cohérence

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

```
{{-flex-adj-|fr}}
'''logicielles''' {{pron||fr}} {{fplur}}
```

→ `logicielles|Afpfp|logiciel` (prononciation ?)

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

```
{{-adj-|fr}}
{{fr-accord-el|lɔ.ʒi.sjɛl}}
'''logiciel''' {{m}}
```

→ génération de :

- `logiciel|Afpms|logiciel|lɔ.ʒi.sjɛl`
- `logiciels|Afpmp|logiciel|lɔ.ʒi.sjɛl`
- `logicielle|Afpfs|logiciel|lɔ.ʒi.sjɛl`
- `logicielles|Afpfp|logiciel|lɔ.ʒi.sjɛl`

Information lacunaire, redondance et cohérence

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

```
{{-flex-nom-|fr}}
'''arrivages''' {{pron|a.ʁi.vaʒ|fr}}
# Pluriel d' '''[[arrivage]]'''.
```

→ arrivages|Nc ?p|arrivage|a.ʁi.vaʒ (genre ?)

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

```
{{-nom-|fr}}
'''arrivage''' {{pron|a.ʁi.vaʒ|fr}} {{m}}
```

→ arrivage|Ncms|arrivage|a.ʁi.vaʒ

Information lacunaire, redondance et cohérence

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

```
{{-flex-nom-|fr}}
'''arrivages''' {{pron|a.ʁi.vaʒ|fr}}
# Pluriel d' '''[[arrivage]]'''.
```

→ arrivages|Nc ?p|arrivage|a.ʁi.vaʒ (genre ?)

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

```
{{-nom-|fr}}
'''arrivage''' {{pron|a.ʁi.vaʒ|fr}} {{m}}
```

→ arrivage|Ncms|arrivage|a.ʁi.vaʒ

⇒ arrivages|Ncmp|arrivage|a.ʁi.vaʒ (inférence du genre)

Information lacunaire, redondance et cohérence

Verbes avec paradigmes incomplets

Exemple : *responsabiliser*
manquent les entrées pour les étiquettes : `Vmip1s-`, `Vmip3s-`, `Vmsp1s-`, `Vmsp3s-`, `Vmmp2s-`
→ utiliser un fléchisseur ?

Information lacunaire, redondance et cohérence

Verbes avec paradigmes incomplets

Exemple : *responsabiliser*
manquent les entrées pour les étiquettes : `Vmip1s-`, `Vmip3s-`, `Vmsp1s-`, `Vmsp3s-`, `Vmmp2s-`
→ utiliser un fléchisseur ?
ou attendre le prochain dump ?
(omission due à une erreur dans le wikitexte, corrigée le 24 mai 2013)

Information lacunaire, redondance et cohérence

Verbes avec paradigmes incomplets

Exemple : *responsabiliser*
manquent les entrées pour les étiquettes : `Vmip1s-`, `Vmip3s-`, `Vmsp1s-`, `Vmsp3s-`, `Vmmp2s-`
→ utiliser un fléchisseur ?
ou attendre le prochain dump ?
(omission due à une erreur dans le wikitexte, corrigée le 24 mai 2013)

Oui... mais les verbes défectifs ?

- *pleuvoir :* 9 formes manquantes ;
- *choir :* 10 formes manquantes ;
- *clore :* 12 formes manquantes ;
- *bruire :* 17 formes manquantes ;
- *frire :* 29 formes manquantes.

(sans oublier *neiger*, *reneiger*, *refrire*, *surfrire*, *pluvioter*, *bruinasser*, *etc.*)

Information lacunaire, redondance et cohérence

contredisez

Français [modifier]

Forme de verbe

contredisez /kɔ̃.tʁə.di.ze/

1. Deuxième personne du pluriel de l'indicatif présent de contredire.
2. Deuxième personne du pluriel de l'impératif de contredire.

Conjugaison en français/contredire

Impératif

Présent	
contredis	/kɔ̃.tʁə.di/
contredisons	/kɔ̃.tʁə.di.zɔ̃/
contredites	/kɔ̃.tʁə.dit/

Information lacunaire, redondance et cohérence

contredisez

Français [modifier]

Forme de verbe

contredisez /kɔ̃.tʁə.di.ze/

1 Deuxième personne du pluriel de l'indicatif présent de contredire.

2 Deuxième personne du pluriel de l'impératif de contredire.

Conjugaison en français/contredire

Impératif

Présent	
contredis	/kɔ̃.tʁə.di/
contredisons	/kɔ̃.tʁə.di.zɔ̃/
contredites	/kɔ̃.tʁə.dit/

$$< contredire, Vmmp2p- >= \begin{cases} contredisez|\text{kɔ̃.trə.di.ze} \\ contredites|\text{kɔ̃.trə.dit} \end{cases} ?$$

→ détection des couple <*lemme, étiquette*> avec plusieurs formes ?

Information lacunaire, redondance et cohérence

contredisez

Français [modifier]

Forme de verbe

contredisez /kɔ̃.tʁə.di.ze/

1 Deuxième personne du pluriel de l'indicatif présent de contredire.

2 Deuxième personne du pluriel de l'impératif de contredire.

Conjugaison en français/contredire

Impératif

Présent	
contredis	/kɔ̃.tʁə.di/
contredisons	/kɔ̃.tʁə.di.zɔ̃/
contredites	/kɔ̃.tʁə.dit/

$$< \textit{contredire}, \textit{Vmmp2p}- >= \left\{ \begin{array}{l} \textit{contredisez}|\text{kɔ̃.trə.di.ze} \\ \textit{contredites}|\text{kɔ̃.trə.dit} \end{array} \right. ?$$

→ détection des couple <*lemme, étiquette*> avec plusieurs formes ? mais :

<payer, Vmip1s> → paie|pɛ, paye|pɛj

<asseoir, Vmip1s> → assois|a.swa, assieds|a.sje

<végéter, Vmif1s> → végéterai|ve.ʒɛ.tə.ʁe, végèterai|ve.ʒɛ.tə.ʁe

(orthographe avant/après réforme 1976)

Quelques vérifications complémentaires...

Longueur des formes/longueur des lemmes, terminaisons verbales

```
Vmn----    ((er)|(ir)|(ïr)|(re))
Vmps-sm    [ûéiïtus]
Vmps-pm    [ûéiïtus]s
Vmps-sf    [ûéiïtus]e
Vmps-pf    [ûéiïtus]es
Vmpp---    ant
Vmip1s-    [esxi]
Vmip2s-    [sx]
Vmip3s-    [aedct]
Vmip1p-    ((ons)|(sommes))
Vmip2p-    ((ez)|(tes))
Vmip3p-    ((ent)|([fsv]?ont))
Vmif1s-    rai
Vmif2s-    ras
Vmif3s-    ra

etc.
```

Caractérisation. . .

. . . et pas évaluation

- comparaison possible avec d'autres ressources
- mais pas d'étalon
  (sinon, on ne constuirait pas un autre lexique. . .)
- tâche : laquelle, comment ?
  étiquetage : corpus d'évaluation ? validation manuelle ?
  ne renseignera que sur l'impact du lexique sur l'étiquetage

Comparaisons à venir

- couverture
- transcriptions phonémiques

Taille

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

Comparaisons à venir

- uniquement noms, verbes, adjectifs et adverbes (pas de mots grammaticaux, pas de locutions)
- uniquement sur les formes graphiques simples

Couverture inter-lexiques

	Lexique	BDLex	Lefff	Morphalou	GLÀFF
Lexique		26,0	25,2	22,5	9,0
BDLex	76.0		79,9	70,4	29,0
Lefff	79,5	86,3		72,3	30,0
Morphalou	79,6	85,4	81,2		32,0
GLÀFF	**84,8**	**93,3**	**90,2**	**85,67**	

(en % de formes fléchies catégorisées)

Couverture Lexiques/Corpus («taille utile»)

Projection des 5 lexiques sur les corpus :

- Frantext 20e, 515 romans du XX^{e}s. (30 millions de mots)
- LM10, archives 1991-2000 du quotidien *Le Monde* (200 millions de mots)
- Wikipédia2008 (260 millions de mots)
- FrWaC [Baroni et al., 2009] (1,6 milliard de mots)
  corpus de pages Web du domaine `.fr`

Utilisation de TreeTagger pour :

- segmenter les corpus ;
- filtrer les catégories syntaxiques N, V, Adj, Adv.
  (pour les formes connues ET inconnues de TreeTagger)
  catégories ensuite ignorées : on s'intéresse aux formes graphiques

On évalue, pour différents seuils de fréquence, le % de formes couvertes par les lexiques

Couverture Lexiques/Corpus («taille utile»)

Seuil : fréquence ≥		1	10	100	1000
Frantext	Nb formes	145 437	43 919	10 767	1 376
	Lexique	66,8	→ **96,9**	→ **99,2**	→ **99,3**
	BDLex	70,9	95,7	99,1	99,2
	Lefff	71,9	96,2	99,1	98,9
	Morphalou	73,9	96,0	98,5	97,1
	GLÀFF	→ **76,9**	96,7	98,8	98,8
LM10	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	→ **98,9**
	Morphalou	39,1	80,3	93,3	97,5
	GLÀFF	→ **45,2**	→ **86,2**	→ **96,5**	98,7
Wikipédia	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	→ **16,4**	→ **55,5**	→ **83,2**	→ **96,1**
FrWaC	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	→ **13,1**	→ **45,4**	→ **76,4**	→ **92,8**

Vocabulaires spécifiques

	Taille vocab. spécifique	Nombre de formes attestées			
		Frantext	LM10	Wikipédia	FrWaC
Lexique	1 509	866	863	1 073	1 320
BDLex	3 981	86	521	1 004	1 496
Lefff	11 050	232	1 479	2 214	3 288
Morphalou	26 881	1 171	1 912	3 995	6 425
GLÀFF	665 290	2 811	13 525	29 230	47 549

Formes spécifiques à GLÀFF, attestées dans LM10 :
transversalité, *attractivité*, *brevetabilité*, *diabolisation*, *employabilité*, *anticorruption*, *homophobie*, *institutionnellement*, *hébergeur*, *fatwa*, *indétrônable*

Top-10 des différences des transcriptions phonémiques

Diff.	%	∑ %
ɛ/e	48,2	48,2
ɔ/o	32,2	80,4
o/ɔ	11,0	91,4
y/ɥ	1,8	93,2
ə/ø	1,4	94,6
ə/œ	1,4	96,0
u/w	0,8	96,9
b/p	0,7	97,6
s/z	0,5	98,1
j	0,3	98,4

BDLex/Lexique

Diff.	%	∑ %
ɔ/o	60,0	60,0
ə	14,2	74,2
e/ɛ	6,9	81,1
ɛ/e	5,0	86,1
ɑ/a	4,9	91,0
s/z	1,3	92,3
ə/ø	0,9	93,2
œ/ø	0,5	93,7
i	0,4	94,1
o/ɔ	0,4	94,5

GLÀFF/Lexique

Diff.	%	∑ %
e/ɛ	66,5	66,5
ɔ/o	10,6	77,1
ə	5,9	83,0
o/ɔ	4,4	87,4
ɑ/a	3,8	91,2
ɥ/y	1,6	92,8
œ/ə	1,1	93,9
ø/ə	0,9	94,8
i	0,8	95,6
w/u	0,8	96,4

GLÀFF/BDLex

voyelles moyennes et schwa → 90% des divergences

Exemples de transcriptions divergentes

Diff.	Forme	BDLex	Lexique	GLÀFF	DPF[Martinet and Walter, 1973]
ɛ/e	été	/ɛ.te/	/e.te/	/e.te/	/ete/
s/z	stalinisme	/sta.li.nis,m/	/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	/po,m/	/pɔm/	/pɔm/	/pɔm/
ə/ø/œ	heureux	/ə.ʁø/	/ø.ʁø/	/œ.ʁø/	/ørø, œrø/
y/ɥ	gradué	/gʁa.dy.e/	/gʁa.dɥe/	/gʁa.dɥe/	/gradɥe/, /grɑdɥe/, /gradye/
u/w	jouer inouï	/ʒu.e/ /i.nu.i/	/ʒwe/ /i.nwi/	/ʒwe/ /i.nwi/	/ʒwe/, /ʒue/ /inwi/, /inui/
a/ɑ	pâte	/pa,t/	/pat/	/pɑt/	/pat/, /pɑt/
i,j	riiez	/ʁi.i.je/	/ʁi.je/	/ʁij.je/	-
ə	contenu	/kɔ̃,tə.ny/	/kɔ̃.tə.ny/	/kɔ̃t.ny/	/kɔ̃t(ə)ny/

- opposition s/z : suffixe *-isme*
- emprunts :
  - *shaker* : /ʃɛi.kəʁ/, /ʃɛj.kœʁ/, /ʃɛ.kœʁ/
  - *chili* : /ʃi.li/, /tʃi.li/
  - *ginseng* : /ʒin.sãg/, /ʒin.sãŋ/, /ʒin.sɛŋ/

Accord inter-lexiques

			Transcriptions phonémiques		Syllabation
Ressources		Intersection	Identiques	Comparables	Identiques
BDLex	Lexique	112 439	58,31	96,88	98,92
GLÀFF	Lexique	123 630	79,50	97,81	98,48
GLÀFF	BDLex	396 114	61,72	96,88	98,30

Accord inter-lexiques

			Transcriptions phonémiques		Syllabation
Ressources		Intersection	Identiques	Comparables	Identiques
BDLex	Lexique	112 439	58,31	96,88	98,92
GLÀFF	Lexique	123 630	79,50	97,81	98,48
GLÀFF	BDLex	396 114	61,72	96,88	98,30

Ni Lexique, ni BDLex ne saurait constituer un étalon !

- BDLex : /ɛ.te/ *(été)*, /po,m/ (*pomme*), /poʁt/ (*porte*), /oʁ/ (*or* et *hors*), etc.
- Lexique : /ʃa.sje/ (*châtié*), /kɑ̃.bʁi.jo.le/ et /kɑ̃.bʁi.o.le/ (resp. *cambriolé* et *cambriolés*)

Accord inter-lexiques

			Transcriptions phonémiques		Syllabation
Ressources		Intersection	Identiques	Comparables	Identiques
BDLex	Lexique	112 439	58,31	96,88	98,92
GLÀFF	Lexique	123 630	79,50	97,81	98,48
GLÀFF	BDLex	396 114	61,72	96,88	98,30

Ni Lexique, ni BDLex ne saurait constituer un étalon !

- BDLex : /ɛ.te/ (*été*), /po,m/ (*pomme*), /poʁt/ (*porte*), /oʁ/ (*or* et *hors*), etc.
- Lexique : /ʃa.sje/ (*châtié*), /kɑ̃.bʁi.jo.le/ et /kɑ̃.bʁi.o.le/ (resp. *cambriolé* et *cambriolés*)

Instabilités, e.g. groupe consonantique /s/ + C
(*ministère* /mi.nis.tɛʁ/ vs. *monistique* /mɔ.ni.stik/)
→ objet d'étude !

v1.1.2

Inclusion des mots grammaticaux

- re-parsing du Wiktionnaire
- révision manuelle

Fréquences

V 1.2

Inclusion des fréquences absolues et relatives :

- des formes fléchies catégorisées
- des lemmes catégorisées

dans les corpus :

- Frantext 20^{e}
- LM10
- FrWaC
- +/- bientôt Wikipédia 2013

Merci...

- étiquetage : Talismane [Urieli and Tanguy, 2013]
- lemmatisation des mots inconnus : Nabil

À venir

Un gros Lexique.org à tout faire

Ajout d'informations telles que :

- nb de voisins orthographiques
- nb de voisins phonémiques
- etc.

Interrogation

Comment chercher...

1. Les formes fléchies du verber *'taler'* ?
2. Les adverbes qui ne se finissent pas par /mɑ̃/
3. Les formes nominales et adjectivales, au singulier, qui se terminent par un 's' graphique qui se prononce
4. Les adjectifs contenant la graphie *ill* qui ne se prononce pas [j]
5. Les noms homophones de *'beau'*

Interrogation

Comment chercher...

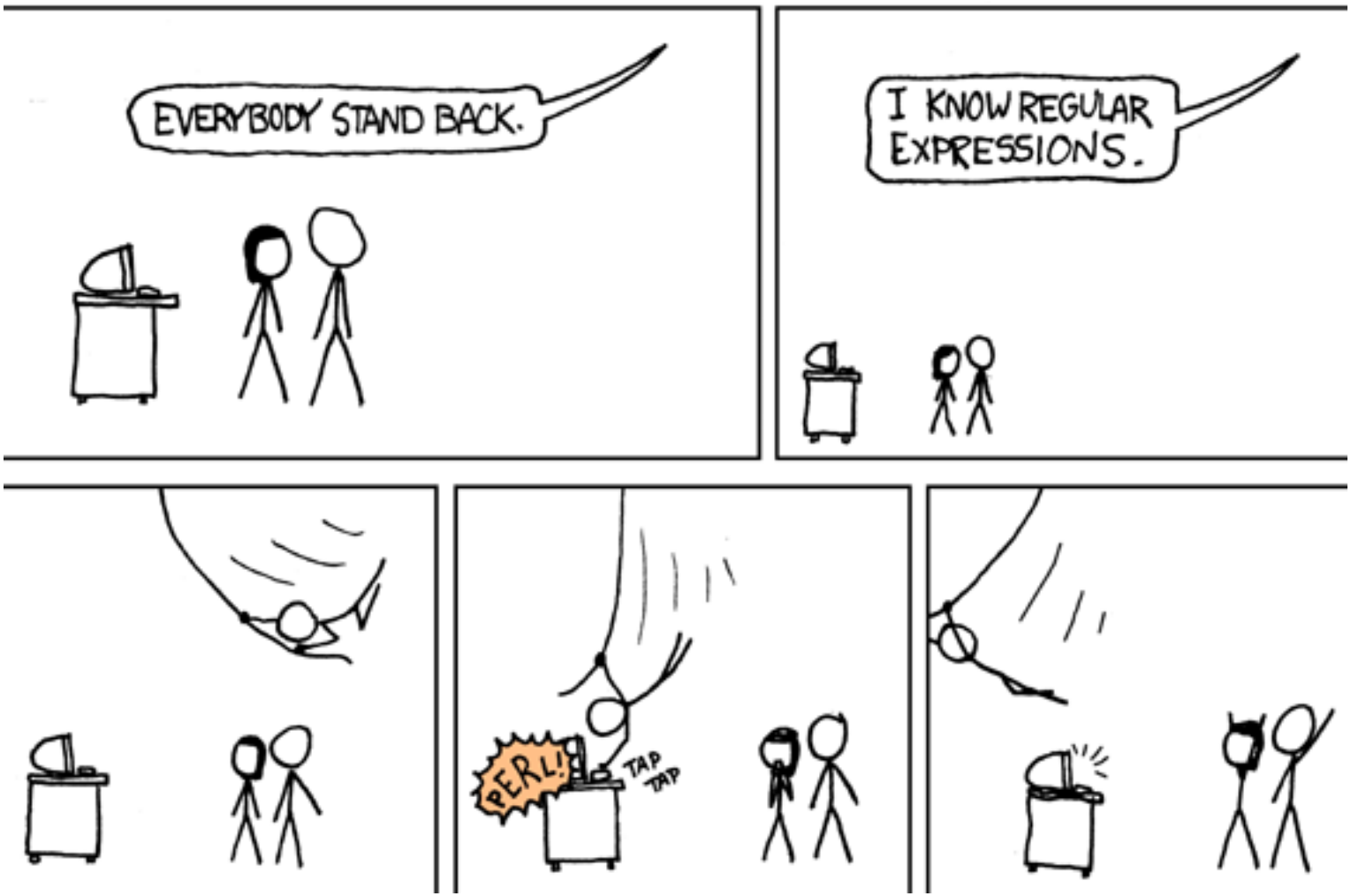
1. Les formes fléchies du verber *'taler'* ?
2. Les adverbes qui ne se finissent pas par /mɑ̃/
3. Les formes nominales et adjectivales, au singulier, qui se terminent par un 's' graphique qui se prononce
4. Les adjectifs contenant la graphie *ill* qui ne se prononce pas [j]
5. Les noms homophones de *'beau'*

Facile ! The grep way

1. `pcregrep '\|taler\|' glaff1.2.txt |cut -d'|' -f1`
2. `pcregrep '\|Rgp\|' glaff-1.2.txt | pcregrep -v 'mA~\|' (| pcregrep -v '||')`
3. `pcregrep '^[^|]*s\|[AN][^|]*s\|[^|]*\|[^|]*s\|' glaff-1.2.txt`
4. `pcregrep '^[^|]*ill[^|]*\|^A[^|]*\|[^|]*\|[^|j]*\|' glaff-1.2.txt`
5.

`pcregrep "^[^|]*\|N[^|]*\|[^|]*\|‘pcregrep '^beau\|' glaff-1.2.txt | tail -1 | cut -d'|' -f4‘[|;]" glaff-1.2.txt`

Ou bien... Le messianisme



https://xkcd.com/208/

Ou alors...

[Search] [Preferences] [Help] [About]

	Not			Not	
Form	☐	s$	IPA	☐	s$
Lemma	☐		SAMPA	☐	
POS	☐	[AN].*s$			

Reset Search

Showing results 1-100 of 379 >

Form	POS	Lemma	IPA
aérobus	Ncms	aérobus	ae.ʁɔ.bys
abies	Ncms	abies	a.bi.ɛs
abraxas	Ncms	abraxas	a.bʁak.sas
acinus	Ncms	acinus	a.si.nys
acropolis	Ncfs	acropolis	a.kʁɔ.pɔ.lis
adénovirus	Ncms	adénovirus	a.de.no.vi.ʁys
édelweiss	Ncms	édelweiss	e.dɛl.vajs
adonis	Ncms	adonis	a.dɔ.nis
ægilops	Ncms	ægilops	e.ʒi.lɔps
éléphantiasis	Ncfs	éléphantiasis	e.le.fɑ̃.tja.zis
albatros	Ncms	albatros	al.ba.tʁos
alios	Ncms	alios	a.ljɔs
alkermès	Ncms	alkermès	al.kɛʁ.mɛs

Demander à GLÀFFi (GLÀFF interface)

Conclusion

Atouts/craintes *a priori*

- + : taille, licence libre
- - : amateurisme, excentricité ?

a posteriori

- couverture effectivement large (et utile !)
- globalement d'une bonne qualité
  . . . et de toute façon : autres ressources de moindre taille (et payantes) également imparfaites
- potentiel pour devenir une ressource de référence ?
  (encore un peu d'effort de nettoyage)
- dans tous les cas : au moins une ressource complémentaire à Lefff et Morphalou (transcriptions + couverture)

Remarque

«on s'en a pas besoin d'avoir un gros lexique, y a plein d'entrées qui servent à rien»

Remarque

«on s'en a pas besoin d'avoir un gros lexique, y a plein d'entrées qui servent à rien»

- enseignement de TALN 2013 : *«la taille, c'est important»*
- oui, mais ça peut introduire de l'ambiguïté

Remarque

«on s'en a pas besoin d'avoir un gros lexique, y a plein d'entrées qui servent à rien»

- enseignement de TALN 2013 : *«la taille, c'est important»*
- oui, mais ça peut introduire de l'ambiguïté

  → `pcregrep -v '\|0\|0\|.*\|0\|0\|.*\|0\|0\|[0-9]+\|[.0-9]+$' psychoglaff.txt`

  1,4M entrées → 337 572 entrées !

Remarque

«on s'en a pas besoin d'avoir un gros lexique, y a plein d'entrées qui servent à rien»

- enseignement de TALN 2013 : *«la taille, c'est important»*
- oui, mais ça peut introduire de l'ambiguïté

  → `pcregrep -v '\|0\|0\|.*\|0\|0\|.*\|0\|0\|[0-9]+\|[.0-9]+$' psychoglaff.txt`

  1,4M entrées → 337 572 entrées !

  Morphalou : 524 725 → 282 379

  Lefff : 472 119 → 281 613

Remarque

«on s'en a pas besoin d'avoir un gros lexique, y a plein d'entrées qui servent à rien»

- enseignement de TALN 2013 : *«la taille, c'est important»*
- oui, mais ça peut introduire de l'ambiguïté

  → `pcregrep -v '\|0\|0\|.*\|0\|0\|.*\|0\|0\|[0-9]+\|[.0-9]+$' psychoglaff.txt`

  1,4M entrées → 337 572 entrées !
  Morphalou : 524 725 → 282 379
  Lefff : 472 119 → 281 613
- attention quand même... FT20e, LM10 et FrWaC ne sont que FT20e, LM10 et FrWaC !

Perspectives

Court terme

- inclusion des mots grammaticaux et locutions
- ajout d'autres types d'information :
  - fréquences en corpus
  - voisinages graphémiques et phonologiques
  - etc.
- détection d'erreurs et révision semi-manuelle
- interrogation en ligne

Perspectives

Court terme

- inclusion des mots grammaticaux ✓
  et locutions 

- ajout d'autres types d'information :
  - fréquences en corpus ✓
  - voisinages graphémiques et phonologiques ✓ 

  - etc.
- détection d'erreurs et révision semi-manuelle 

- interrogation en ligne 

Perspectives

Court terme

- inclusion des mots grammaticaux ✓
  et locutions 

- ajout d'autres types d'information :
  - fréquences en corpus ✓
  - voisinages graphémiques et phonologiques ✓ 

  - etc.
- détection d'erreurs et révision semi-manuelle 

- interrogation en ligne 

- WTFM

Perspectives

Moyen terme

- actualiser WiktionaryX et y intégrer GLÀFF
- évaluer l'apport de GLÀFF sur un analyseur syntaxique (e.g. Talismane) :
  - Lefff tout seul
  - GLÀFF tout seul
  - Lefff + GLÀFF
- exploitation de GLÀFF, dans les domaines suivants (au hasard) :
  - mesure de la similarité morphologique
  - phonotaxe distributionnelle
  - découverte automatique d'espaces thématiques utilisés pour la flexion

DL GLÀFF :

`redac.univ-tlse2.fr/lexiques/glaff.html`

et bientôt accessible : GLÀFFi !

References I


Anton Pérez, L., Gonçalo Oliveira, H., and Gomes, P. (2011).
Extracting Lexical-Semantic Knowledge from the Portuguese Wiktionary.
In *Proceedings of the 15th Portuguese Conference on Artificial Intelligence*, EPIA 2011, pages 703–717. APPIA.


Baroni, M., Bernardini, S., Ferraresi, A., and Zanchetta, E. (2009).
The WaCky wide web : a collection of very large linguistically processed web-crawled corpora.
Language Resources and Evaluation, 43(3) :209–226.


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.


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.

References II


Content, A., Mousty, P., and Radeau, M. (1990).
BRULEX : Une base de données lexicales informatisée pour le français écrit et parlé.
L'Année Psychologique, 90 :551–566.


Encyclopaedia Britannica (2006).
Fatally Flawed : Refuting the Recent Study on Encyclopedic Accuracy by the Journal Nature.


Giles, J. (2005).
Internet Encyclopaedias go Head to Head.
Nature, 438 :900–901.


Gonçalo Oliveira, H. and Gomes, P. (2010).
Onto.PT : Automatic Construction of a Lexical Ontology for Portuguese.
In *Proceedings of 5th European Starting AI Researcher Symposium*, pages 199–211. IOS Press.

References III


Gurevych, I., Eckle-Kohler, J., Hartmann, S., Matuschek, M., Meyer, C. M., and Wirth, C. (2012).
UBY - A Large-Scale Unified Lexical-Semantic Resource Based on LMF.
In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2012)*, pages 580–590.


Krizhanovsky, A. A. (2010).
Transformation of wiktionary entry structure into tables and relations in a relational database schema.
CoRR, abs/1011.1368.


Martinet, A. and Walter, H. (1973).
Dictionnaire de la Prononciation Française dans son Usage Réel.
France Expansion.

References IV

Meyer, C. M. and Gurevych, I. (2012).
OntoWiktionary Constructing an Ontology from the Collaborative Online Dictionary Wiktionary.
In Pazienza, M. T. and Stellato, A., editors, *Semi-Automatic Ontology Development : Processes and Resources*, chapter 6, pages 131–161. IGI Global, Hershey, PA, USA.

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.

New, B. (2006).
Lexique 3 : Une nouvelle base de données lexicales.
In *Verbum ex machina. Actes de la 13e conférence sur le Traitement automatique des langues naturelles*, Louvain-la-Neuve.

References V


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.


Sajous, F., Navarro, E., and Gaume, B. (2011).
Enrichissement de lexiques sémantiques approvisionnés par les foules : le système WISIGOTH appliqué à Wiktionary.
TAL, 52(1) :11–35.

References VI


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.


Sérasset, G. (2012).
Dbnary : Wiktionary as a LMF based Multilingual RDF network.
In *Proc. of the 8th International Conference on Language Resources and Evaluation (LREC)*, Istanbul.


Urieli, A. and Tanguy, L. (2013).
L'apport du faisceau dans l'analyse syntaxique en dépendances par transitions : études de cas avec l'analyseur Talismane.
In *Actes de la 20e conférence sur le Traitement Automatique des Langues Naturelles (TALN'2013)*, pages 188–201, Les Sables d'Olonne, France.

References VII


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.


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.