# Pour une analyse quantitative de là en français parlé : la grammaticalisation revisitée à l'ère des corpus

Sascha Diwersy, Katja Ploog

▶ **To cite this version:**

Sascha Diwersy, Katja Ploog. Pour une analyse quantitative de là en français parlé : la grammaticalisation revisitée à l'ère des corpus. Kommunikationsdynamiken zwischen Mündlichkeit und Schriftlichkeit, ScriptOralia (145), Narr/Francke/Attempto, 2020, Kommunikationsdynamiken zwischen Mündlichkeit und Schriftlichkeit. halshs-03147093

**HAL Id: halshs-03147093**
**https://shs.hal.science/halshs-03147093**

Submitted on 19 Feb 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Pour une analyse quantitative de *là* en français parlé : la grammaticalisation revisitée à l'ère des corpus

Sascha Diwersy (Praxiling, Université Paul Valéry Montpellier) & Katja Ploog (LLL, Université d'Orléans)

## 1. La grammaticalisation des locatifs

Dans le contexte de notre étude, la grammaticalisation sera considérée comme une spécialisation progressive de formes d'origine pronominale dans une fonction de déterminants. Cette trajectoire, attestée pour l'émergence de l'article dans l'histoire du français, semble reconduite par la cliticisation de *là* en français moderne, en français parlé en particulier. L'objectif de cette étude, préliminaire, est d'examiner l'emploi de *là* dans son usage en français parlé. L'accessibilité de corpus oraux d'envergure, comme le corpus ESLO, permet désormais la confrontation des modèles théoriques aux données d'usage, même si, comme nous le verrons, l'exploitation de ces données n'est pas toujours aisée. L'annotation du corpus et la quantification des profils projetés soulève en effet des questions méthodologiques que l'analyse doit affronter.

### 1.1 La grammaticalisation en français

L'intégration de l'article défini au GN[1] s'est développée progressivement en français depuis la juxtaposition en latin des démonstratifs au nom. En ancien français, l'emploi du déterminant n'est pas systématique et de (plus) nombreuses insertions entre article et nom sont observées, ils fonctionnent en contexte d'ellipse ou de coordination avec un autre paradigme accompagnant le nom. Les paradigmes démonstratifs pouvaient se combiner à d'autres paradigmes de détermination, possessive par exemple, et la distinction entre déterminant et pronom n'était pas aussi nette en latin qu'elle ne l'est aujourd'hui en français (Combettes 2001, Marchello-Nizia 2001). Par conséquent, la grammaticalisation des déterminants doit être considérée sur fonds de l'évolution plus générale du français :

> Le fait qu'une forme comme l'article défini soit une création par rapport au latin classique, ou que les démonstratifs du français soient le résultat d'une « recomposition » des formes latines, n'interdit pas, nous semble-t-il, la mise en parallèle du syntagme nominal et du syntagme verbal (...) ; l'effacement des marques casuelles est toutefois un phénomène général et profond, dont les conséquences ne peuvent être limitées à la distinction des catégories majeures. (Combettes 2001:2)

En effet, les corrélations entre différents aspects de la linéarisation ont été mis en évidence en typologie (Greenberg, 1963 ; Lehmann 1982), en particulier entre l'ordre de base des constituants majeurs et la place des expansions nominales. Les contraintes liées à la fonction sujet en particulier font comprendre l'émergence de l'article comme phénomène local de la grammaticalisation du cadre syntaxique tout entier, qui conduit à poser le problème au-delà du cadre strictement syntaxique :

> Dans une optique fonctionnelle, il convient de se demander en effet si ce mouvement général a des relations avec d'autres niveaux d'analyse que le niveau morphosyntaxique ; qu'en est-il par exemple de l'influence des facteurs pragmatiques et textuels, en particulier des modifications que l'on peut percevoir dans le domaine de la sémantique référentielle, ou de l'évolution qui affecte la structure informationnelle de l'énoncé ? (Combettes 2001:9)

- La compréhension de l'articulation entre langue et discours en dépend... Car si les changements linguistiques sont, de manière générale, induits par l'usage, leur motivation peut être supposé résoudre des "problèmes communicatifs" (Detges & Waltereit 2016), autrement dit, elle renvoie aux enjeux de la construction du sens en interaction. L'un des processus structurels impliquant la routinisation d'un tel solutionnement est la grammaticalisation. Le concept a été promu pour expliquer la constitution d'unités ou de paradigmes grammaticaux au sens strict (Lehmann 1982). Son application à des constructions de portée excédant le cadre propositionnel pose le problème à la fois du modèle de syntaxe emprunté (et des unités d'analyse), et celui de la constitution du sens en contexte. Ces problèmes sont particulièrement flagrants dans les corpus oraux. Sans avoir fait ses preuves encore parmi les concepts

1 GN : groupe nominal, GPrep : groupe prépositionnel, GAdv : groupe adverbial ; DEM : démonstratif ; PRO : pronom ; IND : indéfini ; DEF : défini.

d'analyse du domaine, la notion de *pragmatisation* (*pragmaticalization* : Diewald 2011) vise à appréhender les cas où un glissement fonctionnel s'applique à un segment discursif qui "déborde" du cadre propositionnel.

## 1.2 Le système de détermination en français

Les études diachroniques mettent en exergue l'évolution des caractéristiques sémantiques et pragmatiques subie au cours de la grammaticalisation, dont, notamment, les valeurs déictiques et anaphoriques des pronoms et l'opposition de divers types de démonstratifs (Schøsler 2001, Selig 1992, Marchello-Nizia 1995). La grammaticalisation avancée de l'article en français moderne a été analysée par Schøsler (2001). En fondant son propos sur une taxonomie ±défini (relatif au statut informationnel) et ±référentiel, l'auteure observe l'emploi de plus en plus systématique de l'article, qui va de pair avec la neutralisation progressive de la valeur de définitude. Dès lors, l'article marque principalement les catégories grammaticales inhérentes au nom, *i.e*. nombre et genre (compensant la perte des déclinaisons caractéristiques du latin) ; tout comme en latin – et contrairement à l'ancien français – les combinaisons de valeurs -défini référentiel et +défini +référentiel ne sont pas formellement opposées. Enfin, la définitude est progressivement pris en charge par le *là* en position postnominale, qui se combine désormais au paradigme anténominal. Schøsler argumente en faveur de cette conclusion avec l'observation des données de Blanche-Benveniste (2000) qui laissent envisager *là* comme marquage « intersubjectif » entre une donnée connue du locuteur mais potentiellement nouvelle pour l'interlocuteur :

1. comme ils font les autres gens-là qui veulent rentrer le soir (Blanche-Benveniste 2000).

Le locuteur signale par *là* qu'il a en tête un groupe de gens particulier, qui ne sont cependant pas introduits dans le contexte amont, ce qui conduit à la détermination a postériori par la relative.

## 1.3 Le cas de *là* : problématique

La grammaticalisation de *là* en position post-nominale est allé de pair, historiquement, avec la spécialisation du démonstratif dans la fonction de déterminant depuis l'ancien français.

Si la fonction sémantique consistait initialement à compenser la neutralisation de la valeur proximale présente dans l'ancien paradigme démonstratif du latin, la fonction actuelle en retour de l'érosion morpho-phonologique (*semantic bleaching*) est multiple : *là* fonctionne comme marqueur de définitude ; dans ce cas, son utilisation avec des noms à valeur indéfinie ou sa combinaison avec un article indéfini est impossible. Par ailleurs, il revient à *là* la fonction d'un marqueur de "clôture", de fin de syntagme. Ces emplois, particulièrement caractéristiques de l'oralité conceptionnelle, posent régulièrement des problèmes d'interprétation structurelle concernant sa portée dans les hiérarchies syntaxiques. A minima, il revient au *là* enclitique le rôle d'un marqueur de segmentation, dont les aspects rythmiques sont intuitivement repérés, mais pour lequel les études détaillées font défaut.

La trajectoire de grammaticalisation engagée par *là* semble des plus classiques (cf. Heine & Kuteva 2002 ; Kriegel 2003), de nombreuses études (p.ex. Baker 2003 ; Daff 1998 ; Detges & Waltereit 2009 ; Italia 2006 ; Vincent 1993 ; Ludwig & Pfänder 2003 ; Wiesmath 2003) en ont explicité les aspects structuraux pour différentes variétés du français et des langues créoles à base française. Cependant, les analyses quantitatives des données orales sont encore très peu développées. L'objectif de cette recherche sera l'étude exploratoire des caractéristiques combinatoires de *là* en français parlé, ayant pour objectif la remise en perspective des évolutions envisagées dans la littérature. Nous verrons que l'un des problèmes majeurs est l'interprétation structurelle même des formes relevées dans les corpus. Il n'est pas rare en effet de noter l'instabilité notoire des emplois :

> « Il semble alors que *là* ne peut intervenir que sur les entités thématisées ou thématisables dans le discours, et qu'il y a une différence référentielle entre les noms marqués par *là* et les noms marqués par un déterminant préposé, en ce que *là* confère le trait [+présupposé] au nom qu'il marque. Cependant, l'absence du marqueur *là* dans des exemples similaires dont le référent est déjà introduit dans le co-texte (sa bouche, leur bouche) va à l'encontre de cette interprétation. Il est pour cette raison difficile de formaliser une distinction sémantique entre les deux paradigmes en question qui expliquerait leur distribution respective dans le texte. Même si dans la plupart des cas il est possible de conférer une valeur stable à *là* (valeur spécifique maximale), il ne semble

> pas possible, pour le moment, d'articuler une différence sémantique entre les syntagmes det + nom + *là* d'une part, et Ø + nom + *là* d'autre part. (Knutsen & Ploog 2005:50)

- ou encore de « det + nom », pourrait-on ajouter : car si la description se réfère à un usage périphérique (celui d'Abidjan), la co-existence en synchronie de différentes étapes de grammaticalisation semble la règle plutôt que le cas d'exception. Enfin, comme l'indique le terme de *glissement sémantique*, les différentes interprétations possibles sont liées dans un continuum. Une délimitation catégorielle des emplois serait un non-sens dans la saisie descriptive de la grammaticalisation.

# 2. Corpus d'étude

## 2.1 ESLO-MD

Le corpus ESLO-MD (« microdiachronique ») est un corpus oral de français hexagonal composé d'un million de mots (environ 80h enregistrées) issus d'interactions parmi les corpus ESLO1 (1968-1971) et ESLO2 (2008-2019). Les données d'études sont équilibrées entre les deux périodes, en veillant à une répartition équilibrée également entre les genres interactionnels (conférences, repas, entretiens). L'échantillon est constitué comme suit :

| **ESLO-MD** | **Conférences** | **Repas** | **Entretiens** | **TOTAL** | **Nombre de mots** |
|---|---|---|---|---|---|
| ESLO1 | 192 mn | 196 mn | 2042 mn | 2430 mn | 453298 |
| ESLO2 | 186 mn | 201 mn | 2034 mn | 2421 mn | 521931 |
| TOTAL | 378 mn | 397 mn | 4076 mn | 4851 mn | 975229 |

Si l'on décline la taille de ESLO-MD en termes du nombre des sous-échantillons ainsi que des mots-occurrences, on obtient la ventilation suivante par genre et sous-corpus :

| | **Conférences** | | **Repas** | | **Entretiens** | |
|---|---|---|---|---|---|---|
| **Sous-corpus** | **Echantillons** | **Mots** | **Echantillons** | **Mots** | **Echantillons** | **Mots** |
| ESLO1 | 2 | 32866 | 4 | 40665 | 30 | 384279 |
| ESLO2 | 5 | 30432 | 8 | 40864 | 31 | 452293 |
| TOTAL | 7 | 63298 | 12 | 81529 | 61 | 836572 |

*Tableau 1: Nombre de sous-échantillons et de mots-occurrences par genre et sous-corpus*

Les considérations méthodologiques liées aux variables sociologiques qui ont présidé au choix de l'échantillon ESLO-MD sont détaillées dans Abouda & Skrovec (2018). Si les deux sous-corpus ne sont pas sociologiquement représentatifs ni même comparables au sens strict, les méthodologies de recueil le sont dans une large mesure, ce qui rend possible la documentation d'une temporalité intermédiaire des dynamiques linguistiques, entre changement et variation synchronique, dans un intervalle de 40 ans.

Le corpus ESLO dans son intégralité est librement accessible en ligne[2]. Les transcriptions de l'échantillon ESLO-MD ont été enrichies avec une annotation morpho-syntaxique automatique par TreeTagger (Schmid 1994), sous TXM, outil d'analyse textométrique libre (Heiden *et al*. 2010)[3]. Nous reproduisons les extraits tels qu'ils apparaissent dans la transcription ESLO.[4]

2 http://eslo.huma-num.fr (consulté le 24/11/19)
3 http://portal.textometrie.org/demo/ (consulté le 24/11/19)
4 La convention est consultable en ligne : http://eslo.huma-num.fr/index.php/pagetranscription (consulté le 24/11/19).

## 2.2 Limitation du champ

Notre étude est basée sur une concordance que nous avons créée au moyen de TXM. Comme le codage orthographique dans le corpus avec ou sans trait d'union montre des inconsistances, liées ou non à l'ambivalence sémantique de *là*, nous avons extrait ses occurrences à partir de son lemme (tel qu'annoté par TreeTagger). Les occurrences ont été annotées sans tenir compte de la variabilité de la transcription. En vue de l'analyse détaillée selon les différents critères syntaxiques, sémantiques et énonciatifs, que nous allons exposer dans la section 3, nous avons procédé à un échantillonnage aléatoire en retenant un tiers des 5134 occurrences obtenues au départ, ce qui nous donne une concordance de travail comprenant au bout du compte 1711 occurrences.

L'inventaire des catégories structurelles sous 3. fera apparaitre de nombreux types d'ambiguïtés structurelles. Certaines de ces ambiguïtés sont réelles et incitent à opérer avec des catégories "floues" ; dans d'autres cas, il s'agit d'un artéfact induit par la représentation écrite des données dynamiques. Une désambiguïsation par la prosodie était théoriquement possible pour ce corpus, dont le signal est disponible. Les outils d'annotation utilisés n'en permettaient cependant pas un traitement outillé, et l'annotation manuelle n'était pas envisageable au vu de la quantité de données. Pour cette étude exploratoire, ces (nombreux) cas ont fait l'objet d'une double annotation.

# 3. Approche qualitative : catégories d'annotation

## 3.1 Enclise

Un aspect de la grammaticalisation est la perte d'autonomie formelle, marquée, dans le cas de *là*, par la clise, *i.e.* son rattachement phonologique (prosodique) au constituant qui le précède (amont). Une étude prosodique devra établir la clitisation effective, car, bien souvent, la position finale permet une double lecture si l'on se base sur la seule transcription. Pour l'heure, nous excluons de l'étude les occurrences où *là* est syntaxiquement associé au contexte aval :

2. par exemple comme chez vous il y en a trois je pense les travaillistes les libéraux et puis euh les conservateurs mais j'estime que **là ça va** mais chez nous alors vous avez les communistes vous avez les démocrates vous avez les centristes (1502; ESLO1_ENT_149_C, YR399, 1:02:42)

Dans la même perspective, les marqueurs de structuration, très fréquents et apparaissant souvent dans des collocations (*alors là, donc là*) ou configurations syntaxiques récurrentes (*c'est là*), ont été écartés des considérations.

Les contextes syntaxiques, sémantiques et lexicaux sont très hétérogènes et témoignent de la productivité de différents parcours de grammaticalisation engagés. Parmi ces occurrences, il s'agira de focaliser sur les emplois clitiques, dans l'objectif d'établir les contours de l'érosion formelle. A ce titre, les locatifs PP/AP doublés de *là* posent problème, en ce qui concerne non seulement leur interprétation sémantico-référentielle, comme précédemment évoqué, mais également leur interprétation structurelle (cf. ci-dessous 3.2).

## 3.2 Emplois locatifs vs. non-locatifs

Le processus de grammaticalisation suppose, typiquement, le développement de contextes d'emplois abstraits, ou de moins en moins concrets. Afin de cerner plus précisément les contours de la grammaticalisation, l'étude des emplois non locatifs sera privilégiée. Nous entendons par emplois locatifs ceux qui réfèrent à un lieu dans l'espace concret. Pour établir le caractère locatif d'une occurrence, les tests de questionnement (*où ?*) et de substitution (*ici / là-bas*) ont été utilisés :

3. non bah tu veux prendre ma banane qui est juste **là** ? (5062: ESLO2_REP_22_C, LOCH2, 0:19:52)

*la banane qui est où ? / la banane qui est juste ici*

La dynamique de grammaticalisation est établie, dans le paradigme adverbial, notamment par l'emploi de *là* avec une référence temporelle. Dans de nombreux cas, *là* réfère à un objet ou état de choses du discours même, où la notion de lieu (*topos*) prend une dimension plus abstraite :

4. une certaine idée de la littérature conforme des provocations des transgressions on peut aussi estimer au contraire qu'il s'agit **là** d'un signe de temps potentiellement fécond susceptible de produire une nouvelle synthèse entre la théorie et la pratique entre la sensibilité et l'abstraction (2017: ESLO2_CONF_4CPMEb_C, Marcos Eymar, 0:19:32)
5. essaye au moins bah non mais il faut toujours essayer eh ben non pas quand c'est chaud c'est dangereux **ce que tu dis là** c'est dangereux je vais aller à l'hôpital ma main c'est très bon c'est très salé vraiment je pense pas que je (4939; ESLO2_REP_11_C, repas_03_01_loc03, 0:04:58)

Si les deux exemples qui précèdent (4. et 5.) illustrent des emplois distinctement non locatifs, d'autres occurrences, d'apparence locative, se situent sur un continuum représentationnel ontologique/discursif :

6. c'est rapproché au début y avait pas de terrains de construction hm hm hm c'est vrai mais nos terrains étaient en bordure justement **là** et quand on s'est marié et on a été exproprié trois fois (3450; ESLO2_ENT_1024_C, GC24, 0:04:38)
7. (l')aînée qui a fait des études commerciales elle elle pou- elle pourrait les faire maintenant à Orléans oui maintenant y a une chambre de commerce **là** euh (3022; ESLO2_ENT_1015_C, ch_AC7, 0:06:51)

Lorsque *là* succède à un syntagme prépositionnel locatif, le caractère locatif de *là* n'est souvent pas avéré :

8. vu il y a encore des des maisons là qu'ils sont en train de j'ai pas vu construire et de **sur l'avenue là** y a encore deux maisons (3085; ESLO2_ENT_1015_C, EW15, 0:40:42)

Dans cet exemple, *là* peut être interprété, *a priori*, comme redondance syntaxique (GPrep suivi de GAdv), ou comme marqueur du GN/GPrep. Car même si le lieu réfère à l'espace concret, la référence de *là* est endophorique. En particulier, ce sont les combinaisons du clitique avec un locatif adverbial (ici, *là*) qui font douter du caractère locatif du *là* clitique :

9. euh tiens moi quand j'ai eu mon appendicite et puis j'ai été ouvert **ici là** euh on mettait de la Bétadine mais c'est pour euh recouvrir ça désinfecte hein (V-la_345; ESLO2_REP_14_C, repas_04_01_loc01, 0:16:10)

Si l'on imagine que la mention *ici* s'accompagne peut-être d'une désignation gestuelle (interprétation locative), le clitique lui-même semble davantage se rapporter au contexte de parole qu'au lieu désigné.

Dans d'autres cas, la référence d'apparence locative, circonstancielle, est celle du discours même. Pour ces complois, endophoriques, l'interprétation comme locatif ou non locatif est souvent discutable. Nous noterons qu'une catégorisation binaire, certaine, comme étant ±[locatif] s'avère discutable dans de nombreux cas.

## 3.3 Types de GN

Dans 449 des 1711 occurrences cibles *là* succède à un GN. Ces exemples se répartissent sur différentes constructions de déterminants, respectivement, DEM (et PRO), IND, et DEF:

10. par exemple en revenant de Bidarray je ne sais pas si vous connaissez **cette chanson-là** (2027: ESLO2_CONF_5FLb_C, François Legouy, 0:07:23)
11. enfin si il en a eu une mais on la compte pas **celle-là** (5025: ESLO2_REP_21_C, repas_14_01_loc01, 0:13:38)
12. Claire m'a envoyé euh l'offre de poste euh pour **ton amie là** le elle peut postuler hein (4822: ESLO2_REP_01_02_C, INC1, 0:10:37)

13. ah oui pour euh pour euh avoir aussi de la langue euh voilà c'est un peu un échange c'est un peu **un échange là** (3636: ESLO2_ENT_1025_C, LF25FEM, 0:40:13)

14. c'est quand il a supprimé **la retraite là** il a supprimé la retraite du combattant (1639: ESLO1_ENT_160_B, 4003, 0:49:18)

Les constructions sans adjectif ou pronom démonstratif, en particulier celle avec le déterminant défini (14) représentent, dans la perspective de la grammaticalisation (cf. Schøsler 2001), l'étape la plus avancée du processus, étant donné que *là* semble prendre en charge la valeur de définitude/spécificité plus que de pointer un lieu à proprement parler et constituer ainsi un véritable déterminant nominal.

Bien que peu nombreuses, les occurrences impliquant un déterminant anténominal indéfini montrent une convergence intéressante, à savoir, qu'elles présentent toutes des traces marquées du travail de formulation, comme les disfluences, reformulations et répétitions dans l'élaboration du GN :

15. la cérémonie du Panthéon euh culte auto-orchestré enfin par Jack Lang euh euh qui euh résonne effectivement comme un une euh je dirais **une résurgence là** (2005: ESLO2_CONF_2JGb_C, Jean Guarrigues, 0:43:20)

Le cas résonne avec celui où *là* se rapporte à un GN distant, notamment dans un contexte d'expansion, de type proposition relative en particulier :

16. je vais demander euh à Pauline parce que hm j'ai vu la voiture que vite fait mais oui c'est une cinq portes c'est le euh hm deuxième modèle de Clio là **celle qui est un petit peu arrondie là** euh celle qu'on voit partout partout en ville (V-là_73; ESLO2_REP_14_C, repas_04_01_enqMM, 0:11:49)

La difficulté de prise en compte des GN expansés est liée à la formulation de la requête : concrètement, il faut évaluer le nombre d'unités (qui composent la relative) entre le GN et le *là* final, puis éliminer le "bruit" dû à des constructions n'étant pas des expansions nominales. Tout en se servant des fonctionnalités de TXM, et plus généralement du moteur de requête CQP (*Corpus Query Processor*, Evert & Hardie 2011), le décompte reste essentiellement manuel.

Une nouvelle hypothèse émerge alors : dans le contexte où *là* succède à un constituant nominal, sa portée ne se limite pas toujours, ou pas nécessairement, à celui-ci, comme pourrait l'illustrer le contexte de l'exemple précédemment cité sous 16, où la tête nominale, déjà marquée par *là*, est suivi d'une expansion, elle aussi marquée par *là*, qui crée un effet de redondance. Le rôle de *là* serait ainsi de renvoyer l'interlocuteur à des connaissances partagées:

17. chose qu'il a fait qui était qui a fait beaucoup de mal chez nous anciens combattants c'est quand il a supprimé **la retraite là** il a supprimé la retraite du combattant (1639 : ESLO1_ENT_160_B, 4003, 0:49:18)

Cette hypothèse pose, en retour, un problème substantiel à l'interprétation structurelle, précisément, concernant la portée de *là* dans les nombreux contextes où la référence reste floue, le plus souvent ancrée dans l'énonciation même : ces occurrences se caractérisent par la prévalence de pronoms interlocutifs (avec des sujets de 1e ou 2e personne), et/ou une modalité exclamative ou interrogative :

18. oui c'était l'intention bien sûr et c'est une question de justice et d'égalité (...) et puisque enfin c'était une sorte de privilège euh oui je suppose **c'est un privilège qu'on accorde là** l'école libre vous savez quand on parle à des gens qui sans approfondir la question ils vous répondront tout de suite et ben celui qui veut que ces enfants aillent à l'école libre il a qu'à payer voilà c'est ça (V-la_2; ESLO1_ENT_160_B, 4003, 0:41:56)

19. à se débrouiller sur tout ce qu'il faut c'est avoir une marchandise valable et un bon emplacement oui oui **vous êtes bien placé là** ? (676: ESLO1_ENT_045_C, OU, 0:03:07)

Dans ces exemples, le recours à *là* permet de lier l'énoncé qui le précède plus fortement à l'énonciation.

## 4. Analyse quantitative

Cette section est consacrée aux résultats du traitement statistique appliqué aux données qui relèvent de la distribution fréquentielle de *là* dans les constellations[5] qui impliquent, à différents niveaux de granularité, la dimension diachronique construite (certes de manière minimaliste) par la mise en contraste des sous-corpus ESLO1 et ESLO2, ainsi que la dimension différentielle des genres interactionnels (cf. l'aperçu donné en section 2.1). La méthode utilisée pour caractériser statistiquement les divergences distributionnelles observables est celle du calcul des spécificités proposée par Lafon (1980)[6]. L'intérêt de l'approche est de révéler, pour chaque partie du corpus concernée, les emplois qui se démarquent par une fréquence d'apparition atypiquement élevée ou faible par rapport à leur fréquence théorique relevant d'un modèle de distribution probabiliste, en l'occurrence hypergéométrique. Dans les deux cas de figure, on parlera respectivement d'emplois sur- ou sous-représentés. Quant aux emplois dont la fréquence observée est proche de la fréquence théorique, ils seront qualifiés de banals.

Si l'on se place dans un premier temps au niveau de différenciation le moins développé, qui concerne la seule dimension diachronique, on constate qu'avec un score de spécificité de 12,43, *là* (tous emplois confondus, tableau 2 ci-après) est sur-représenté[7] dans ESLO2, ce qui indique une augmentation significative de sa fréquence depuis le tournant des années 1960/1970, représenté par ESLO1.

Ce constat se nuance lorsqu'on différencie la distribution de *là* entre ESLO1 et ESLO2 par genre interactionnel. Comme le montre le tableau 1, c'est notamment dans les entretiens que *là*, qui s'avère nettement sur-représenté par rapport à ce genre dans ESLO2, a connu une forte progression.

| | **CONF** | | **ENT** | | **REPAS** | |
|---|---|---|---|---|---|---|
| **Mot** | **ESLO1** | **ESLO2** | **ESLO1** | **ESLO2** | **ESLO1** | **ESLO2** |
| *là* | 1,1903 | -1,1903 | **-16,9691**[8] | **16,9691** | -0,4263 | 0,4263 |

*Tableau 2: Scores de spécificité calculés pour là dans ESLO1 et ESLO2 avec différenciation préalable par genre*

Si nous distinguons les occurrences de *là* par type d'emploi (locatif [LOC], ambigu [LOC / NON-LOC], non-locatif [NON-LOC], cf. section 3.1), les observations que nous venons de faire se confirment en ce sens que chaque emploi pris pour lui-même a augmenté en passant de ESLO1 à ESLO2 (cf. tableau 3) et que cette augmentation concerne notamment le genre des entretiens (cf. tableau 4).

| **Emploi** | **ESLO1** | **ESLO2** |
|---|---|---|
| LOC | **-7,3378** | **7,3378** |
| LOC / NON-LOC | **-12,4251** | **12,4251** |
| NON-LOC | **-2,4953** | **2,4953** |
| *#RESTE* | *12,4272* | *-12,4272* |

*Tableau 3: Scores de spécificité des emplois de là dans ESLO1 et ESLO2*

| | **CONF** | | **ENT** | | **REPAS** | |
|---|---|---|---|---|---|---|
| **Emploi** | **ESLO1** | **ESLO2** | **ESLO1** | **ESLO2** | **ESLO1** | **ESLO2** |
| LOC | 0,2477 | -0,2477 | **-8,9974** | **8,9974** | -0,2947 | 0,2947 |

5 Le terme de *constellation* renvoie, dans le contexte de cette étude, au croisement de deux variables, *i.e.* linguistique et extralinguistique.

6 Les calculs ont été effectués au moyen du package R textometry (Heiden et al. 2010), qui est disponible à l'adresse https://CRAN.R-project.org/package=textometry et que nous avons utilisé en ligne de commande, et non pas à partir de l'interface graphique de TXM.

7 Dans la suite, nous parlons de sur-représentation lorsque le score de spécificité observé est égal ou supérieur à 2,0, et de sous-représentation lorsque ce score est inférieur ou égal à -2,0. Conformément aux pratiques établies par les travaux lexicométriques faisant appel à la méthode du calcul des spécificités, nous considérons tout score situé dans l'intervalle allant de -2,0 à 2,0 comme statistiquement non significatif, c'est-à-dire banal.

8 Dans les tableaux que nous utilisons dans cette contribution pour donner un aperçu des scores de spécificité, les chiffres négatifs ou positifs surlignés en gras sur fond gris indiquent une sous- ou sur-représentation significative du mot ou de l'emploi en question. Les chiffres non surlignés indiquent, quant à eux, une fréquence d'apparition banale.

| | | | | | | |
|---|---|---|---|---|---|---|
| LOC / NON-LOC | -0,1367 | 0,1367 | **-12,3838** | **12,3838** | -0,613 | 0,613 |
| NON-LOC | 1,2805 | -1,2805 | **-4,2181** | **4,2181** | -0,2907 | 0,2907 |
| *#RESTE* | *-1,1903* | *1,1903* | *16,9691* | *-16,9691* | *0,4263* | *-0,4263* |

*Tableau 4: Scores de spécificité des emplois de là dans ESLO1 et ESLO2 avec différenciation préalable par genre*

La restriction des calculs à la seule constellation des emplois de *là* fait pourtant apparaître un détail important : rapportés à eux-mêmes et non pas à l'ensemble des occurrences des différents sous-corpus, les emplois connaissent des tendances opposées, les locatifs et les ambigus étant en progression au détriment des emplois non-locatifs, comme l'indique le tableau suivant :

| **Emploi** | **ESLO1** | **ESLO2** |
|---|---|---|
| LOC | **-2,0738** | **2,0738** |
| LOC / NON-LOC | **-6,5466** | **6,5466** |
| NON-LOC | **7,8935** | **-7,8935** |

*Tableau 5: Scores de spécificité des emplois de là dans ESLO1 et ESLO2 avec exclusion du reste des occurrences des sous-corpus*

A en juger les chiffres du tableau 6, donnés ci-dessous, c'est encore le genre des entretiens qui est l'épicentre du changement en termes fréquentiels.

| | **CONF** | | **ENT** | | **REPAS** | |
|---|---|---|---|---|---|---|
| **Emploi** | **ESLO1** | **ESLO2** | **ESLO1** | **ESLO2** | **ESLO1** | **ESLO2** |
| LOC | -0,416 | 0,416 | **-2,1844** | **2,1844** | 0,2729 | -0,2729 |
| LOC / NON-LOC | -0,185 | 0,185 | **-5,1983** | **5,1983** | -0,564 | 0,564 |
| NON-LOC | 0,4502 | -0,4502 | **7,2451** | **-7,2451** | 0,4158 | -0,4158 |

*Tableau 6: Scores de spécificité des emplois de là dans ESLO1 et ESLO2 avec différenciation préalable par genre et exclusion du reste des occurrences des sous-corpus*

Pour obtenir une vision plus claire des tendances relevées jusqu'ici, il semble utile de procéder par coupe synchronique en séparant les sous-corpus ESLO1 et ESLO2 et en projetant la constellation respective des emplois à leur distribution par genre.

| | **ESLO1** | | | **ESLO2** | | |
|---|---|---|---|---|---|---|
| **Emploi** | **CONF** | **ENT** | **REPAS** | **CONF** | **ENT** | **REPAS** |
| LOC | -0,8781 | -0,6347 | 1,6297 | -0,8155 | 0,3335 | 0,5166 |
| LOC / NON-LOC | -1,8117 | 1,1831 | -0,2501 | **-2,9002** | **2,8167** | -1,0165 |
| NON-LOC | **2,0793** | -0,3581 | -1,3563 | **3,0279** | -1,9637 | 0,4728 |

*Tableau 7: Scores de spécificité des emplois de là dans ESLO1 et ESLO2 selon les genres, avec séparation par coupe synchronique*

Ce qui ressort de ces chiffres, c'est d'une part la position spécifique qu'occupent les emplois non-locatifs par rapport au genre des conférences dans le cadre de ESLO1. Il s'agit là du seul cas de sur-représentation dans une constellation, qui pour le reste des associations entre genre et type d'emploi se caractérise par une distribution banale. D'autre part, dans le cas du sous-corpus ESLO2, on a affaire à une sous-constellation quasi symétrique qui oppose les emplois ambigus aux emplois non locatifs dans leurs rapports avec les genres interactionnels. Ainsi, là où la sur-représentation des emplois ambigus dans les entretiens contraste avec leur sous-représentation dans les conférences, les emplois non-locatifs gardent leur association préférentielle à ce genre interactionnel alors qu'ils sont à la marge de la sous-représentation par rapport au genre des entretiens (ce dont témoigne le score de spécificité de -1,96, proche du seuil de -2,0). Au vu des tendances diachroniques que nous avons relevées ci-dessus, la progression que nous avons pu constater dans le cas des emplois ambigus dans ESLO2 (cf. tableaux 3 et 5) se reflète avant tout dans la position prédominante qu'ils ont acquise par rapport au genre des entretiens. Quant aux emplois non-locatifs, qui sont en nette régression en comparaison

directe avec les autres types d'emploi (cf. tableau 5), et ceci malgré une augmentation générale de leur fréquence (cf. tableau 3), ils maintiennent leur ancrage dans les conférences, mais ne résistent guère à la poussée des emplois ambigus dans les entretiens, où ils sont en passe de former le pôle négatif en termes d'association préférentielle. Pour ce qui est des emplois locatifs, l'accroissement significatif de leur fréquence qui se manifeste de ESLO1 à ESLO2 semble avoir contribué à la stabilité de leur comportement statistiquement neutre à travers les deux sous-corpus.

Pour terminer la présentation des résultats lexicométriques, nous allons nous intéresser au cas particulier de l'emploi non-locatif de *là* combiné à un GN, ce qui nous renvoie à la question, soulevée par Schøsler 2001 (cf. section 1.2), de la grammaticalisation de *là* en tant que marqueur de définitude. Les tableaux 8 et 9 présentent la distribution entre ESLO1 et ESLO2 des occurrences de *là* en fonction du déterminant avec lequel est construit le GN. Les scores de spécificité respectifs étant calculés, en 8, par rapport à l'ensemble des GN apparaissant dans les deux sous-corpus, en 9, par rapport au seul paradigme des GN auxquels est associé *là*.

| **Construction** | **ESLO1** | **ESLO2** |
|---|---|---|
| DET:def + N + *là* | -1,8732 | 1,8732 |
| DET:dem + N + *là* | -0,3063 | 0,3063 |
| DET:indef + N + *là* | -0,8155 | 0,8155 |
| DET:poss + N + *là* | 0,5021 | -0,5021 |
| PRO:dem + *là* | 0,7918 | -0,7918 |
| *#RESTE:DET+N* | *-1,3102* | *1,3102* |

*Tableau 8: Scores de spécificité des différentes constructions de GN combinées avec là dans ESLO1 et ESLO2*

| **Construction** | **ESLO1** | **ESLO2** |
|---|---|---|
| DET:def + N + *là* | -1,7943 | 1,7943 |
| DET:dem + N + *là* | 0,6455 | -0,6455 |
| DET:indef + N + *là* | -0,7636 | 0,7636 |
| DET:poss + N + *là* | 0,5546 | -0,5546 |
| PRO:dem + *là* | 0,948 | -0,948 |

*Tableau 9: Scores de spécificité des différentes constructions de GN combinées avec là dans ESLO1 et ESLO2 avec exclusion du reste des occurrences des sous-corpus*

Ces chiffres montrent que, d'une manière générale, la fréquence de chaque construction n'a pas évolué de façon significative d'une période à l'autre. L'on peut retenir, d'un côté, la distribution statistiquement non marquée des GN intégrant un démonstratif, qui forme avec *là* ce qu'il est convenu de catégoriser comme un déterminant composé discontinu, et, de l'autre, l'accroissement, tendant à la sur-représentation, de la fréquence des GN avec article défini auxquels se combine *là*, cas prototypique, relevé par Schøsler 2001, de l'emploi de *là* en tant que marqueur de définitude. Les scores de spécificité, statistiquement non significatifs, ne permettent pas d'établir la progression de la grammaticalisation de *là* dans le domaine de l'actualisation nominale dans l'intervalle ESLO1 / ESLO2.

## 5. Bilan

L'étude statistique révèle la particularité des emplois de *là* dans le genre entretien et une augmentation générale des occurrences de *là* pour lesquelles une catégorisation sémantico-référentielle ou syntaxique claire pose problème. L'augmentation des constructions les plus innovantes (DET:def + N + là), bien que constatée, n'est pas suffisamment massive pour être significative. Quelles conclusions peuvent être tirées de ces constats pour l'observation de la dynamique de grammaticalisation en cours ?

L'augmentation de *là* dans les entretiens de ESLO1 à ESLO2, ainsi la distribution différentielle significative dans ce genre en général pose question : cette différence révèle-t-elle le caractère significatif du genre pour l'observation de la dynamique, ou alors, les méthodes de recueil précises utilisées dans les entretiens diffèrent-elles de manière substantielle entre ESLO1 et ESLO2 ?

De manière générale, les ambiguïtés semblent en progression, sauf dans le genre conférence, monologal au contraire des genres entretien et repas. Or, il semble plausible que les genres dialogaux recourent plus fortement au marquage de l'intersubjectivité, a fortiori, de la spécificité.

Si les emplois innovants exploitent précisément l'espace de la sous-détermination, l'augmentation des occurrences qualifiées d'ambiguës parait indiquer l'affirmation des emplois innovants de *là*, en particulier, associés à un GN non démonstratif, expansé, ou encore à des segments propositionnels.

Sur le plan constructionnel, deux dynamiques inverses peuvent être observées dans les occurrences catégorisées comme non locatives : la tendance au figement lexico-grammatical dans les occurrences en bordure de GN, démonstratives en particulier ; le dépassement des contraintes lexico-grammaticales avec convergence des contextes énonciatifs dans les occurrences où *là* semble avoir une portée plus large.

## 5.1 Emplois non-locatifs à portée étroite et figement lexico-grammatical

L'étude des emplois non-locatifs de *là* permet de relever un grand nombre d'occurrences où il se trouve en bordure de GNs construits avec un déterminant démonstratif qui font partie intégrante d'un complément circonstanciel et dont la tête est pour la plupart constituée par un N abstrait. Une analyse plus ciblée, qui vise le comportement collocationnel de *là* au moyen d'un calcul de cooccurrences[9] appliqué à son co-texte immédiat à gauche (ici restreint à une fenêtre d'un mot), fait apparaître que les noms les plus fortement associés à *là* (en l'occurrence *moment*, *cas*, *époque*, cf. tableau 10A en annexe) relèvent d'une même catégorie fonctionnelle, en ce sens qu'on peut les qualifier avec Schmid (2000:275 sqq.) de « shell nouns » circonstanciels[10]. Nous avons affaire ici à un cas où *là* se trouve étroitement associé à des syntagmes qui se caractérisent par un certain degré de figement lexico-grammatical, et dont la fonction est de contribuer à la structuration textuelle en tant qu'organisateur temporel (*à ce moment-là*, *à cette époque-là*) ou connecteur de thématisation (*dans ce cas-là*). A cet égard, le comportement collocationnel de *là* témoigne d'une pragmatisation liée, en portée étroite, au cadre d'une construction lexico-grammaticale bien circonscrite.

## 5.2 Pragmatisation

Lorsque *là* se combine avec un GN déterminée hors démonstratif, le dénominateur commun des occurrences est sa fonction de borne de clôture intersubjective. Dans cette perspective, la sur-représentation de ces emplois dans les genres dialogaux du corpus semble plausible.

En corollaire, l'annotation de données fortement ancrées dans l'espace énonciatif défie l'approche quantitative. Pour l'ensemble de cette étude exploratoire, il convient noter que l'annotation morphosyntaxique de ESLO-MD issue du traitement avec TreeTagger ne permet pas, à ce stade, de contrôler les requêtes pour établir des classes d'occurrences consistantes. En effet, les catégories d'analyse restent encore largement à affiner pour profiler des classes plus précises. Le flou catégoriel de *là* en position finale (ou postposée) s'articule autour de trois problématiques :

- la distinction des emplois clitiques de ceux où *là* est un constituant syntaxique à lui seul; une meilleure catégorisation passera par une caractérisation prosodique des très nombreux types d'occurrences où *là* se rapporte à un contexte amont;

- une annotation morphosyntaxique détaillée des occurrences de *là* à portée "large" (GN expansées et non-GN), qui permettra d'argumenter la dynamique de pragmatisation ;

- une étude plus détaillée des locatifs, en particulier des combinaisons de *là* avec un GPrep (*sur l'avenue là*) ou GAdv (*ici-là) : c*ombinant les deux paradigmes formels, ce contexte structurel peut être considéré comme l'environnement d'origine de l'interprétation *origo* des occurrences à portée large.

9 Ce calcul a été effectué au moyen de TXM, qui implémente la méthode lexicométrique mise au point par Lafon (1981).

10 Sous-catégorie de shell nouns non prototypique selon Schmid (ibid.), les shell nouns circonstanciels renvoient aux concepts généraux de SITUATION, LIEU, TEMPS et CONDITION.

Reste que l'analyse qualitative montre de très nombreuses ambiguïtés structurelles. Or la sous-détermination structurelle est un terroir fertile pour la dynamique linguistique, en ce qu'elle offre concrètement un contexte favorable à l'émergence de structures innovantes.

## Références

Abouda, Lotfi/Skrovec, Marie (2018). Pour une micro-diachronie de l'oral : le corpus ESLO-MD, SHS Web Conf, volume 46, 6e Congrès Mondial de Linguistique Française.

Baker, Philip (2003). Quelques cas de réanalyse et de grammaticalisation dans l'évolution du créole mauricien. In: Kriegel, Sibylle (Hrsg.), Grammaticalisation et réanalyse: Approche de la variation créole et française, 111–141.

Blanche-Benveniste, Claire (2000). Transcription de l'oral et morphologie. In Guille, Martine/Kiesler, Reinhard (Hrsg.), Romania una et diversa, Philologische Studien für Theodor Berchem zum 65. Geburtstag, Band 1, Tübingen: Gunter Narr, 61–74.

Combettes, Bernard (2001). L'émergence d'une catégorie morphosyntaxique : les déterminants du nom en français. Abrufbar unter: http://journals.openedition.org/linx/818 (Stand: 30.04.2019)

Daff, Moussa (1998). Petite vitrine syntaxique du français oral en Afrique. In: Queffelec, Ambroise (Hrsg.) Alternances codiques et français parlé en Afrique. Aix-en-Provence : Publications de l'Université de Provence, 107-129.

Detges, Ulrich/Waltereit, Richard (2002). Grammaticalization vs. reanalysis: A semantic-pragmatic account of functional change in grammar. Zeitschrift für Sprachwissenschaft 21:2, 151–195.

Detges, Ulrich/Waltereit, Richard (2016). Grammaticalization and pragmaticalization. In: Fischer, Susann/Gabriel, Christoph (Hrsg.) Manual of grammatical interfaces in romance. Manuals of romance linguistics, Band. 10. Berlin/ Boston: De Gruyter, 635–657.

Diewald, Gabriele (2011). Pragmaticalization (defined) as grammaticalization of discourse functions. Linguistics 49:2, 265-290.

Evert, Stefan/Hardie, Andrew (2011). Twenty-first century Corpus Workbench: Updating a query architecture for the new millennium. In: Proceedings of the Corpus Linguistics 2011 conference, Birmingham, UK. Abrubar unter: http://www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2011/Paper-153.pdf (Stand: 27.02.20)

Greenberg, Joseph H. (Hrsg.) (1963). Universals of Language. Cambridge: MIT Press.

Heiden, Serge/Mague, Jean-Philippe/Pincemin, Béatrice (2010). TXM : Une plateforme logicielle open-source pour la textométrie - conception et développement. In: Bolasco, Sergio/Chiari, Isabella/Giuliano, Luca. 10th International Conference on the Statistical Analysis of Textual Data - JADT 2010, Jun 2010, Rome, Italie. Edizioni Universitarie di Lettere Economia Diritto, 2 (3),1021-1032. <halshs-00549779>

Hine, Bernd/Kuteva, Tania (2002). World Lexicon of Grammaticalization. Cambridge : Cambridge University Press.

Italia, Magali (2006). Le morphème là dans les variétés mésolectales et basilectales en français du Gabon: Le français en Afrique. ROFCAN 21, 281–290.

Knutsen, Anne M./Ploog, Katja (2005). La grammaticalisation de LA en (français) abidjanais : Une marque à tout faire ? In: Kabatek, Johannes/Pusch, Claus D./Raible, Wolfgang. (Hrsg.) Romanistische Korpuslinguistik II: Korpora und diachrone Sprachwissenschaft / Romance Corpus Linguistics II: Corpora and Diachronic Linguistic. Tübingen: Gunter Narr, 469–482.

Kriegel, Sibylle (2003). Grammaticalisation et réanalyse: Approche de la variation créole et française. Paris : CNRS éditions.

Lafon, Pierre (1980). Sur la variabilité de la fréquence des formes dans un corpus. Mots 1, 127–165.

Lehmann, Christian (2002 [1982]). Thoughts on Grammaticalization. 2. Aufl. Erfurt: Seminar für Sprachwissenschaft.

Ludwig, Ralph/Pfänder, Stefan (2003). La particule *là*/LA en français oral et en créole caribéen : Grammaticalisation et contact de langues. In: Kriegel, Sibylle (dir.) Grammaticalisation et réanalyse: Approche de la variation créole et française. Paris : CNRS, 269–284.

Marchello-Nizia, Christiane (1995). L'évolution du français. Paris: Armand Colin.

Schøsler, Lene (2001). Reanalysing Structure: The Modern French Definite Article, its Predecessors and Development. Acta Linguistica Hafniensia 33:1, 91–108.

Schmid, Helmut (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. Proceedings of International Conference on New Methods in Language Processing. Manchester, UK.

Schmid, Hans-Jörg (2000). English Abstract Nouns as Conceptual Shells: From Corpus to Cognition. Berlin / New York: Mouton de Gruyter.

Selig, Maria (1992). Die Entwicklung der Nominaldeterminanten im Spätlatein. Tübingen : Gunter Narr.

Vincent, Diane (1993). Les ponctuants de la langue et autres mots du discours. Québec : Nuit Blanche Editeur.

Wiesmath, Rafaële (2003). La particule *là* dans le parle acadien du Nouveau-Brunswick / Canada. In: Kriegel, Sibylle (dir.), Grammaticalisation et réanalyse. Approche de la variation créole et française, Paris : CNRS éditions, 285–302.

Ich habe ein paar Sachen im Text gefunden aber nicht in der Bibliographie:
- Marchello-Nizia 2001 (s.1)
- Detges & Waltereit 2009 (s.2)
- Heiden et al 2010 (s.3)
- Corpus ESLO 1, ESLO MD und ESLO 2 (sie erscheinen im Fußnoten aber vielleicht sollten sie in der Bibliographie?