

Corpus, ressources et linguistique outillée

présentation et organisation du cours

Karën Fort

karen.fort@sorbonne-universite.fr / <https://members.loria.fr/KFort>



Organisation

Contrat d'enseignement

Cours

La forme, le fond, l'art et la manière

Organisation du cours

- ▶ 13 séances de 2h30, dont :
 - ▶ de la pratique
 - ▶ du cours
 - ▶ des interventions d'industriels du TAL (si c'est possible)
- ▶ évaluation :
 - ▶ contrôle continu - exercices (à rendre à **la fin du TD**)
 - ▶ contrôle continu - présentations (détails plus loin)
 - ▶ examen final

Corpus, ressources et outils

TAL, multilinguisme et grammaires locales

- ▶ très utilisées dans les entreprises de TAL
- ▶ case à cocher dans vos CV
- mais il y a une différence entre compétences et connaissances. . .
- ⇒ comprendre les couches basses

Plan (provisoire) du cours

- ▶ séance 1 : Corpus et droits
- ▶ séance 2 : Encodages
- ▶ séance 3 : TAL, niveaux d'analyse - exercices
- ▶ séance 4 : Universal Dependencies et GREW-match - exercices
- ▶ séance 5 : GREW-match par script (requêtes multilingues) - exercices
- ▶ séance 6 : TXM : introduction et commandes de base
- ▶ séance 7 : TXM : commandes avancées
- ▶ séance 8 : Comment faire une présentation ? TXM (fin) ou Unitex ?
- ▶ séance 9 : Unitex : Introduction à Unitex, recherche de motifs par regex - exercices
- ▶ séance 10 : Unitex : recherche de motifs par grammaires - exercices
- ▶ séance 11 : Unitex : grammaires locales - exercices
- ▶ séances 12 et 13 : Présentations

Présentations : explications

- ▶ Sujet imposé
- ▶ Recherches et préparation d'une présentation et d'une démo
- ▶ Présentées à la classe en 20 minutes + questions, lors de l'avant-dernier et du dernier cours
- ⇒ Participe de la note finale (1/3)
- ⇒ Pourront faire l'objet de questions lors de l'examen

Présentations : attention !



Présentations : étapes évaluées

1. Au bout du premier tiers, document (pdf) détaillant l'organisation prévue :
 - ▶ identification des difficultés (créneaux pour travailler, machines, etc)
 - ▶ rétroplanning pour prévoir la progression du travail
2. Au bout du deuxième tiers (testé en classe) :
 - ▶ outil installé,
 - ▶ connaissance minimale de l'outil
3. Présentation et démo en classe

Organisation

La forme, le fond, l'art et la manière

Le plagiat et ses conséquences

Bibliographie

Citer ses sources

Chacun de mes cours sera :

- ▶ précédé d'une liste des sources principales qui l'ont inspiré
- ▶ émaillé de citations sous la forme [Church, 2011]
- ▶ suivi d'une bibliographie détaillant les citations

Parenthèse sur le plagiat

Que savez-vous du plagiat ?

Parenthèse sur le plagiat

« emprunt d'idée ou de texte à d'autres auteurs sans que la citation soit explicite » (A. Nazarenko, LIPN / Paris 13)

Ce que vous devez faire :

- ▶ citer toutes vos sources
- ▶ mettre des guillemets lors d'emprunts mot à mot
- ▶ synthétiser la pensée de l'auteur (si la citation est longue)

Conséquences du plagiat

- ▶ le plagiat – la copie du travail d'autrui sans citation des sources – est interdit et passible de sanctions.
- ▶ ceci s'applique tant aux textes écrits qu'aux programmes ou aux présentations.
- ▶ le plagiat à l'examen entraîne la note de 0 et/ou la saisine du conseil de discipline (avec des sanctions potentiellement encore plus graves)

Conséquences du plagiat

- ▶ le plagiat – la copie du travail d'autrui sans citation des sources – est interdit et passible de sanctions.
- ▶ ceci s'applique tant aux textes écrits qu'aux programmes ou aux présentations.
- ▶ le plagiat à l'examen entraîne la note de 0 et/ou la saisine du conseil de discipline (avec des sanctions potentiellement encore plus graves)

Pire

le plagiat est un **délit**

« Toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayant cause est un délit. Il en est de même pour la traduction, l'adaptation ou la transformation, l'arrangement ou la reproduction par un art ou un procédé quelconque. »

Article L.122-4 du Code de la propriété intellectuelle.



Church, K. (2011).

A pendulum swung too far.

Linguistic Issues in Language Technology - LiLT, 6.