



Annotation collaborative de corpus pour le TAL présentation et organisation du cours

Karën Fort

karen.fort@sorbonne-universite.fr / <https://www.schplaf.org/kf/>

6 novembre 2020



Présentation

Cours

Contrat d'enseignement

Plan du cours

Annotation collaborative de corpus pour le TAL

Descriptif

Avec l'essor de l'apprentissage supervisé, la formalisation linguistique, au sein du TAL, s'est (largement) déplacée des grammaires locales à l'annotation manuelle. Or, celle-ci reste un processus mal maîtrisé. Ce cours présente les enjeux de l'annotation manuelle de corpus pour le TAL, les problématiques liées à sa formalisation et à son évaluation et les solutions proposées, notamment la myriadisation (crowdsourcing), en particulier via des jeux ayant un but (GWAP).

Organisation du cours

- ▶ 8 séances de "cours" :
 - ▶ lire et annoter le cours **avant** : voir diapo suivante
 - ▶ visio aux heures prévues (8h30-11h) :
<https://rendez-vous.renater.fr/karenfort>
- ▶ évaluation : contrôle continu intégral
 - ▶ devoirs à rendre (notés)
 - ▶ présentation à faire (notée), en présentiel j'espère...

Hypothes.is : mise en place (une fois)

- ▶ s'inscrire sur <https://hypothes.is>
- ▶ installer l'extension ou le marque page pour votre navigateur
- ▶ rejoindre le groupe : <https://hypothes.is/groups/GLGX48V9/m2-annotationcollab>

Hypothes.is : Utilisation (toutes les semaines)

- ▶ se connecter au cours (j'enverrai un lien)
- ▶ l'annoter en tant que membre du groupe, **dans le groupe**

The screenshot displays a web browser window with the Hypothes.is interface. The address bar shows the URL: https://via.hypothes.is/https://zombiludik.org/master/Unitex2_gra.... The document being viewed is titled "Corpus, ressources et linguistique outillée : graphes et grammaires locales dans Unitex" by Karën Fort. The text "graphes" is highlighted in green. The right sidebar shows an annotation by "karen_fort" on the word "graphes", dated May 11. The sidebar also includes a "Page Notes" section and a "Annotations" section. The bottom right corner of the document shows "1 / 43".

Plan du cours

1. introduction et motivations
2. formats et outils
3. accords inter-annotateurs
4. solutions
5. *crowdsourcing* : typologie(s), légendes et réalités
6. Jeux ayant un but (*GWAP*) et gamification
7. dimensions de complexité de l'annotation