

Un projet FNS

Ferdinand Online

Knowledge engineering models and tools
for the digital scholarly publishing of
manuscripts

Gilles Falquet, Claire Forel, Luka Nerima,
Giuseppe Cosenza

Université de Genève

Ferdinand de Saussure

- ❖ Scientifique genevois, 1857 - 1913
 - ❖ enseigne le cours de linguistique générale (CLG) à l'Université de Genève, considéré comme fondateur de la linguistique moderne
- ❖ mais
 - ❖ a très peu publié (rien sur le CLG)
 - ❖ nous a légué beaucoup de notes manuscrites



Buts du projet

❖ Un outil pour les linguistes, *facile à utiliser*, pour la

1. visualisation

2. transcription

3. annotation

des manuscrits de F. de Saussure

Visualisation: le manuscrit au centre !

Bienvenue : massimo

Déconnexion Mon compte

Ferdinand de Saussure - Archives numériques

http://saussure.com/ressource/test

ms_fr_03951_10_f030v_031

ms_fr_03951_10_f031v_032

ms_fr_03951_10_f032

ms_fr_03951_10_f031v_032

Annotations Textuelles Annotations Conceptuelles Transcriptions

De L'anti-historie du langage.

L'usage d'un objet double, mais double d'une façon qui semblerait inextricable, si nous ne recourions à une comparaison.

Dans une partie d'échecs, n'importe quelle position donnée a pour caractère singulier éminent d'être affranchie des antécédents. Donnée "historique", c'est-à-dire qu'il n'est pas "plus ou moins" indifférent, mais totalement indifférent qu'on en soit arrivé à cette telle position par une voie ou par une autre; en sorte que celui qui connaît tout ce qui s'est passé depuis le commencement de la partie a suivi toute la partie, n'a pas plus de lumière la moindre plus petite lumière de plus.

❖ Premier avantage -> éviter la manipulation physique des manuscrits

Transcription – pourquoi c'est si important ?

- ❖ ~ 50'000 pages manuscrites, jamais publiées, déposées dans les bibliothèques de Genève (BGE), Paris et Harvard
- ❖ pas de catalogue des manuscrits
- ❖ Solution ?
 - ❖ Reconnaissance d'écriture par ordinateur (OCR) ?
Impossible sur ce genre de manuscrit (pour le moment) !

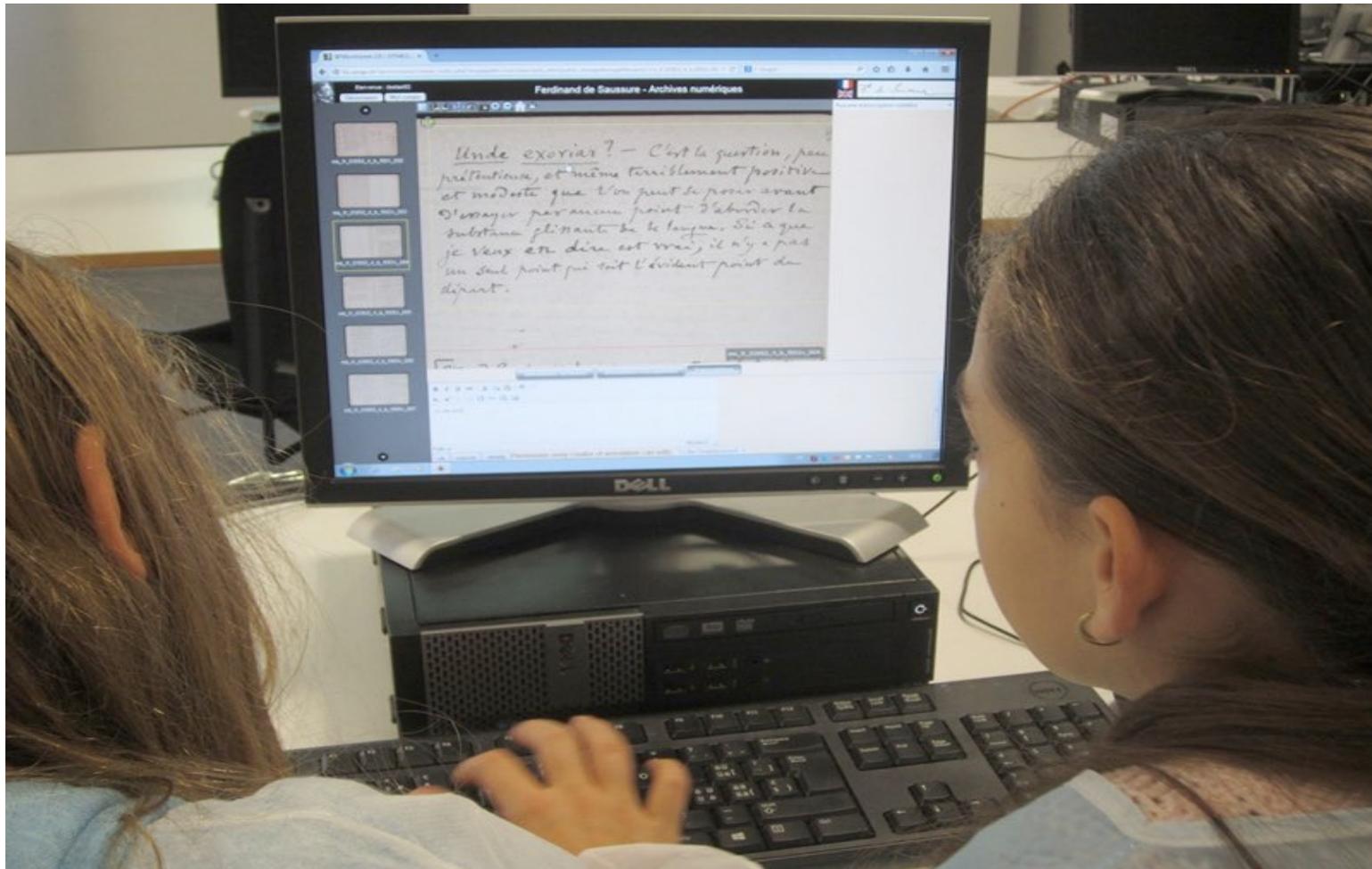
Transcription et OCR

Ces conclusions n'ont d'abord une portée
générale que parce qu'on insinue tacite-
ment que bien naïfs seraient ceux
qui ne verraient pas qu'à toute époque
en réalité, aussi bien qu'à la nôtre, un

Transcription et OCR

du roi
le roi
X et que ~~précédent~~
~~des~~ ~~précis~~, ou ~~la position~~ ^{la situation}
l'état
ne serait
du mot X
pas la même qu'elle est.
travaux
On ne compare rien, on
comporte une description et une
appréciation mathématique,
mais elle ne comporte pas
des dissertations flottantes,
Dans la science qui a l'honneur d'être la
Nous voyons en plein a priori, c'est-à-dire que
la diversité successive des combinaisons linguistiques
(dits états de langue) qui sont amenés par l'accident
sont éminemment comparables à la diversité des
situations d'une partie d'échecs. Or chacune de ces situations
d'avance classable, selon ^{si la} ~~la~~ ^{théorie} ~~théorie~~ ^{ou} ~~des~~ ^{des} ~~signes~~ ^{signes}.

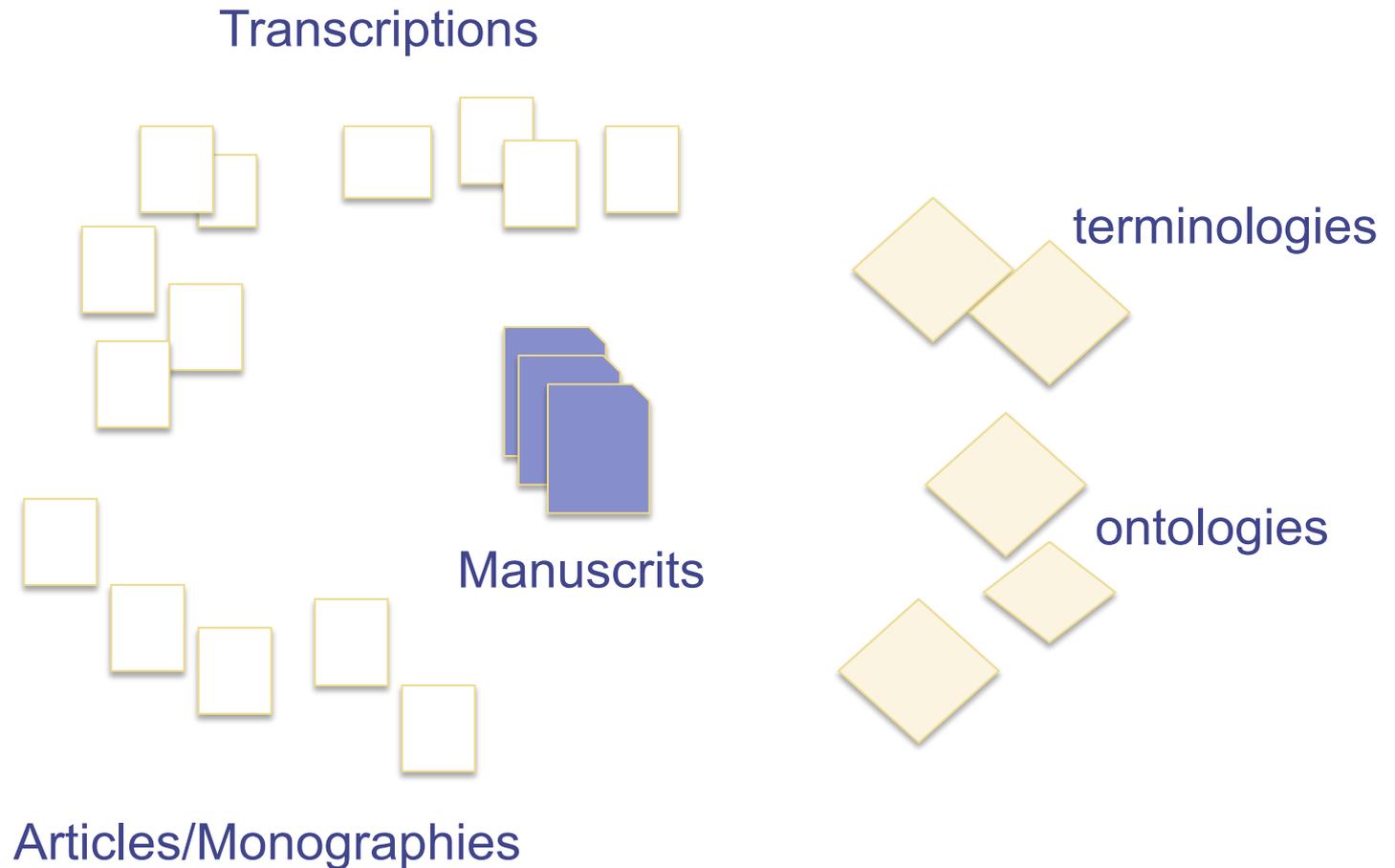
Le crowd sourcing



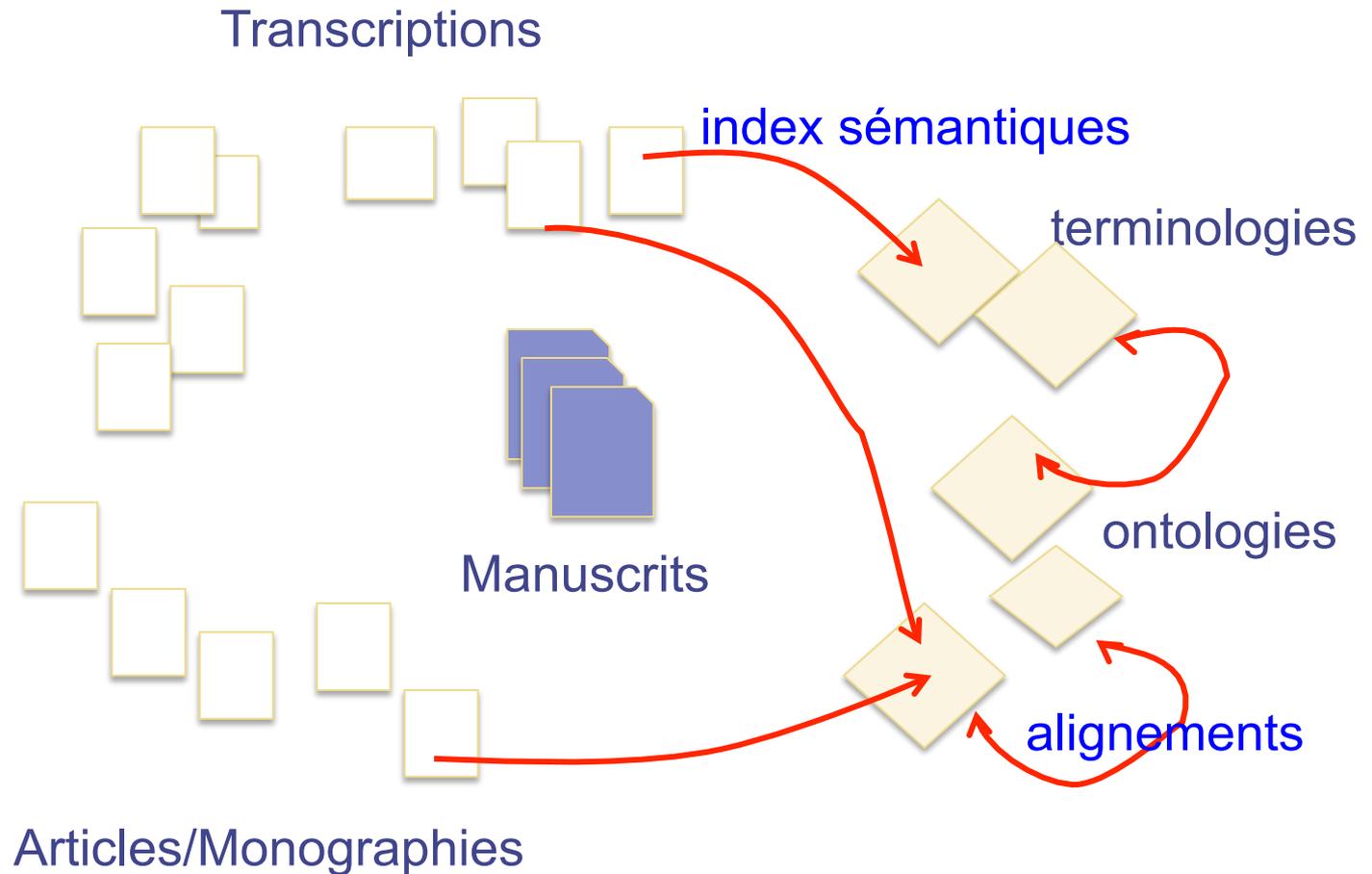
Transcription – résultats attendus

- ❖ recherche dans les manuscrits
- ❖ traitement automatique des langues appliqué aux manuscrits:
 - ❖ catalogue
 - ❖ terminologie
 - ❖ concepts <-> contenu
 - ❖ manuscrits <-> documents publiés
 - ❖ calcul de similarité entre documents

L'édition numérique comme un réseau de connaissance



L'édition numérique comme un réseau de connaissances



L'édition numérique comme un réseau de connaissance

