



PROPOSITION DE STAGE

Année 2021



Laboratoire L3i

Sujet de stage :

Analyse de cartes linguistiques : Association des points d'enquêtes et des éléments linguistiques

Résumé du travail proposé :

Le stage sera réalisé dans le cadre du projet ANR ECLATS. Ce projet a pour objectif de contribuer à la valorisation et l'analyse des documents cartographiques anciens. Il s'intéresse plus particulièrement à l'Atlas Linguistiques de la France (ALF). Les recherches menées au Laboratoire ont permis de développer différentes approches pour séparer les éléments graphiques et textuels présents dans ces cartes. Plus précisément, les algorithmes permettent d'obtenir la position des noms de département, des points d'enquête représentés par un nombre et les éléments phonétiques (cf. figure 1).

L'objectif du stage de développer une méthode permettant d'associer le point d'enquête à son élément phonétique correspondant.

Le travail sera réalisé en collaboration avec le laboratoire LIRIS de Lyon ainsi que le LIG et le GIPSA Lab de Grenoble.

Mots clés :

Traitements d'images, segmentation, reconnaissance de texte, symboles et éléments phonétiques, relations spatiales

Informations complémentaires :

Encadrant(s) : Mickael Coustaty ; Jean-Christophe Burie, Véronique Eglin

Equipe :

- Images et Contenus
- Dynamique des systèmes et adaptativité
- Modèle et Connaissance

Domaine d'application stratégique :

- E-éducation
- Environnement et développement durable
- E-culture
- Valorisation de contenus numériques

Cadre de coopération : ANR ECLATS

Date de début du stage : entre le 1^{er} février et le 1 avril 2021

Durée du stage : 3 à 5 mois

Financement : ANR ECLATS

Contexte de l'étude :

La dialectologie s'intéresse à l'étude des traits linguistiques caractéristiques des langues à tradition orale comme les parlers locaux, appelés patois ou, encore, dialectes. Ces traits linguistiques peuvent être de nature très différente - phonétique, morphosyntaxique, lexicale, sémantique ou prosodique - et évoluent dans un espace géographique donné, dans le temps et au contact de la société. Pour étudier les parlers locaux, la dialectologie s'est spécialisée dans la constitution de corpus de données descriptives, collectées via une méthodologie d'enquête qui repose sur des questionnaires, sur le choix de réseaux de points linguistiques et d'informateurs. Le traitement et l'analyse des données de terrain se fait au moyen de supports cartographiques, sur lesquels sont portées sous forme de points les localités enquêtées et les formes linguistiques collectées en transcription phonétique. À chaque concept exprimé par une entrée lexicale sous la forme d'un titre de carte en français est associée une et une seule carte sur laquelle figurent toutes les formes dialectales transcrites phonétiquement désignant le concept en question. L'Atlas Linguistique de France fut le premier du genre : après une campagne de terrain menée entre 1897 et 1901 dans 639 localités, J. Gilliéron et E. Edmont publient les résultats de 1902 à 1910. L'ALF concerne le domaine des dialectes gallo-romans de France, ainsi que d'une partie de la Belgique, de la Suisse, de l'Italie, et déborde également sur le domaine catalan de France et, dans un volume spécifique, sur la Corse. La publication de l'œuvre, en format papier, comporte 35 fascicules, réunissant en 12 volumes, 1920 cartes géolinguistiques qui présentent une photo instantanée de la situation dialectale de la France à la fin du XIXe siècle.

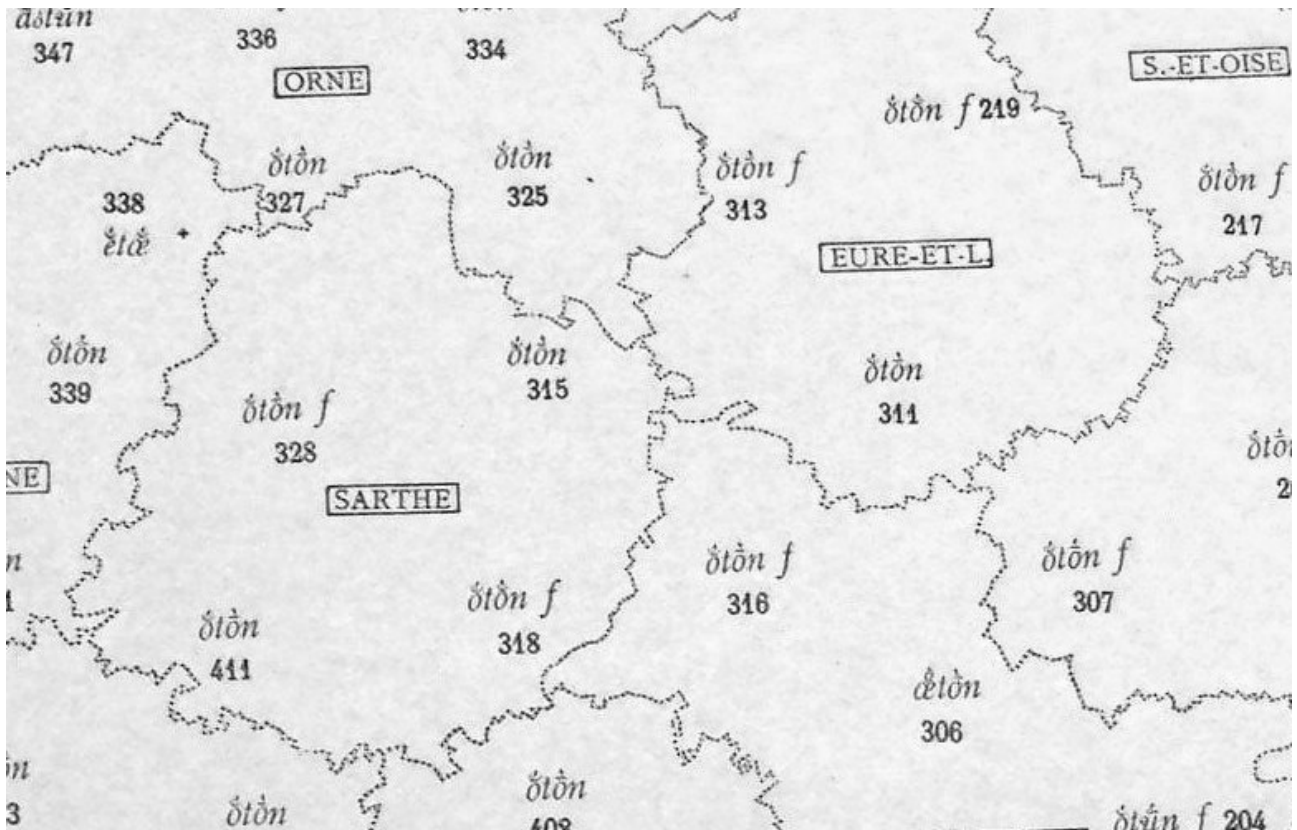


Figure 1 : Extrait d'une carte de l'ALF : prononciation du mot « automne »

L'objectif du projet ECLATS est de développer des outils d'analyse automatique de ces cartes géolinguistiques afin de nourrir un système d'information géographique, accessible via le site web [CartoDialect](#). Les données pourront ainsi être étudiées de façon plus efficace par les linguistes et dialectologues pour comprendre l'évolution de la langue.

Description du sujet :

L'objectif du stage sera dans un premier temps de s'approprier les travaux déjà réalisés et qui permettent de localiser les points d'enquêtes représentés par un nombre (comme les points

numérotés 411, 318, 311) sur la figure 1. A proximité de ces points d'enquête se trouve la transcription en phonétique de l'entrée lexicale de la carte. Sur la figure 1, l'entrée lexicale est le mot « automne ». À côté de chaque nombre on trouve la transcription phonétique du mot « automne ». Cette transcription correspond à la façon dont le mot « automne » est prononcé dans la localité correspondant au point d'enquête.

L'objectif du stage sera de proposer une approche permettant d'associer automatiquement le point d'enquête et la transcription phonétique correspondante, si elle est inscrite sur la carte. Sur la figure 1, la tâche semble simple, mais parmi les 1920 cartes, la longueur et la position de certaines transcriptions peut générer des ambiguïtés qu'il faudra tenter de lever. Une simple règle consistant à prendre l'élément le plus proche ne sera pas toujours applicable. Une analyse plus globale des relations spatiales et la comparaison entre plusieurs cartes pourra être envisagée pour résoudre le problème.

Ce travail sera réalisé dans le cadre d'un projet interdisciplinaire. Des échanges avec des linguistes et des dialectologues pourront avoir lieu pour bien comprendre les caractéristiques de l'alphabet phonétique utilisé. En fonction de l'état d'avancement du projet, le ou la stagiaire sera également amené.e à participer aux réunions pour faire part de son travail à l'ensemble des partenaires du projet.

Prérequis et contraintes particulières :

Le.la candidat.e doit :

- être actuellement en master 1 d'informatique ou justifier de compétences équivalentes
- de préférence, avoir suivi quelques cours de traitement et d'analyse d'images
- avoir un certain goût pour la recherche (étude bibliographique possible)
- avoir un bon niveau de programmation
- avoir un bon niveau d'anglais lu et écrit

Le.la candidat.e intéressé.e enverra une lettre de motivation, ainsi qu'un CV détaillé, aux deux encadrants du stage, avant le 1^{er} mars 2021. Des auditions seront organisées la première quinzaine du mois de mars.

Références bibliographiques :

- Dang Quoc B., Muzzamil L. M., Coustaty M., Nayef N., De Tran C., Ogier J.-M. (2014), A multi-layer approach for camera-based complex map image retrieval and spotting system, *4th International Conference on Image Processing Theory, Tools and Applications (IPTA) IEEE*, 2014
- Guyomard J., Thome N., Cord M., Artières T., Contextual Detection of Drawn Symbols in Old Maps, in *9th International Conference on Image Processing, ICI 2012, Orlando*, 837-840, 2012
- Luo H, Agam G., Dinstein I. (1995), Directional Mathematical Morphology Approach for Line Thinning and extraction of Character Strings from Maps and Lines Drawings, in *Proc. of 3rd International Conference on Document Analysis and Recognition (ICDAR)*, Montreal, Canada :257-260, 1995

Contacts :

Email : mcoustat@univ-lr.fr, jcburie@univ-lr.fr