Constitution d‘une collection de corpus oraux:

Une méthodologie pour l’étude des langues parlées en Afrique de l’Ouest

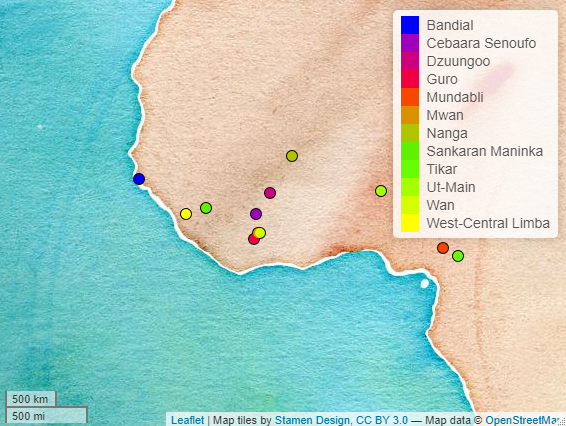
Cette étude présente un cadre méthodologique pour la constitution d’un corpus multilingue comparatif avec traductions juxtalinéaires, composé des neuf langues parlées en Afrique de l’Ouest.

[un image d‘un chercheur sur le terrain]



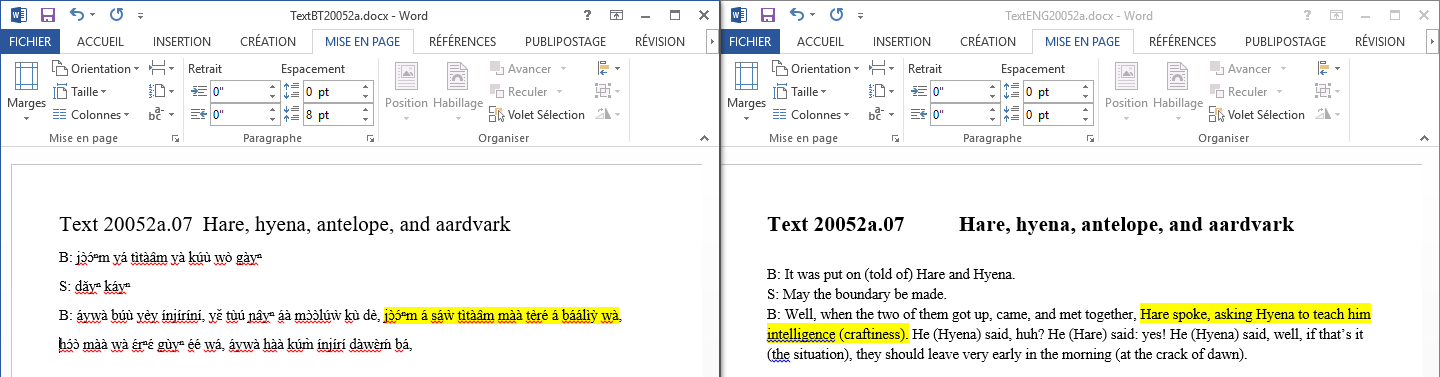
Les enregistrements audio et vidéo inclus dans le corpus ont été recueillis par plus d‘une dizaine de chercheurs au Burkina Faso, Cameroun, Côte d’Ivoire, Mali, Nigéria, Sénégal et Togo.

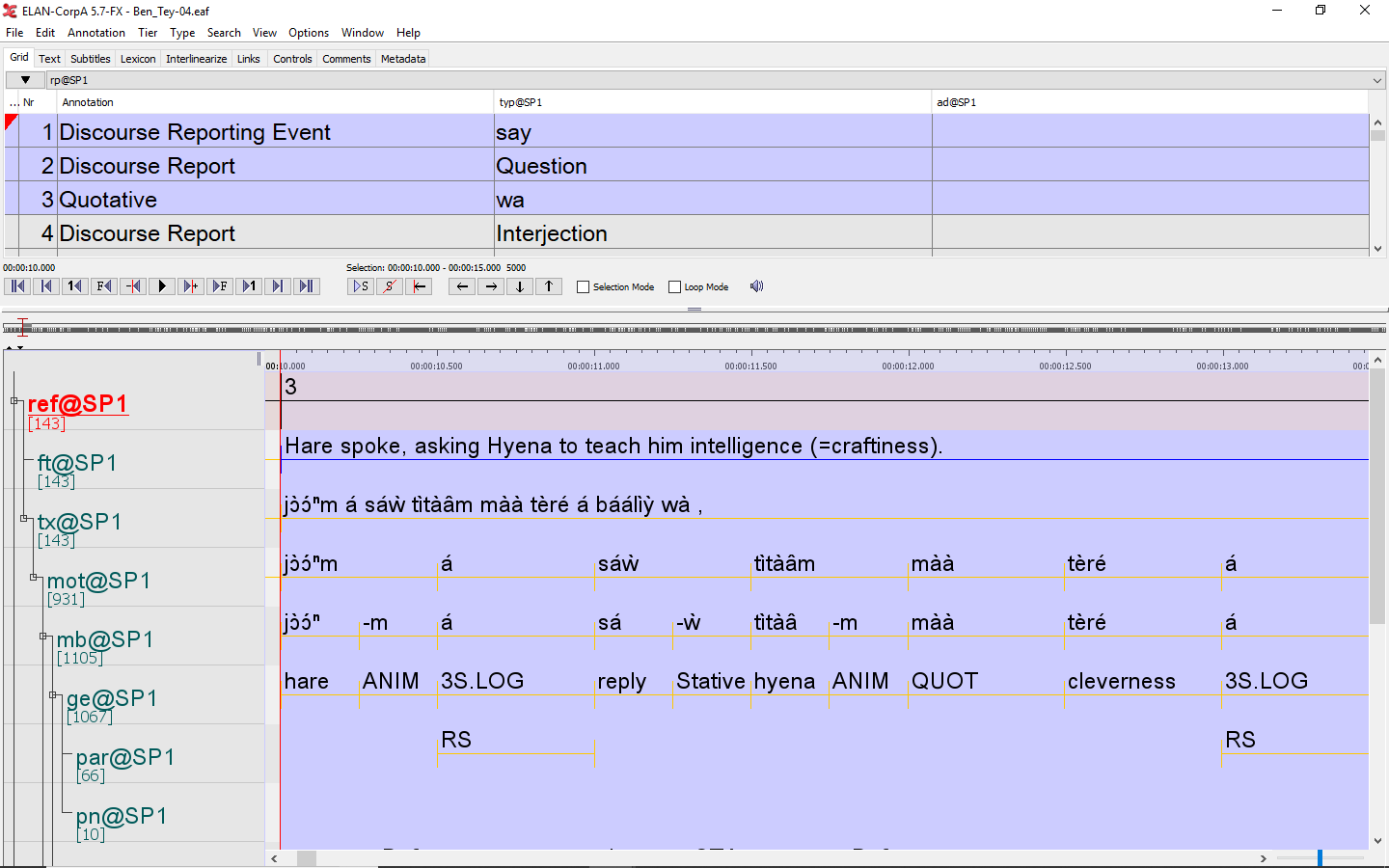
[une carte avec les pointes]

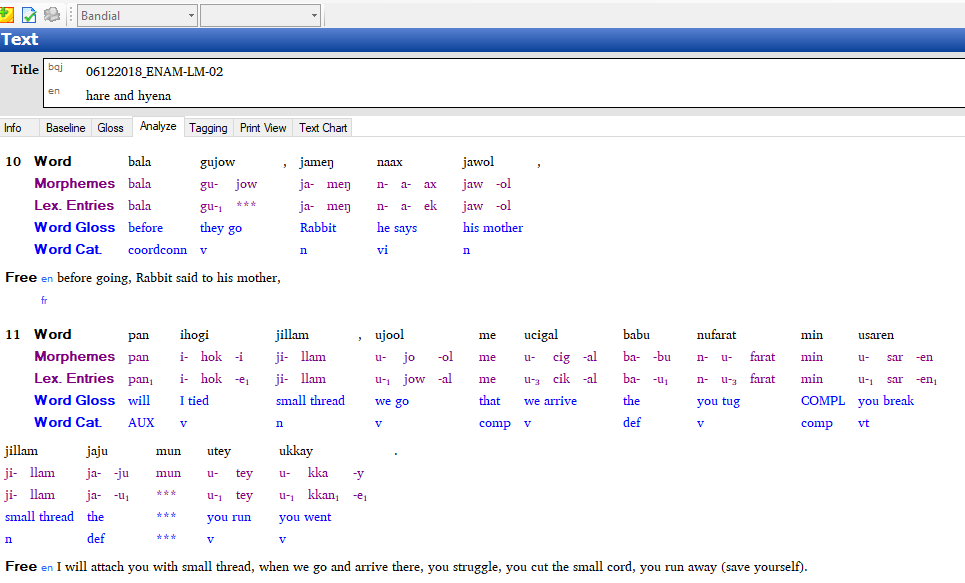


Un but spécifique de cette recherche est d’examiner quelles sont les similitudes et les différences dans le cadre de contes traditionnels.

[vignettes avant: Word, FLEx, Toolbox]







Ces données dans différents formats n’étant pas exploitables dans l’état, nous les avons fait converger vers un format commun. Nous avons choisi le logiciel ELAN pour sa capacité à importer (et exporter) des donnés dans ces différents formats. Plus spécifiquement, la version ELAN-CorpA (Chanard, 2018) qui permet d’annoter des textes à partir d’un lexique.

[un exemple concrète de fiche ELAN (les mêmes contenues que les vignettes)]

L’objectif de notre projet, « Discourse Reporting in African Storytelling », est d’étudier les constructions du *discours rapporté* dans les différentes langues et de mettre en évidence leurs similitudes et différences.

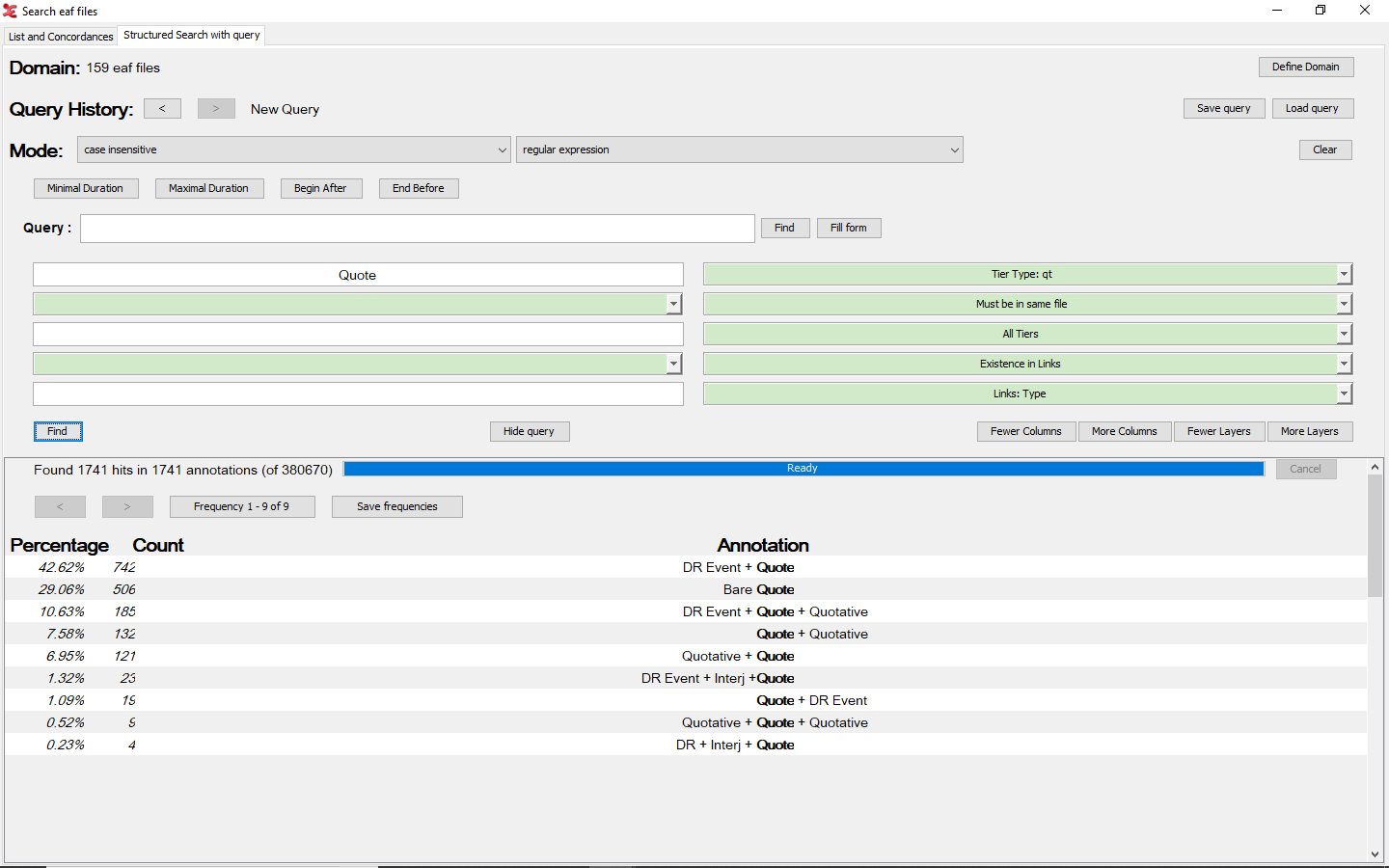
[un exemple d’une phrase avec *discours rapporté* (glosé - traduit)]

Les annotations morphosyntaxiques étaient nécessaires mais pas suffisantes pour l’étude du discours rapporté (cf. Spronck et Nikitina (2019): le discours rapportéconstitue-t-il une catégorie syntactique ?). Par conséquent, les lignes d’annotation suivantes étaient ajoutées: RP (reported phrase), TYP (type of the reported phrase), QT (quoted type), PAR (participant).

[un exemple concrète de fiche ELAN (avec le template)]

Cette approche nous permet de faire des requêtes pour, par exemple, voir la distribution des types syntaxiques des constructions du discours rapporté.

[un exemple de requête dans tout les langues (par exemple)]



Bien que les langues inclues dans le corpus ne sont pas apparentées, les contes de plusieurs langues de l’Afrique occidentale ont une tradition dans laquelle le lapin et l’hyène jouent des rôles importants. Ainsi, nous pouvons chercher dans le corpus les caractéristiques intéressantes qui ne seraient pas comparables autrement, pour examiner la grammaire et les marqueurs de discours qui ont des rôles similaires.

La Figure 1 donne l’exemple d’une recherche de langues Africaines qui ne sont pas apparentées, et qui ont été documentées et compilées par des auteurs différents, mais qui ont toutes une manière similaire de raconter des contes d’animaux.

La structure d’un corpus comparatif en ELAN CorpA

comme celui qui est illustré dans la Figure 1, a l’avantage de pouvoir facilement parcourir

de nombreuses annotations avec Regular Expressions pour ELAN (Mosel, 2015), qui sont

aussi associées aux enregistrements et meta-data. Une recherche de cet échantillon a révélé

que les mots les plus utilisés sont ‘lapin’ et ‘dire’, donc, les exemples dans (1) sont de deux

langues Dogon documentées par l’auteur et (2) de base de données de langue Goemai, parlée

en Nigeria (Hellwig, 2003) et Tabaq, parlée en Sudan (Dimmendaal and Hellwig, 2013). Ces

exemples démontrent la comparabilité de contextes similaires, bien que les données soient de

sources complètement différentes. La présentation donnera aussi une illustration instructive

sur la façon de créer un corpus multilingue utilisant les méthodes linguistiques FLEx et

ELAN CorpA, et d’exporter les résultats vers d’autres logiciels.

<https://www.free-power-point-templates.com/presentation-poster-templates/>

<https://lift2019.sciencesconf.org/>

La taille maximum des poster est A0 (portrait ou paysage). Les langues des Journées scientifiques seront l’anglais et le français. Pour les posters, si votre poster est en français, merci de prévoir un résumé en anglais et/ou une version anglaise à distribuer sur feuille A3 ou A4.