



Nancy, 21 novembre 2023

Claire Gardent, CNRS/LORIA, Nancy

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
- Parallèle au **GdR TAL (INS2I)**, avec des points de convergence

LIFT

Historique

Les grandes étapes

- Janvier 2017 : Appel à manifestation d'intérêt

Les grandes étapes

- Janvier 2017 : Appel à manifestation d'intérêt
- Décembre 2017 - Janvier 2018 : Rédaction collaborative

Les grandes étapes

- Janvier 2017 : Appel à manifestation d'intérêt
- Décembre 2017 - Janvier 2018 : Rédaction collaborative
- Décembre 2018 : Notification d'acceptation

Les grandes étapes

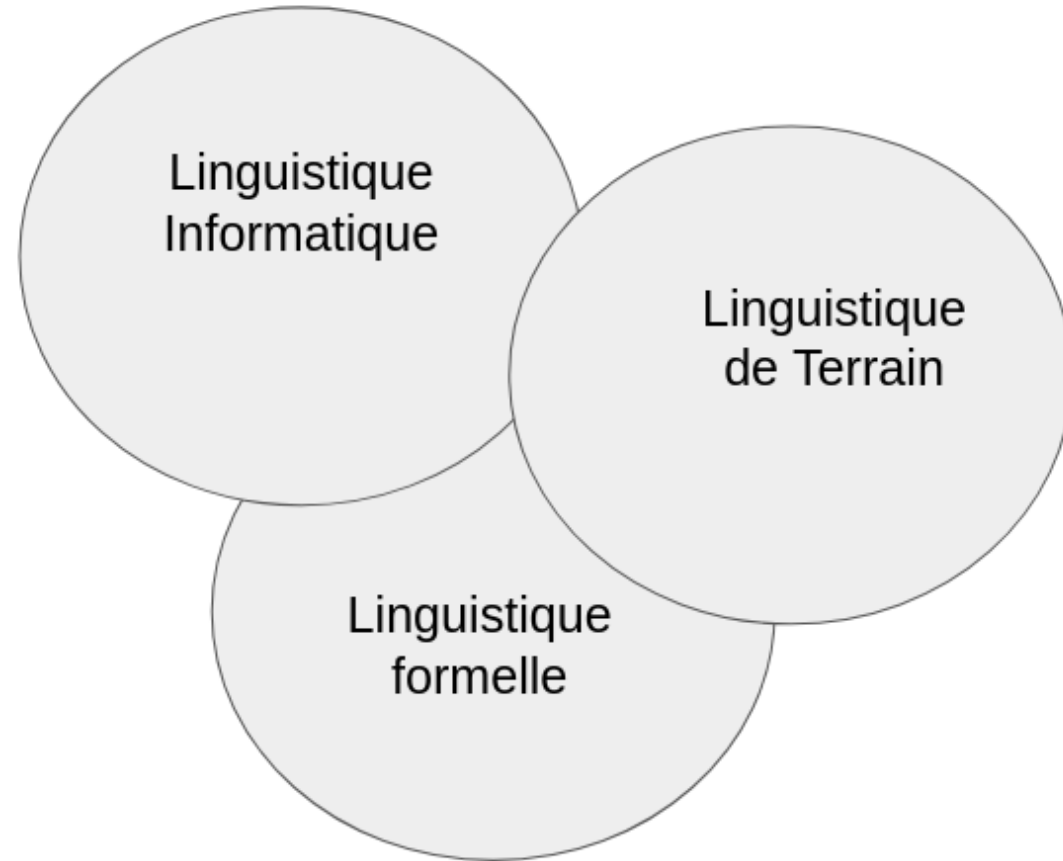
- Janvier 2017 : Appel à manifestation d'intérêt
- Décembre 2017 - Janvier 2018 : Rédaction collaborative
- Décembre 2018 : Notification d'acceptation



LIFT

Favoriser les interactions entre trois communautés

Objectif



Thématiques

Linguistique informatique / Linguistique Formelle

- Découverte d'invariants et de généralisations linguistiques à partir de données (UD Treebanks, données de terrain, etc.)
- Quels outils, ressources et techniques TAL pour faciliter le travail des linguistes et des linguistes de terrain?

Linguistique formelle / Linguistique informatique

- Critères linguistiques et création de batteries de tests linguistiques pour l'évaluation des systèmes de TAL
- Que disent les approches neuronales sur les langues naturelles ?
- Quelle place pour les grammaires formelles dans la modélisation informatique des langues du monde ?

Thématiques

Linguistique de terrain / Linguistique informatique

- *méthodes d'apprentissage faiblement, semi ou non supervisées* pour des données de taille réduite provenant de langues peu dotées, peu écrites ou non documentées
- *méthodes d'expansion automatique des données* permettant d'appliquer des méthodes d'apprentissage automatique à des données de petite taille
- utiliser et adapter les *méthodes symboliques (grammaires computationnelles, automates à états finis)* pour l'analyse et la génération de langues peu dotées
- exploiter des méthodes d'apprentissage automatique pour *créer des moteurs de gloses* permettant de *minimiser les prétraitements* ou pour en *extraire des connaissances linguistiques* .

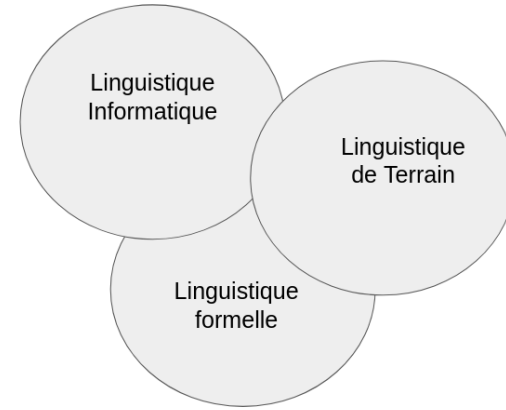
Thématiques

Science ouverte

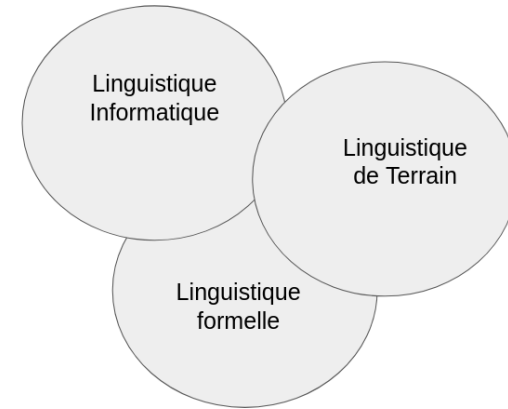
- collecter les *données* de travail des membres du GDR dans un *espace partagé*
- *favoriser le partage et la réutilisabilité* des données linguistiques (batteries de tests linguistiques, données orales et écrites collectées par les linguistes de terrain)
- faciliter et standardiser les processus d'*annotation* 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 le *partage des données*

LIFT

Journées d'études



- Juin 2019, Paris
Journée ouverte
- Novembre 2019, Orléans
Journées LIFT
Emily Bender, Sabrina Benjaballah, Steven Bird, Michel Jacobson
- Décembre 2020, Paris (virtuel)
Journées LIFT
Syntaxe et langues de terrain,
Grammaires formelles
Katya Aplonova, Lionel Clément, Annie Forêt, Kim Gerdes, Sylvain Loiseau, Aleksandra Miletic
- 01 Juin 2021 (Virtuel)
Séminaire ILFC
Interactions between Computational and Formal Linguistics *Juan Luis Gastaldi, Koji Mineshima, Maud Pironneau, Marie-Catherine de Marneffe, Jacob Andreas, Olga Zamaraeva*



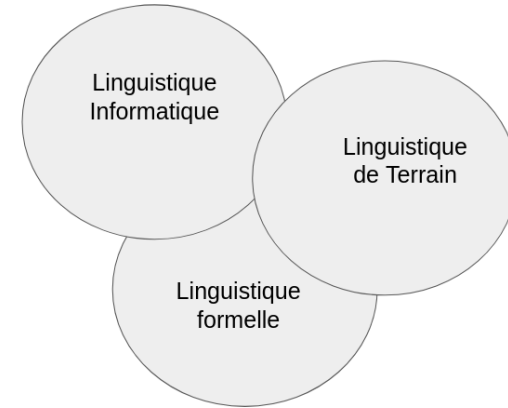
- Décembre 2021, Grenoble
Journées LIFT
Laurent Besacier, Alda Mari, Daan van Esch, Matti Miestamo
- Novembre 2022, Marseille
Journées LIFT+TAL
"Ressources", "Tâches et langues peu dotées"
Antonios Anastasopoulos, Rachel Bawden, Karën Fort, Alexis Michaud

Seminaire Mensuel ILFC (Virtual)

Interactions entre linguistiques formelles et Computationnelles

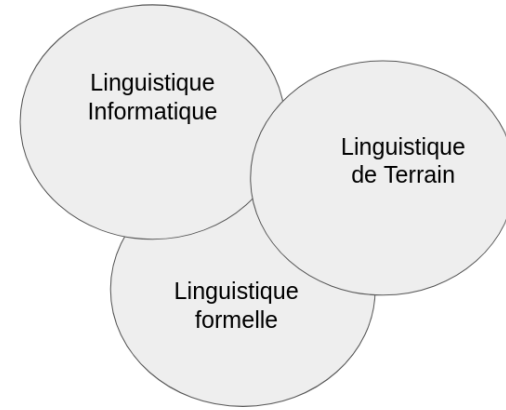
Timothée BERNARD et Grégoire WINTERSTEIN

- 2021/10/12: Christopher Potts (Stanford University)
Causal Abstractions of Neural Natural Language Inference Models
- 2021/11/16: Alex Lascarides (University of Edinburgh) Situated Communication
- 2021/12/14: Lisa Bylinina (Bookarang, Netherlands) Polarity in multilingual language models (en commun avec Alexey Tikhonov)
- 2022/01/18: Johan Bos (Rijksuniversiteit Groningen) TBA



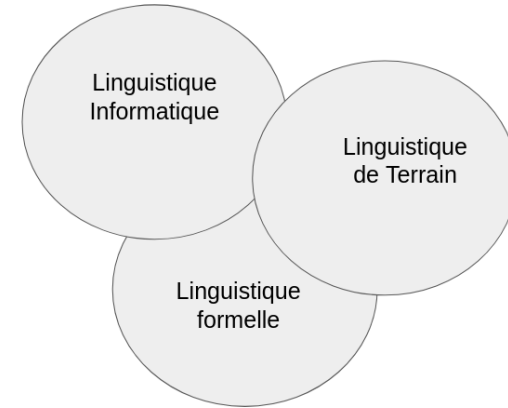
LIFT

Ecoles d'été



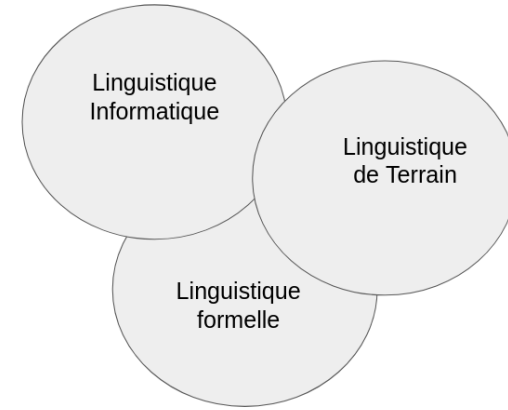
Ecoles d'été

- Python4NLP
26-20 aout 2019, Nancy
- Python4NLP
15-19 juin 2020, Nancy (ANNULE)
- Ecole thématique CNRS "Annotations"
juillet 2019, Roscoff (ANNULE)
- Python4NLP
20-21 mai 2022, Nancy
- Ecole thématique CNRS "Annotations"
30 Mai-3 juin 2022, Banyuls



Novembre 2021
Archivage de données, Orléans
Tutoriels et Travaux Pratiques

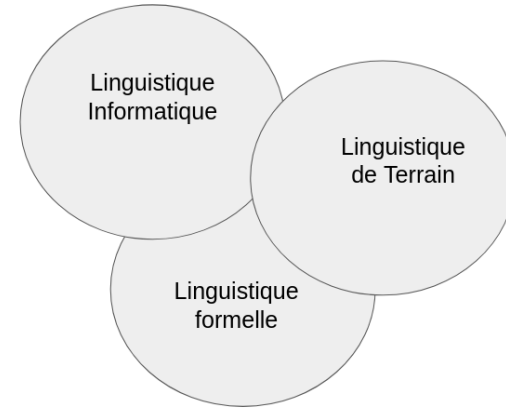
- Dépôt, archivage et diffusion
- Description des données et référencement
- Articulation entre données, outils et publications



Des données à archiver ? Contactez Alexis Michaud!

LIFT

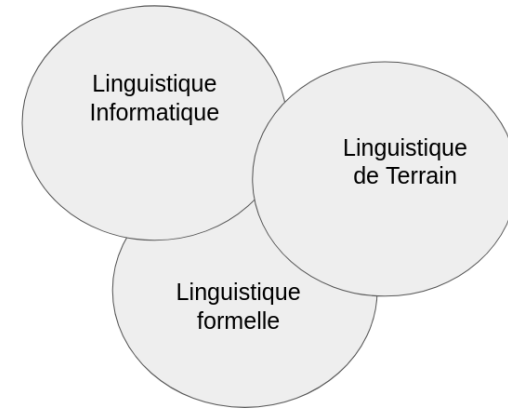
Co-
encadrements



LIFT

Co-encadrements

- LACITO/LIG/Université de Lorraine
Alignement temporel entre transcriptions et audio de données de langue japhug
- LLL/LORIA/Université de Lorraine
Identification automatique du Créole
- LACITO/Université de Lorraine
Recognizing lexical units in low-resource language contexts with supervised and unsupervised neural networks



Linguistes: Proposez des sujets pour les Master TAL!

LIFT

Comment participer

Propositions d'action au fil de l'eau (2K euros)

<https://gdr-lift.loria.fr/appele-a-propositions-dactions-2020/>

- Lancer un groupe de travail, de lecture
- Organiser une journée d'étude
- Lancer une tâche partagée
- Partager une ressource (corpus, grammaire, lexique ...)
- Organiser un hackaton, annotaton etc.
- Poster sur la liste :
lift_members@inria.fr
- Twitter : @GdrLift

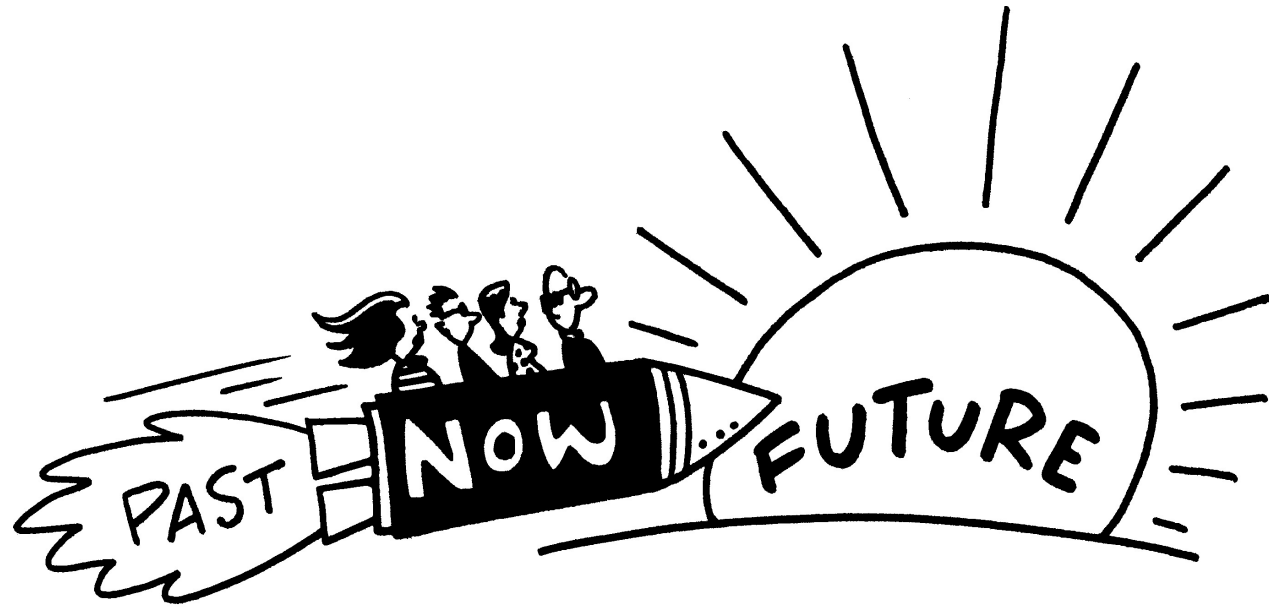


***Toute action
permettant de
promouvoir les
thématiques du GdR***

Site web : <https://gdr-lift.loria.fr/>

Twitter : @GdrLift

Mailing List : lift_members@inria.fr



LIFT est mort, vive LIFT

Contact LIFT 2 : Karen.fort@loria.fr