SLIDE 2

- la transcription puis l'annotation des documents permet de mettre en place des traitements automatiques sur les sources pour générer des données quantitatives (séries temporelles de rémunérations, informations sur le temps dédié à telle ou telle tâches...)
- l'annotation peut se faire de manière automatique grâce au TAL, mais elle doit passer par une phase préparatoire et itérative de modélisation de l'annotation et des tests manuels de cette annotation.
- l'annotation mannuelle réalisée en amont permet de produire plusieurs éléments nécessaires à l'annotation automatique :
  - un guide d'annotation qui permet de comprendre comment cette annotation est appliquée au texte (important pour des réutilisation des corpus annotés au-delà du projet, pas seulement dans le cadre de Time Us)
  - des index qui permettent de fournir un vocabulaire au programme chargé de faire l'annotation automatique
  - des échantillons annotés qui servent de modèles pour le programme et de données d'entraînement

---

SLIDE 3

Le guide d'annotation rassemble des consignes pour réaliser une annotation de qualité, par exemple ici la délimitation des zones de texte annotées

---

SLIDE 4

Il liste les tags ou balises prévues pour la campagne d'annotation avec des règles pour leur application et une série d'illustrations et de contre-exemples.

SLIDE 5

- informations dont l'annotation dans le texte a été jugée utile pour Time Us :
  - informations sur les rémunérations (passages indiquants des rémunérations, types de rémunération, montants, tâches, produits, et durées ou rythmes de travail)
  - différentes informations que l'on peut considérer comme des entités :
    - informations sur les métiers (ensemble des termes pour désigner les métiers du textile)
    - noms de personnes
    - noms de lieux
    - noms d'établissements ou d'organisations
  - informations supplémentaires qui permettent de compléter l'annotation (traitement TAL pour

liaison avec les bons éléments)
    - statuts matrimoniaux
    - types de rémunération
    - adresses
  - des informations contextuelles : dates, références à des documents externes (pour préciser les rémunérations ou identifier des sources supplémentaires)

SLIDE 6

15 éléments d'annotation listés dans le guide d'annotation réalisé et disponible sur le site timeusage.paris.inria.fr. + un éléments pour signaler passages problématiques à faire traiter par des experts

---

SLIDE 7

Tous ces éléments ont des équivalents TEI pour obtenir des corpus annoter de bonne qualité, qui font partie des productions finales du projet Time Us.

SLIDE 8

Solution d'annotation : Transkribus. - A les avantages de Transkribus (accès aux différentes versions successives, collaboratif, set de tags identique) + système d'annotation prévu par Transkribus et pris en charge pour l'export "TEI" (XML). - Mais export de mauvaise qualité (traitement complexe nécessaire pour obtenir une TEI de bonne qualité) et les inconvénients de Transkribus (lenteur, manque de documentation).

SLIDE 9

Il y a pour ainsi dire eu 2 phases d'annotations : - 1ere : série 9M5 annotée par Romane Ducrocq : mais avant modélisation de l'annotation et avant guide donc corrections nécessaires. Mais a permis d'identifier certaines difficultés d'annotation. - 2eme : série 9M5 par Marie Lauricella, et HH par Hugues Serveau : - BUT = confronter un guide fait à partir de 9M5 (= 1890s, rapports de police) à un autre type de formulation des informations (HH = prudhomme 1771) ; - identifier les imprécisions du guide, les difficultés à appliquer les règles (par exemple : distinguer le concept de la mention réelle d'un nom) - identifier les manques dans le modèle (besoin d'ajouter des tags comme "recensement/quantité", "contexte de production des documents" ...

SLIDE 10

- Transkribus remplit-il suffisamment son rôle pour l'annotation ou faut-il trouver un autre outil (TXM ?) ?
- comment mieux communiquer le guide ? (un tutoriel sous forme de mini exercices d'annotation ?)
- quels tags supplémentaires sont nécessaires ? Les choix fonctionnent-ils pour l'annotation automatiques ?
- reste à mettre en place une configuration du wiki qui permet de visualiser le résultat des transcriptions et annotations