Le registre comme continuum. L'interaction entre proximité de l'échange et hiérarchie d'accessibilité

Pierre Larrivée *

¹Normandie Université, Université de Caen Normandie, CRISCO, Esplanade de la Paix, CS 14032, 14032 Caen cedex 5, France

Résumé. Les registres ont été envisagés comme des entités discrètes, en particulier dans les approches qui voient le vernaculaire comme obéissant à des règles distinctes du français normé (Massot et Rowlett 2013). C'est au débat concernant la relation discrète ou continue entre les registres que contribue ce travail. Il propose de nouvelles données quant à l'impact du registre d'un échange sur le comportement des marqueurs de fonctions syntaxiques. En effet, dans certaines configurations comme les subordonnées relatives, les marqueurs de différentes fonctions (*qui*, *que*, *où*) suivent une hiérarchie d'accessibilité au sens de Keenan et Comrie (1977) : le relatif sujet est par exemple le plus fréquent de tous les relatifs utilisés. Cependant, selon des travaux récents, tout en restant dominante, la prépondérance relative du marqueur sujet diminue à mesure qu'augmente le degré de proximité de l'échange. Cette corrélation originale entre poids des marqueurs par fonction et degré de proximité des échanges (au sens de Koch et Oesterreicher 2001) est testée pour les pronoms clitics. Les données dégagées confirment qu'un échange de proximité entraîne un écart moins important entre les marqueurs des fonctions. La conséquence en est que les registres constituent une réalité non pas discrète, mais continue.

Abstract. The continuous reality of register. Correlations between the accessibility hierarchy and the proximity of exchanges. Registers have been envisaged as discrete entities, especially by approaches for which the vernacular register follows different grammatical rules from normative practice (Massot and Rowlett 2013). The discrete or continuous relation between registers is what this paper addresses. It puts forward new data regarding the impact of register on syntactic function markers. In some configurations such as relative clauses, the quantitative behavior of syntactic function markers (subject, object, oblique relative pronouns) follows an accessibility hierarchy (in the sense of Keenan and Comrie 1977): thus, subject relative pronouns are the best represented. However, following recent work, the preponderance of the subject decreases as the level of proximity of the exchange increases. This original correlation between the quantitative behavior of syntactic function markers and the proximity of exchanges, that apparently also applies to clefts, is tested for

* Corresponding author : pierre.larrivee@unicaen.fr

clitic pronouns. The data demonstrate that a proximity exchange correlates to less quantitative disparity between the function markers. This implies that register is a continuous rather than discrete reality.

1 Introduction

Le rapport entre syntaxe et registre¹ a été diversement apprécié. Longtemps, dans les langues de culture prescriptive, la pratique vernaculaire, spontanée et non préparée, a été considérée comme une version désarticulée de la pratique normée (Blanche-Benveniste 2010 : 78), par un étrange renversement qui fait d'une pratique subsidiaire l'entier de la langue. Avec l'émergence de la sociolinguistique, et particulièrement des travaux quantitatifs (Labov 1972), s'est posée la question de la relation entre les pratiques vernaculaire et normée. Cette question a été reprise pour le français dans le travail de Massot et Rowlett (2013) selon qui le français est une langue diglossique. Cela suggère que le registre vernaculaire obéit à des règles syntaxiques (au moins partiellement) distinctes de la pratique normée. Cette conclusion semble pourtant intuitivement trop forte. En effet, si certaines variables sont réservées à certains registres (le relatif *lequel*, la question partielle avec inversion de la pratique normée), la majorité sont partagées (le relatif *qui*, la question partielle *in situ*). Lefeuvre et Rossi-Gensane (2016) dans leur synthèse de la grammaire vernaculaire de l'interrogation montrent que si différences il y a, elles se trouvent surtout dans les taux d'usage. Si le taux d'usage d'un ensemble largement partagé d'options grammaticales est ce qui différencie les registres, plutôt que la forme des options elles-mêmes, force serait donc de constater que les registres forment non pas des ensembles disjoints, mais un continuum.

S'inscrivant au cœur de ces débats, ce travail a pour but d'évaluer l'impact du registre sur le comportement syntaxique. On se concentre sur de configurations syntaxiques où les formes critiques se différencient notamment par leur fonction syntaxique. Les pronoms relatifs du français se distinguent ainsi par la fonction qu'ils occupent dans la proposition qu'ils introduisent. Dans cette configuration, les pronoms relatifs s'organisent en une hiérarchie d'accessibilité selon Keenan et Comrie (1977), et la fonction sujet dominant la hiérarchie fait que le pronom *qui* est de très loin le plus fréquent. Cependant, un travail récent (Larrivée et Skrovec 2019) observe que le rapport quantitatif entre les membres de la hiérarchie se modifie selon le degré de proximité de l'échange telle que définie par Koch et Oesterreicher (2001). La proportion du sujet diminue dans les échanges caractérisés par un plus grand degré de proximité. La vérification de cette corrélation étonnante, et qui aurait des conséquences théoriques importantes pour la conceptualisation du rapport entre dimensions structurales et dimensions d'usage du langage, est l'objet de ce travail, qui pose la question de savoir si cette corrélation s'étend également aux pronoms personnels clitiques, où les formes sont aussi définies par leur fonction.

L'article est organisé comme suit. La section 2 fait l'état de travaux récents qui établissent l'impact du registre sur les marqueurs de fonctions organisés par la hiérarchie d'accessibilité. L'évaluation de cet impact est faite selon une méthode de corpus exposée dans la section 3. Les résultats sur le poids relatif des marqueurs de fonction pour les pronoms personnels clitiques sont présentés dans la section 4, qui confirme qu'un échange de proximité entraîne une disparité quantitative moins grande entre les fonctions de la hiérarchie d'accessibilité. L'importance de ces résultats est discutée dans la section conclusive, qui en envisage les cause.

2 État de la question

La hiérarchie d'accessibilité est une hypothèse typologique proposée par Edward Keenan et Bernard Comrie en 1977. Elle est élaborée à partir de l'étude des subordonnées relatives à travers les langues. Elle table sur le caractère distinct des relatives parmi les subordonnées, qui au moins dans certaines langues se structurent par un pivot qui varie dans sa forme selon la fonction qu'il occupe dans la subordonnée – ce qui n'est pas le cas des conditionnelles ou des concessives par exemple. Cela est manifeste en français dans la distinction entre *qui*, *que*, *où* et *dont*.

- (3) Le chalet *qui* est devant le lac.
- (4) Le chalet *que* les pins protègent du vent.
- (5) Le chalet *où* ils habitent.
- (6) Le chalet *où* elle a écrit ses mémoires.
- (7) Le chalet *dont* le toit menace de s'effondrer.

Les pronoms relatifs ne sont pas substituables car chacun marque la fonction qu'exercerait l'antécédent dans la subordonnée, savoir celles de sujet, d'objet, d'objet indirect, de complément oblique et de complément du nom respectivement. C'est dans cet ordre que Keenan et Comrie placent les fonctions des relatives. Cet ordre définirait une hiérarchie d'accessibilité, où chaque fonction est plus accessible que celles qui la suivent, et moins accessible que celles qui la précèdent.

- (8) Accessibility Hierarchy :
SU > DO > IO > OBL > GEN > OCOMP

Ainsi, le relatif à fonction de sujet est le plus accessible, plus que celle d'objet, qui prime sur l'objet indirect, qui domine le complément oblique, qui surpasse le génitif, lequel prend le pas sur le relatif comparatif. Cette hiérarchie se vérifie dans les comportements typologiques. Les langues qui ont une construction relative pour l'objet indirect en ont une pour l'objet direct. Autrement dit, la relativisation d'une fonction moins accessible entraîne celle des fonctions plus accessibles, mais non l'inverse : ainsi, la fonction la moins accessible de la hiérarchie, représentée en urhobo (Keenan & Comrie 1975 : 75, leur exemple (29))

- (9) oshale na l- i Mary rho n- o
man the that Mary big than him

n'est pas attestée en français.

- (10) ?* L'homme que Mary est plus gros que

En outre, la hiérarchie se manifesterait dans la facilité de traitement et, comme l'a montré la littérature ultérieure, l'acquisition : les relatifs plus accessibles sont plus faciles à produire et à comprendre, et s'acquièrent plus tôt en langue maternelle et seconde (voir par exemple Larrivée 2018 et les références qui y sont fournies).

Ce cadre est utilisé par Larrivée et Skrovec (2019) pour explorer les pronoms relatifs en français vernaculaire. La question que les auteurs posent est de savoir si cette variété suit la même hiérarchie que le français normé. Il pourrait se faire que la généralisation de Keenan et Comrie ne vaille que pour les variétés normées des langues, qui souvent sont les seules

dont rendent compte les descriptions sur lesquelles reposent les comparaisons typologiques. Le français vernaculaire est exploré sur la base du corpus ESLO. Ce corpus représente un ensemble de situations d’interactions orales. Le nombre de chaque relatif y est relevé, et reflète la hiérarchie d’accessibilité.

Tableau 1. Relevé brut du nombre d'occurrences des pronoms relatifs du corpus ESLO2 étiqueté (900 000 mots)

Pronom relatif	Nombre d'occurrences	Pourcentages
<i>qui</i>	7623	69 %
<i>que</i>	2372	21.5 %
<i>où</i>	841	7.6 %
<i>dont</i>	125	1.2 %
<i>XquelX</i>	57	0.5 %
<i>quoi</i>	20	0.2 %
Totaux	11038	100 %

En effet, le relatif le plus accessible *qui* représente un plus grand nombre d’occurrences que *que*, *que* plus que *où*, *où* que *dont*, et *dont* que les composés de *lequel*. En outre, *qui* d’une part, *que* et *où* d’autre part, et *dont* et les composés de *quel* représentent des taux croissants de dysfluences (à environ 3% pour le premier, 11% pour les deux suivants et entre 14% et 18% pour les deux derniers). Hormis *quoi*, qui pour des raisons historiques est utilisé uniquement dans des fonctions indirectes en français, les marqueurs suivent les comportements quantitatifs et qualitatifs prédits par la hiérarchie.

Le travail fournit également des données sur le comportement des relatives dans le contexte des clivées. La raison en est que dans l’acquisition de la langue seconde, l’emploi en clivées est un contexte critique, où les apprenants avancés surgénéralisent l’emploi de la relative (Welcomme 2013). Les relatives devraient par leur poids respectif également y suivre la hiérarchie d’accessibilité. L’examen dans le corpus des relatifs dans les constructions clivées² *c’est X qu- P*, présentative *il y a X qu- P* et semi-clivée *ce qu- P, c’est X* donne les résultats suivants.

Tableau 2. Relevé brut du nombre d'occurrences des pronoms relatifs dans une clivée du corpus ESLO2 étiqueté (900 000 mots)

Pronom relatif	Nombre d'occurrences	Pourcentages
<i>qui</i>	234	71,5 %
<i>que</i>	49	15 %
<i>où</i>	39	11,9 %
<i>dont</i>	2	0,6 %
<i>XquelX</i>	3	0,8 %
Totaux	327	100

On le voit, les ordres de grandeur sont respectés, avec la grande majorité des occurrences représentées par *qui* et *que*, *où* livrant l’essentiel du reste, *dont* et les composés de *quel* faisant acte de présence.

En lien comme les clivées avec la structure informationnelle, les dislocations correspondent à des syntagmes généralement nominaux ou prépositionnels qui se trouvent détachés d’une proposition, où ils sont repris par un anaphorique (Berrendonner 2015). Ces syntagmes sont susceptibles de représenter différentes fonctions. La question qui se pose est de savoir si les fonctions représentées par les dislocations démontrent une asymétrie quantitative, et si cette asymétrie reflète la hiérarchie considérée. Une réponse peut être trouvée dans deux études quantitatives récentes. D’une part, dans sa synthèse de l’emploi

des dislocations, Blasco-Dulbecco (2019) recense les fonctions des dislocations dans le corpus CorpAix de français vernaculaire. Les résultats sont présentés dans le tableau suivant, adapté du tableau original.

Tableau 3. Répartition DG / DD selon la fonction syntaxique

	CorpAix	
	<i>À gauche</i>	<i>À droite</i>
<i>Sujets</i>	76,6%	65,9%
<i>COD</i>	9%	16%
<i>Complément prépositionnel</i>	9,4%	11,6
<i>Adjoints</i>	5%	6,84%

Les pourcentages des fonctions représentées par les dislocations gauche et droite sont très largement comparables. Dans les deux cas, les dislocations sujet constituent la vaste majorité des emplois. Les compléments directs et les indirects se disputent la deuxième place en position gauche, le premier distançant le second à droite. La dernière place est occupée par les adverbiaux. On notera que le décompte semble distinguer les compléments prépositionnels sous-catégorisés par le verbe (sous la catégorie *complément prépositionnel*, de type A *Clermont-Ferrand, j’y habite depuis toujours*) et ceux qui ne le sont pas (de type A *Clermont-Ferrand, j’y ai acheté un chalet*), ce qui n’était pas le cas auparavant pour les fonctions du pronom relatif *où* par exemple.

En outre, le travail sur les relatives de Larrivée et Skrovec (2019) montre que si la forme de la hiérarchie se maintient, le poids relatif des termes de la hiérarchie varie selon le degré de proximité des échanges. La proximité est entendue au sens de Koch & Oesterreicher (2001) qui réfère à la planification du discours et à sa formalité, lesquels s’associent typiquement à une plus grande surveillance des formes de l’expression par le locuteur et au choix de variables normées. Un degré différencié de proximité est démontré par les échanges contenus dans le corpus ESLO2. La proximité est intuitivement le plus grande pour le sous-corpus *24 heures*, réunissant les productions d’une étudiante pendant une journée, et diminue progressivement avec les *repas*, les *entretiens* semi-dirigés, les *assemblées générales* et les *conférences*. Cette dimension n’affecte pas l’existence de la hiérarchie d’accessibilité, comme en témoignent les nombres d’occurrences fournis dans tableau suivant, de Larrivée et Skrovec (2019).

Tableau 4. Nombre d'occurrences des relatifs par genre

	24h		Repas		Entretiens		Assemblées générales		Conférences		Total
<i>qui</i>	123	65,1 %	189	70,8 %	242	70,1 %	456	77 %	495	78,9 %	1505
<i>que</i>	53	28 %	53	19,9 %	61	17,7 %	85	14,4 %	65	10,4 %	317
<i>où</i>	11	5,8 %	22	8,2 %	35	1 %	36	6,1 %	36	5,7 %	140
<i>dont</i>	1	0,5 %	3	1,1 %	3	0,9 %	10	1,7 %	21	3,3 %	38
<i>lequel</i> et <i>composés</i>	1	0,5 %	0	0 %	4	1,2 %	5	0,8 %	10	1,6 %	20
<i>Total</i>	189		267		345		592		627		2020

Par contre, le poids relatif des items indiqué par les pourcentages évolue avec le degré de distance communicative. Cette évolution est régulière avec *qui*, qui passe de 65 % à 79 %, et, à l’inverse, avec *que* de 28 % à 10 %. Une régularité est détectable pour *dont* qui passe de 0,5 % à 3,3 % et pour *lequel*, de 0,5 % à 1,6 %, avec une baisse inattendue de 1,1 % dans *Repas* à 0,9 % dans *Entretiens*. Le comportement quantitatif de *où* ne démontre pas de progression particulière. Les irrégularités sont cependant le fait d’items ayant moins de

50 occurrences par sous-corpus. La relation demeure entre le poids respectif des marqueurs de fonction de la hiérarchie et la proximité de l'échange.

Malheureusement, cette dernière étude ne fournit pas les données concernant les emplois en clivées par sous-corpus. De telles données figurent dans une étude récente de Davide Garassino (2019). Ce dernier compare le comportement des clivées du français sous différents rapports, y compris la fonction syntaxique de l'élément au foyer. Il se fonde sur des sources orales, les discours faits au parlement européen, et des sources écrites, le *Monde diplomatique*. Même si ces échanges relèvent de situations formelles, elles permettent néanmoins d'évaluer une différenciation de proximité. Les résultats sont fournis dans le tableau suivant, adapté de Garassino.

Tableau 5. Relevé du nombre et % du focus des clivées par fonctions dans deux corpus

	<i>Europarl</i>		<i>Le Monde</i>	
	Nombre d'occurrences	Pourcentages	Nombre d'occurrences	Pourcentages
<i>Sujet</i>	37	37	54	59
<i>Objet</i>	7	7	7	8
<i>Complément prépositionnel</i>	47	47	27	29
<i>Adverbial</i>	9	9	4	4
<i>Totaux</i>	100	100	92	100

Les pourcentages ne correspondent pas aux attentes en termes de hiérarchie d'accessibilité, puisque les compléments prépositionnels sont plus nombreux que les objets directs, et plus que tout autre catégorie à l'oral, sujet compris. Il faut cependant souligner que ces disparités peuvent provenir de distorsions dues au nombre relativement faible d'occurrences. Un seuil inférieur à 200 occurrences reste problématique, comme on le reverra dans la prochaine section.

Pour ce qui est des dislocations, Raikovic & Skrovec (2019) les étudient dans cinq types d'entretiens de la base ESLO2, et en quantifient différentes dimensions. Une de ces dimensions est la fonction syntaxique de l'élément dans la proposition qui fait anaphore à la dislocation. Agrégeant les chiffres à travers les cinq sous-corpus, ils fournissent les pourcentages suivants, qui sont similaires à ceux de Blasco-Dulbecco.

Tableau 6. Nombre d'éléments disloqués par fonction syntaxique

	ESLO2	
<i>Sujets</i>	707	84,9 %
<i>Complément non prépositionnel</i>	71	8,5 %
<i>Complément prépositionnel</i>	24	2,9 %
<i>Complément locatif</i>	15	1,8 %
<i>Autres</i>	15	1,8 %
	832	

Ce que les auteurs montrent en outre est que plus le type d'échange relève de la distance communicative, plus la proportion des dislocations gauche est écrasante. Les dislocations fournissent donc un exemple supplémentaire de respect de la hiérarchie d'accessibilité, et de variation selon le registre dans l'écart quantitatif entre chaque fonction de cette hiérarchie.

Les études montrent donc un rapport entre variation diaphasique et poids relatif des fonctions syntaxiques représentées. On peut se demander si le même impact peut être établi pour d'autres phénomènes justiciables d'une hiérarchie syntaxique. Parmi ces phénomènes

figurent les pronoms personnels clitiques par exemple. On sait, comme le rappelle Meisner (2010), que pour les sujets, 97% sont représentés par des clitiques dans les échanges informels et 81% dans les échanges formels. L'impact du registre sur des marqueurs de fonctions syntaxiques est donc manifeste, et on serait en droit de s'attendre à retrouver la variation régulière dégagée pour les relatives. Ce sont donc ces clitiques que nous examinons dans leur rapport entre poids relatif des marqueurs et proximité de l'échange. Les choix méthodologiques de l'étude, qui cherchent à rendre les résultats aussi comparables que possibles à ceux de Larrivée et Skrovec (2019) pour les relatives, sont présentés dans la prochaine section.

3 Méthode

Ce travail examine les rapports entre registre et taux d'usage des marqueurs syntaxiques de certaines configurations supposées soumises à la hiérarchie d'accessibilité. Ces configurations comprennent le paradigme des pronoms personnels clitiques. En effet, le tableau de ces formes distingue, malgré les syncrétismes, une forme sujet (*il* par exemple), une forme objet (*le*), un objet indirect (*lui*) et une oblique (*y*) (Choi-Jonin et Lagae 2016). Si elles sont soumises à la hiérarchie d'accessibilité, on s'attendrait à ce que chacune de ces formes soit quantitativement mieux représentée que les suivantes. Si cette hiérarchie est corrélée à la variation de registre observée, on s'attendrait à ce que l'écart entre les formes se réduise avec le plus grand degré de proximité de l'échange.

La vérification de ces attentes pose de nombreux défis méthodologiques. D'abord, il faut pouvoir contrôler le registre représenté par des échanges contrastés sous ce rapport. Des différenciations de registre ne sont pas offertes par tous les corpus, et ce sont donc des corpus différents auxquels il faudrait normalement recourir, corpus différents dont les conditions de constitutions divergentes peuvent entraver la comparaison. Heureusement, ESLO2 présente des sous-corpus caractérisés par une différenciation de registres. Cette ressource a été constituée dans les années 2000 pour faire pendant à un premier ensemble réuni dans les années 1960 par des sociolinguistes britanniques. Sont retenus trois sous-corpus, qui représentent trois positions sur le continuum entre proximité et distance communicative, tel que démontré par Larrivée et Skrovec (2019). Le sous-corpus 24h présente la captation auto-enregistrée des échanges d'une étudiante pendant une journée ; un des cinq fichiers est disponible sur le site, pour une séquence de deux heures. Les entretiens sont des interactions décrites comme « discussions en face à face entre un chercheur et un locuteur témoin à partir d'une trame d'entretien » ; 78 fichiers sont accessibles sur le site, représentant 150 heures. Les conférences enfin sont des exposés scientifiques monologiques, que ne décrit pas plus avant le site.

Des défis supplémentaires sont posés par l'investigation des clitiques. D'une part, les formes présentent une homonymie grammaticale notable : *le* est à la fois un pronom personnel clitique et un déterminant défini ; *lui* un pronom personnel clitique et un pronom personnel tonique. Cela ne serait que peu problématique si le corpus disponible en ligne était annoté morphosyntaxiquement, ce qui n'est malheureusement pas le cas. Des emplois figés sont en outre susceptibles de brouiller les résultats : *y* intervient dans l'expression très fréquente (*il*) *y* a. Face au nombre important d'occurrence obtenues pour les sous-corpus *Entretiens* et *Conférences* à la requête des formes sous la fonction « mot exact » dans le moteur de recherche du site ESLO, j'ai dans un premier temps cherché à travailler par sondage pour les formes homonymes. Les 100 premières occurrences de chaque forme dans chaque sous-corpus ont été annotées manuellement ; le pourcentage constitué par le nombre de formes recherchées parmi ces 100 occurrences est appliqué au nombre total de formes pour donner un total présumé ; par exemple, sur les 5876 *il* bruts dans les entretiens, 90% des 100 premières occurrences correspondent au pronom personnel clitique sujet, ce qui livre une projection de 5288, projection indiquée par l'astérisque dans le tableau suivant.

De même pour les emplois productifs de *y*. Les résultats de cette démarche sont fournis par le tableau suivant.

Tableau 7. Projection du nombre et % de clitiques dans 3 sous-corpus d’ESLO2

	24h		Entretiens		Conférences	
<i>il</i>	83	71,6 %	5288*	89,8 %	234 *	89,3 %
<i>le</i>	12 *	10,3 %	149*	2,5 %	6*	2,3 %
<i>lui</i>	17	14,7 %	379*	6,4 %	21	8 %
<i>y</i>	4	3,4 %	70*	1,2 %	1*	0,4 %
<i>Total</i>	11 6		5886		262	

Il présente deux types de résultats inattendus. D’une part, l’objet indirect *lui* est mieux représenté que l’objet *le*, et ce à travers les trois types d’échanges. D’autre part, les échanges *Entretiens* représenteraient, si on se fie au pourcentage de clitiques sujet, une distance communicative plus grande que les *Conférences*, alors que c’est l’attente contraire à laquelle donnaient lieu les chiffres concernant les relatifs dans le tableau 4. Ces anomalies jettent un doute sur la démarche par sondage. Elles rappellent en outre qu’un seuil de 100 occurrences est insuffisant. Cela m’a amené à envisager une démarche alternative de relevé exhaustif.

Pour des raisons de comparabilité avec Larrivée et Skrovec (2019), les sous-corpus utilisés précédemment sont maintenus. Ce sont les versions de ces corpus disponibles sur le portail ESLO2 qui sont utilisées, afin de favoriser la réplication du protocole et la vérification des résultats. Afin d’obtenir un nombre équipollent de résultats, au regard du nombre très important d’occurrences dans *Entretiens* et *Conférences* face à leur relative rareté dans *24h*, j’ai retenu le texte complet des transcriptions dont j’ai copié et collé le texte dans un document de traitement de texte, représentant un volume de 190 pages pour chaque sous-corpus. (Il s’agit des transcriptions de 24h Conf_1249, des Entretiens Ent_1001 et Ent_1002 et des Conférences Conf_1240, Conf_1241 et Conf_1243). J’ai ensuite cherché et annoté les formes individuelles des pronoms personnels dits clitiques. Comme les relatives, ces clitiques sont en principe sujets à la hiérarchie d’accessibilité puisqu’ils présentent des pronoms se distinguant par leur fonction syntaxique, de sujet (*qui, il*), d’objet (*que, le*), d’objet indirect et de compléments non-sous-catégorisés (*où, dont et lequel ; lui et y*). Ce sont les marqueurs généraux de troisième personne, le sujet *il*, l’objet direct *le*, les objets indirects *lui* et *y* qui sont retenus. De ces deux derniers, on suppose que *lui* signale surtout des fonctions obliques et *y* des fonctions adverbiales. Seules les formes de *y* hors de la construction très fréquente (*il*) *y a* sont comptabilisées. Les formes féminines et plurielles ont été incluses pour assurer une quantité suffisante de résultat. Les résultats de cette démarche sont exposés dans la prochaine section.

4 Résultats

Cette étude s’intéresse au rapport entre le comportement quantitatif des marqueurs organisés dans une hiérarchie d’accessibilité et le registre de l’échange. Un tel rapport a été observé pour les relatives du français, selon lequel la dominance du marqueur sujet diminue avec le degré de proximité de l’échange. S’observe-t-il pour d’autres phénomènes grammaticaux liés par la dite hiérarchie comme les pronoms personnels clitiques ? Les

formes annotées des clitiques de troisième personne en fonction sujet, objet, objet indirect et locatif dans le même ensemble continu de données textuelles de chaque sous-corpus d’ESLO2 différencié selon le degré de proximité de l’échange livre les données suivantes.

Tableau 8. Nombre et % de clitiques dans des transcriptions de trois sous-corpus d’ESLO2

	24h		Entretiens		Conférences	
<i>il(s), elle(s)</i>	261	65,6 %	221	83 %	204	73,3 %
<i>le, la, les</i>	100	25,6 %	34	12,8 %	57	20,5 %
<i>lui, leur</i>	29	7,3 %	3	1,1 %	16	5,8 %
<i>y</i>	8	2 %	8	3 %	1	0,4 %
<i>Totaux</i>	398		266		278	

Ces résultats correspondent partiellement aux attentes. Globalement, les chiffres relatifs montrent qu’est respectée la hiérarchie, sauf pour les deux dernières cellules des Entretiens, avec un nombre faible d’occurrences. Par contre, la dominance du sujet ne suit pas la progression attendue. Si le sujet est la fonction dominante dans les échanges de moindre proximité, en se fondant sur les observations précédentes à propos des relatives rapportées dans le tableau 4, on s’attendrait à ce que cette dominance soit optimale dans les Conférences. Or, cette dominance est pour les clitiques moins grande dans les Conférences que dans les Entretiens. Cela pose évidemment question. On ne peut attribuer ce comportement à un trop faible nombre d’occurrences, puisqu’on est bien au-delà des 200 occurrences pour chaque entretien. Il se pourrait que les transcriptions utilisées soient moins marquées du registre attendu que l’ensemble plus large utilisé par Larrivée et Skrovec (2019). C’est ce que j’ai voulu vérifier en comptant les relatives dans les transcriptions utilisées ici. Les chiffres résultants sont donnés par le tableau suivant (qui n’inclut pas les 4 emplois obliques *qui*, c’est-à-dire 1 dans les Entretiens et 3 dans les Conférences).

Tableau 9. Nombre et % de pronoms relatifs dans des transcriptions de trois sous-corpus d’ESLO2

	24h		Entretiens		Conférences	
<i>qui</i>	42	41,2 %	68	53,1 %	354	70,5 %
<i>que</i>	54	52,9 %	40	31,3 %	103	20,5 %
<i>où</i>	5	4,9 %	17	13,3 %	23	4,6 %
<i>dont</i>	1	1%	0	0	16	3,2 %

-quel-	0	0	3	2,3 %	9	1,8 %
quoi	0	0	0	0	0	0
Totaux	102		128		502	

On est donc face à un paradoxe. Un sous-ensemble des sous-corpus *Entretiens* et *Conférences* ne donne pas les mêmes niveaux de registre que l’ensemble des sous-corpus. Bien sûr, rien n’exclut à priori qu’une partie d’un sous-corpus puisse relever d’un degré de proximité (ou de distance) différent du corpus entier. Cela nous ramène à la difficile question de savoir comment évaluer le registre d’un échange. Cette question peut trouver une réponse dans les propositions de Coveney (2016). Il note que les pronoms d’adresse utilisés devraient se montrer révélateurs du registre d’un échange. On s’attend en effet à ce que dans un échange de proximité soit désigné l’interlocuteur par *tu* plutôt que par *vous*. On sait de même que le vernaculaire préfère le *on* au *nous*. Ainsi, les quatre formes ont été relevées dans les transcriptions retenues des trois sous-corpus.

Tableau 10. Nombre et % d'occurrences de clitiques de trois sous-corpus d’ESLO2

	24h	Entretiens	Conférences
<i>tu</i>	192	6	189
<i>on</i>	74	144	189
<i>vous</i>	27	51	28
<i>nous</i>	27	11	27
Totaux	320	212	433

Ce décompte présente des formes brutes. Il devrait strictement exclure les formes ne servant pas d’adresse à un interlocuteur singulier, et les *nous* et *vous* qui ne sont pas sujet. Néanmoins, les chiffres suggèrent que les *Entretiens* retenus relèvent d’un registre de proximité plus grand que les *Conférences*. En effet, alors que le rapport des *vous* aux *tu* est de 8,5 fois dans les *Entretiens* retenus, il est de 6,8 fois dans les *Conférences* choisies. Cela se compare à des résultats bruts de 252 *tu* et 27 *vous* dans toutes les *Conférences* du site ESLO2 (pour un ratio de 9,3), et de 3277 *tu* et 9311 *vous* dans tous les *Entretiens* (pour un ratio de 2,8), alors que ce ratio est de 0,1 dans l’entier de 24h (avec 252 *tu* et 27 *vous*). Autrement dit, la différence entre *Entretiens* et *Conférences* dans le tableau 8 et le tableau 4 tient effectivement au registre des transcriptions retenus par rapport à l’ensemble des transcriptions. La corrélation entre registre et écart des marqueurs de fonction qu’on avait établi pour les relatifs tient pour les clitiques.

5 Conclusions

Revisitant la question de la nature catégorique ou continue du registre, ce travail explore l'impact du registre d'un échange sur le comportement des marqueurs de fonctions syntaxiques liés par une hiérarchie d'accessibilité. Il se base sur l'observation récente selon laquelle la dominance des marqueurs plus accessibles, et en particulier le sujet, diminue graduellement avec la plus grande proximité de l'échange. La chose est démontrée pour les pronoms relatifs dans des sous-corpus d'ESLO2 différenciés par leur valeur de proximité au sens de Koch et Oesterreicher (2001) : *qui* est de moins en moins dominant à mesure que les échanges sont plus vernaculaires (Larrivée et Skrovec 2019). Cette corrélation est évaluée dans cet article qui examine le comportement des clitics dans une partie des mêmes sous-corpus. Les nouvelles données valident la corrélation : globalement, le clitique sujet est numériquement moins dominant dans les échanges de plus grande proximité. Cette corrélation entre comportement quantitatif des marqueurs de fonctions syntaxiques organisés par une hiérarchie et le registre d'un échange engage à envisager le registre comme un ensemble continu.

Du travail se dégage un certain nombre d'observations de portée générale pour la recherche à venir sur la dimension fréquentielle de l'usage réel en relation avec les paramètres structurels.

-Le registre pourrait bien n'être pas une dimension séparée du structurel. Alors qu'on a tendance à séparer les facteurs externes et les facteurs internes, la recherche présentée ici suggère un lien plus intime entre les deux dimensions.

-La fréquence des phénomènes linguistiques ne peut être évaluée en deçà d'un certain nombre d'occurrences. Ce travail suggère qu'à moins de 200 tokens, les risques de distorsion sont trop importants pour permettre des conclusions fermes.

-Partant, l'étude de phénomènes grammaticaux peut difficilement s'accomplir sans des ressources annotés syntaxiquement et pleinement interrogeables. Pour le français, le développement de telles ressources librement accessibles nous manque encore cruellement.

Références bibliographiques

- Berrendonner, A. (2015). Constructions disloquées. *Encyclopédie grammaticale du français*, encyclogram.fr
- Blanche-Benveniste, Cl. 2010. *Approches de la langue parlée en français*. Gap : Ophrys.
- Blasco-Dulbecco, M. 2019. Les dislocations en français contemporain vernaculaire. P. Larrivée et F. Lefeuvre (dirs). *Projet Fracov*. <http://www.univ-paris3.fr/francais-contemporain-vernaculaire-fracov>
- Choi-Jonin, I. et V. Lagae. (2016). Les pronoms personnels clitics. *Encyclopédie grammaticale du français*, encyclogram.fr
- Coveney, A. (2016). *La quête du vernaculaire dans l'étude de la variation grammaticale*. Congrès Mondial de Linguistique Française - CMLF 2016. Paris : Institut de Linguistique Française.
- Garassino, D. (2019). French cleft sentences across genres. What is (not) prototypical? Présentation à la Journée d'étude *When Data challenges Theory. Non-Prototypical, Unexpected and Paradoxical Evidence in the Field of Information Structure*, Freiburg, 16.02.2019. Diaporama, 60 diapositives.
- Gadet, F. (1996). Niveaux de langue et variation intrinsèque. *Palimpseste* 10, 17-40.
- Gadet, F. (1995). Les relatives non standard en français parlé, le système et l'usage. *Études romanes* 34, 141-162.
- Keenan, E. L. et B. Comrie. (1977). Noun Phrase accessibility and Universal Grammar. *Linguistic Inquiry* 8,1, 63-99.

- Koch, P. et W. Oesterreicher. (2001). *Gesprochene Sprache und geschriebene Sprache / Langage parlé et langage écrit*. G. Holtus, M. Metzeltin, Ch. Schmitt (dirs). *Lexikon der Romanistischen Linguistik*, Bd. I/2. Tübingen: Niemeyer. 584-627.
- Labov, W. 1972. *Sociolinguistic patterns*. Philadelphie: University of Pennsylvania Press.
- Larrivée, P. 2018. Complexité et relatives en français vernaculaire langue maternelle et langue seconde. Catrine Bang-Nilsen (dir.). Numéro spécial « Acquisition du français », *Synergies Pays Scandinaves* 13, 65-77. <http://gerflint.fr/Base/Paysscandinaves13/Paysscandinaves13.html>
- Larrivée, P. et M. Skrovec. 2019. Variation en français vernaculaire : le cas des relatives. *L'Information grammaticale* 163, 21-28.
- Lefeuve, F. et N. Rossi-Gensane. (2015), *Interrogation*. P. Larrivée et F. Lefeuve (dirs). *Projet Fracov*. <http://www.univ-paris3.fr/francais-contemporain-vernaculaire-fracov>
- Massot, Benjamin et Paul Rowlett. 2013. Le débat sur la diglossie en France: aspects scientifiques et politiques. *Journal of French Language Studies* 23,1, 1-16.
- Meisner, C. 2010. A corpus analysis of intra- and extralinguistic Factors triggering *ne*-deletion in phonic French. F. Neveu, V. Muni Toke, J. Durand, T. Klingler, L. Mondada et S. Prévost (dirs). *Congrès Mondial de Linguistique Française - CMLF 2010*. Paris : Institut de Linguistique Française.
- Raickovic L. et M. Skrovec. 2019. Les dislocations dans les ESLOs : valeurs informationnelles en interaction. Diaporama, 27 diapositives.
- Welcomme, A. 2013. Les subordonnées relatives à l'oral en français langue étrangère et en français langue maternelle. *Mémoires de la Société Néophilologique* LXXXVI, 97-136.

¹ La difficulté de définir la notion de registre est notoire (voir par exemple Gadet 1996). La perspective que j'adopte est celle de Koch et Oesterreicher.

² Permettant de recueillir un plus grand nombre d'occurrences, la prise en compte de différentes formes de clivée répond néanmoins à la définition d'une structure mettant en relief un foyer en le séparant de la proposition, et dont le contenu est communiqué par le foyer réintégré à la proposition : dans C'est l'épidémie qui m'inquiète, Il y a l'épidémie qui m'inquiète, Ce qui m'inquiète c'est l'épidémie, il s'agit de diversement mettre en relief un aspect de la même proposition, L'épidémie m'inquiète.