

Linguistique et traitement automatique des langues : quels outils pour quel public... et par quel(s) informaticien(s) ?

« *Les outils pour la recherche* »
Capitoul, Toulouse, 20 juin 2019

Franck Sajous

CLLE-ERSS, CNRS & Université de Toulouse 2



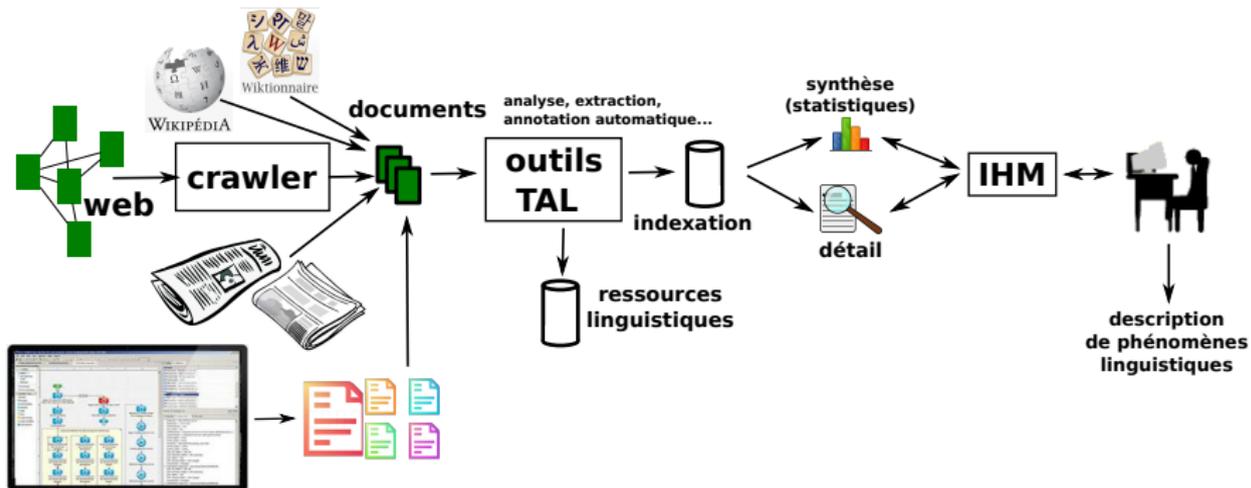
Les outils : quels besoins dans quel contexte ?

Quelles équipes d'informaticiens ?

UMR CLLE ~ 200 personnes

- composante LTC (psychologie cognitive, ergonomie du travail)
- composante **ERSS (linguistique)**
Recherche fondamentale, discipline descriptive
→ développement d'applications pour la collecte, l'observation et l'analyse de phénomènes langagiers
 - CARTEL, une équipe de traitement automatique des langues (TAL)
Discipline expérimentALE à visée (+/- ?) applicative
→ statistiques, calcul scientifique
- 10 ITAs, 1 BIATSS

L'informatique pour outiller la linguistique



Un outil générique pour l'observation linguistique : le concordancier

AntConc, développé par Laurence Anthony (Univ. Waseda, Japon)

AntConc 3.5.0 (Windows) 2017

File Global Settings Tool Preferences Help

Corpus Files

- BROWN1_A.TXT
- BROWN1_B.TXT
- BROWN1_C.TXT
- BROWN1_D.TXT
- BROWN1_E.TXT
- BROWN1_F.TXT
- BROWN1_G.TXT
- BROWN1_H.TXT
- BROWN1_I.TXT
- BROWN1_K.TXT
- BROWN1_L.TXT
- BROWN1_M.TXT
- BROWN1_N.TXT
- BROWN1_P.TXT
- BROWN1_R.TXT

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits 279

| Hit | KWIC | File |
|-----|---|-----------|
| 1 | 650 absorbed from the written word. A fundamental source of | BROWN1_A. |
| 2 | out of it with éclat, in a word a man who creates G50 1: | BROWN1_G |
| 3 | to make it apply to the wrong word. A verse R05 0460 familiar | BROWN1_R. |
| 4 | 40 0950 of never having read a word about art, but there was ne | BROWN1_A. |
| 5 | odbye forever. She never said a word about the fifty dollars. K18 | BROWN1_K. |
| 6 | tual one-digit or two-digit index word address in the range J69 1 | BROWN1_J. |
| 7 | digit J69 1890 or two-digit index word address in the range 3-94. | BROWN1_J. |
| 8 | mbled by C14 0840 putting one word after another. Mr& Sansor | BROWN1_C. |
| 9 | he Sioux, refused to say a harsh word against him. F22 1770 He | BROWN1_F. |
| 10 | might lead to devices like a 5000-word alphabetized dictionary J3 | BROWN1_J. |
| 11 | I owe it all to them>'. The word also K07 0520 made him I | BROWN1_K. |
| 12 | 016 1150 you, he who hears my word, and believes him who ser | BROWN1_D |
| 13 | ses to symbolic J69 0280 index word and electronic switch nam | BROWN1_J. |
| 14 | nglish sentence and the Martian word and felt closer grokking. M | BROWN1_M |

Total No. 15
Files Processed

Search Term Words Case Regex Search Window Size 50

word Advanced

Start Stop Sort Show Every Nth Row 1

Kwic Sort Level 1 1R Level 2 2R Level 3 3R Clone Results

Un outil générique pour l'observation linguistique : le concordancier

AntConc, développé par Laurence Anthony (Univ. Waseda, Japon)

AntConc 3.5.0 (Windows) 2017

File Global Settings Tool Preferences Help

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits 279

| Hit | KWIC | File |
|-----|---|-----------|
| 1 | 650 absorbed from the written word. A fundamental source of | BROWN1_A. |
| 2 | out of it with éclat, in a word a man who creates G50 1: | BROWN1_G |
| 3 | to make it apply to the wrong word. A verse R05 0460 familiar | BROWN1_R. |
| 4 | 40 0950 of never having read a word about art, but there was ne | BROWN1_A. |
| 5 | odbye forever. She never said a word about the fifty dollars. K18 | BROWN1_K. |
| 6 | tual one-digit or two-digit index word address in the range J69 1 | BROWN1_J. |
| 7 | igit J69 1890 or two-digit index word address in the range 3-94. | BROWN1_J. |
| 8 | mblmed by C14 0840 putting one word after another. Mr& Sansor | BROWN1_C. |
| 9 | he Sioux, refused to say a harsh word against him. F22 1770 He | BROWN1_F. |
| 10 | might lead to devices like a 5000-word alphabetized dictionary J3 | BROWN1_J. |
| 11 | I owe it all to them>". The word also K07 0520 made him I | BROWN1_K. |
| 12 | 016 1150 you, he who hears my word, and believes him who ser | BROWN1_D |
| 13 | ses to symbolic J69 0280 index word and electronic switch nam | BROWN1_J. |
| 14 | nglish sentence and the Martian word and felt closer grokking. M | BROWN1_M |

Total No. 15
Files Processed

Search Term Words Case Regex
word
Advanced
Search Window Size 50

Start Stop Sort Show Every Nth Row 1

Kwic Sort
 Level 1 1R Level 2 2R Level 3 3R

Clone Results

application standalone et (petits) corpus sur poste utilisateur

Un outil générique pour l'observation linguistique : le concordancier

BaTelÖc : Basa Textuala per la lenga d'Öc

[Acùelh] [Causada del còrpus] [Cèrca simpla] [Cèrca avançada] [Ajuda] [Projècte] [Contacte]

Cèrca simpla

Cercar un mot : lo

sensibla la caissa [?]
a a e e i i l l o o u u g g u u A A E E i i l l o o u u

Resultats 150 / 10000+ >

Exportar : [Tède (txt)] [Taula (csv)]

| | | | |
|--------------------|---|----------|--|
| PeiyotesDe la p.] | | 1 | Casimir de Camèra, lo mèste deus Escarts, a Labordèra, qu'èra de |
| PeiyotesDe la p.] | vòler hèter. La vèlha, Janòt de Latruèvera, lo batutsaire de Carboèras, qu'ave installat tot l'aparohatge | | |
| PeiyotesDe la p.] | , Bertran ! Qu'as pro netejat. Ça i prèner lo café, * Bertran, timidòt, qu'entrè hons | | |
| PeiyotesDe la p.] | , dab un beròi davant de popas. Qu'ave lo gèsto mordent, com si qu'auquaren la tarritava. Bertran | | |
| PeiyotesDe la p.] | de mau hum, e fobrer que s'bevò lo café shens diser amen mei. Ua pòrta que s' | | |
| PeiyotesDe la p.] | aves perdut dormir dras mèi. Hes pas lo dia i ce he Casimir, Espia, Bertran, | | |
| PeiyotesDe la p.] | he Casimir, Espia, Bertran, aci qu'avem lo vailet. Qu'arribè ger ser. Que va càler | | |
| PeiyotesDe la p.] | arribè ger ser. Que va càler que l'armushes lo trin-tran. Que ve n'avienerez, behidha. * | | |
| PeiyotesDe la p.] | pajar los sacs tau solèr de haut. Que sera lo ton tribah. Paula, que'u vas amushar per | | |
| PeiyotesDe la p.] | plan, mes vos n'as pas parier. Dab lo cot, que s'i cau avisar. Aquò | | |
| PeiyotesDe la p.] | requirit tà l'ahreda, Janòt de Latruèvera qu'estè lo darer. Que semblava mau desvelhat, dab lo peu | | |
| PeiyotesDe la p.] | estè lo darer. Que semblava mau desvelhat, dab lo peu long e despientat. Mes qu'èra beròi dròtle | | |
| PeiyotesDe la p.] | a com de fuèlh. Paula que'u servi lo café e que | | |
| PeiyotesDe la p.] | bevè s'hem ser'n dir. Quan avò ferit, lo gogol que's | | |
| PeiyotesDe la p.] | , Qu'èran aquis un quinzant, vienguts de tot lo veniat tad a | | |
| PeiyotesDe la p.] | pas tròp lèu. veiat ? Vinceng que segòti lo cap com un | | |
| PeiyotesDe la p.] | l'un e de l'aut e que sabè diser lo son mot qu'èra | | |
| PeiyotesDe la p.] | auto e que s'abò diser lo son mot quan avò lo parat. A Vin | | |
| PeiyotesDe la p.] | podosse estar vailet de paisan. * Un còp hèt lo servic, que | | |
| PeiyotesDe la p.] | soi pas est e a les escòlas. Qu'èr sonque lo comitat, e | | |
| PeiyotesDe la p.] | sòus en suenhant l'un e l'aut e, mes lo monde que | | |
| PeiyotesDe la p.] | 'us bocè las màges gòteras e que'us adobè drin lo dehenz tà g | | |
| PeiyotesDe la p.] | qu'èra era la dauna. Tad era, lo hih qu'èra | | |
| PeiyotesDe la p.] | qu'aqueas duas hermas qu'èran traccassat tà'us tirar lo ben ... Mes | | |
| PeiyotesDe la p.] | trobèn estavinda. Mor | | |
| PeiyotesDe la p.] | D | | |
| PeiyotesDe la p.] | shens verg | | |
| PeiyotesDe la p.] | i avit | | |
| PeiyotesDe la p.] | es hòla o g | | |
| PeiyotesDe la p.] | laur | | |
| PeiyotesDe la p.] | lo brabant ? Urban | | |
| PeiyotesDe la p.] | crompat lo trac | | |
| PeiyotesDe la p.] | miedia. Mes en | | |
| PeiyotesDe la p.] | sap | | |
| PeiyotesDe la p.] | Escota | | |
| PeiyotesDe la p.] | mans a la | | |

Cèrca avançada

Forma 1 : es - [?]

sensibla la caissa [?]
a a e e i i l l o o u u g g u u A A E E i i l l o o u u
Interval [?] Min : 3 Max : 9

Forma 2 : còrten - [?]

sensibla la caissa [?]
a a e e i i l l o o u u g g u u A A E E i i l l o o u u

Resultats 150 / 4033 >

Exportar : [Tède (txt)] [Taula (csv)]

| | |
|---|---|
| La lantana, la montanha, un jaubèta que | Un mat estivenc, au moment qui lo dia e s'apèlhava de son aubeta vermelha darèr lo cim de la montanha, ua jaubèta que |
| La lantana, la montanha, un jaubèta que | hasè son passet d'òri sus l'ambèra deu gave, au parcan de Lahontan. Que n'èra de la vita au melid punt qui l'arrosa quan s'aparia ad |
| La lantana, la montanha, un jaubèta que | arànçat gaudaments. Mes de la sòla deu pè que rasà de tròp près la carèra deu volader, de fagion que cadè dab un gran |
| La lantana, la montanha, un jaubèta que | Vièhòra. E quan lo pescador estavini e s'arrevrà, un gorgolh que l'armushà solet on s'èra ahonada ua creatura deu bon Dieu. |
| La lantana, la montanha, un jaubèta que | Com l'eslènbre e hen la nublà, eth que henò l'algia blua e a maugrat |
| La lantana, la montanha, un jaubèta que | drin, la gòsta qu'arreviscòla, e au cap de mieja òra, que caminava la pescador los sons l'espada deu salvador. D'ua vet nauticosa, on |
| La lantana, la montanha, un jaubèta que | s'endevinava lo bonur de viver, que gorguejava a l'aurèlha deu pescador los sons grans mercès. - Mes n'ac disis pas ad arès, e |
| La lantana, la montanha, un jaubèta que | pregava era, la mia mat, bonè com lo pan, |
| La lantana, la montanha, un jaubèta que | de viver, que gorguejava a l'aurèlha deu pescador los sons grans mercès. - Mes n'ac disis pas ad arès, e pregava era, la mia mat, bonè com lo pan, que m'aleudaja desempuish qui peu purnèr còp èi vet la lator deu dia : ua de mosca que se'n la halarè tau pais on lueishen las estèlas. Que'm brombarèi de |
| La lantana, la montanha, un jaubèta que | Jo ton corau, e jamei la mia amistat tà un autè òmi ne serà mèger que la qui jòt balhà a tu. - Que serèi mut com lo peish qui aciu e pesquì, Jo tarben que l'aimarèi còu darer bohet de la darèra òra ; e si atau eì lo ton voler, que seras, tu, la mia molher aimada. - Que m' |
| La lantana, la montanha, un jaubèta que | l balhà a tu. - Que serèi mut com lo peish qui aciu e pesquì. Jo tarben que l'aimarèi còu darer bohet de la darèra òra ; e si atau eì lo ton voler, que seras, tu, la mia molher aimada. - Que m'apèri Zabelina de Lator. Davant Nost Duana de Dabet qui n'as èi tant ajudat, que hèt lo segrament d |

Un outil spécifique (GlàffOLI)

pour interroger un lexique particulier (Glàff)

<http://redac.univ-tlse2.fr/glaffoli/>

Résultats 1-100 of 231 >

| Forme | POS | Lemme | IPA |
|--------------------|-------|-------------------|-------------------|
| achilletalonesque | Afpfs | achilletalonesque | a.ʃil.ta.lɔ̃.nesk |
| achilletalonesque | Afpms | achilletalonesque | a.ʃil.ta.lɔ̃.nesk |
| achilletalonesques | Afpfp | achilletalonesque | a.ʃil.ta.lɔ̃.nesk |
| achilletalonesques | Afpmp | achilletalonesque | a.ʃil.ta.lɔ̃.nesk |
| ancillaire | Atpts | ancillaire | ɑ̃.si.lɛʁ |
| ancillaire | Afpms | ancillaire | ɑ̃.si.lɛʁ |

Forme: CONTIENT | | | IPA: REGEX: ^[\^]]+\$

Lemme: CONTIENT | | | æ ø η œ œ ð ʒ ə ɛ ɛ ɥ ɸ ʒ ɔ ɛ

POS: COMMENCE PAR | Afp | SAMPA: CONTIENT | | | Chercher

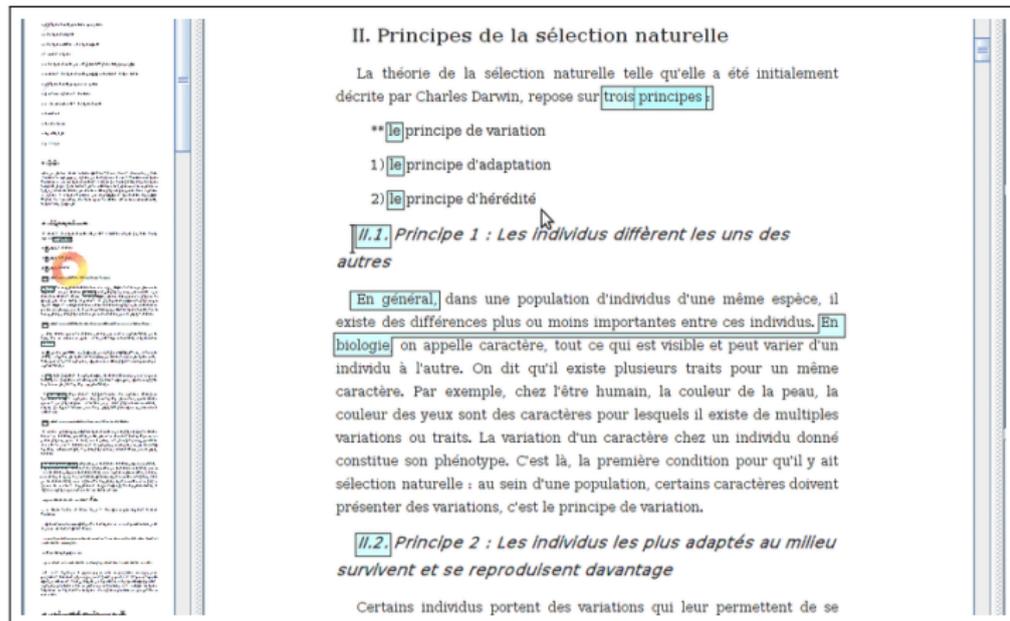
Lexiques/dictionnaires constitués à partir du Wiktionnaire

- analyse du *dump* (code wiki) : Perl
- crawl et analyse des pages HTML : Java (+API HtmlUnit)
- ressources libres disponibles en téléchargement (<http://redac.univ-tlse2.fr/lexiques/>)

Interface en ligne

- SGBD (MySQL)
- Java/Tomcat

Projet ANR AnnoDis (2008-2011) – CLLE, IRIT, GREYC



II. Principes de la sélection naturelle

La théorie de la sélection naturelle telle qu'elle a été initialement décrite par Charles Darwin, repose sur **trois principes** :

- ** **le** principe de variation
- 1) **le** principe d'adaptation
- 2) **le** principe d'hérédité

III.1. Principe 1 : Les individus diffèrent les uns des autres

En général, dans une population d'individus d'une même espèce, il existe des différences plus ou moins importantes entre ces individus. **En biologie** on appelle caractère, tout ce qui est visible et peut varier d'un individu à l'autre. On dit qu'il existe plusieurs traits pour un même caractère. Par exemple, chez l'être humain, la couleur de la peau, la couleur des yeux sont des caractères pour lesquels il existe de multiples variations ou traits. La variation d'un caractère chez un individu donné constitue son phénotype. C'est là, la première condition pour qu'il y ait sélection naturelle : au sein d'une population, certains caractères doivent présenter des variations, c'est le principe de variation.

III.2. Principe 2 : Les individus les plus adaptés au milieu survivent et se reproduisent davantage

Certains individus portent des variations qui leur permettent de se

Un phénomène linguistique plus complexe : les structures énumératives

Glozz, interface graphique d'annotation

The screenshot displays the Glozz 1.0.0-beta software interface. The main window shows a text document with various linguistic annotations. On the left, a vertical sidebar contains a tree view of the document's structure. The main text area is annotated with red and green boxes and lines, indicating different linguistic units and their relationships. A 'Unit style' dialog box is open in the foreground, showing a tree view of the document's structure. The 'Alignement' dialog box is also open, showing a grid of annotations for three accented units (Accenté 0, 1, 2) across different levels (Unit, Phrase, Prénom, Verbe). The 'Units' panel on the right lists the following units: Nom, Prénom, Verbe, Sujet, and Paragraphe. The 'Relations' panel lists: Anaphore, Elaboration, sujet, and complément. The 'Schemas' panel lists: Schema-Anaphorique, Schema-Elaboratif, and Intra-Discoursive. The 'Alignement' dialog box shows a grid of annotations for three accented units (Accenté 0, 1, 2) across different levels (Unit, Phrase, Prénom, Verbe). The 'Unit style' dialog box shows a tree view of the document's structure. The 'Alignement' dialog box shows a grid of annotations for three accented units (Accenté 0, 1, 2) across different levels (Unit, Phrase, Prénom, Verbe). The 'Unit style' dialog box shows a tree view of the document's structure. The 'Alignement' dialog box shows a grid of annotations for three accented units (Accenté 0, 1, 2) across different levels (Unit, Phrase, Prénom, Verbe).

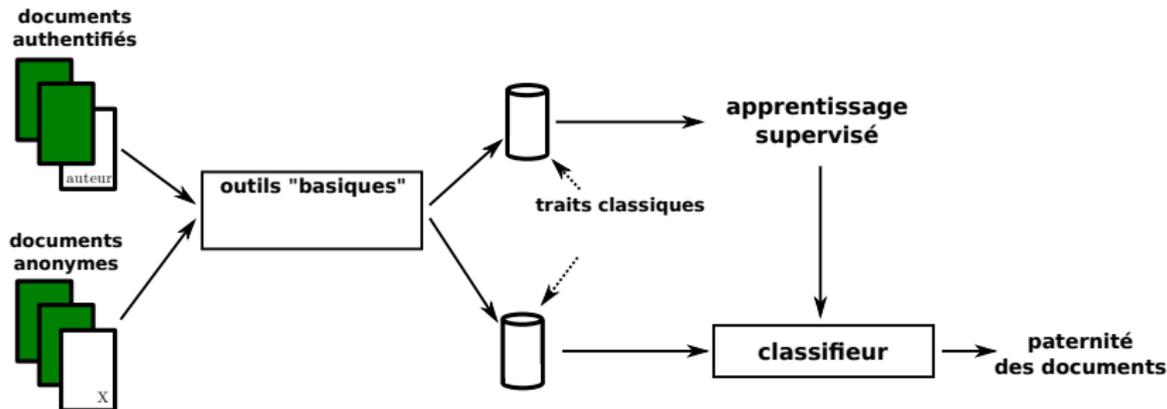
Un phénomène linguistique plus complexe : les structures énumératives

Glazz, interface graphique d'annotation

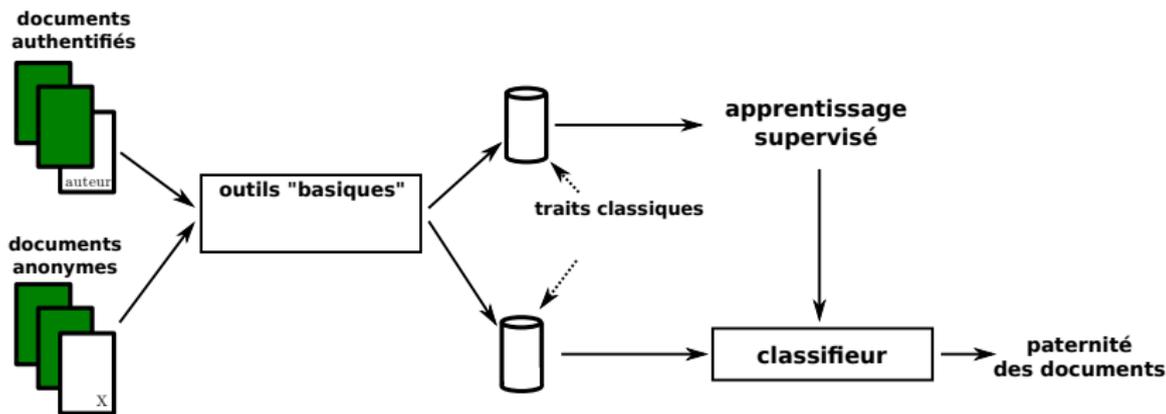
The screenshot displays the Glazz software interface, which is used for linguistic annotation. The main window shows a text document with various annotations (red boxes, lines, and arrows) highlighting specific linguistic features. A sidebar on the left contains a tree view of the document structure. A panel on the right lists 'Units' (Nom, Pronom, Verbe, etc.), 'Relations' (Anaphore, Elaboration, etc.), and 'Schemas' (Schema-Anaphorique, etc.). A 'Alignement' window is open, showing a grid with three axes (Accedat 0, 1, 2) and various colored markers. A 'Unit style' dialog box is also visible, showing options for 'Type name', 'Name', 'Background-color', and 'Style'.

développée par 1 doctorant et 1 MCF pendant 3 ans (et plus)

Un exemple de tâche de TAL : l'attribution d'auteurs



Un exemple de tâche de TAL : l'attribution d'auteurs



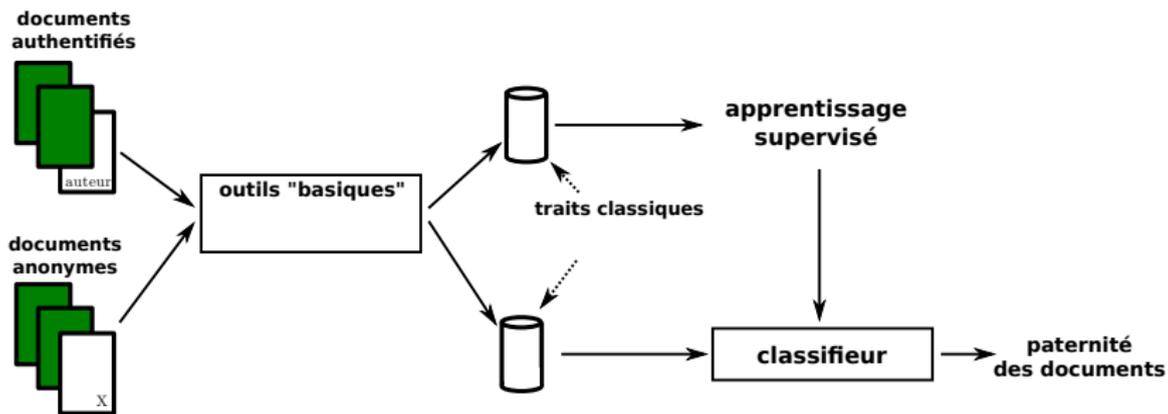
Le gloubi-boulga est la nourriture préférée de Casimir, le dinosaure imaginaire de l'émission télévisée pour enfants l'Île aux enfants. Il s'agit en fait d'un gâteau réputé comme étant immangeable et que seul le Casimir adore [...]

fréquence de mots

23 de
14 la
12 .
9 du
7 le
7 et
6 un
6 gloubi-boulga

...

Un exemple de tâche de TAL : l'attribution d'auteurs

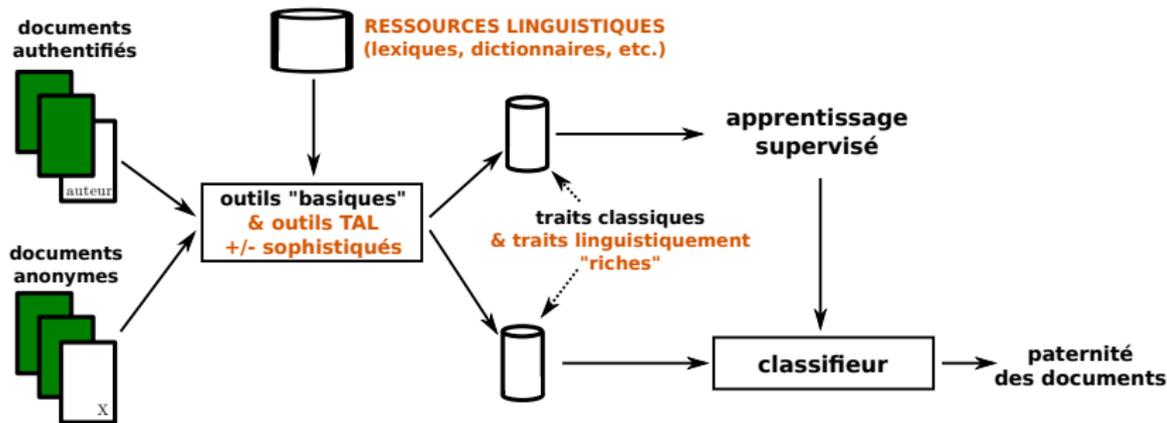


Le gloubi-boulga est la nourriture préférée de Casimir, le dinosaure imaginaire de l'émission télévisée pour enfants l'Île aux enfants. Il s'agit en fait d'un gâteau réputé comme étant immangeable et que seul le Casimir adore [...]

ngrammes de caractères

30 _de
27 de_
20 es_
16 e_d
16 ___
14 la_
14 _la
13 oul
...

Un exemple de tâche de TAL : l'attribution d'auteurs



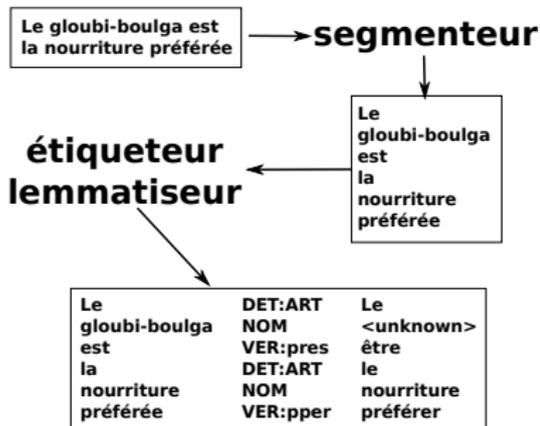
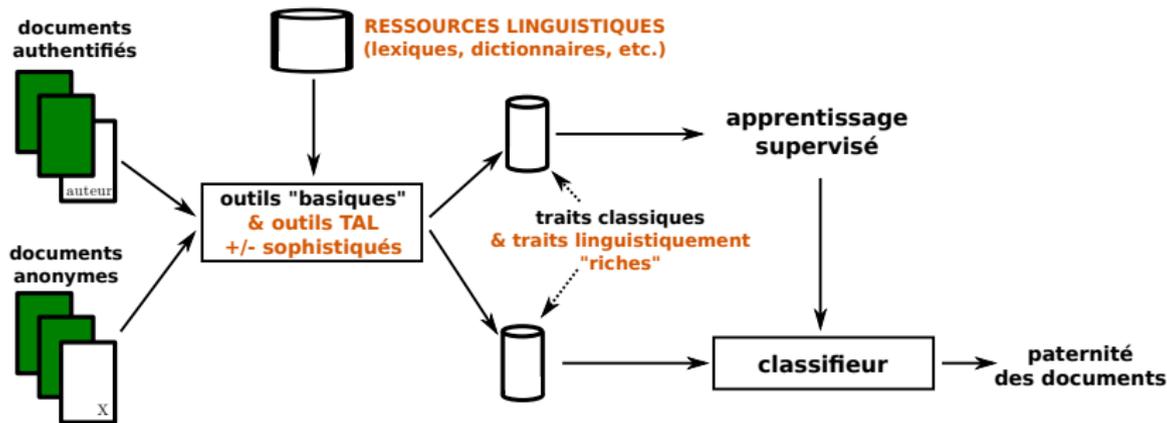
Le gloubi-boulga est la nourriture préférée de Casimir, le dinosaure imaginaire de l'émission télévisée pour enfants l'Île aux enfants. Il s'agit en fait d'un gâteau réputé comme étant immangeable et que seul le Casimirus adore [...]

POS

| | | |
|----|------|----------|
| 87 | NOM | (22.83%) |
| 57 | PRP | (14.96%) |
| 48 | PUN | (12.59%) |
| 41 | VER | (10.76%) |
| 35 | DET | (9.18%) |
| 26 | NAM | (6.82%) |
| 25 | ADJ | (6.56%) |
| 14 | SENT | (3.67%) |

...

Un exemple de tâche de TAL : l'attribution d'auteurs

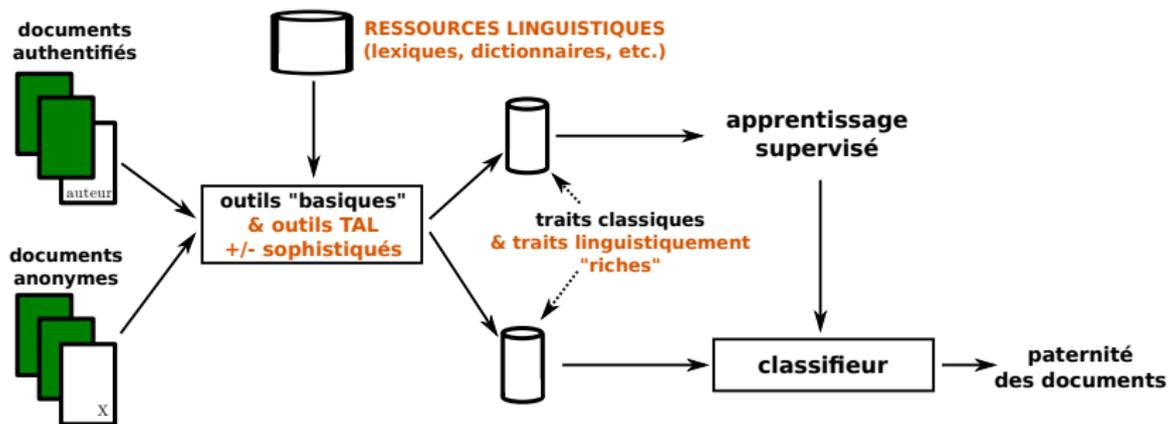


POS

| | | |
|----|------|----------|
| 87 | NOM | (22.83%) |
| 57 | PRP | (14.96%) |
| 48 | PUN | (12.59%) |
| 41 | VER | (10.76%) |
| 35 | DET | (9.18%) |
| 26 | NAM | (6.82%) |
| 25 | ADJ | (6.56%) |
| 14 | SENT | (3.67%) |

...

Un exemple de tâche de TAL : l'attribution d'auteurs

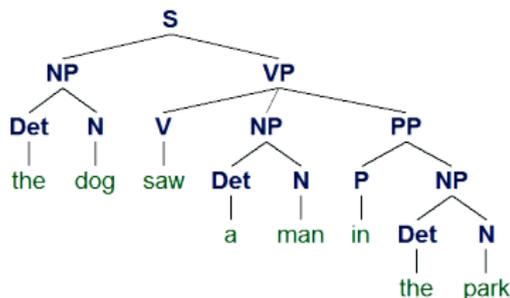
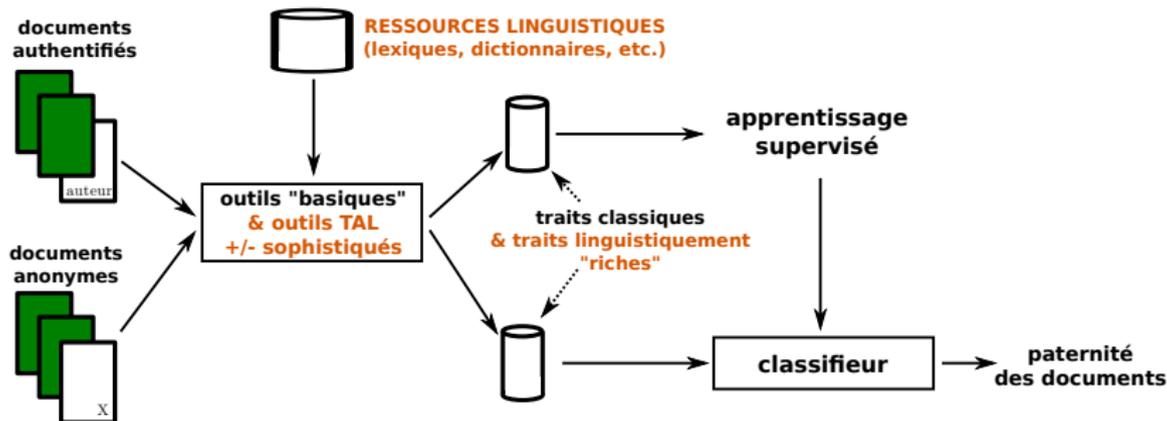


NAM DET:ART NOM VER:pres DET:ART NOM VER:pper
PRP NAM PUN DET:ART NOM ADJ PRP DET:ART NOM
VER:pper PRP NOM DET:ART NOM PRP:det NOM SENT
PRO:PER PRO:PER VER:pres PRP NOM PRP DET:ART
NOM ADJ KON VER:ppre ADJ KON KON ADJ DET:ART
NOM VER:pres SENT [...]

ngrammes de POS

13 PRP-DET-NOM
13 NOM-PRP-NOM
9 PRP-NOM-VER
9 DET-NOM-PRP
8 VER-PRP-DET
6 VER-PRP-NOM
6 PUN-DET-NOM
...

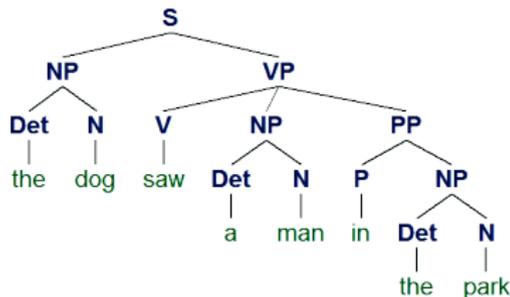
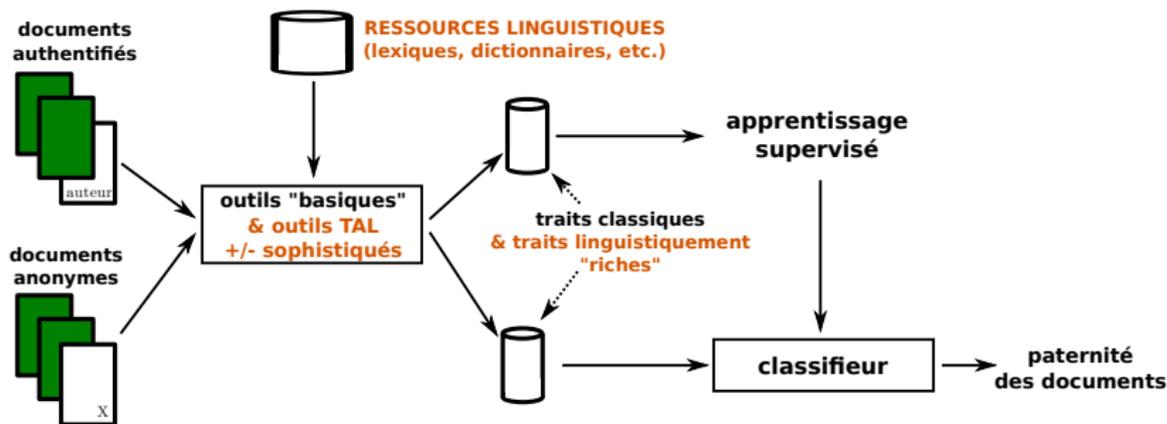
Un exemple de tâche de TAL : l'attribution d'auteurs



Traits linguistiquement riches

- degré de généricité (abstrait vs concrets), de polysémie
- type d'erreurs ortho/gram
- variantes UK/US
- profondeur/complexité syntaxique
- cohésion lexicale
- "subject shift" (*I suggest that we ...*)
- ponctuation, smileys, etc.

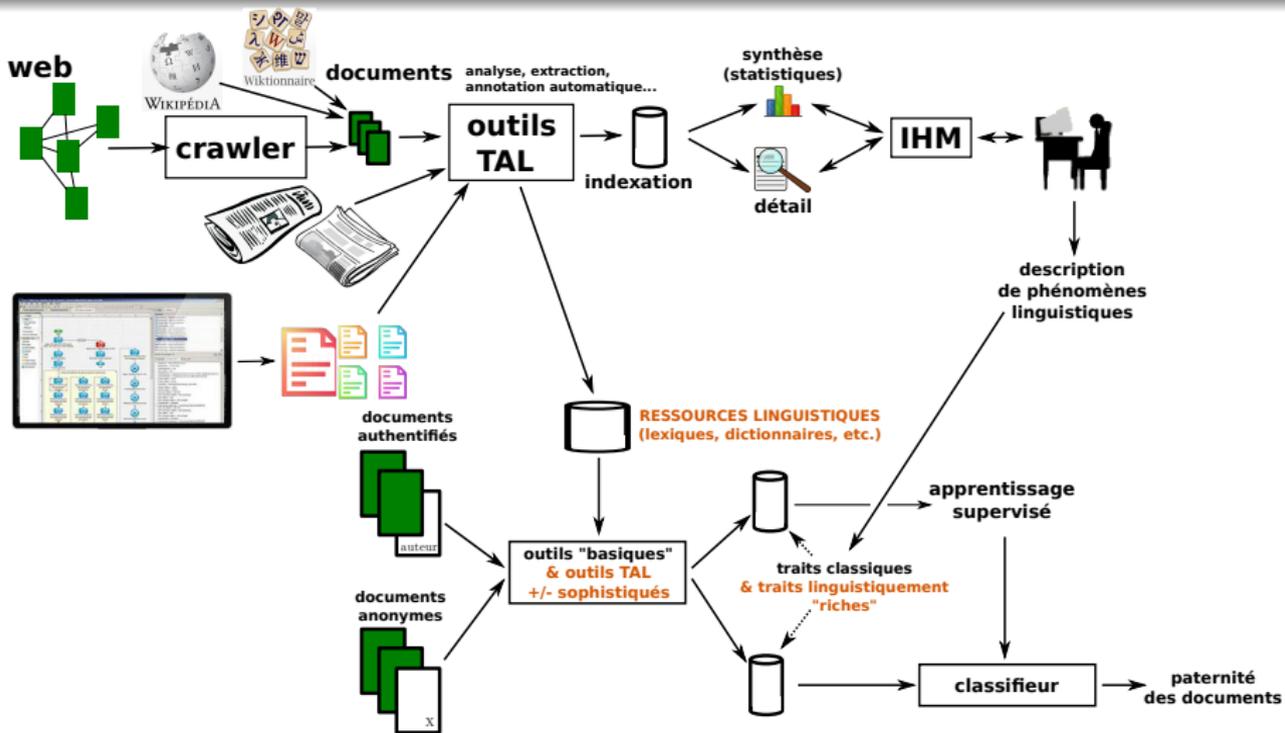
Un exemple de tâche de TAL : l'attribution d'auteurs



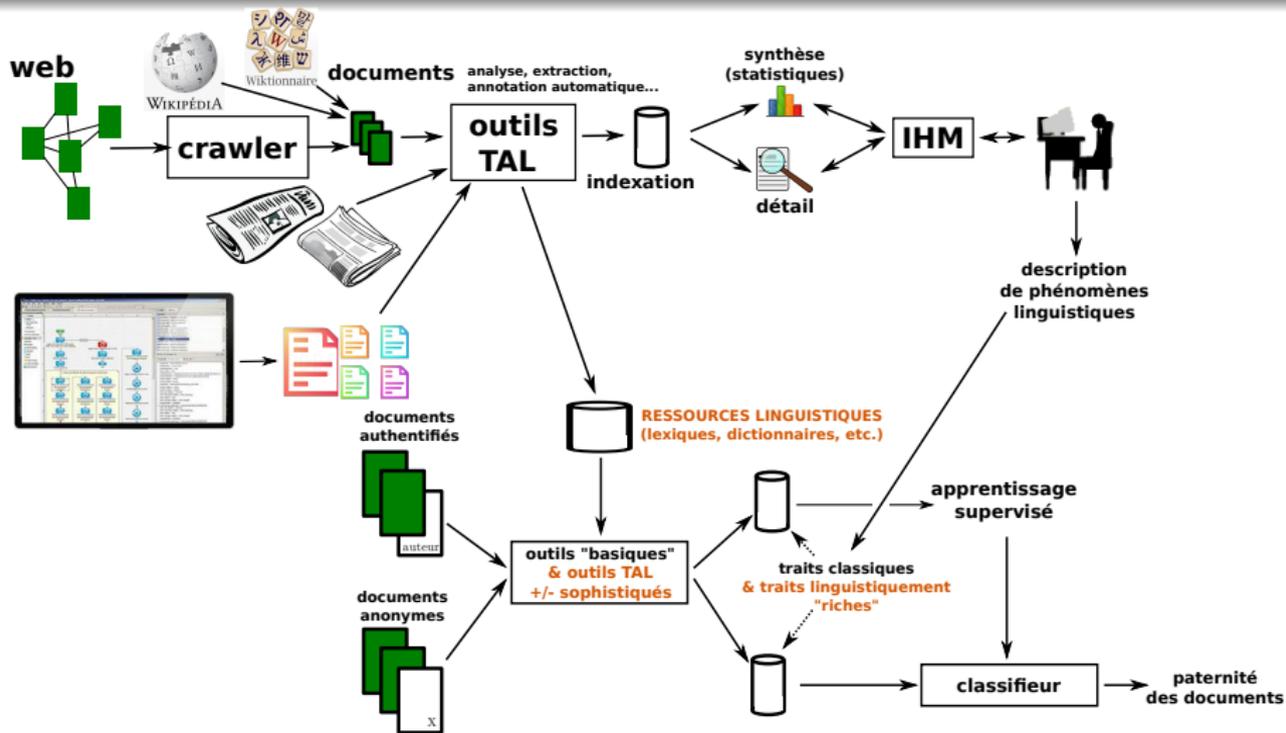
Outils existants et dev. *ad hoc*

- acteurs hétérogènes → outils hétérogènes. Outils et bibliothèques existants + dev. Perl, Python, Java
- analyseur syntaxique : 1 doctorant (développeur pro. depuis 20 ans) pendant 3 ans
- apprentissage : Java (Maxent) + Python (gensim) + R

Besoins informatiques partout...



Besoins informatiques partout...



TAL

Discipline expérimentale par essence,
dév. de nouveaux outils = but

Linguistique

Pas ou peu d'outils *off-the-shelf*
Pb de manipulation des données

Quelle équipe informatique (pour quels besoins et quel contexte) ?

- 2003 : 56 membres d'une UMR de linguistique → 1 ITA informaticien

Quelle équipe informatique (pour quels besoins et quel contexte) ?

- 2003 : 56 membres d'une UMR de linguistique → 1 ITA informaticien
- 2019 : 200 membres, dont 110 linguistes

Quelle équipe informatique (pour quels besoins et quel contexte) ?

- 2003 : 56 membres d'une UMR de linguistique → 1 ITA informaticien
- 2019 : 200 membres, dont 110 linguistes → 1 ITA informaticien

Quelle équipe informatique (pour quels besoins et quel contexte) ?

- 2003 : 56 membres d'une UMR de linguistique → 1 ITA informaticien
- 2019 : 200 membres, dont 110 linguistes → 1 ITA informaticien

Entre temps, quelques développements sans ITA

| Outil | Développeur(s) | Affiliation | Pérennisation |
|--|---|-------------|---|
| Yakwa , concordancier Client (Java)/Serveur (CGI Perl !) | 1 MCF informaticien-linguiste | CLLE | abandonné (incompatibilité après montée en version de Perl sur serveur) |
| Prox , calcul de marches aléatoires et interface en ligne | contractuels | CLLE | maintenance en pointillé, hébergement perdu, etc. |
| Glozz , interface d'annotation | 1 MCF + 1 doct. | GREYC | succès : encore utilisée et maintenue |
| Talismane , analyseur syntaxique | 1 doct. développeur professionnel, reparti depuis | CLLE | encore utilisé, mais quelle pérennisation ? |
| Parcolab , concordancier multilingue en ligne | 1 prestataire privé | Liban | durera tant que les crédits seront récurrents... |

Quelle équipe informatique (pour quels besoins et quel contexte) ?

- 2003 : 56 membres d'une UMR de linguistique → 1 ITA informaticien
- 2019 : 200 membres, dont 110 linguistes → 1 ITA informaticien

Informatique scientifique : pour conclure (ou pas)...

- Tendances actuelles
 - retours de mode : AI, réseaux de neurones
 - augmentation de la **volumétrie**, montée en puissance des **besoins en calcul** (*big data, deep learning...*)
 - **compétences à mobiliser** : prog. réseau, dev. objets, scripts, shells, admin, SGBD, IHM, statistiques, machine learning, HPC

Quelle équipe informatique (pour quels besoins et quel contexte) ?

- 2003 : 56 membres d'une UMR de linguistique → 1 ITA informaticien
- 2019 : 200 membres, dont 110 linguistes → 1 ITA informaticien

Informatique scientifique : pour conclure (ou pas)...

- Tendances actuelles
 - retours de mode : AI, réseaux de neurones
 - augmentation de la **volumétrie**, montée en puissance des **besoins en calcul** (*big data, deep learning...*)
 - **compétences à mobiliser** : prog. réseau, dev. objets, scripts, shells, admin, SGBD, IHM, statistiques, machine learning, HPC
+ phonologie, morphologie, sémantique, discours, pragmatique, TAL

Quelle équipe informatique (pour quels besoins et quel contexte) ?

- 2003 : 56 membres d'une UMR de linguistique → 1 ITA informaticien
- 2019 : 200 membres, dont 110 linguistes → 1 ITA informaticien

Informatique scientifique : pour conclure (ou pas)...

- Tendances actuelles
 - retours de mode : AI, réseaux de neurones
 - augmentation de la **volumétrie**, montée en puissance des **besoins en calcul** (*big data, deep learning...*)
 - **compétences à mobiliser** : prog. réseau, dev. objets, scripts, shells, admin, SGBD, IHM, statistiques, machine learning, HPC
+ phonologie, morphologie, sémantique, discours, pragmatique, TAL
 - L'« excellence », la « *compétitivité internationale* » ne passent :
 - ni par un recours massif à des CDD (projets)
→ quels profils ? Quelle pérennisation ?
 - ni par la mutualisation des personnels
→ quelles connaissances disciplinaires ?
- ...mais par le recrutement de permanents affectés aux Unités

Quelle équipe informatique (pour quels besoins et quel contexte) ?

- 2003 : 56 membres d'une UMR de linguistique → 1 ITA informaticien
- 2019 : 200 membres, dont 110 linguistes → 1 ITA informaticien

Informatique scientifique : pour conclure (ou pas)...

- Tendances actuelles
 - retours de mode : AI, réseaux de neurones
 - augmentation de la **volumétrie**, montée en puissance des **besoins en calcul** (*big data, deep learning...*)
 - **compétences à mobiliser** : prog. réseau, dev. objets, scripts, shells, admin, SGBD, IHM, statistiques, machine learning, HPC
+ phonologie, morphologie, sémantique, discours, pragmatique, TAL
- L'« *excellence* », la « *compétitivité internationale* » ne passent :
 - ni par un recours massif à des CDD (projets)
→ quels profils ? Quelle pérennisation ?
 - ni par la mutualisation des personnels
→ quelles connaissances disciplinaires ?

...mais par le recrutement de permanents affectés aux Unités même si cela semble moins « *tendance* »...