

L'anonymisation au service du repérage conceptuel dans le contexte de la SLA

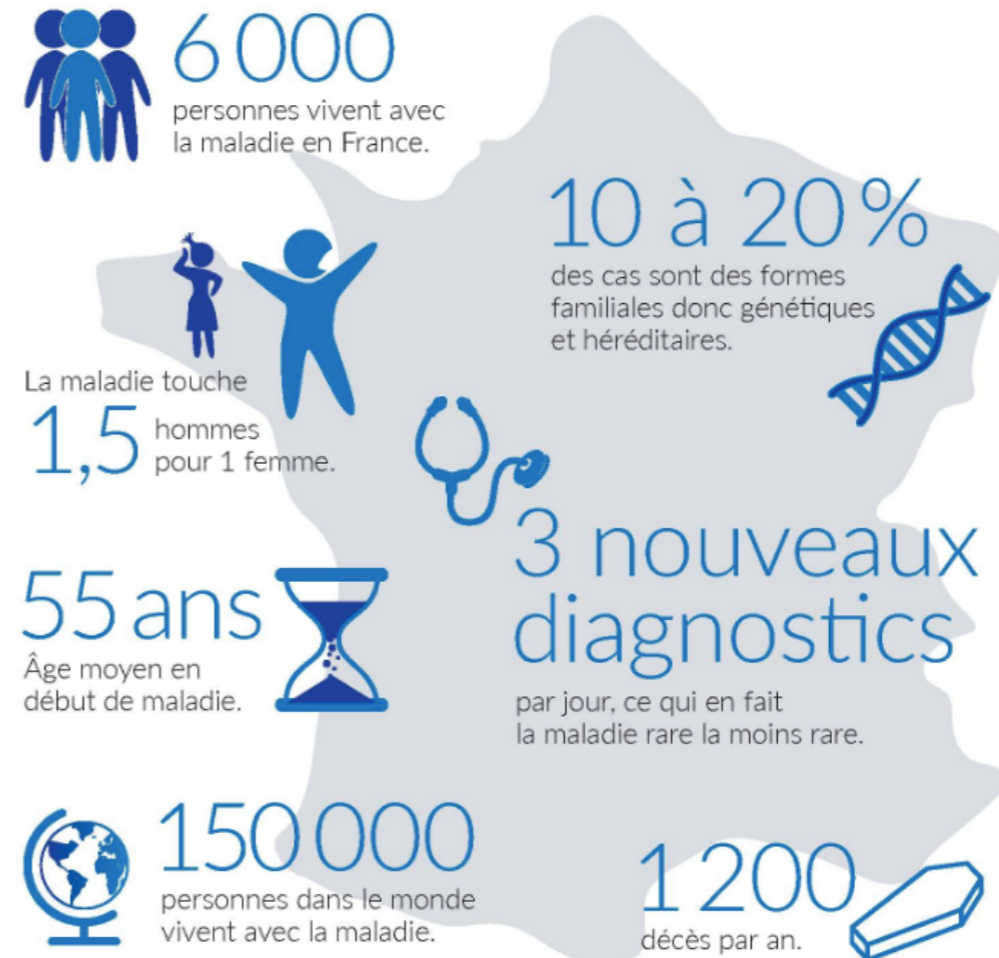
Sonia Cardoso, Jean Charlet

Pr. Vincent Meininger, Dr Grabli David, Marie Christine Jaulent
Luis Felipe Melo Mora, Xavier Aimé



Contexte

- La Sclérose Latérale Amyotrophique
 - Maladie Neurodégénérative
 - Grave décès dans les 3 à 5 ans après le diagnostic (*source : INSERM*)
 - Rare 800 nouveaux cas/ans (prévalence 5 à 8 / 100 000 *source : INVS 2016*)
 - Conséquences médicales et sociales
- Prise en charge spécifique, à Paris :
 - Centre expert SLA
 - Réseau SLA IDF, centre ressource et de coordination de parcours de santé, intervention auprès de nombreuses structures (sociales, médicales, médico-sociales)
 - 92% des patients ayant une SLA en IDF sont inclus dans le réseau



Base évènementielle

Contenu : Ensemble de transcription des échanges (écrit, oraux) :

- des demandes
- des actions de coordination
- de données médicales (compte rendu), sociales, environnementales

Caractéristiques des évènements :

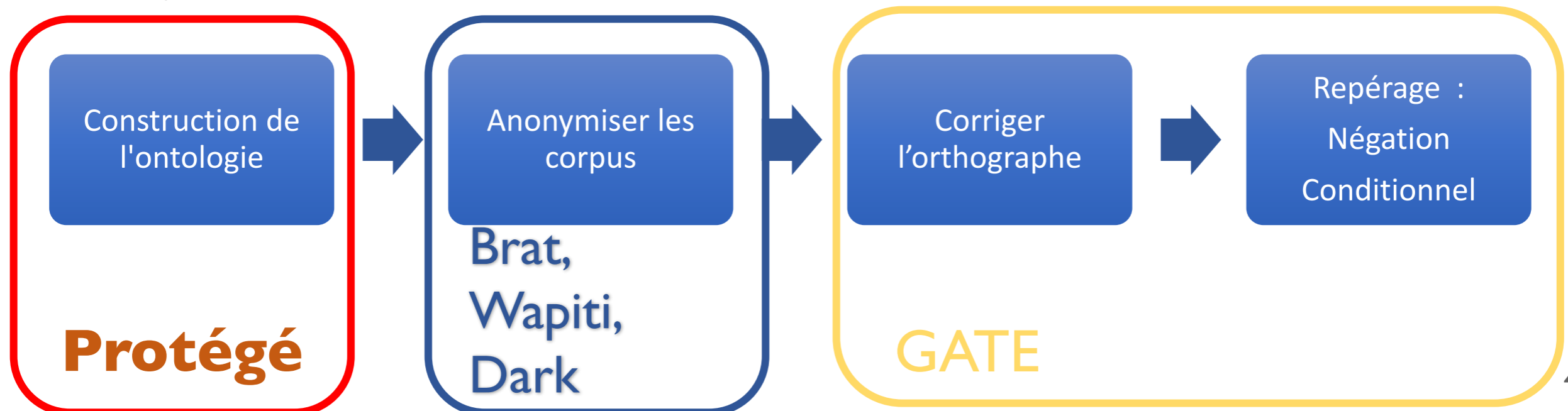
- Polymorphes
- Non structurés
- Chronologiques
- Nominatifs
- Complets et riches d'informations

2310 Dossiers, 35 621 évènements.

DUPONT BRIGITTE: rappel fait au patient au sujet de son dossier MDPH Information prise: il a bien le dossier MDPH du **99** donné par **Marie José pierre**, mais n'a pas de certificat medical du **DR CERVEAU**. Appel fait a **MME JEANNE** pour l'informer de cet echange: message laisse sur sa messagerie

Hypothèse et Méthodologie

- **Hypothèse de recherche** : l'analyse du corpus :
 - Décrire et comprendre les parcours de santé des patients SLA
 - Identifier s'il existe des indicateurs de ruptures
- **Objectif** : améliorer les parcours de santé et diminuer les hospitalisations (évitables)
- **Méthodologie** :
 - Utilisation d'une ontologie du domaine
 - Utilisation des outils de TALN pour annoter les corpus
 - Etapes intermédiaires nécessaires :



Construction de l'ontologie

- Modélisation des actions de coordinations par l'utilisation des script de Schank.
- Catégorisation des concepts : **Agents, Processus, Objets, Etat.**
- Intégration de concept du domaine :
 - **médical (neurologie)**
 - **social**
 - **environnemental**
 - **médico-social**
 - **de coordination**
- Recherche de l'ensemble des synonymes et acronymes, utilisés dans les corpus pour définir un même concept (*Prescription médicale, PM, Ordonnance*).



Anonymisation

Anonymiser CNIL : retirer suffisamment d'éléments pour que la personne concernée ne puisse plus être identifiée.

Définition et catégorisation des entités à anonymiser (22 catégories)

Agents

- Patient
- Neurologue
- Entourage
- ...

Structures

- Hôpitaux
- Associations
- Structures Médico-Sociales

Lieux

- Villes
- Départements
- Pays

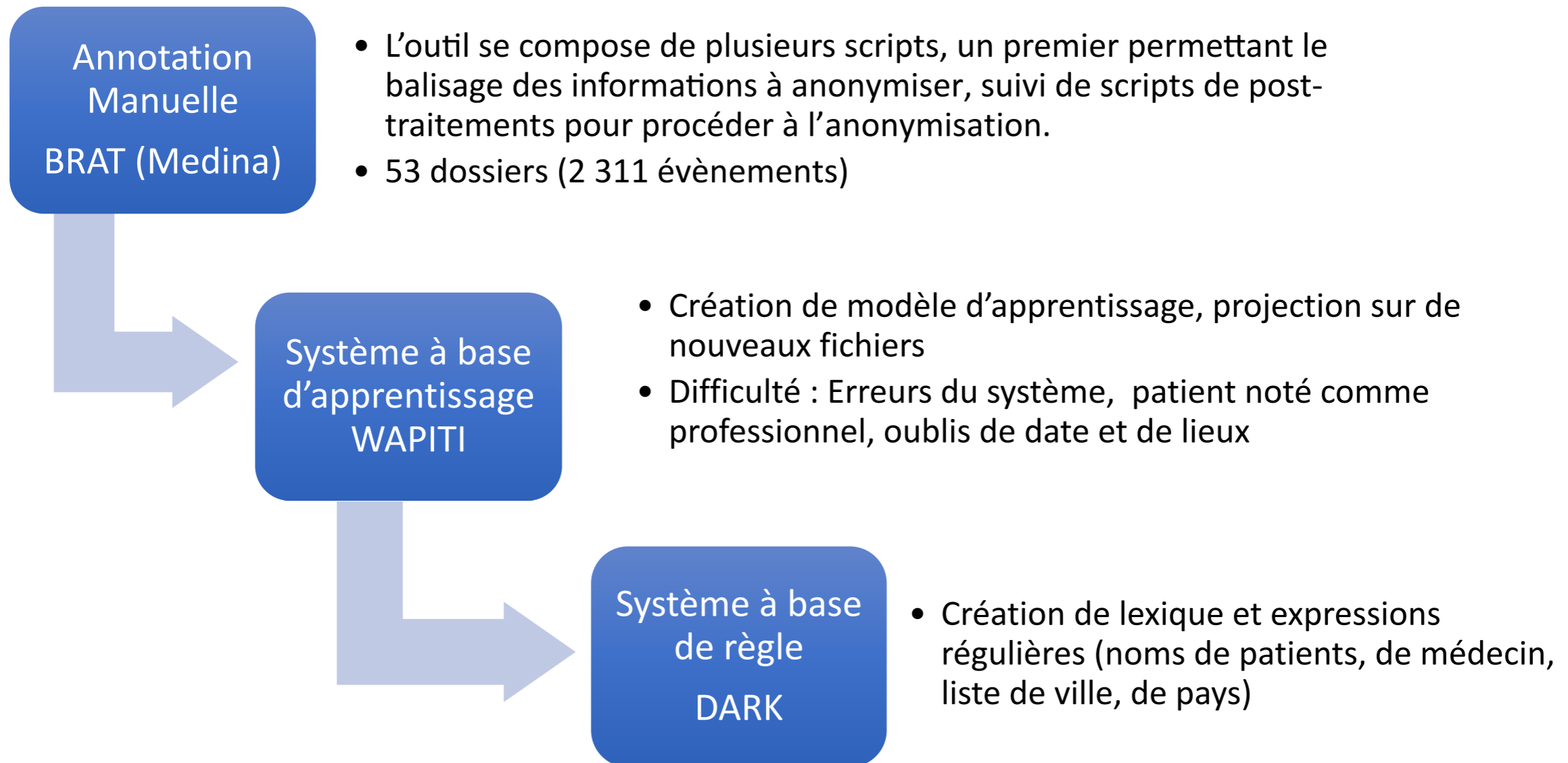
Identifiants Numériques

- N° de téléphone
- N° de Dossier
- ...

Dates

- Format numérique
- Format alphabétique

Outils (LIMSI) utilisés et méthodologie d'anonymisation

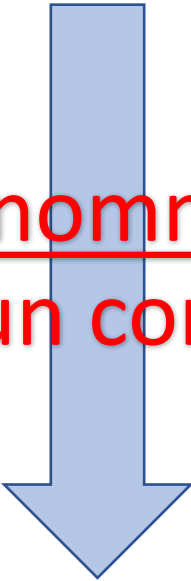


Résultat des 2 systèmes couplés :

Rappel de 0,81 ; Précision de 0,90 ; F-mesure de 0,85

DUPONT BRIGITTE: rappel fait au patient au sujet de son dossier MDPH Information prise: il a bien le dossier MDPH du **99** donné par **Marie José pierre**, mais n'a pas de certificat medical du **DR CERVEAU**. Appel fait a **MME JEANNE** pour l'informer de cet echange: message laisse sur sa messagerie

Transformation d'une entité nommée en catégorie sémantique faisant référence à un concept de l'ontologie.



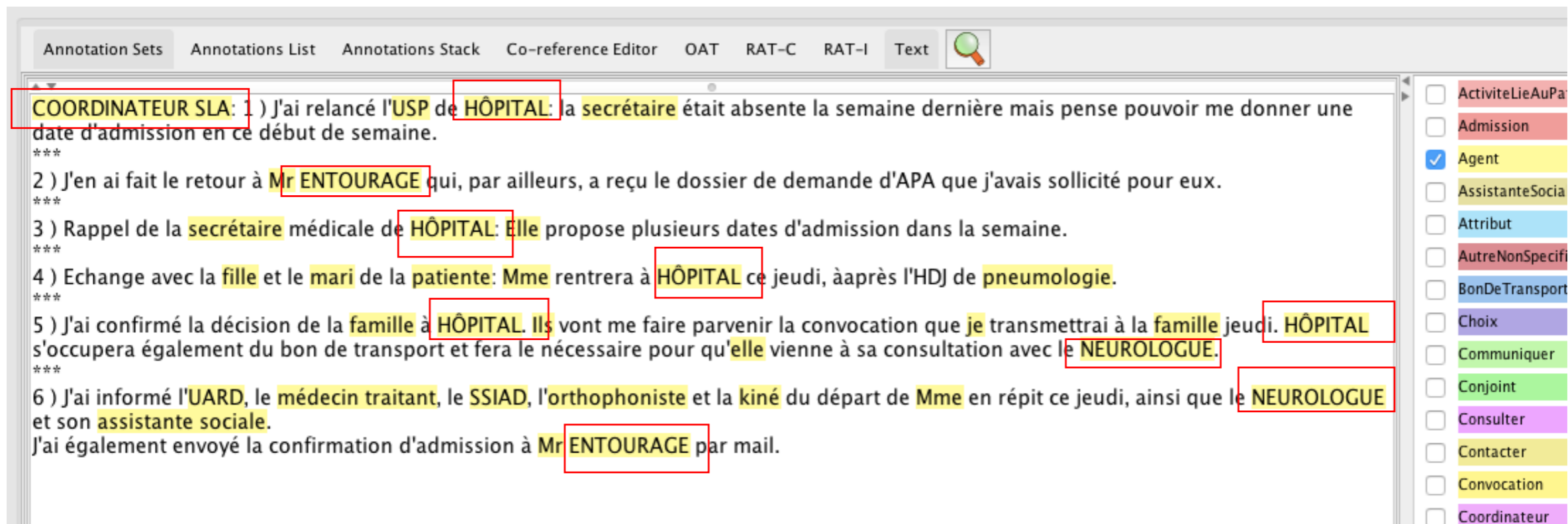
Coordinatrice SLA: rappel fait au patient au sujet de son dossier MDPH Information prise: il a bien le dossier MDPH du Département donné par Assistante-Sociale, mais n'a pas de certificat medical du Neurologue. Appel fait a Assistante-Sociale pour l'informer de cet echange: message laisse sur sa messagerie

Apport de l'anonymisation

- **Amélioration du repérage** des « *agents* » acteurs dans le parcours de santé
 - **Compréhension des interactions** et actions des « *agents* »
 - **Qui sollicite** le réseau , qui sont les demandeurs : le patient, l'entourage, les soignants de ville
 - **Qui informe** d'une dégradation de la situation clinique ou familiale (épuisement de l'aidant)
 - Pour **quel motif** : médical, social, environnemental
- Permet de créer des corpus utilisables pour des recherches

Utilisation de la plateforme GATE

- Gate utilise l'ontologie du domaine comme ressource lexicale
- Création d'une chaîne de traitement pour améliorer la qualité du corpus (Correction d'orthographe).
- Création d'une chaîne de traitement pour l'annotation.



The screenshot displays the GATE software interface with the following text and annotations:

COORDINATEUR SLA: 1) J'ai relancé l'USP de HÔPITAL: la secrétaire était absente la semaine dernière mais pense pouvoir me donner une date d'admission en ce début de semaine.

2) J'en ai fait le retour à Mr ENTOURAGE qui, par ailleurs, a reçu le dossier de demande d'APA que j'avais sollicité pour eux.

3) Rappel de la secrétaire médicale de HÔPITAL: Elle propose plusieurs dates d'admission dans la semaine.

4) Echange avec la fille et le mari de la patiente: Mme rentrera à HÔPITAL ce jeudi, à après l'HDJ de pneumologie.

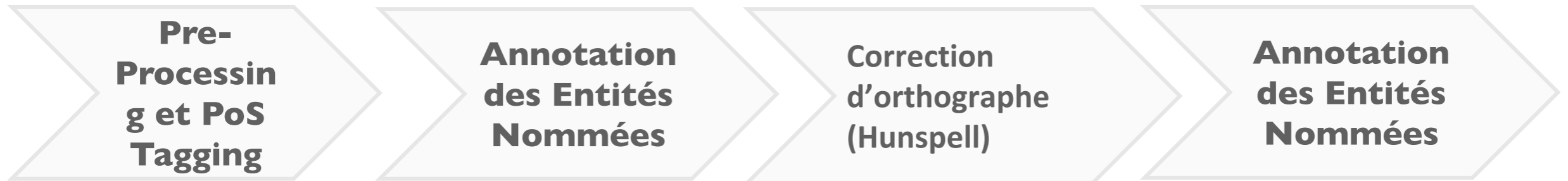
5) J'ai confirmé la décision de la famille à HÔPITAL. Ils vont me faire parvenir la convocation que je transmettrai à la famille jeudi. HÔPITAL s'occupera également du bon de transport et fera le nécessaire pour qu'elle vienne à sa consultation avec le NEUROLOGUE.

6) J'ai informé l'UARD, le médecin traitant, le SSIAD, l'orthophoniste et la kiné du départ de Mme en répit ce jeudi, ainsi que le NEUROLOGUE et son assistante sociale.
J'ai également envoyé la confirmation d'admission à Mr ENTOURAGE par mail.

On the right side, a list of ontology classes is shown with checkboxes:

- ActiviteLieAuPa
- Admission
- Agent
- AssistanteSocia
- Attribut
- AutreNonSpecifi
- BonDeTransport
- Choix
- Communiquer
- Conjoint
- Consulter
- Contacter
- Convocation
- Coordinateur

Chaine de traitement de GATE : Correction orthographique



■ Basé sur:

- Les suggestions de **Hunspell** (dictionnaire le plus souvent utilisé. Ex: LibreOffice, Firefox).
- Une mesure de similarité (**Optimal string alignment distance** pondérée par la disposition des touches du clavier).

■ Apporte une **amélioration** du repérage des entités nommées.

Problèmes

■ *“Real word spelling errors”*

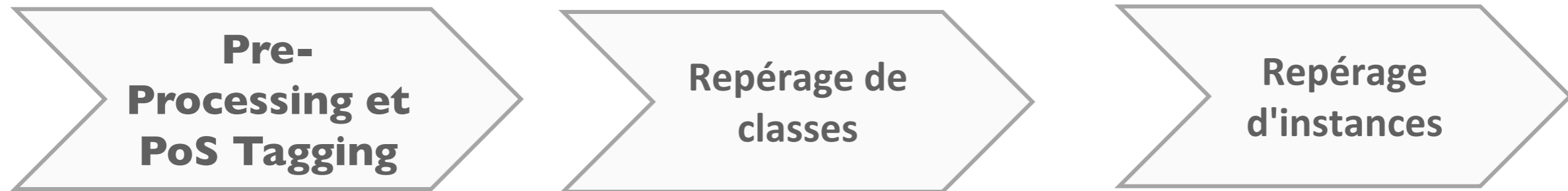
Erreurs qui se produisent quand un utilisateur tape (par erreur) un **mot correctement orthographié** quand un **autre mot était destiné**.

■ *“Non real word spelling errors”*

Erreurs qui se produisent quand un utilisateur tape (par erreur) un **mot incorrectement orthographié**.

■ *Ensemble des termes spécifiques (marque de machine, de fauteuil)*

Chaine de traitement de GATE : Annotation des entités nommées



- Pour palier le bruit des différentes **formes lexicales** (féminin, pluriel,...) : **Lemmatization** (Tree Tagger)
- Repérage des **concepts** de l'ontologie : "PrefLabel", "altLabel"
- Repérage des **instances** de concept : "classURI"
- Repérage de la **plus longue entité**.

■ Problèmes : Ambiguïté

Une chaîne peut représenter un concept de l'ontologie ou un mot en français : « **ces** » peut se référer

« ***l'adjectif démonstratif*** » ou bien « ***Carte Européenne de Stationnement*** »

PERSPECTIVES

- Processus itératif continu d'enrichissement des différents modules (noms des patients, médecins, structures...)
- Développement de l'extraction et du repérage des relations :
 - Extraction de **patrons séquentiels**.
- Intégration de la négation et du conditionnel
- Orientation vers une ontologie modulaire
- Transposition des outils développés à d'autres bases de coordination (Parkinson et Traumatisme Crânien)
- Évaluation

Merci de votre attention

Références

TALN

- H. Cunningham, et al. Getting more out of biomedical documents with GATE's full lifecycle open source text analytics. Plos Comput Biol. 2013.
- Deléger, L. et al. Detecting negation of medical problems in french clinical notes. IHI 2012
- Harkema, H. et al. ConText: An algorithm for determining negation, experiercer, and temporal status from clinical reports. Journal of Biomedical Informatics. 2009.
- Abbe, A. et al. Text mining applications in psychiatry: A systematic literature review. Int J Methods Psychiatr Res. 2015.
- Ben Abacha, A. et al. Automatic extraction of semantic relations between medical entities: Application to the treatment relation. Semantic Mining in Biomedicine 2010.
- Béchet, N. et al. Sequential pattern mining to discover relations between genes and rare diseases. CBMS 2012.
- Thompson, P. et al. Customised OCR correction for historical medical text. Proceedings of DigitalHeritage 2015.
- **Parcours de Santé et Ontologie**
 - Bachimont B., Isaac A. & Troncy R. (2002) Semantic Commitment for Designing Ontologies: A Proposal. In A. Gomez- Pérez & V. Benjamins, Eds., 13th International conference on knowledge Engineering and Knowledge Management (EKAW'02), volume 2473 of Lecture Notes in Artificial Intelligence, p.114-121, Sigüenza, Espagne: Springer Verlag.
 - Bondenreider O. (2008) Biomedical Ontologies in Action: Role in Knowledge Management, Data Integration and Decision Support. Yearbook of medical informatics, p 67-69
 - Cardoso S., Aimé X., Melo Mora L-F., Jaulent M-C., Grabli D., Meininger V., Charlet J. (2016) Les ontologies pour aider à comprendre les parcours de santé dans le cadre des maladies neurodégénératives. Conférence: IA & Santé 2016 - Deuxième Atelier sur l'Intelligence Artificielle et la Santé, At Montpellier
 - Charlet J., Declerck G., Dhombres F., Gayet P., Miroux P. et Vandebussche P-Y. (2012) Construire une ontologie médicale pour la recherche d'information : problématiques terminologiques et de modélisation. In : Szulman S., coordinateur. Actes des 23es Journées Ingénierie des Connaissances, Paris, France, 27-29 juin, p. 33-48.
 - Charlet J., Bachimont B., Mazuel L., Dhombres F., Jaulent M. et Bouaud J. (2012) OntoMenelas : motivation et retour d'expérience sur l'élaboration d'une ontologie noyau de la médecine. Technique et Science Informatiques, 31(1).
 - Charlet J., Darmoni S -J. (2015) Knowledge Representation and Management. From Ontology to Annotation. Findings from the Yearbook 2015 Section on Knowledge Representation and Management. Yearbook of Medical Informatics, p 134-136.
 - Cordesse V., Sidorock F., Schimmel P., Holstein J., Meininger V. (2015) Coordinated care affects hospitalization and prognosis in amyotrophic lateral sclerosis: a cohort study. BMC Health Services Research
 - Drame K., Diallo G., Delva F., Dartigues J-F., Mouillet E., Salamon R., Mouglin F. (2014) Reuse of termino-ontological ressources and text corpora for building a multilingual domain ontology: An application to alzheimer's disease. Journal of Biomedical Informatics 48, 171-182
 - Grouin C. (2013) Anonymisation de documents cliniques : performances et limites des méthodes symboliques et par apprentissage statistique, PhD Thèse. Université Pierre et Marie Curie, Paris, France.
 - Grosjean, J; Merabti, T; Dahamna, B; Kergourlay, I; Thirion B; Soualmia LF & Darmoni, SJ. (2011) Health Multi-Terminology Portal: a semantics added-value for patient safety. Patient Safety Informatics - Adverse Drug Events, Human Factors and IT Tools for Patient Medication Safety, Studies in Health Technology and Informatics, Volume 166, Pages 129-138.
 - Gomez-Perez, A., Fernandez M. and De Vicente A.J. (1996) "Towards a Method to Conceptualize Domain Ontologies", ECAI-96 Workshop on Ontological Engineering, Budapest.
 - Lavergne T., Cappé O., Francois Y. (2010) Pratical VeryLarge Scale crfs. Proceedings the 48th Annual Meeting of the Association for Computational Linguistics, 504-513. Uppsal, Sweden
 - Popoejoy LL., Khalilia M-A., Popescu M., Galambos C., Lyons V., Rantz M., Hicks L., Stetzer F., (2014) Quantifying care coordination using natural language processing and domain-specific ontology. Journal of the American Medical Informatics Association, Volume 22, p 93-103.
 - Pilone D., Pitman N. (2006) UML 2 en concentré. Edition O'Reilly, Paris.
 - Richard M., Aimé X., Krebs M.-O. & Charlet J. (2013) Au-delà du DSM : les ontologies comme aide aux classifications descriptives psychiatriques ? 2e édition du Symposium sur l'Ingénierie de l'Information Médicale, Jul 2013, Lille, France.