



Orléans, 28 novembre 2019

Claire Gardent, CNRS/LORIA, Nancy

Denis Paperno, Utrecht University, Pays-Bas

Qu'est ce que
c'est?

LIFT

Linguistique Informatique, Formelle et de Terrain

Un Groupement de Recherche (GdR) du CNRS :

- de l'Institut des Sciences Humaines et Sociales (INSHS)
- 133 membres, 28 laboratoires
- ciblé sur l'interaction entre linguistique informatique, linguistique formelle et linguistique de terrain
- Parallèle au **pré-GdR TAL**, avec des points de convergence et un recouvrement important en termes de personnes

Qu'est ce que
c'est?

Objectifs

Rassembler les trois communautés

Favoriser l'émergence de nouvelles méthodes

- Pour les linguistes: *vérification automatique des modèles formels*
- Pour les linguistes de terrain : *faciliter le processus de collecte, d'analyse et de modélisation des données*
- Pour les linguistes informaticiens : développer de *nouvelles techniques nécessitées par l'analyse linguistique* et des *méthodes non ou faiblement supervisées* pour l'analyse des langues peu dotées, peu écrites ou non documentées

Collaborer avec le pre-GDR TAL

- notamment sur le développement de *méthodes faiblement supervisées* pour les langues peu dotées.

Historique

- Proposition

Les grandes étapes

- 11/17 : Préparation
Hamida Dermidache (DAS INSHS), Claire Gardent, Denis Paperno
- 05/12/17 : Appel à manifestation d'intérêt
- Décembre 17 - Janvier 2018 : Rédaction collaborative
- 26/01/2018 : Soumission de la proposition
- 26/02/2018 : Collecte des signatures
- 14/06/2018 : Acceptation informelle par le comité de direction du CNRS
- 19-21/06/2018 : Examen du dossier par la section 34
- 12/12/2018 : Notification d'acceptation

Historique

- Proposition

- Thématiques

Axe 1 : Extraction automatique de généralisations linguistiques

Berthold Cysmann, Thierry Poibeau

- promouvoir l'utilisation des méthodes de TAL dans la *découverte d'invariants et de généralisations linguistiques*
- stimuler le *développement de méthodes informatiques permettant l'extraction de généralisations* à partir de données (UD Treebanks, données de terrain, etc.)

Axe 2 : Linguistique et évaluation

Karën Fort, Guillaume Wisniewski

- proposer des *critères linguistiques pour évaluer*, de façon automatique, la qualité linguistique des résultats produits par les systèmes de TAL.
- construire, à partir des recherches conduites en linguistique formelle, des *batteries de tests linguistiques* qui permettent d'évaluer les systèmes de TAL

Historique

- Proposition

- Thématiques

Axe 3 : Outils de collecte et d'analyse pour les linguistes

Alexis Michaud, Tatiana Nikitina et Angélique Amélot

- *promouvoir l'utilisation des outils, ressources et techniques TAL* qui peuvent faciliter le travail des linguistes et en particulier des linguistes de terrain
- *concevoir de nouveaux outils* et accompagner leur développement et leur déploiement (e.g.,n LIG-AIKUMA Grenoble/Darwin U.)

Historique

- Proposition

- Thématiques

Axe 4 : Données et défis partagés pour une science ouverte

Valentin Vydrine

- collecter les *données* de travail des membres du GDR dans un *espace partagé*
- *favoriser le partage et la réutilisabilité* des données linguistiques (batteries de tests linguistiques, données orales et écrites collectées par les linguistes de terrain)
- faciliter et standardiser les processus d'*annotation* des données
- définir et organiser des *défis scientifiques communs*
- définir des *thématiques de travail communes* comme par exemple, le traitement multi niveaux (oral, lexicale, morphologie, syntaxe) d'une langue non documentée.
- promouvoir, dans la communauté linguistique, une culture de publication où le *partage des données* devient un élément intégral de la communication scientifique, et où la réutilisation des données est reconnue par citation

Historique

- Proposition

- Thématiques

Axe 5 : Linguistique informatique pour les langues peu dotées ou non documentées

Laurent Besacier, Emmanuel Schang et Anaïs Lefeuvre-Halftermeyer

- développer des *méthodes d'apprentissage faiblement, semi ou non supervisées* pour des données de taille réduite provenant de langues peu dotées, peu écrites ou non documentées
- concevoir des *méthodes d'expansion automatique des données* permettant d'appliquer des méthodes d'apprentissage automatique à des données de petite taille
- utiliser et d'adapter les *méthodes symboliques (grammaires computationnelles, automates à états finis) pour l'analyse et la génération de langues peu dotées* (afin par exemple de valider une grammaire et de tester sur- et sous-génération)
- exploiter des méthodes d'apprentissage automatique pour *créer des moteurs de gloses* permettant de minimiser les prétraitements.

Historique

- Proposition

- Avril 2019
Mise en ligne du [site web](https://gdr-lift.loria.fr/)
<https://gdr-lift.loria.fr/>
Compte Twitter : @GdrLift
Listes : lift_members@inria.fr, lift_resp@inria.fr

- Thématiques

- Juin 2019, Paris
[Journée ouverte](#)

- Événements

- Aout 2019, Nancy
[Ecole d'été Python4NLP](#)
- Novembre 2019, Orléans
[Journées inaugurales](https://lift2019.sciencesconf.org/)
<https://lift2019.sciencesconf.org/>
[Emily Bender](#), U. of Washington, USA. Computational Syntax.
[Sabrina Benjaballah](#), U. Nantes. Morpho-phonologie, Langues afroasiatiques et germaniques.
[Steven Bird](#), Charles Darwin U., Australie. Evolution of the world's languages.
[Michel Jacobson](#), Huma-Num. Archivage des données.

Groupes de travail

Langues et Variations

- Regards croisés sur les variations : vers une formalisation ?
Contacts : Alice Millour et Karèn Fort
 - confronter les expériences de variations (en linguistique, linguistique de terrain, TAL), qu'elles soient diachroniques, dialectales, orthographiques, code-switching, de registre familier/formel, un recouvrement de variations, etc.
 - initier une discussion sur comment ces variation(s) entre dans (ou interfère avec) le périmètre d'intérêt du GDR (entre autre la collecte de corpus, l' extraction de généralisations linguistiques, l'évaluation de systèmes de TAL).

Contractions, Alternances phonologiques et morphologiques

- Catalogue cross-lingue des alternances, modélisation formelle et informatique
Contact : Emmanuel Schang

2020

Perspectives

- Ecole d'été Python4NLP
- Ecole thématique CNRS "Annotations"
Proposition soumise le 14/06/2019
- Groupes de travail
 - Glosses : standardisation et exploitation automatique
 - Méthodes automatique pour la transcription de l'oral de langues peu documentées
 - ...
- *Votre proposition ?*

Participation

Engagez vous !

- S'abonner à la liste des membres
Mail à : sympa_inria@inria.fr
Sujet : subscribe_lift_members@inria.fr Prénom Nom
Laissez le corps du message vide.
- Nous suivre sur Twitter : [@GdrLift](https://twitter.com/GdrLift)
- Remplir le formulaire de contact : <https://gdr-lift.loria.fr/contact/>
- Parler avec l'un des responsables
- Lancer un groupe de travail, de lecture
- Organiser une journée d'étude, une tâche partagée, la création d'une ressource (corpus, grammaire, lexique ...)

Responsables

Angelique Amelot	angelique.amelot@sorbonne-nouvelle.fr
Laurent Besacier	Laurent.Besacier@imag.fr
Berthold Crysmann	berthold.crysmann@gmail.com
Karën Fort	karen.fort@sorbonne-universite.fr
Claire Gardent	claire.gardent@loria.fr
Anais Halftermeyer	anais.halftermeyer@univ-orleans.fr
Alexis Michaud	alexis.michaud@vjf.cnrs.fr
Tatiana Nikitina	tavnik@gmail.com
Denis Paperno	d.paperno@uu.nl
Thierry Poibeau	thierry.poibeau@ens.fr
Emmanuel Schang	emmanuel.schang@univ-orleans.fr
Valentin Vydrine	vydrine@gmail.com
Guillaume Wisniewski	guillaume.wisniewski@limsi.fr

Merci à

Emmanuel Schang, Alexis Michaud

Emilie M. Bender, Sabrina Benjaballah, Steven Bird and Michel Jacobson

Vous