DEMC Journal

N° 4 | 2026
2026

Etude discursive d’un dispositif d’égalisation langagière : expérimentation méthodologique avec ELAN et R pour une approche multi-scalaire et quantitative

Marie VENIARD *Maître de conférences HDR*
CEPED
University of Paris
Alioscha MASSEIN *Ingénieur d'études*
'' Maison des Sciences sociales et des Humanités - Lyon St-Étienne
Justine LASCAR *Ingénieur de Recherche*
ICAR UMR 5191
Interactions, Corpus, Apprentissages, Représentations, Lyon

Édition électronique :
URL :
https://demc-journal.org/articles/revue-4/3354-etude-discursive-dun-dispositif-degalisation-langagiere-experimentation-methodologique-avec-elan-et-r-pour-une-approche-multi-scalaire-et-quantitative
Date de publication : 30/09/2026

CertiScience© *Certifié évalué par les pairs*

Cette publication est sous licence **CC BY-NC-ND** (Attribution - No commercial - No derivatives).

Pour **citer cette publication** : VENIARD, M., MASSEIN, A., LASCAR, J. (2026) Etude discursive d’un dispositif d’égalisation langagière : expérimentation méthodologique avec ELAN et R pour une approche multi-scalaire et quantitative. *DEMC Journal*, (4). https://doi.org/10.34745/

Cet article, qui s’inscrit dans le projet VALELANG, rend compte d’une expérimentation d’analyse multi-scalaire de données orales pour une démarche d’analyse du discours, discipline qui se pratique dans un va-et-vient constant entre micro (marques linguistiques), meso (genre, séquence textuelle) et macro (idéologie, univers de croyance, doxa). Nos travaux antérieurs sur les pratiques langagières militantes au sein de la cause des étrangers ont permis de mettre au jour la saillance de la valeur de l’autonomie et sa centralité dans la réflexivité dont font preuve les membres du collectif sur leurs actions (langagières ou non). Sur la base de travaux articulant valeurs, genres textuels et des sphères d’activité telles que l’éducation, le militantisme ou encore la communication numérique, on fait ici l’hypothèse que l’autonomie de la prise de parole politique se manifeste à travers un dispositif communicatif spécifique, basé sur l’intention politique d’égaliser les interactions. Après une définition de la notion de dispositif et une description qualitative et quantitative de celui-ci, nous proposons ici une expérimentation à même de rendre compte de sa mise en œuvre et de son actualisation interactionnelle d’un point de vue quantitatif et multi-scalaire en s’attachant à restituer une propriété centrale des interactions : la temporalité. Cette expérimentation exploite avec le langage de programmation R une annotation du corpus faite sur ELAN. Nous présenterons successivement les données et la démarche, puis les différents calculs et visualisations possibles, enfin, les premiers résultats portant sur l’actualisation du dispositif en termes de distribution de la parole et de modalités interactionnelles.

This article, which is part of the VALELANG project, reports on an experiment in multiscalar analysis of oral data for discourse analysis - a discipline that operates in a constant back-and-forth between the micro (linguistic markers), the meso (genre, textual sequence), and the macro (ideology, belief systems, doxa). Our previous work on activist language practices within the cause of immigrants has revealed the prominence of the value of autonomy and its centrality to the reflexivity demonstrated by members of the collective regarding their actions (whether linguistic or not). Building on research that links values, text genres, and spheres of activity such as education, activism, and digital communication, we hypothesize here that the autonomy of political speech manifests itself through a specific communicative dispositif, based on the political intention to equalize interactions. After defining the concept of a “dispositif” and providing a qualitative and quantitative description of it, we propose an experiment designed to account for its implementation and interactional actualization from a quantitative and multi-scalar perspective, focusing on capturing a central property of interactions: temporality. This experiment utilizes the programming language R to

analyze corpus annotations created in ELAN. We will first present the data and methodology, then the various calculations and visualizations that are possible, and finally, the initial results regarding the refinement of the model in terms of speech distribution and interactional modalities.

Mots-clés :
Interaction, Discours, Analyse quantitative

Introduction

Cet article, qui s'inscrit dans le projet VALELANG[1], rend compte d'une expérimentation d'analyse multi-scalaire de données orales pour une démarche d'analyse du discours, discipline qui se pratique dans un va-et-vient constant entre micro (marques linguistiques), meso (genre, séquence textuelle) et macro (idéologie, univers de croyance, doxa). Nos travaux antérieurs sur les pratiques langagières militantes au sein de la cause des étrangers ont permis de mettre au jour la saillance de la valeur de l'autonomie et sa centralité dans la réflexivité dont font preuve les membres du collectif sur leurs actions (langagières ou non) (voir Veniard 2022a et b, Veniard 2024). Sur la base de travaux articulant valeurs, genres textuels et des sphères d'activité telles que l'éducation, le militantisme ou encore la communication numérique (Oger 2008, Him-Aquilli 2018, Atifi et al. 2007), on fait ici l'hypothèse que l'autonomie de la prise de parole politique se manifeste à travers un dispositif communicatif spécifique, basé sur l'intention politique d'égaliser les interactions. Après une définition de la notion de dispositif et une description qualitative et quantitative de celui-ci, nous proposons ici une expérimentation à même de rendre compte de sa mise en œuvre et de son actualisation interactionnelle d'un point de vue quantitatif et multi-scalaire en s'attachant à restituer une propriété centrale des interactions : la temporalité. Cette expérimentation exploite avec le langage de programmation R une annotation du corpus faite sur ELAN. Nous présenterons successivement les données et la démarche, puis les différents calculs et visualisations possibles, enfin, les premiers résultats portant sur l'actualisation du dispositif en termes de distribution de la parole et de modalités interactionnelles.

1. Question de recherche : dispositif, genre et norme

L'expérimentation méthodologique proposée ici est fortement corrélée à la problématique de la recherche, la manifestation de l'autonomie dans la parole militante et à une notion centrale, celle de dispositif, notion entendue ici dans le sens de dispositif de communication.

1.1. Un dispositif d'égalisation langagière

On entend ici le dispositif comme un agencement d'éléments humains ou matériels réalisés en fonction d'un but à atteindre (Appel *et al*., 2010, Auboussier *et al.*, 2023). La dimension stratégique des dispositifs est centrale (Dodier, Barbot 2016) : le dispositif a vocation à constituer une solution à un problème. On postule ici que cette « recherche de solution » de la part des acteurs s'exerce de manière située, et est donc sensible aux évènements, aux contraintes matérielles et aux contraintes communicationnelles, y compris les contraintes langagières telles que le genre discursif. Or, au sein des mobilisations des militants sans-papiers, un problème émerge avec constance, depuis longtemps (Blanc-Chaléard, 2014), celui de la préservation de l'autonomie des militants sans-papiers dans le cadre d'une lutte d'alliance avec des militants soutiens. La question de l'autonomie est donc corrélée à celle de l'égalité entre les catégories de militants. Afin de parvenir à l'égalité, ces derniers se livrent à un travail d'égalisation langagière (*equalizing work*, Carlsen *et al.,* 2022, Veniard 2024).

Les dispositifs communicationnels d'égalisation langagière (voir les nombreuses études de cas dans Ferron, Née, Oger, 2022) varient en fonction du médium (réseaux sociaux, ouvrage imprimé, brochure, documentaire sonore, performance artistique …), de l'activité des acteurs (associatifs, militants, artistes …), du genre textuel dans lequel ils viennent à s'incarner (portrait à visée de témoignage, réunion politique, dispositif participatif), des langues privilégiées (langue première ou seconde, traduction éventuelle). Bien sûr, ces dispositifs ne garantissent pas à eux seuls une juste répartition de la parole (Koliopanos 2015 ; Pereira Da Silva 2022) et peuvent mener à une injonction à la parole (Carrel, 2015). L'objectif de la recherche est d'examiner leur fonctionnement, pas de statuer sur leur efficacité – mais sans fermer la possibilité que cette description puisse participer à une amélioration des pratiques.

Le dispositif étudié ici s'actualise au sein d'un genre interactionnel, celui de la réunion militante politique (sous-genre du macro-genre de la réunion, voir Etienne, André, Divoux, 2022 pour une définition). Il s'agit de la réunion hebdomadaire du collectif de militants sans-papiers, à laquelle assistent quelques militants soutiens, dans le cadre de l'alliance politique évoquée plus haut. La mixité des catégories de participants impacte les modalités de mise en œuvre de la réunion, quant aux langues utilisées par exemple, mais aussi par la présence d'une réflexivité sur le risque de rapports de pouvoir susceptibles de s'exercer, par la parole notamment, au bénéfice des militants soutiens. Ainsi, au sein de cette réunion, le dispositif de régulation des échanges vise à proposer une réponse à la question des conditions pratiques de suspension, de limitation voire d'inversion provisoire du monopole de la parole publique légitime (Ferron, Oger, Née 2022, p. 286).

Ces réunions sont marquées par la volonté de mettre en œuvre une communication plus égalitaire, ce qui se traduit, dans cette situation, par l'engagement des participants à réaliser le couple de normes méta-communicationnelles « ne pas parler à la place des premiers concernés » / « parler pour soi-même » (voir Veniard, 2022b pour la dialectique de ces deux normes). La volonté d'égalisation se traduit donc, pour les militants soutiens, par la volonté de laisser la parole aux militants sans-papiers et pour ces derniers, de s'approprier une parole politique pour décider par eux-mêmes de la manière de mener les actions.

Pour résumer, on explore ici l'hypothèse selon laquelle le dispositif de communication – les modalités interactionnelles qui se manifestent dans la réunion – est dépendant du discours méta-communicationnel – et des normes langagières associées – sur l'autonomie dans la mobilisation politique. Les caractéristiques interactionnelles et sociolinguistiques de la réunion ont servi de base à l'annotation.

1.2. Caractéristiques des réunions militantes du corpus

Les caractéristiques du dispositif de communication qui président au déroulement de l'interaction dans ces réunions peuvent être brièvement résumées ici :

- deux catégories de participants : les militants « sans-papiers » (entre 20 et 60) et des militants « avec papiers » qui les soutiennent et prennent part à un même collectif (entre 4 et 6) (voir l'annotation MONOLOG-ACTEUR) ;

- les réunions se caractérisent par la construction et le maintien d'un foyer d'attention commun, mais, du fait de leur caractère polylogal, le cadre participatif peut varier : si la règle est la prise de parole adressée à l'ensemble des participants (parole collective), tout polylogue comporte la possibilité de schismes, c'est-à-dire d'échange entre un groupe restreint de participants, plusieurs schismes pouvant évoluer vers une extinction de la parole collective, c'est-à-dire au brouhaha (voir l'annotation CADRE PARTICIPATIF) ;

- les réunions sont plurilingues et comportent des traductions des prises de parole : afin de garantir un égal accès de tous aux informations et à la participation aux échanges et aux décisions, le collectif pratique la traduction, du français vers (et inversement) le soninké et le pulaar, deux langues d'Afrique de l'Ouest parlées au Mali, au Sénégal et en Mauritanie, pays d'origine de la grande majorité des militants sans-papiers du collectif de l'enquête (voir l'annotation LANGUE) ;

- les réunion respectent, bien qu'assez lâchement, les caractéristiques des réunions militantes, notamment la division du travail de gestion de la parole en divers rôles méta-interactionnels (animateur, preneur de tour de parole, etc.). Ce niveau est trop micro-interactionnel pour être pris en compte dans cette annotation qui cherche à rendre compte des caractéristiques générales des réunions.

A travers le dispositif communicatif qui configure leur déroulement et assure leur rattachement à une histoire des pratiques d'égalisation, ces réunions sont prises dans une organisation qui dépasse le *hic et nunc* de la situation de communication.

2. Données : nature et démarche de traitement

Les données sont tirées d'une enquête ethnographique auprès d'un collectif de militants sans-papiers et d'un collectif de militants soutiens (Veniard, 2024). Cette enquête a permis le recueil et la constitution d'un ensemble de données hétérogènes (enregistrements de réunion, entretiens, notes de terrain, textes). Le travail d'annotation présenté ici a porté sur les enregistrements de la réunion hebdomadaire du collectif de militants sans-papiers.

2.1. Données et construction du corpus

Les réunions ont été enregistrées sur une période de huit mois en 2019, avec l'accord du collectif. Un sous-corpus de quatre réunions successives (mars-avril 2019) a été constitué. Ces réunions ont été sélectionnées car elles constituent la préparation à un évènement marquant, un rendez-vous avec le Ministre de l'Intérieur pour envisager les possibilités de régularisation.

Ces données sont des données sensibles pour deux raisons, d'abord parce que certains acteurs sont vulnérables du fait de leur précarité administrative ; ensuite parce que les données portent sur un sujet sensible, des opinions politiques. Un travail de pseudonymisation, des noms des personnes, des organisations, ainsi que des lieux pouvant permettre l'identification des personnes, a donc été nécessaire[2]. Pour ces raisons, ce type de données n'est pas éligible à un dépôt sur des bases de données ouvertes. Par ailleurs, les données orales qui sont l'objet du présent article ne sont pas séparables de l'enquête ethnographique qui a permis leur recueil. Par conséquent, ces données n'ont donc pas vocation à être exploitées selon une approche qui serait strictement quantitative. On commente plus bas la question de la taille de l'échantillon.

2.2. Etapes du traitement

Le sous-corpus qui fait l'objet de cette étude comporte 4 réunions, de respectivement 2h08, 1h42, 1h51 et 55 min (5h56min au total). Les enregistrements ont été transcrits sous Word, pseudonymisés (noms de personnes et de lieux permettant l'identification) et traduits, pour les parties en pulaar et soninké, par des traducteurs professionnels. Assez rapidement, le choix du logiciel ELAN[3] s'est imposé pour l'annotation du corpus : il s'agit d'un logiciel d'aide à l'analyse, un CAQDAS[4] qui permet d'aligner le signal audio ou vidéo et sa transcription et d'annoter des fichiers multimédias. Il permet d'associer un nombre illimité d'annotations textuelles, organisées en lignes (désormais les « tiers »), à des fichiers audio ou vidéo (fig. 1).

Pour résumer brièvement les étapes :

1. un sous-corpus de quatre réunions a été transcrit, traduit[5] et pseudonymisé

2. une grille d’annotation a été élaborée et implémentée avec ELAN

Cette grille a été exportée vers R, au format csv, ce qui permet d’appliquer des méthodes quantitatives sur les annotations.

En parallèle, le corpus des transcriptions a été chargé dans un concordancier simple (Antconc). On peut ainsi naviguer entre différentes échelles de construction du sens : le signal audio, l’exploration du texte de la transcription et l’approche quantitative avec R tirée des annotations d’ELAN.

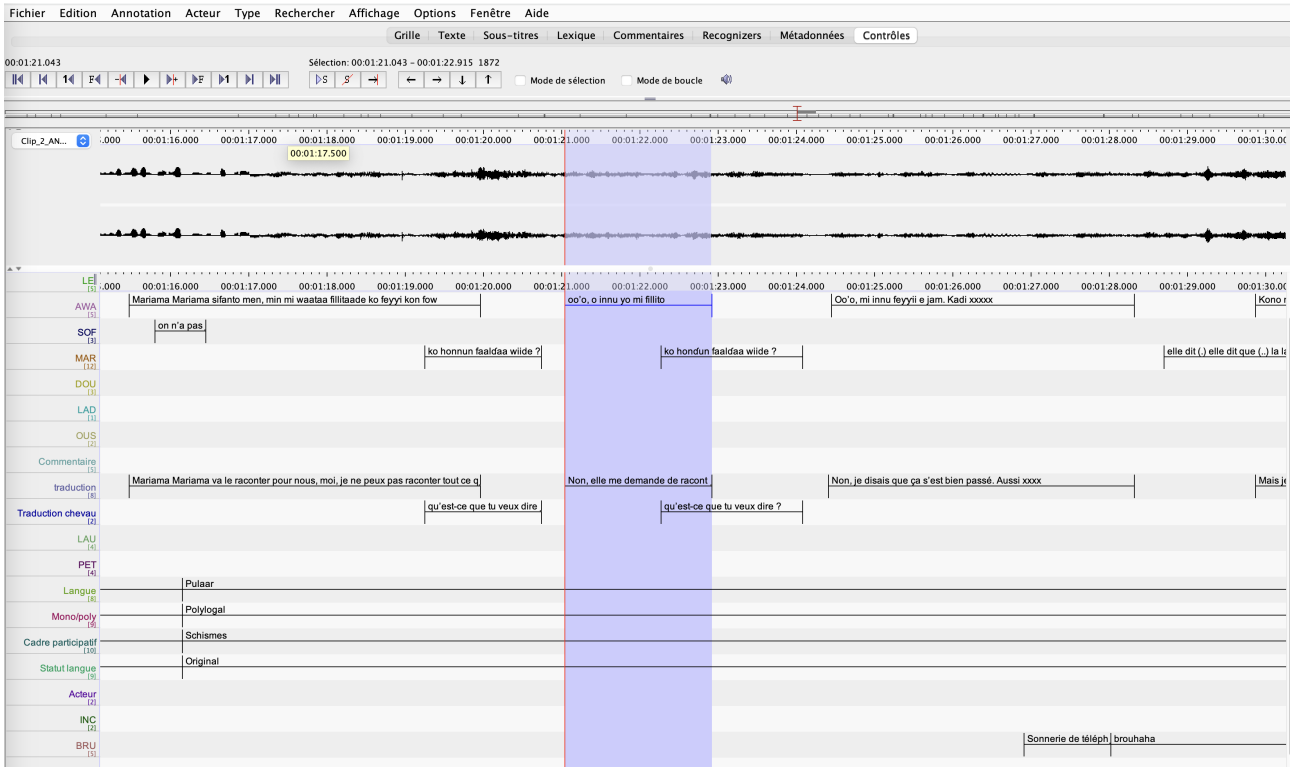


Figure 1. Interface du logiciel ELAN, extrait de transcription alignée et annotée

Chaque annotation peut prendre la forme d’une phrase, d’un mot, d’un commentaire, d’une traduction ou d’une description d’un élément observé dans le média. Ces annotations sont organisées sur différentes couches appelées « niveaux », qui peuvent être reliées entre elles de manière hiérarchique.

2.3. Annotation

Après une phase de test, un jeu d'annotations[6] de six lignes d'annotation, nommées « tiers » dans ELAN (équivalent à des balises textométriques) a été implémenté : LANGUE, MONO/POLY, CADRE PARTICIPATIF, STATUT LANGUE, MONOLOG-ACTEUR et MONOLOG-INDIV.

La tier LANGUE annote la langue des échanges collectifs, c'est-à-dire de ceux qui constituent le foyer d'attention commun (*vs* des schismes). Elle comporte les modalités suivantes : français, soninké, pulaar et plurilingue. On relève la langue qui domine l'échange/la prise de parole. Ainsi, une prise de parole en français avec un énoncé en peul sera annotée comme un segment en français. En cas de schismes, on note par défaut plurilingue car il est impossible d'écouter ces moments-là dans le détail et les participants recourent dans ce cas fréquemment à une langue vernaculaire. La plupart des locuteurs comprennent deux langues (français et soninké ou pulaar), quelques rares comprennent les trois langues, et certains, dont les soutiens, n'en comprennent qu'une.

Le changement de langue s'accompagne nécessairement d'une restriction de l'auditoire (parce qu'il comprend ou non cette langue) et donc d'un changement du cadre participatif. Cependant, tant que la parole reste adressée à un groupe de locuteurs, nous l'avons considérée comme collective.

La tier CADRE PARTICIPATIF vise à annoter le foyer d'attention. Elle comporte trois modalités : Parole collective, Parole collective+schisme(s) et Schismes. S'il existe un foyer d'attention commun, on annotera Parole collective, le groupe est alors orienté vers un seul fil de conversation. Ces échanges collectifs peuvent se tenir dans les différentes langues de la réunion – la tier Parole collective n'implique pas que tout le monde comprenne l'interaction (cf plus haut), mais qu'un foyer collectif d'attention existe pour un groupe. A l'opposé, lors des schismes multiples, le foyer d'attention est fragmenté en de multiples conversations qui empêchent la parole collective, ou interviennent en son absence. Au milieu, se trouvent des situations mixtes, dans lesquelles on observe à la fois une parole collective, mais aussi des schismes, impactant plus ou moins le déroulé de la parole collective sans toutefois l'empêcher.

Les chevauchements ne peuvent être pris en compte dans cette annotation, ce qui constitue nécessairement une simplification de l'oralité des interactions. Si un locuteur entame un chevauchement pendant un monologue, le segment sera compté dans la prise de parole « chevauchée » jusqu'au basculement éventuel vers la prise de parole du locuteur ayant lancé le chevauchement. Les segments annotés comme polylogaux intègrent naturellement des chevauchements.

La tier MONO/POLY vise justement à identifier la nature des interactions, selon qu'il s'agit d'une une prise de parole de type monologale ou d'un polylogue. Ce dernier cas englobera deux situations qui pourront être distinguées par un croisement avec la tier CADRE PARTICIPATIF : soit un échange polylogal dans le cadre d'une parole collective, soit dans le cadre de schismes.

La tier STATUT LANGUE renvoie à l'activité de traduction déployée systématiquement pendant l'évènement communicatif qu'est la réunion. Elle comporte les variantes : Original, Traduction et Original non-traduit. Un segment annoté Original est traduit peu après dans la temporalité de l'interaction, tandis qu'un segment annoté « Original non traduit » est un échange, une prise de parole qui – *de facto* ou après négociation – ne sera pas traduit dans la suite de l'interaction.

Les deux dernières tiers ne visent que les segments monologaux, prises de parole originales ou non traduites et traductions. La traduction a été considérée comme un monologue, donc une prise de parole propre, pour trois raisons : d'une part, parce que, du point de vue de l'adresse prendre la parole pour traduire est bien une prise de parole collective (telle que définie plus haut) ; ensuite, parce que, sur le plan du contenu, les traductions comportent fréquemment des séquences de prise de parole au sens propre (soit que le traducteur complète la prise de parole d'un autre locuteur après l'avoir traduite, ou qu'il complète sa propre prise de parole dans le cas de l'auto-traduction) ; enfin, certains militants sans-papiers ne s'expriment en réunion que par le biais de la traduction, soit en tant que traducteur désigné pour la réunion, soit en tant que traducteur occasionnel. Mettre de côté d'emblée ces prises de parole de traduction aurait eu pour résultat de minorer les comportements interactionnels de ces militants sans-papiers. Plus généralement, des travaux de recherche ont montré que cette activité participait pleinement au déroulement de l'interaction (Ticca et Traverso 2015).

La tier MONOLOG-ACTEUR sert pour les monologues uniquement et vise à rendre compte des deux catégories de locuteurs qui s'expriment : Militants Sans-papiers et Militants Soutiens. Dans les polylogues, les locuteurs sont mixtes et difficile à identifier dans les schismes, par conséquent, ces segments n'ont pas été annotés de ce point de vue.

Enfin, la tier MONOLOG-INDIV sert à indiquer les pseudonymes des locuteurs. Elle pourra servir, par la suite, de base à une grille de correspondance (intégrant des données de genre, de degré de connaissance du militantisme, etc.)

3. Potentiel quantitatif

L'annotation des données offre ainsi la possibilité d'adopter un regard macroscopique sur les interactions via la quantification de l'information. Les annotations d'ELAN peuvent être exportées au format csv, et ensuite importées et manipulées dans des logiciels statistiques, comme R et RStudio par exemple[7]. Nos tiers et la durée des modalités qui les composent servent alors respectivement de variables et de quantités dans une approche quantitative des réunions. Nous prenons ici les informations temporelles en tant que variable numérique qui représente une quantité, et l'ensemble des autres informations en tant que variables qualitatives, des catégories. On peut ensuite par exemple, calculer pour chaque modalité des variables les temps parlés en langue française, leur durée, et comparer cette durée aux autres langues parlées.

Cette approche amène dans un premier temps un regard nouveau sur les données : la nécessité d'avoir des catégories normées et régulières dans l'analyse quantitative permet de repérer quelques incohérences et de corriger les erreurs d'annotation. C'est ensuite la question de la pertinence des catégories, de leur construction et des éventuels recodages qui se pose, pour créer des catégories significatives statistiquement, mais aussi par rapport au terrain. Enfin, dans la perspective de comparer des informations, comme par exemple le temps parlé en langue français en fonction de la catégorie de locuteur, il faut bien faire attention, sur ELAN, à l'alignement des annotations dans le temps, afin d'avoir des informations comparables, et non manquantes. Nous avons pris conscience de l'importance de la temporalité dans l'annotation pour rendre comparable les variables les unes aux autres : là où l'annotation sur ELAN n'était pas précise au millième de seconde près, nous avons dû faire en sorte de l'ajuster pour couvrir très précisément les temps de chaque modalité afin d'abord de bien respecter les quantités attribuées à chacune des modalités et d'ensuite rendre comparables les variables et les modalités entres elles.

3.1. A quoi servent les calculs et les visualisations ?

Pour le montrer plus simplement, le logiciel ELAN nous donne une sortie sous forme de tableur, que l'on peut manipuler à des fins d'analyses, et qui prend cette forme :

Description: df [490 × 8]

TAG <chr>	start_time <S3: POSIXct>	start_sec <dbl>	end_time <S3: POSIXct>	end_sec <dbl>	length <S3: POSIXct>	length_sec <dbl>	CODE <chr>
Langue	2023-05-01 00:00:00	0.31	2023-05-01 00:00:35	35.16	2023-05-01 00:00:35	34.84	français
Langue	2023-05-01 00:00:35	35.16	2023-05-01 00:00:43	42.59	2023-05-01 00:00:07	7.43	français
Langue	2023-05-01 00:00:51	51.37	2023-05-01 00:02:46	165.59	2023-05-01 00:01:54	114.22	plurilingue
Langue	2023-05-01 00:02:46	165.59	2023-05-01 00:05:50	350.39	2023-05-01 00:03:05	184.80	français
Langue	2023-05-01 00:05:50	350.39	2023-05-01 00:06:11	370.71	2023-05-01 00:00:20	20.31	plurilingue
Langue	2023-05-01 00:06:11	370.71	2023-05-01 00:06:30	389.57	2023-05-01 00:00:19	18.86	pulaar
Langue	2023-05-01 00:06:30	389.57	2023-05-01 00:06:33	392.82	2023-05-01 00:00:03	3.25	pulaar
Langue	2023-05-01 00:06:33	392.82	2023-05-01 00:08:11	490.70	2023-05-01 00:01:38	97.87	pulaar
Langue	2023-05-01 00:08:11	490.70	2023-05-01 00:08:27	507.36	2023-05-01 00:00:17	16.67	plurilingue
Langue	2023-05-01 00:08:27	507.36	2023-05-01 00:13:49	828.67	2023-05-01 00:05:21	321.30	soninké
Langue	2023-05-01 00:13:49	828.67	2023-05-01 00:15:44	944.44	2023-05-01 00:01:56	115.78	pulaar
Langue	2023-05-01 00:15:44	944.44	2023-05-01 00:16:04	964.22	2023-05-01 00:00:20	19.78	plurilingue
Langue	2023-05-01 00:16:04	964.22	2023-05-01 00:20:41	1240.67	2023-05-01 00:04:36	276.44	français
Langue	2023-05-01 00:20:41	1240.67	2023-05-01 00:21:09	1268.80	2023-05-01 00:00:28	28.13	plurilingue
Langue	2023-05-01 00:21:09	1268.80	2023-05-01 00:23:05	1385.22	2023-05-01 00:01:56	116.42	pulaar

Tableau 1 : Données issues d'ELAN

Si on prend les lignes du tableau ci-dessus, la colonne « TAG » nous donne la variable, et la colonne « CODE » nous donne la modalité de cette dernière. Les autres informations sont relatives au temps (début, fin et durée d'une séquence pour un CODE dans un TAG). Ainsi, sur la question de la langue, on peut obtenir l'ensemble des séquences pour chacune des langues parlées dans les réunions. De cette manière, on peut extraire des sous-parties du tableau afin d'observer les variables une à une : on peut ainsi déployer la plupart des outils de la statistique descriptive univariée.

On comprend intuitivement la nécessité d'avoir une approche rigoureuse dans l'annotation des données en amont de la phase d'analyse pour ce type de manipulation : si l'on ne respecte pas une dénomination des modalités strictes, ou que l'on n'annote pas précisément les interactions, on se retrouve soit avec beaucoup de « nettoyage » des données, soit avec des informations erronées. De plus, le tableau fourni à l'origine n'est pas formaté pour permettre la comparaison des informations, nous devons le « pivoter » afin d'avoir en ligne les temps de l'annotation, en colonne les variables, et dans les cellules les modalités. Cela implique de découper le temps dans un référentiel commun pour chaque variable dans ELAN, et notamment de s'ajuster sur le découpage le plus petit.

18 × 11

start_time <S3: POSIXct>	end_time <S3: POSIXct>	length <S3: POSIXct>	length_sec <dbl>	Langue <chr>	Mono/Poly <chr>	Cadre Participatif <chr>	Statut langue <chr>	Monolog-Acteur <chr>	Monolog-individu <chr>
2023-05-01 01:09:33	2023-05-01 01:11:50	2023-05-01 00:02:17	137.07	français	monologal	parole collective	original	soutien	AGN
2023-05-01 01:11:50	2023-05-01 01:12:07	2023-05-01 00:00:18	17.51	plurilingue	polylogal	parole collective+schisme(s)	original non traduit	NA	NA
2023-05-01 01:12:07	2023-05-01 01:13:44	2023-05-01 00:01:37	96.71	pulaar	monologal	parole collective	traduction	sans-papier	MAR
2023-05-01 01:13:44	2023-05-01 01:14:19	2023-05-01 00:00:35	35.00	pulaar	polylogal	parole collective	original non traduit	NA	NA
2023-05-01 01:14:19	2023-05-01 01:16:51	2023-05-01 00:02:32	152.10	plurilingue	polylogal	parole collective+schisme(s)	original non traduit	NA	NA
2023-05-01 01:16:51	2023-05-01 01:18:14	2023-05-01 00:01:23	83.15	français	monologal	parole collective+schisme(s)	original	sans-papier	MAR

Tableau 2 : Tableau « pivoté » des données

C'est ce travail d'hybridation de nos pratiques qui a permis d'identifier ces nécessités pour coller à l'exploitation quantitative des données. Deux pistes sont alors explorables : des études univariées (une variable) et multivariées (plusieurs variables) des données.

3.2. Visualiser pour représenter

Pour représenter nos informations, il faut choisir entre les quantités ou la temporalité. Dans le premier cas, pour visualiser les quantités par modalité de variable, nous utilisons des diagrammes en barre pour représenter l'information. Sur les statistiques univariées, il s'agit soit de montrer les quantités, soit les proportions par réunion, ce qui permet de comparer la durée des modalités en fonction de ces dernières. On complète ces visualisations d'un tableau qui nous donne les durées pour chaque modalité.

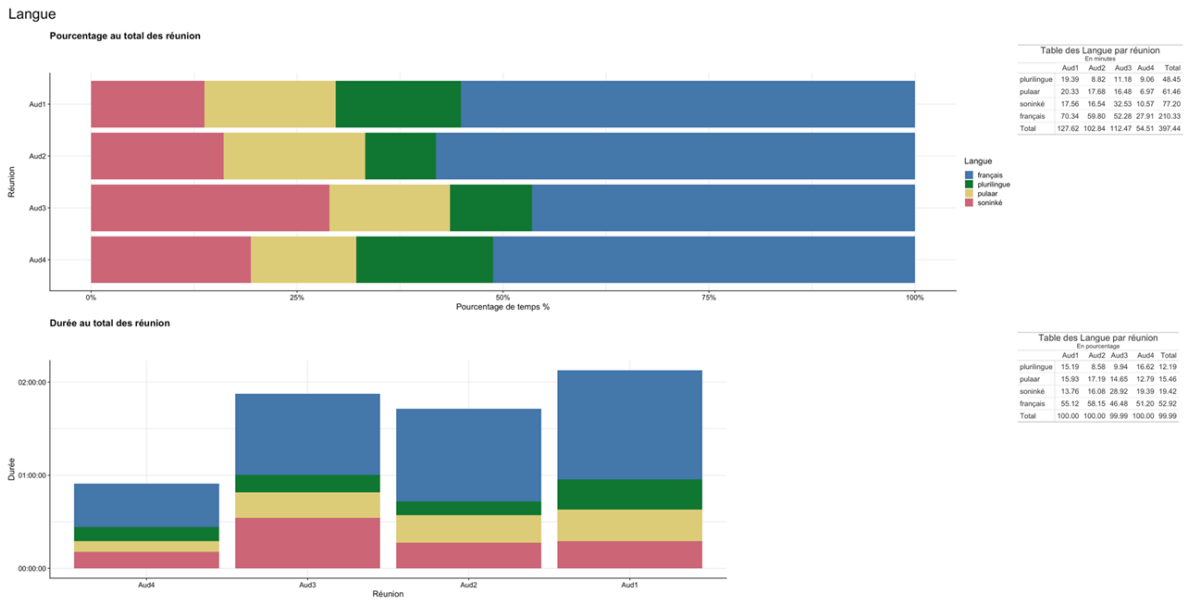


Figure 2. Visualisations de calculs statistiques univariés

On peut jouer sur plusieurs paramètres dans la visualisation, notamment les couleurs qui sont projetées, qui peuvent nous donner un niveau d'information supplémentaire, comme c'est le cas sur les individus.

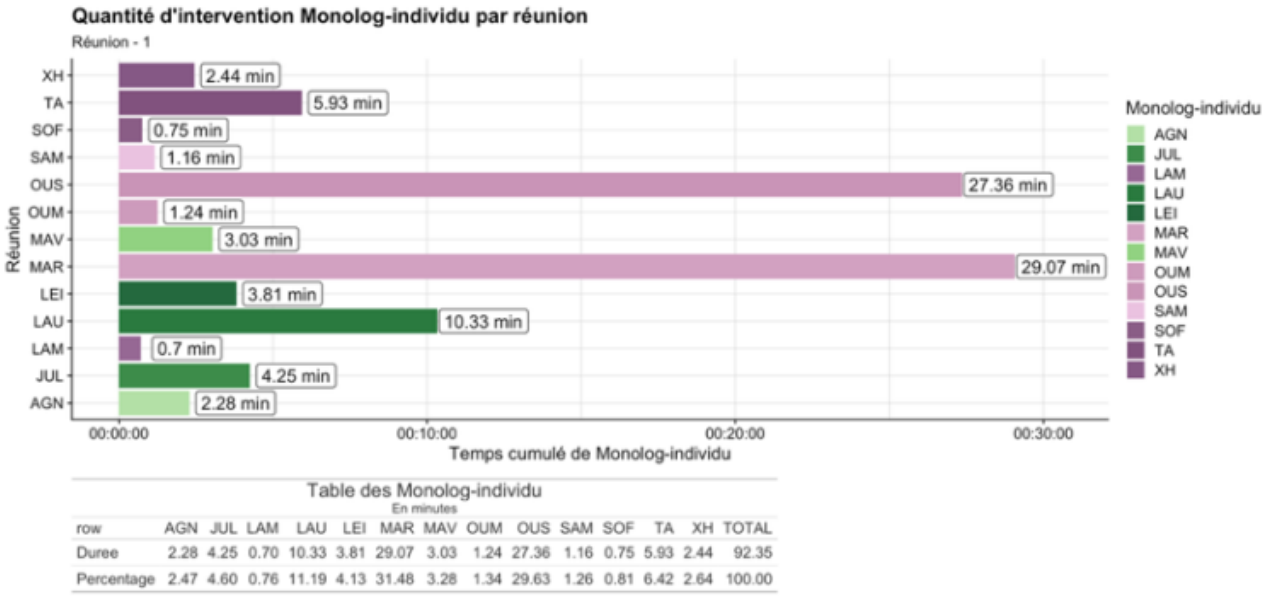


Table des Monolog-individu

En minutes

row	AGN	JUL	LAM	LAU	LEI	MAR	MAV	OUM	OUS	SAM	SOF	TA	XH	TOTAL
Duree	2.28	4.25	0.70	10.33	3.81	29.07	3.03	1.24	27.36	1.16	0.75	5.93	2.44	92.35
Percentage	2.47	4.60	0.76	11.19	4.13	31.48	3.28	1.34	29.63	1.26	0.81	6.42	2.64	100.00

Figure 3. Représentation des strates d'informations

Plusieurs informations sont présentes ici : d'abord, les quantités par individu, suivi de la valeur associée convertie en minutes (pour plus de lisibilité qu'en seconde ou qu'en heure). On utilise ensuite la couleur pour signifier la catégorie de chacun des individus dans le graphique : soit « soutien », soit « sans-papier ».

3.2.1. Graphique temporel

On peut aussi se servir du temps comme une variable à représenter dans nos graphiques, non plus comme une quantité, mais comme une durée. On peut ainsi s'intéresser à l'enchaînement des modalités dans le temps de la réunion :

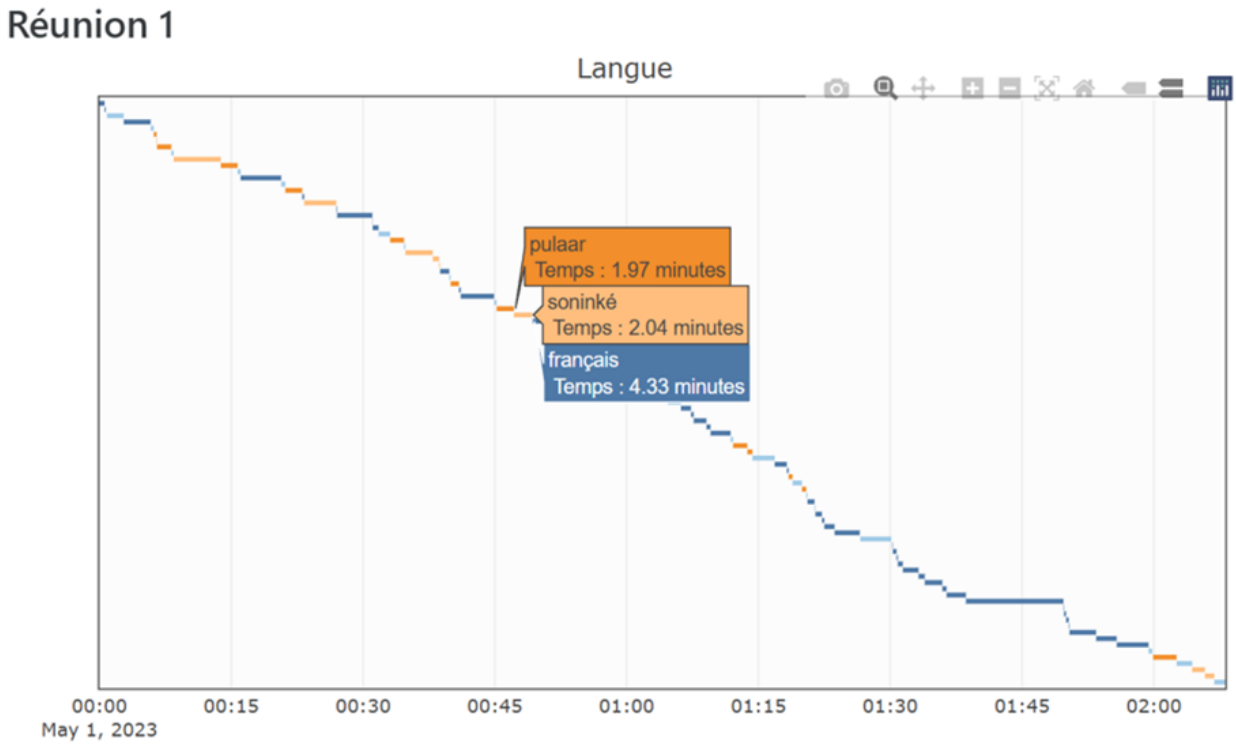


Figure 4. Représentation temporelle de la tier Langue

De cette manière, on ne représente pas uniquement des quantités, mais on cherche à observer l'évolution dans le temps d'une réunion sur une variable donnée. On ne s'intéresse plus tant à la quantification des résultats qu'à la représentation graphique du déroulé d'une réunion. Une nouvelle fois, cela nécessite une attention particulière dans l'annotation pour que les durées soient strictement bien « traduites » dans les graphiques.

3.2.2. Graphique bivarié

On a pu s'intéresser à la description chiffrée et graphique d'une variable jusqu'ici, c'est-à-dire à des statistiques univariées. Mais qu'en est-il de pouvoir comparer, croiser, deux variables entre elles afin de questionner plus en profondeur le corpus de réunions ?

On peut en effet croiser les informations après la production du tableau pivoté mentionné plus haut. Le temps devient ainsi l'individu statistique de notre tableau et est « typé » par les modalités des variables descriptives de ce temps, des interactions qui se déroulent au cours des réunions. La question reste cependant ouverte sur la manière de découper le temps pour l'annotation : est-ce que l'annotation lié au temps de parole permet la comparaison, malgré des longueurs différentes les unes entre les autres ? Ou

faut-il découper le temps de manière régulière sur une réunion, afin de garder une unité de temps commune ? La question se pose surtout si l'on souhaite décorréler les variables du temps auxquels les modalités sont assignées. On ne peut pas en l'état produire une mesure du Khi2 pour vérifier la significativité statistique de nos catégories car elles ne sont pas indépendantes du temps, et les 'individus' (c'est-à-dire dans notre cas les temps d'interaction) ne sont pas similaires par leur durée, ce qui rend la comparaison difficile si on ne travaille pas sur le temps.

On peut tout à fait croiser des informations dans des tableaux de contingence en prenant le temps comme paramètre de comparaison. On peut ainsi représenter graphiquement des durées annotées avec plusieurs informations :

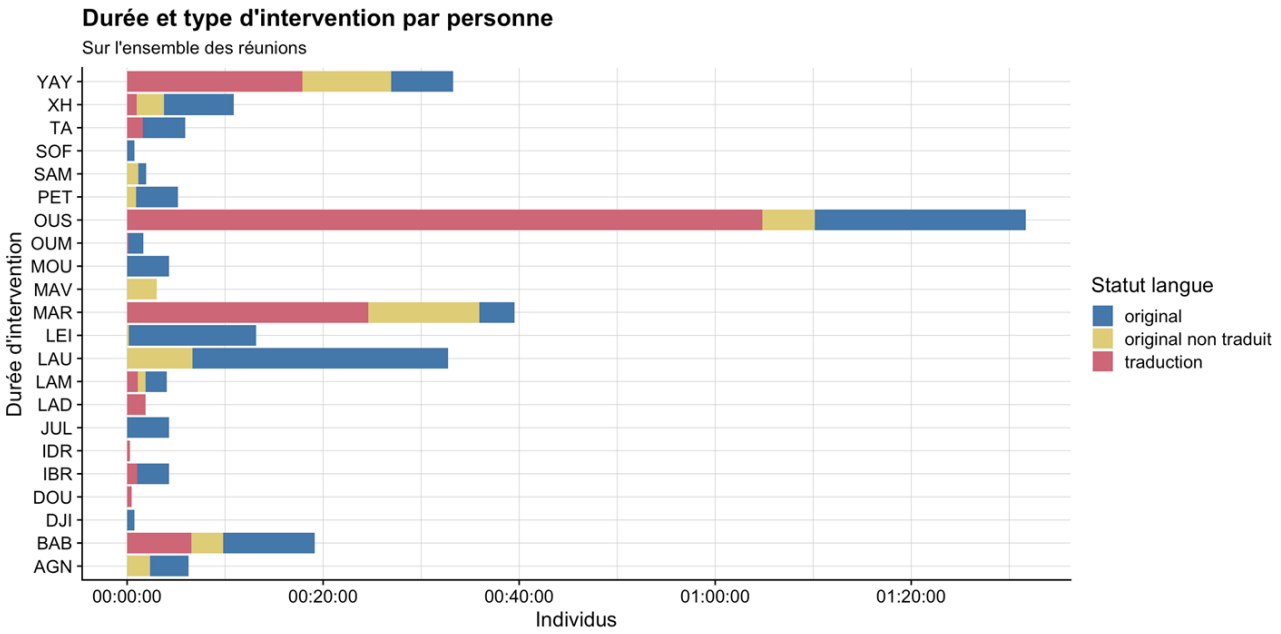


Figure 5. Visualisations de calculs statistiques bivariés (temps d'intervention et tier STATUT-LANGUE)

Ici, on peut représenter à la fois le temps d'intervention de chaque personne, mais aussi la tier STATUT-LANGUE associée à ce temps d'intervention (sur l'ensemble des réunions). En plus d'avoir l'information sur les personnes qui interviennent le plus souvent, on a également l'information de la nature de la prise de parole qu'elles utilisent sur ce temps d'intervention. On pourrait comparer la tier STATUT-LANGUE par individu avec des pourcentages relatifs aux individus, ce qui nous permettrait de mieux voir les types d'intervention par locuteur :

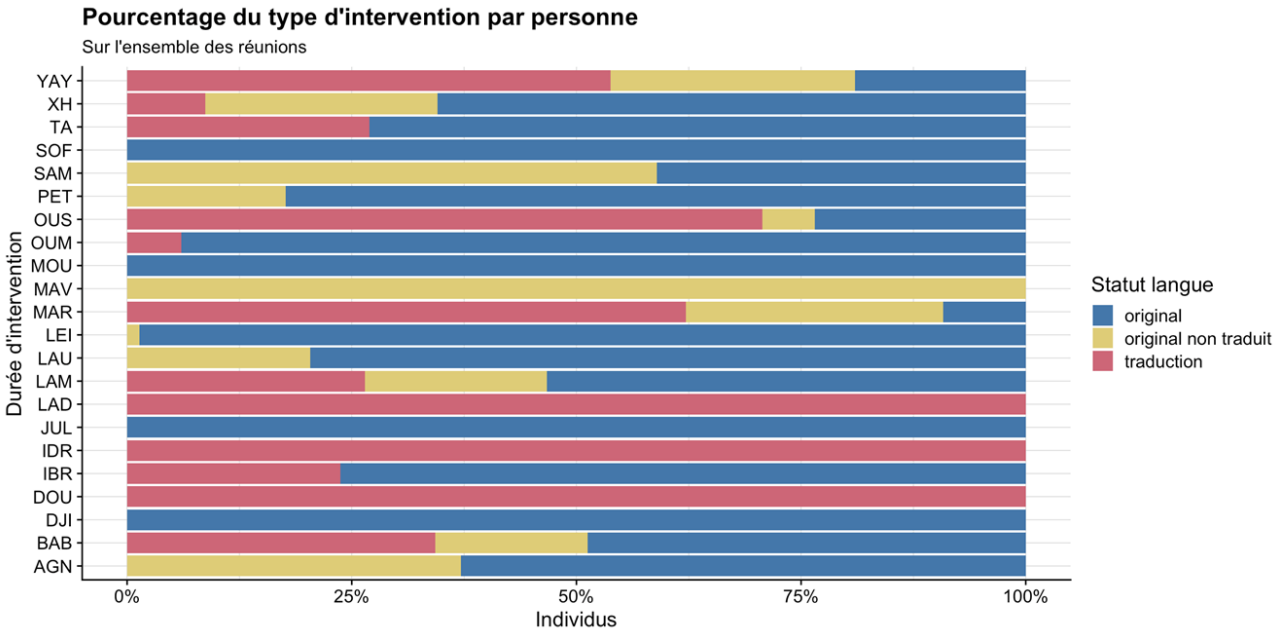


Figure 6. Visualisations de calculs statistiques bivariés (pourcentages de temps d'intervention et tier STATUT-LANGUE)

3.3. Précaution interprétative

Il convient de rappeler qu'on ne peut pas tout faire avec les données. Il faut voir ce qui est comparable dans les données, et il nous semble que des temps qui n'ont pas les mêmes longueurs ne peuvent être comparés de manière qualitative, au sens statistique du terme, car l'individu qui est comparé (le temps de la réunion) n'est pas fixe, ne correspond pas à une unité délimitée et donc comparable. De plus, on ne peut bien sûr pas faire l'hypothèse de l'indépendance des observations entre elles, dans la mesure où les prises de parole sont la conséquence des interventions précédentes, peuvent être des traductions de la parole précédente, et peuvent être prononcées par une même personne.

L'intérêt principal de la démarche est alors de décrire les informations avec des outils graphiques, des tableaux, des durées interprétés comme des quantités. De fait, on pourrait s'intéresser à calculer des indicateurs en fonction de nos catégories, comme par exemple des moyennes, des médianes, écarts-types, en fonction de nos catégories. Mais on se retrouve rapidement confronté à la taille de l'échantillon : le nombre de réunion est trop peu important pour pouvoir tirer des généralités sur celles-ci ; on ne peut pas calculer l'espérance sur 4 éléments et ainsi produire des informations normalisées qui nous permettrait de comparer, d'inférer des résultats.

Si l'on se penche plutôt sur ce qu'il se passe à l'intérieur des réunions, on pourrait essayer de comparer des quantités comme par exemple la durée des interventions sur une variable en fonction de sa position dans l'espace de la réunion, comme la première ou seconde moitié. Mais la tâche reste descriptive et relative à la réunion étant donné, une nouvelle fois, la taille de l'échantillon. On peut ajouter que la constitution du corpus lui-même n'a pas été réalisée dans le but de constituer un corpus contrôlé pour permettre une analyse quantitative dans la mesure où un protocole spécifique n'a pas été mis en place pour systématiser la collecte de manière aléatoire et en quantité suffisante.

Ce n'était pas l'objectif de la démarche initiale, qui s'inscrit en analyse du discours, avec une enquête par observation-participation d'inspiration ethnographique. Toutes ces précautions n'empêchent pas cependant de considérer qu'une démarche quantitative est possible sur ce type de données qui sont traditionnellement exploitées de manière qualitative. Nous pensons avoir démontré que c'est le cas et donnons dans la section suivante quelques pistes de réflexion sur les précautions à mettre en place pour permettre l'exploitation des données. La visualisation des informations à l'aide de graphiques permet d'apporter un autre regard sur les interventions et sur les réunions : elle permet de prendre de la distance avec le matériel d'origine et d'apporter un niveau de description qui ne serait pas possible sans ces outils.

4. Analyse : genre discursif, modalités interactionnelles et autonomie politique

Dans cette section, nous donnerons quelques résultats d'analyse afin d'illustrer notre démarche multi-scalaire[8] en nous focalisant sur des paramètres généraux de la situation de communication, tels que les langues utilisées, les locuteurs ou les groupes de locuteurs prenant la parole ; la comparaison d'un ensemble de réunions ou d'une réunion par rapport à l'ensemble ; et enfin la séquentialité d'une réunion.

4.1. Approche par le temps de parole

L’annotation offre une vision globale de la prise de parole dans ce corpus de réunions et permet d’objectiver certaines conduites langagières. Tout d’abord, le dispositif mis en place, dispositif d’égalisation censé favoriser la pleine participation des militants sans-papiers par la limitation de la place interactionnelle des militants soutiens, a un effet visible quant à la durée de la prise de parole des sans-papiers (plus de 3h30 sur les 4 réunions), qui dépasse largement celle des soutiens (environ 1h sur le même corpus).

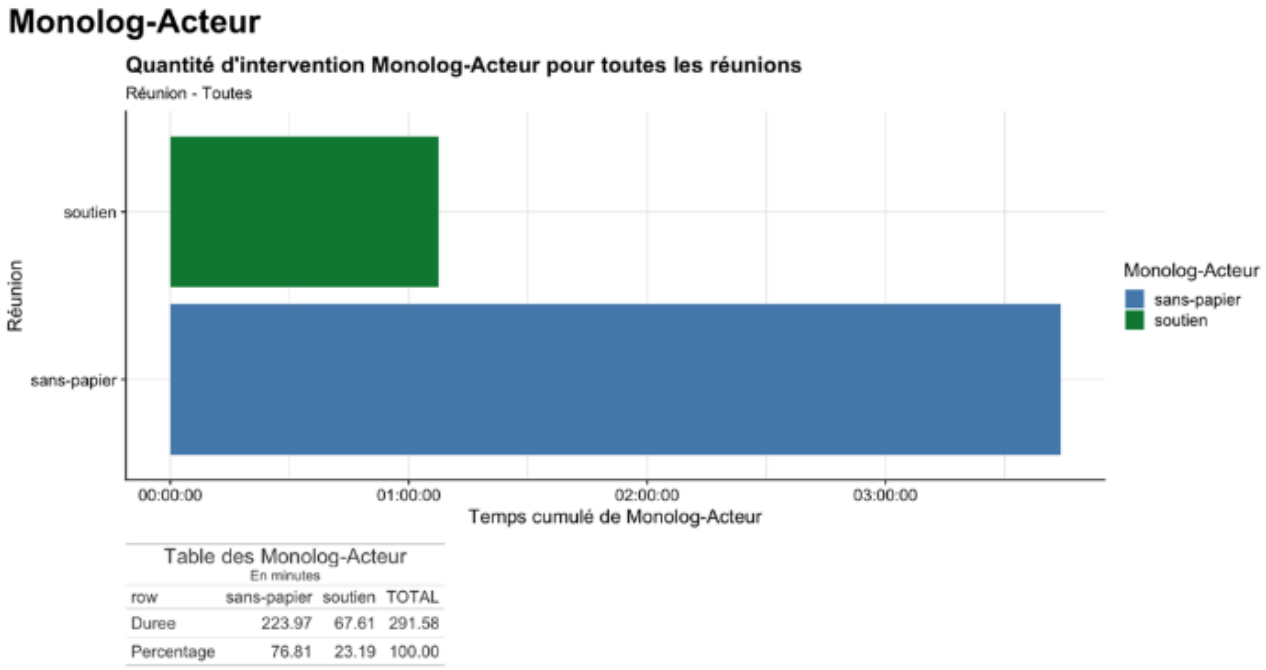


Table des Monolog-Acteur
En minutes

row	sans-papier	soutien	TOTAL
Duree	223.97	67.61	291.58
Percentage	76.81	23.19	100.00

Figure 7 : Quantité d’intervention par catégorie de locuteur sur l’ensemble des réunions

Ce résultat global apparait comme une régularité au sein du corpus, puisque sur les 4 réunions, la proportion de prise de parole des deux groupes d’acteurs est stable, les militants sans-papiers occupant entre 73 et 82% de la durée totale de la réunion.

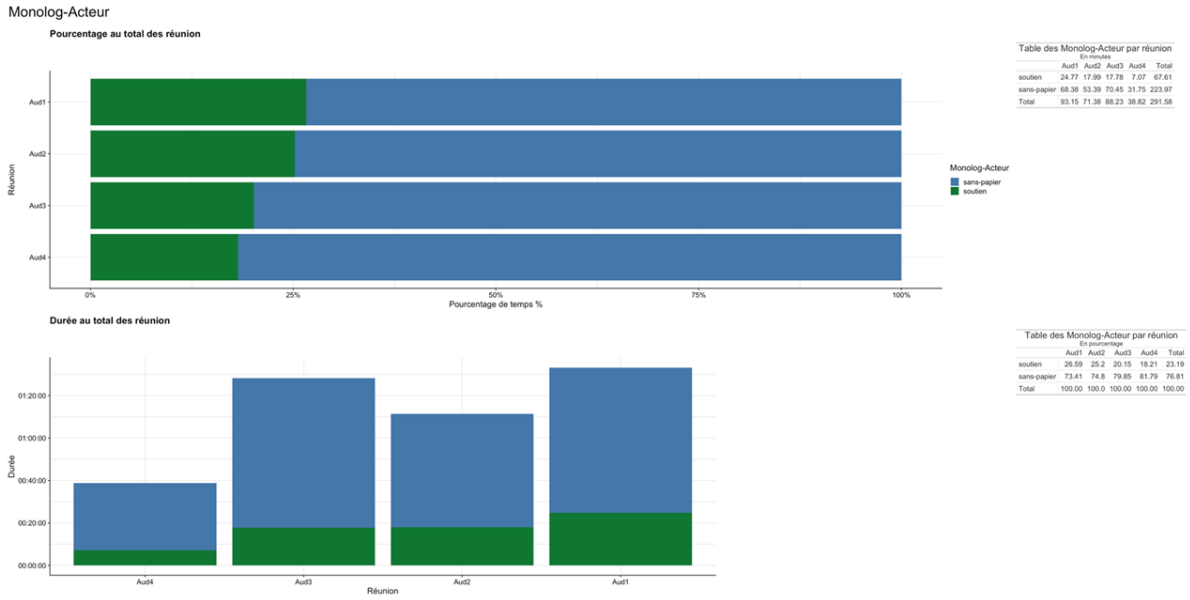


Figure 8 : Quantité et proportion d'intervention par catégorie de locuteur selon la réunion

Ce résultat n'est pas si étonnant dans la mesure où les soutiens ont une conscience réflexive explicite de la nécessité de laisser la parole aux sans-papiers (voir plus haut et Veniard, 2024). La prise en compte de la temporalité des réunions permet d'affiner, et de relativiser, ce constat.

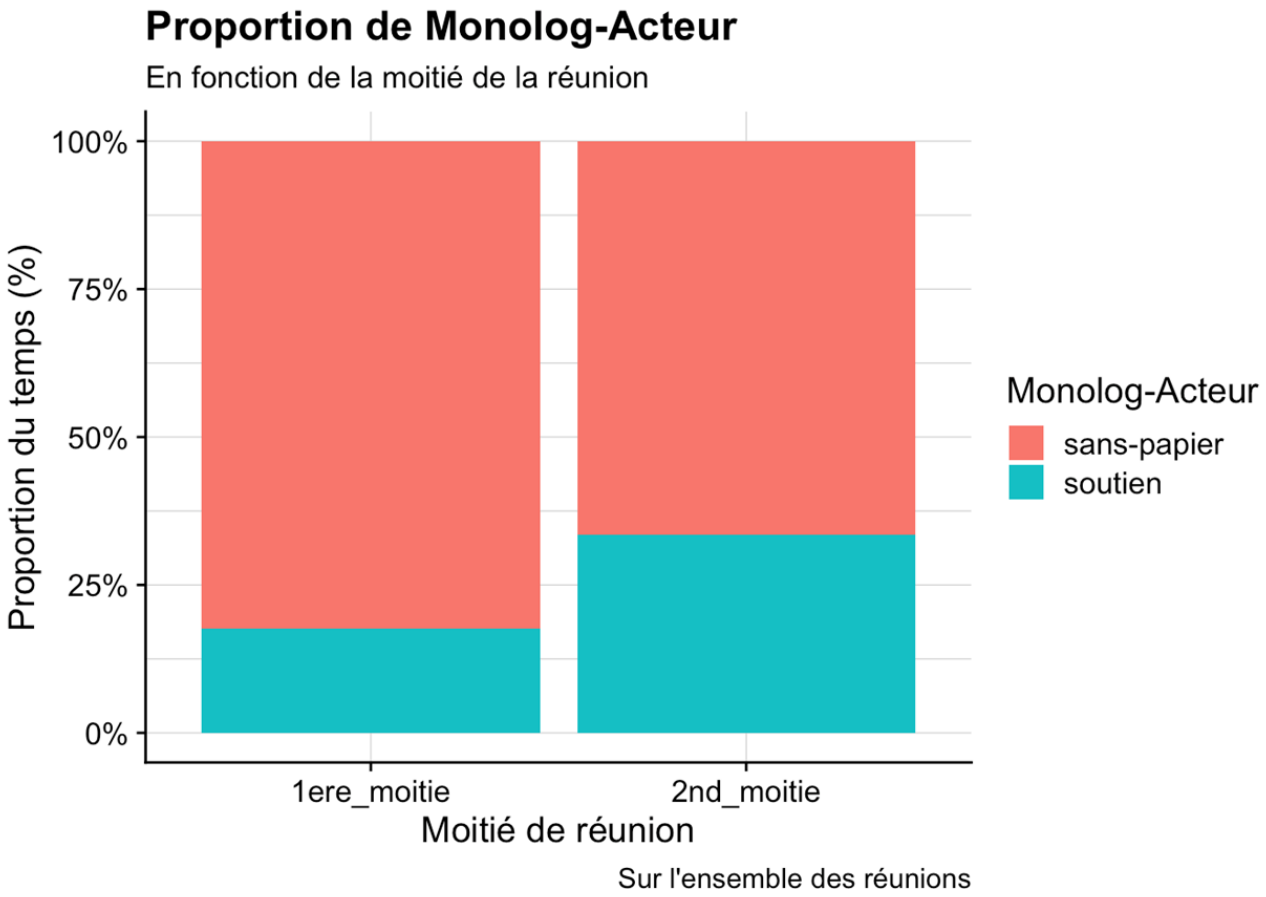


Figure 9 : Proportion d'intervention en fonction des catégories de locuteurs (somme des réunions)

En comparant la durée des prises de parole selon la division arbitraire des réunions en deux moitiés, on constate que les militants soutiens prennent plus la parole pendant les 2èmes moitiés de réunion, ce qui montre que la norme de restriction de leur parole n'est pas appliquée de manière systématique tout au long de la réunion.

4.2. Importance des séquences monologales

L'annotation permet de caractériser ces interactions militantes observables dans la réunion hebdomadaire d'un collectif de quartier. La réunion se présente comme une alternance entre des séquences monologales et des séquences polylogales.

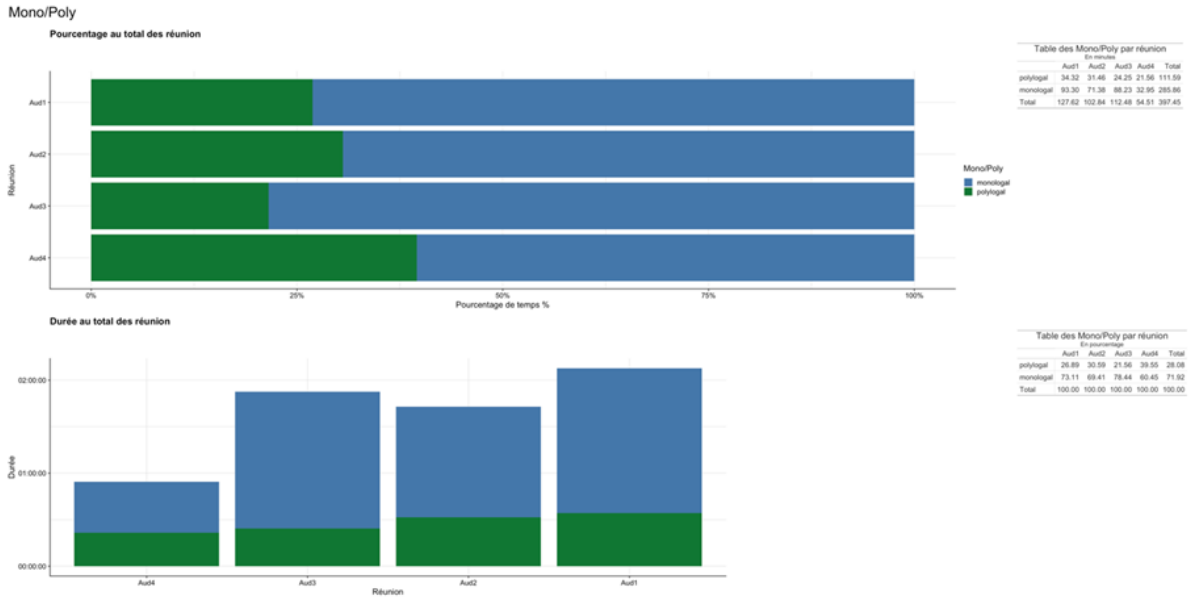


Figure 10: Quantité et proportion d’intervention monologale ou polylogale

Les séquences monologales, catégorisées par les acteurs comme des « prises de parole », constituent entre 60 et 78% du temps de réunion, occupant ainsi un rôle structurant.

4.3. Des réunions plurilingues

Du fait du plurilinguisme des militants sans-papiers et pour que la réunion soit comprise de tout le monde, elle se déroule en français (langue-pivot, c’est-à-dire à partir de laquelle et vers laquelle s’effectuent les traductions), avec traduction en pulaar et soninké, deux langues d’Afrique de l’Ouest. Les locuteurs parlant les trois langues, ou le couple pulaar-soninké, sont très peu nombreux. De ce fait, la compréhension reste, pour la grande majorité des militants, partielle.

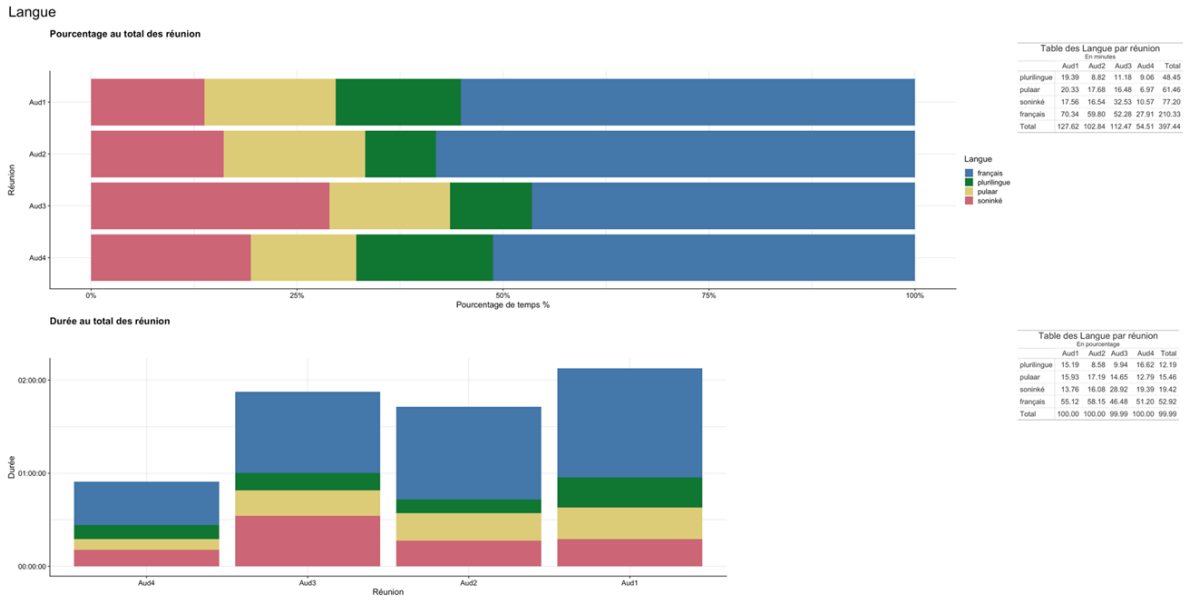


Figure 11 : Proportion d'utilisation des langues en fonction de la réunion

Le français est parlé 52% du temps de la parole monologale en réunion, le pulaar (16%) et le soninké (20%) représentent à eux deux 35%, le temps restant étant un temps plurilingue qui correspond à des moments de polylogues. Ces moments sont difficiles à annoter du point de vue de la langue puisque des schismes interactionnels (perte d'un foyer d'attention commun au profit d'échanges en petits groupes) peuvent émerger, dans d'autres langues que la langue du foyer d'attention, ceci dans l'une des trois langues officielles de la réunion, ou encore dans une autre langue.

Le français, s'il n'est pas la seule langue parlée, conserve tout de même un statut central dans les prises de parole monologales, notamment parce qu'il constitue une langue véhiculaire entre les locuteurs de pulaar et de soninké.

4.3. Temporalité de l'interaction et pistes interprétatives

Dans la perspective d'une approche multi-scalaire de la construction du sens, notre démarche méthodologique propose de mettre en perspective ces résultats quantitatifs avec le déroulement temporel de l'interaction et avec la contextualisation tirée de l'ethnographie.

Cette mise en perspective s'opère à travers une comparaison des réunions, en

concentrant nos remarques sur la réunion n°4 qui présente différents écarts par rapport aux tendances mentionnées plus haut.

Table des Mono/Poly par réunion

En minutes

	Aud1	Aud2	Aud3	Aud4	Total
polylogal	34.32	31.46	24.25	21.56	111.59
monologal	93.30	71.38	88.23	32.95	285.86
Total	127.62	102.84	112.48	54.51	397.45

Table des Mono/Poly par réunion

En pourcentage

	Aud1	Aud2	Aud3	Aud4	Total
polylogal	26.89%	30.59%	21.56%	39.55%	28.08%
monologal	73.11%	69.41%	78.44%	60.45%	71.92%
Total	-	-	-	-	100%

Tableau. 3 : Temps d’intervention en fonction du type d’intervention

Alors que les trois premières réunions comportent seulement, chronologiquement, 27, 31 et 22 % d’échanges polylogaux, conformément à la tendance générale, la réunion 4 en présente presque 40% (voir tab. 3).

La répartition des choix de langues, elle, est conforme à la tendance (51% des monologues en français, voir fig. 11). Mais on observe que le pulaar cesse d’être utilisé environ 20 minutes après le début de la réunion, et le soninké à environ 30 minutes, pour laisser place au français.

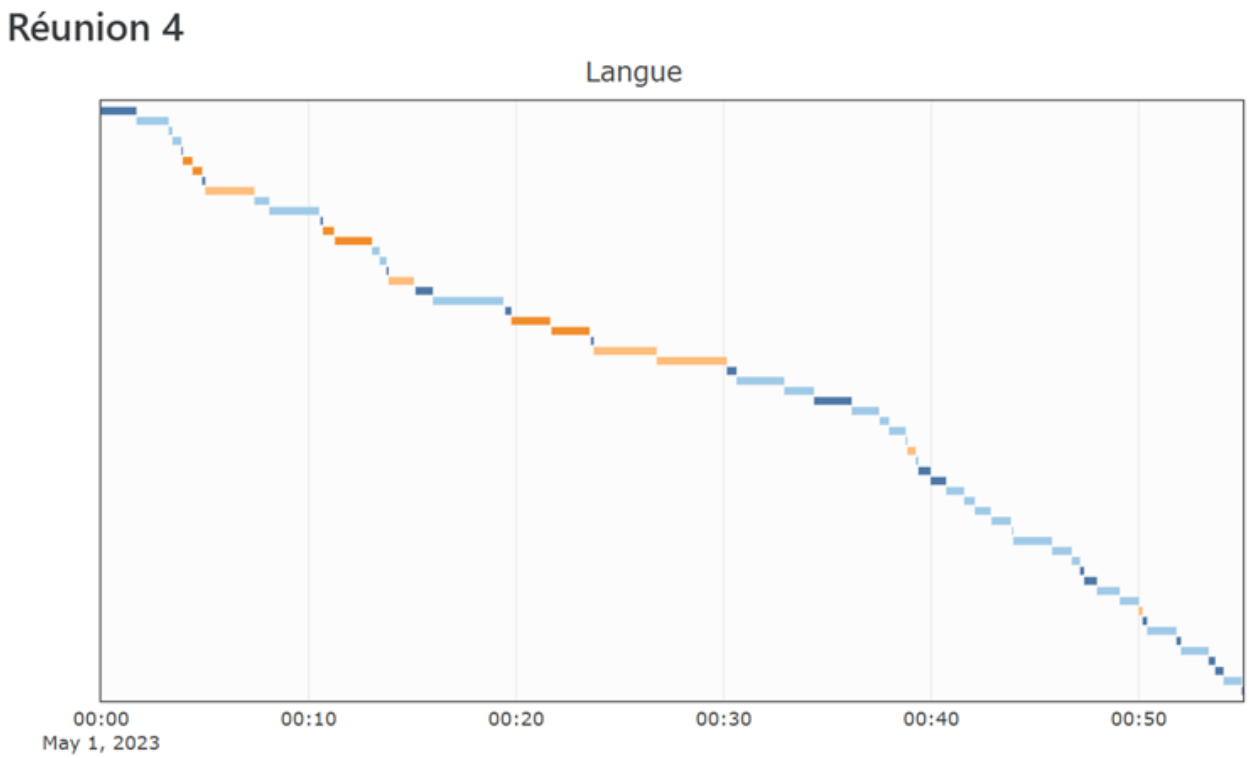


Figure 12 : Distribution temporelle du pulaar (orange foncé), soninké (orange clair), du français (bleu clair) et des sections plurilingues (bleu foncé) (réunion n°4)

Cette représentation temporelle (fig. 12) complète ce que montre le diagramme (fig. 11) : le français est surtout utilisé en seconde partie de réunion. Un autre graphe temporel, représentant la distribution des locuteurs (fig. 13), fournit l'indication que ces prises de parole en français sont assurées principalement par des militants sans-papiers.

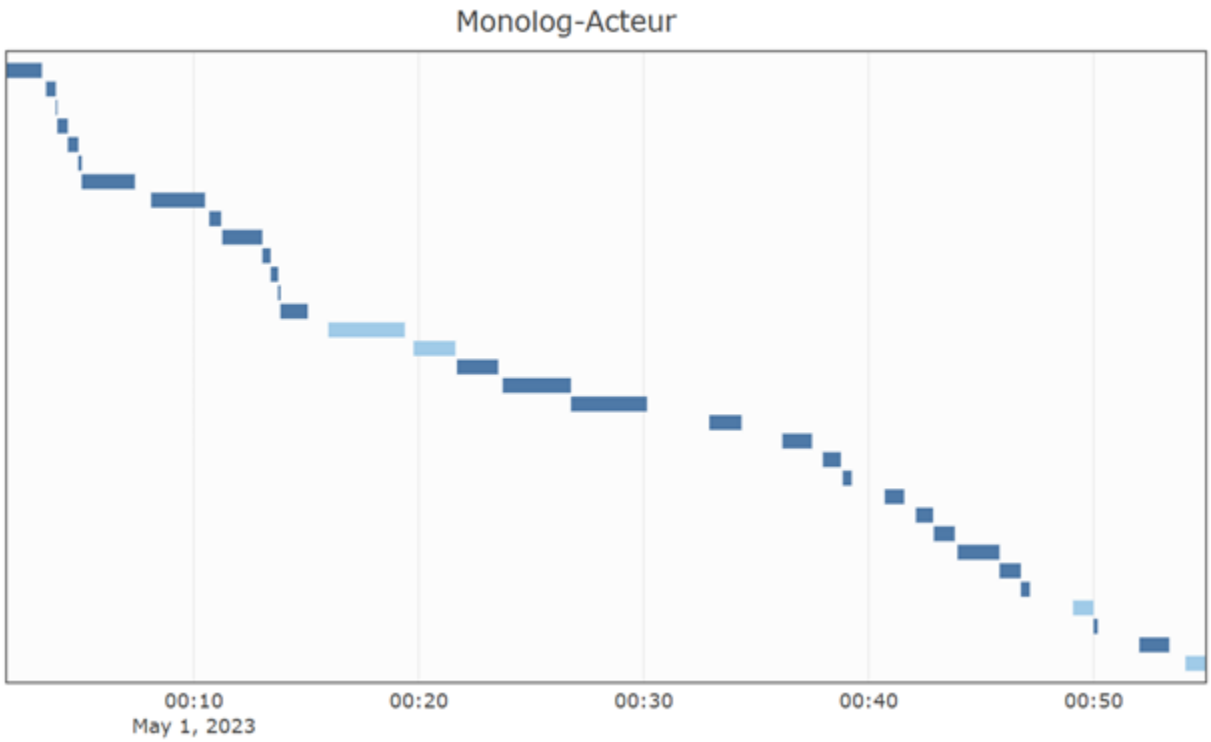


Figure 13 : Distribution temporelle des catégories de locuteurs pour les prises de parole monologales (réunion n°4) (sans-papiers en bleu foncé, soutiens en bleu clair)

Plus précisément, la traduction est progressivement abandonnée pendant cette réunion (fig. 14).

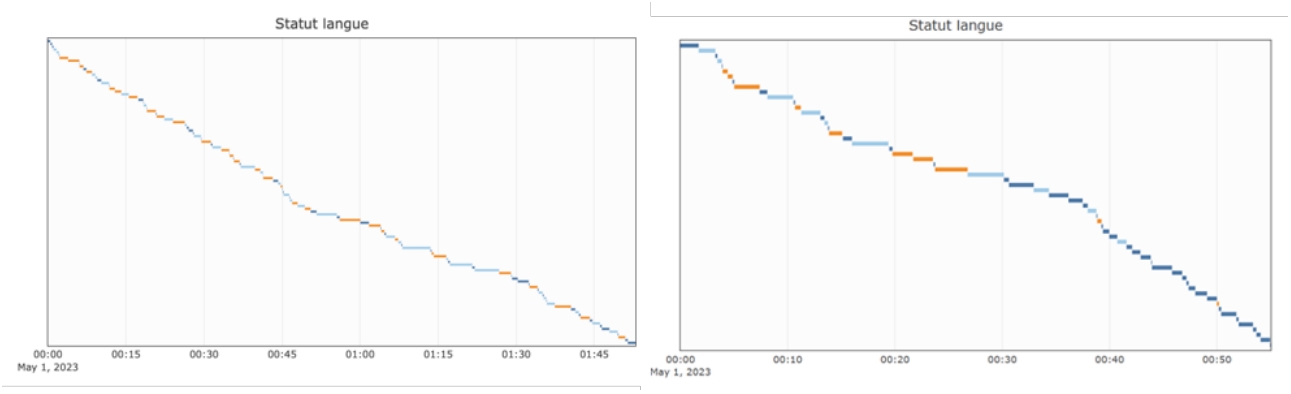


Figure 14 : Distribution temporelle du statut des prises de parole : originale (en bleu clair), originale non traduite (en bleu foncé) et traduction (en orange) (réunion n°3 à gauche et réunion n°4 à droite)

Il faut convoquer des données sémantiques d’un côté et ethnographiques de l’autre

pour expliquer ce changement linguistique et communicationnel. En consultant la transcription et en réécoutant cette réunion, on constate que, lors de la réunion n°4, le groupe doit choisir un porte-parole, activité qui se prête plus aux échanges qu'aux monologues. Un premier porte-parole – locuteur de soninké –, a déjà été désigné dans un autre cadre militant, à l'échelle de la ville. Le choix s'oriente progressivement vers un locuteur de l'autre des langues en usage dans le collectif, et donc vers un locuteur de pulaar. Ce processus se poursuit au moyen de la langue commune à tous les militants sans-papiers, le français. L'enquête permet de compléter la compréhension de l'interaction en apportant l'information que ces porte-paroles doivent participer à une réunion avec le Ministre de l'Intérieur, ce qui constitue un enjeu particulièrement important pour le collectif de militants sans-papiers. Cette importance, on en fait l'hypothèse, est susceptible de motiver un changement dans le dispositif. Le dispositif se révèlerait donc sensible à l'évènementiel.

Conclusion

Nous avons proposé dans cet article une approche langagière « multi-scalaire » des données orales, sachant qu'elle peut être multi-scalaire sans être toujours nécessairement quantitative comme on a pu le voir avec les graphes temporels qui restituent la temporalité d'une interaction unique mais sans quantification. La combinaison d'une annotation de données orales sous ELAN et d'une analyse par R permet une approche quantitative du corpus, à la fois pour explorer le corpus de manière globale *à différentes échelles* (ce que ne permet pas ELAN) selon une approche inductive et pour appuyer des résultats, et renforcer une argumentation, dans une démarche plus déductive.

On a fait l'hypothèse que le dispositif d'égalisation langagière nait de l'articulation d'un genre interactionnel et de normes méta-communicationnelles militantes relatives à l'organisation de la parole – normes militantes générales (succession des prises de parole) et normes spécifiques aux mobilisations des sans-papiers (traduction, travail d'égalisation co-réalisé). La description des réunions selon différents paramètres (locuteurs, langues, modalités interactionnelles en particulier) montre une stabilité du dispositif dans les quatre réunions du corpus, ce qui reste toutefois à confirmer sur un corpus plus important. Pour résumer les résultats, on peut dire que quantitativement, les militants sans-papiers s'expriment plus ; constat tempéré par le fait que les militants soutiens prennent plus la parole dans la seconde moitié des réunions ; que malgré le plurilinguisme de la réunion, le français est la langue pivot, notamment parce qu'elle permet l'interaction entre groupes linguistiques (langue véhiculaire). S'il est difficile d'affirmer que ce dispositif résous le problème qui l'a fait naître, on peut conclure qu'il a

des effets quantifiables visibles sur l'actualisation des interactions. Cependant la question de la répartition de la parole ne peut pas être envisagée sans prendre en compte la dimension temporelle du genre interactionnel, ce dont nous avons donné une illustration.

L'analyse permet de mettre au jour le poids du paramètre situationnel (événementiel) dans des variations observées localement à l'échelle d'une réunion. Nous souhaitons également mettre en avant l'importance des connaissances tirées de l'enquête ethnographique (18 mois sur le terrain) pour éclairer l'interprétation. Il serait impossible, dans l'approche contextualisée qui est la nôtre, de se satisfaire du seul quantitatif. Parmi les pistes de travail envisagées, il reste à explorer le niveau du locuteur, ce qui fera l'objet de travaux ultérieurs.

Si les données de cette recherche, du fait de leur caractère sensible, ne sont pas compatibles avec la science ouverte, les scripts ont été déposés sur HAL (voir note 1), notre démarche – c'est-à-dire l'export vers R d'une annotation réalisée sur ELAN – est donc reproductible. Au-delà des sciences du langage, cette démarche peut intéresser toute personne qui travaille sur des données audio ou vidéo, dans des champs aussi divers que les sciences de l'information et de la communication, les sciences politiques, ou la sociologie. Par ailleurs, nous souhaitons mettre en avant le caractère collectif de ce travail, qui met en jeu nos diverses compétences, dans un processus de recherche interdisciplinaire et interprofessionnel.

Références bibliographiques

Appel, V., Boulanger, H., & Massou, L. (dir.). (2010). Les dispositifs d'information et de communication : Concept, usages et objets. De Boeck Supérieur.

Atifi, H., Mandelcwajg, S., & Marcoccia, M. (2007). The maxim of quantity in computer-mediated communication. Dans E. Weigand (dir.), Proceedings of the 11th International Conference on Dialogue Analysis (IADA 2007) (p. 205–216). IADA Online Series.

Auboussier, J., Doytcheva, M., Seurrat, A., & Tatchim, N. (2023). La diversité en discours : Contextes, formes et dispositifs. Mots. Les langages du politique, 131, 9–26.

https://doi.org/10.4000/mots.30824

Blanc-Chaléard, M.-C. (2014). Les travailleurs immigrés en quête d'autonomie. Dans M. Pigenet & D. Tartakowsky (dir.), Histoire des mouvements sociaux en France (p. 521–532). La Découverte.

Carlsen, H. B., Doerr, N., & Toubøl, J. (2022). Inequality in interaction: Equalising the helper–recipient relationship in the refugee solidarity movement. VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations, 33(1), 59–71.

Carrel, M. (2015). Faire participer les habitants ? Citoyenneté et pouvoir d'agir dans les quartiers populaires. ENS Éditions.

Dodier, N., & Barbot, J. (2016). La force des dispositifs. Annales. Histoire, Sciences Sociales, 71(2), 421–448. https://doi.org/10.1353/ahs.2016.0064

Étienne, C., André, V., & Divoux, A. (2022). Interagir en réunion de travail : De l'étude des pratiques aux ressources didactiques. SHS Web of Conferences, 138, article 06011. https://doi.org/10.1051/shsconf/202213806011

Ferron, B., Née, É., & Oger, C. (dir.). (2022). Donner la parole aux « sans-voix » ? Construction sociale et mise en discours d'un problème public. Presses universitaires de Rennes.

Ferron, B., Oger, C., & Née, É. (2022). Mise en discours des problèmes publics et mise en problème des discours publics : Conclusion générale. Dans B. Ferron, É. Née, & C. Oger (dir.), Donner la parole aux « sans-voix » ? Construction sociale et mise en discours d'un problème public (p. 283–303). Presses universitaires de Rennes. https://doi.org/10.4000/13t5i

Him-Aquilli, M. (2018). Prendre la parole sans prendre le pouvoir : Réflexivité, discours et interactions dans les assemblées générales anarchistes et/ou autonomes [Thèse de

doctorat, Université Paris Descartes].

Koliopanos, Y. (2015). Surveiller et « embellir » : Les écrits des prostitué.e.s et des travailleur.se.s du sexe à l'aune de l'encadrement discursif. Between, 5(9).

Oger, C. (2008). Le façonnage des élites de la République : Culture générale et haute fonction publique. Presses de Sciences Po.

Pereira da Silva, C. (2022). Accès à la parole, consentement et réception contrainte : Informer les publics « défavorisés » sur le dépistage du cancer. Dans B. Ferron, É. Née, & C. Oger (dir.), Donner la parole aux « sans-voix » ? Construction sociale et mise en discours d'un problème public (p. 211–218). Presses universitaires de Rennes. https://doi.org/10.4000/13t2t.

Ticca, A. C., & Traverso, V. (2015). Interprétation, traduction orale et formes de médiation dans les situations sociales : Introduction. Langage et société, 153(3), 7–30. https://doi.org/10.3917/ls.153.0007

Veniard, M. (2022a). « Ne pas parler à la place des premiers concernés ». Étude d'une pratique politique d'éthique langagière chez des militants politiques pour les droits des étrangers en France. Nouvelle Revue Synergies Canada, 15, 1–16. https://doi.org/10.21083/nrsc.v2022i15.6526

Veniard, M. (2022b). « Ne pas parler à la place des premiers concernés » : Questionnements méthodologiques autour de la variation dialogique d'un impératif langagier dans le milieu des militants pour les droits des étrangers en France. Glottopol, 36. https://doi.org/10.4000/glottopol.1698

Veniard, M. (2024). D'une approche lexico-discursive au tournant praxéologique de l'analyse du discours : Un itinéraire de recherche [Habilitation à diriger des recherches, Université Paris Cité].

Contributions des auteurs

CRediT (https://credit.niso.org/): Conceptualisation : MV; Préparation des données : MV, AM, JL, QA; Analyse formelle : MV; Méthodologie : MV, AM, JL; Présentation visuelle : AM; Rédaction - ébauche du manuscrit initial : MV, AM; Rédaction - relectures et corrections : MV, AM, JL.

Notes

[1] Projet VALELANG (Valeurs, éthique et langage : entre actualisation et réflexivité), financé par le Labex ASLAN, https://aslan.universite-lyon.fr/projet-valelang--362166.kjsp?RH=1525438355903 . Les auteurs remercient le LABEX ASLAN (ANR-10-LABX-0081) de l'Université de Lyon pour son soutien financier dans le cadre du programme "Investissements d'Avenir" (ANR-11-IDEX-0007) de l'Etat Français géré par l'Agence Nationale de la Recherche (ANR).

[2] Sur ELAN, les données identifiantes ont été annotées, puis un script a permis la déformation sonore automatique de ces segments.

[3] ELAN (Version 7.0) [Computer software]. (2025). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan"

[4] Computer-Assisted Qualitative Data Analysis Software

[5] Merci à Idrissa Konté (traducteur soninké-français) et à Afrilangues pour la traduction, à Denis Creissels (Pr. émérite en linguistique, spécialiste notamment du soninké) et à Abdourahmane Diallo (professeur à l'INALCO, spécialiste du pulaar) pour la révision de certains passages. La traduction des données a été financée avec le soutien de la faculté Sociétés & Humanités d'Université Paris Cité , de l'Institut Convergences Migration et du Labex Aslan (projet VALELANG).

[6] Nous remercions ici Quentin Ancarola, qui a contribué au travail d'annotation.

[7] Le code qui a permis de traiter les données et de créer les graphiques est mis à disposition sur HAL à cette adresse : https://hal.science/hal-05648023

[8] La composante ethnographique et micro-interactionnelle de l'enquête est développée dans d'autres publications (Veniard 2022a et b). Le plurilinguisme de la réunion décrite quantitativement dans le présent article fait l'objet d'un travail spécifique avec Anna Marie Diagne (Université Cheikh Anta Diop, Dakar, Sénégal), axé sur l'analyse de l'interaction et l'argumentation.