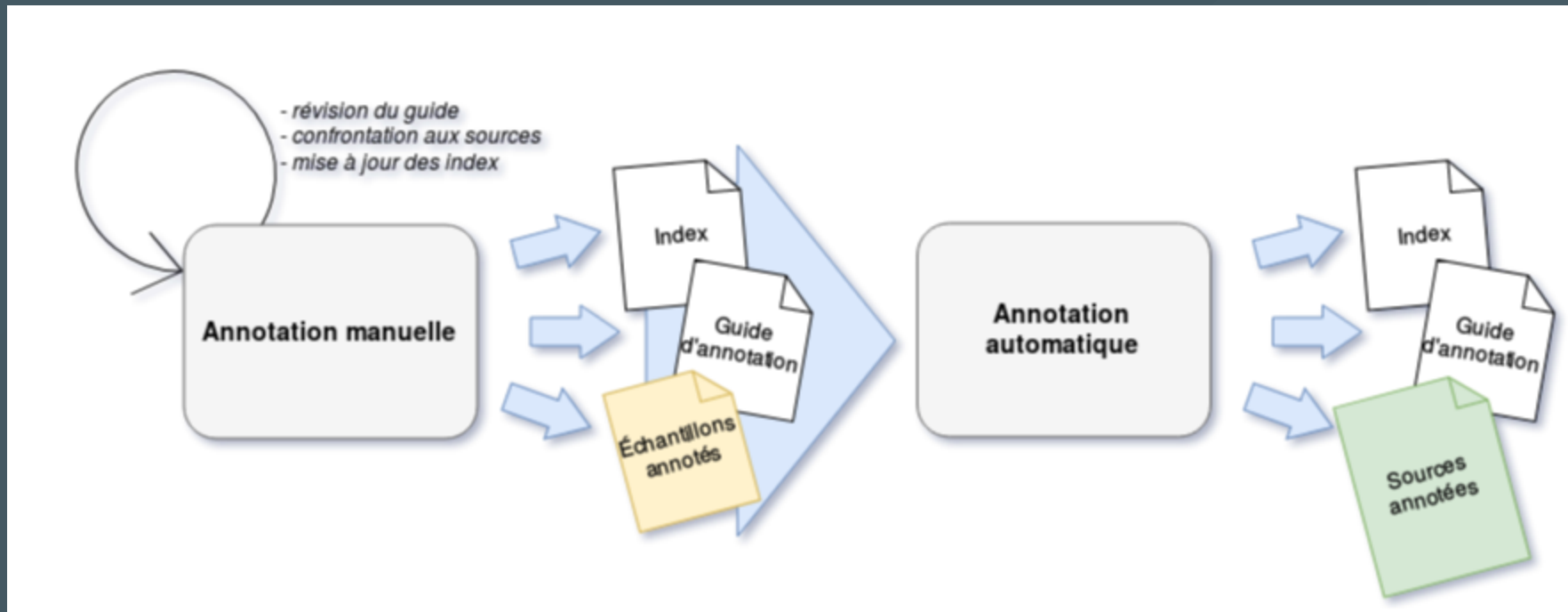


Annotation des sources

- Annoter une source qualitative rend possible un **traitement quantitatif**.







- Une phase d'annotation manuelle est nécessaire pour la mise en place de l'annotation automatique.

Sélectionner le texte [\[modifier\]](#)



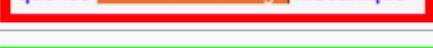
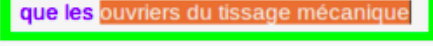
Lors de l'application d'un tag dans Transkribus, il est fondamental de respecter quelques règles afin d'optimiser l'annotation des textes.

1. La portion de texte sélectionnée ne doit pas inclure les signes de punctuations et les espaces situés de part et d'autre du mot ou du groupe de mots annoté. Si la ponctuation se trouve à l'intérieur de la sélection, dans une locution ou dans une adresse par exemple, elle fait bien légitimement partie du groupe de mot sur lequel porte l'annotation.

2. La portion de texte sélectionnée ne doit pas inclure les déterminants précédent le mot ou le groupe de mots annoté.

	La sélection inclut uniquement le texte sur lequel doit porter l'annotation.
	Cette sélection n'est pas bonne car elle intègre un déterminant inutile dans le groupe de mots annoté.
	Cette sélection n'est pas bonne car elle inclut l'espace qui précède le mot sur lequel porte l'annotation.
	Cette sélection n'est pas bonne car elle inclut la virgule qui suit le mot sur lequel porte l'annotation.

3. La portion de texte sélectionnée doit être la plus correcte possible : elle ne doit pas inclure plus de mots que nécessaires, mais elle doit contenir l'ensemble des éléments du groupe de mots qui constituent l'entité identifiée.

	Cette sélection n'est pas bonne car elle ne prend qu'un seul mot en compte alors qu'il fait partie d'un groupe de mots formant une entité.
	Cette sélection n'est pas bonne car elle ne prend pas tous les mots du groupe formant une entité.
	Cette sélection est correcte car elle prend en compte la totalité de l'expression formant une entité et permet d'avoir une dénomination la plus précise possible.
	Cette sélection est correcte car elle prend en compte la totalité de l'expression formant une entité ; on perdrait des informations en n'intégrant pas "ouvriers" dans la sélection, de même si on ignorait "du tissage mécanique".

"Organization" [\[modifier\]](#)

Règle [\[modifier\]](#)

Le tag "organization" doit être utilisé pour annoter les institutions, les entreprises ou encore les associations.

Lorsqu'un groupe de mot désignant une organisation contient un nom de métier ou une occupation, on utilise également le tag "TU_occupation" sur la portion de texte concernée. En revanche, on n'annote pas le nom d'une personne dans le nom d'une organisation.

La désignation d'une organisation peut avoir plusieurs variantes, par exemple : "Maison Gindre" peut aussi être appelée "Gindre et Cie" ou encore "usine Gindre".

En fonction du contexte, une organisation peut servir à désigner un lieu plutôt que l'organisation elle-même. Autant que possible, on prendra en compte le contexte.

Exemples [\[modifier\]](#)

N°	Annotation correcte	Commentaires
1	Une gréviste de la <u>Maison Gindre</u>	L'annotation de l'organisation porte sur "Maison Gindre" ; on n'annote pas "Gindre" comme un nom de personne.
2	Le <u>siège du syndicat des ouvriers tisseurs et similaires</u> , rue d'Austerlitz 8, ...	On distingue ce qui relève de la dénomination d'un lieu (ici : "siège du syndicat des ouvriers tisseurs et similaires") et ce qui relève de l'organisation (ici : "syndicat des ouvriers tisseurs et similaires"). En outre, le nom de cette organisation comprend un nom de métier, que l'on annote avec le tag "TU_occupation" : "ouvriers tisseurs".
3	L'entrevue a eu lieu cette après-midi à 3 heures au <u>Palais du Commerce</u> .	Il est possible d'utiliser les tags "place" et "organization" simultanément. Dans ce contexte cependant, "Palais du Commerce" est avant tout utilisé pour désigner un lieu, on peut donc se contenter de l'annoter avec le tag "place" uniquement.

Quelles sont les informations que l'on souhaite annoter ?

- informations sur les rémunérations
- différentes entités (*personnes, métiers, lieux, ...*)
- informations sur les entités (*status matrimoniaux, type de rémunérations, ...*)
- informations contextuelles (dates)

Les 15 éléments d'annotation

date	lieu	adresse
organisation	personne	statut matrimonial
métier	références documentaires	heure
rémunération	type de rémunération	montant
tâche	produit	durée/rythme

- en plus : signalement des passages difficiles

Equivalents TEI

<date>	<placeName>	<adress>
<orgName>	<persName>	<rs type='status'>
<rs type='occupation'>	<bibl>	<time>
<rs type='revenue'>	<rs type='revenue' subtype='...'>	<measure type='sum'>
<rs type='task'>	<rs type='product'>	<rs type='duration'>

Annoter avec Transkribus : les tags

Points positifs :

- éviter des problèmes de migration et la multiplication des logiciels/outils
- le texte taggé peut être exporté au format XML, dans un fichier qui contient les tags et l'alignement texte/image

Points négatifs :

- l'export XML n'est pas de bonne qualité
- manque de documentation et de transparence

Retours sur les 2 phases d'annotation

1. Première phase d'annotation (Romane D.)

- annotations à corriger

2. Deuxième phase d'annotation (Marie L, Hugues S.)

- difficulté : distinguer concept et nom
- besoin : ajouter des tags

Pour la suite !

- Changer de logiciel pour l'annotation ? (TXM)
- Dynamiser le guide d'annotation ?
- Améliorer le modèle d'annotation et le guide
- Visualiser les annotations sur le wiki