



HAL
open science

Transcrire : les bonnes pratiques des linguistes

Olivier Baude

► **To cite this version:**

Olivier Baude. Transcrire : les bonnes pratiques des linguistes. Journées d'étude de l'Association des Amis de Jacques Lacan L'interprétation dans la cure et dans la transcription des textes, 2005, Paris, France. halshs-01162548

HAL Id: halshs-01162548

<https://shs.hal.science/halshs-01162548>

Submitted on 11 Jun 2015

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.

L'interprétation dans la cure et dans la transcription des textes

Transcrire: les "bonnes pratiques" des linguistes

La relation forme écrite / forme orale de la langue est source de débats aux lourds enjeux théoriques depuis plus de trente ans chez les linguistes. Il ne s'agit pas ici de retracer l'ensemble de ces discussions mais de présenter et de mettre en perspective une initiative actuelle, émanant d'une communauté scientifique, qui s'est concrétisée par la rédaction du *Guide des bonnes pratiques pour la constitution, l'exploitation, la conservation et la diffusion des corpus oraux*. Les réflexions et les propositions contenues dans le Guide, tout autant que l'originalité de la démarche qui associe linguistes, juristes et conservateurs dans un travail réflexif sur la méthodologie de la transcription interpellent les différentes disciplines qui sont confrontées à la question de la mise à l'écrit de productions orales dans toute leur complexité.

La relation oral/écrit chez les linguistes.

Pour comprendre pourquoi les linguistes réfléchissent actuellement aux "bonnes pratiques" de transcription, il convient de situer la place de la langue orale et de sa relation à l'écrit dans le domaine de la linguistique. En effet, jusqu'au début des années 80, la forme orale de la langue n'intéressait que très peu des linguistes qui n'y recouraient que par défaut (quand la forme écrite était absente), soit pour l'étude des phénomènes proprement sonores (phonétiques, phonologiques, prosodiques,...), soit pour la description des langues et parlers dits exotiques (langues à tradition orales, parlers régionaux, et même le parler des jeunes enfants qu'on classait souvent parmi les objets qui n'accédaient pas entièrement au statut d'une parole structurée).

Il y a sûrement de nombreuses explications au fait que les linguistes ne souhaitent pas "entendre" la langue sous sa forme la plus concrète. Parmi leurs nombreuses réticences deux sont particulièrement déterminantes dans le contexte français : la prédominance du français écrit comme norme et la constitution d'un domaine scientifique fondée sur la dichotomie langue/parole. En effet, la France a une relation très particulière à la variété et à la diversité linguistique (rappelons que les premières enquêtes en France sur le parler des français aboutiront aux propositions contenues dans le rapport de l'abbé Grégoire de 1794 *sur la nécessité et les moyens d'anéantir les patois et d'universaliser l'usage de la langue française*. A partir de là, seul le français, qui plus est sous sa forme écrite (la plus normée), sera reconnu. Epistémologiquement l'étude de l'oral sera associé à la dialectologie, au recueil des patois et à la description des langues sans tradition écrite. Quand au début du XXe siècle Ferdinand Brunot parcourt la France avec les premiers instruments permettant d'enregistrer la voix humaine (les phonographes) il s'agit de créer les "Archives de la parole" avec la vocation de recueillir des patois et des variétés régionales et du français de tous les jours à des fins "d'étude, d'archivage et d'analyse de parlers d'hommes et de femmes parlant comme à l'auberge ou à la fontaine". Cependant cette initiative ne continuera pas et l'enregistrement de l'oral restera surtout lié à une démarche folklorique et/ ou muséographique. De fait, aux "archives de la parole" commencées en 1911, succéderont en 1927 les "archives de la parole et du geste", puis la "phonothèque nationale" en 1938 qui sera totalement incorporée au "département de l'audiovisuel" lors de la création de la Bibliothèque Nationale de France en 1995. Ces archives orales n'intéressent que fort peu les linguistes : la référence reste le français écrit comme le souligne Claire Blanche-Benveniste *"Mais qui s'intéresse au français parlé? (...) peu de gens y voient un objet légitime d'étude (même chez les linguistes) pour bon nombre de ceux-ci la langue parlée c'est bon pour l'exotisme; la description de la langue parlée vaut pour les dialectes et les patois du français ; elle vaut aussi pour les langues sans écritures dites "exotiques" ; mais pas pour une langue de culture comme le français"*.

La deuxième raison de ce manque de considération de l'oral chez les linguistes provient de la constitution même de cette discipline scientifique. En effet, on a l'habitude d'associer la linguistique moderne au *Cours de linguistique général* de Ferdinand de Saussure et aux trois dichotomies (synchronie/diachronie, signifiant/signifié, et surtout langue/parole). Or la lecture de cette dernière par les linguistes provoquera le rejet de la description des variations en général et des productions orales en particulier qui se retrouvent associées à l'étude de la parole et donc constamment exclues du champ d'une science vouée à la

recherche d'invariants structurés en système. Le structuralisme tout comme la grammaire générative Chomskyenne se fondent alors, sur une méthodologie qui exclut les manifestations de l'oral pour se concentrer sur la forme normée de l'écrit -les grammaires par exemple restent fondamentalement construites sur des exemples provenant de l'écrit- et sur des données écrites -il existe actuellement des grandes bases de données en ligne contenant du français écrit mais aucune ne prend en compte les données orales-.

Ce détour épistémologique fournit plusieurs éléments pour une réflexion sur l'enjeu théorique de la relation forme écrite/forme orale. Premièrement, Il y a un dédain de l'oral important dans le domaine de la linguistique mais aussi dans l'espace public plus large qui a pour conséquence de rendre inconcevable une représentation de l'oral sous une autre forme que la norme de l'écrit (l'oral restant considéré comme une forme *primitive* de la langue). Deuxièmement les théories linguistiques ont longtemps renforcées cette idée et l'écrit constitue encore aujourd'hui les données de base de la grande majorité des linguistes. Troisièmement ce manque de considération de l'oral a des effets considérables sur les techniques et sur la définition même de la transcription. En effet, jusqu'aux travaux de Claire-Blanche Benveniste commencés à la fin des années 70, la transcription était très peu théorisée et se résumait à une simple alternative : *transcription phonétique* ou *adaptation à l'écrit*.

Cette présentation diachronique et ces repères historiques précisent ce que les linguistes ont apportés (et n'ont pas apportés) à une théorie de la transcription au cours du XXe siècle. Au début du XXe, au moment même où l'enregistrement de la voix devient possible, la linguistique se structure autour d'une éviction du sujet et de l'oral au nom de la dichotomie langue/parole. La transcription est alors souvent réduite à un outil technique pour les phénomènes purement sonores ou comme première étape de description des langues non écrites (API). La réflexion théorique date en France de la fin des années 70 (1977 : création du Groupe aixois de recherche en syntaxe et publication du premier numéro de la revue Recherche sur le français parlé) et il faut attendre 1986 et l'ouvrage *le français parlé : transcription et édition* de Claire Blanche-Benveniste et Colette Jeanjean pour que la question semble tranchée. Il est très intéressant de faire le lien entre cette chronologie et celle des grands intellectuels nés au début du siècle (tel que Jacques Lacan) qui ne seront donc pas en mesure de profiter du travail des linguistes sur la transcription.

Transcription des corpus oraux

Depuis quelques années la situation a considérablement évoluée, d'une part le travail théorique sur la transcription a produit des résultats significatifs, la linguistique de corpus s'est considérablement développée et le travail sur des masses de données a remplacé l'intuition du chercheur et, d'autre part l'oral a conquis une part de légitimité. Nous assistons à un tournant majeur qui donne une place de première importance à l'oral et aux transcriptions. Depuis peu les initiatives se multiplient et c'est en toute logique qu'une réflexion sur la place de l'oral s'est élaborée. Elle s'est développée autour de trois questions principales : Transcrire quoi? Transcrire comment? Transcrire pourquoi?

Transcrire quoi? Cette première question n'est pas à négliger tant les linguistes rassemblent sous le terme de corpus oraux des objets très divers. On entend généralement par corpus oraux une collection ordonnée d'enregistrements sonores et de leurs transcriptions, cependant ces enregistrements peuvent être très hétérogènes (interview de personnalités, enregistrement médiatique, récit de vie, conversation d'anonymes, liste de mots prononcés par des panels de locuteurs, enregistrement de scènes de la vie quotidienne, conte traditionnel, conversation téléphonique, demande de renseignements à un guichet, discours politique, etc). Cette définition en terme de corpus oraux fut une révélation pour les linguistes, qui pour bon nombre d'entre eux (re)découvrirent le concept de "locuteur". En effet, l'oral c'est aussi un *contenu*, un *locuteur* une *personne* qui a des *droits*, et il faut alors se poser de nouvelles questions : que peut-on enregistrer, transcrire, diffuser? et sous quelles conditions?

L'hétérogénéité des corpus oraux accroissent la complexité de la situation : les grandes enquêtes actuelles (comme par exemple le projet Phonologie du Français Contemporain) donnent lieu à de nombreux enregistrements qui sont ensuite transcrits, catalogués, indexés mais aussi annotés, étiquetés, etc. Ils offrent ainsi la possibilité de travailler sur la prononciation, les liaisons, le lexique, la syntaxe de la phrase, mais aussi sur le contenu même...., Doit-on conserver ces enregistrements? Peut-on alors les diffuser (par l'internet par exemple et dans ce cas qui est propriétaire de quoi)? Les réponses ne sont pas

évidentes pour des objets aux double statut : "objets scientifiques" mais aussi "patrimoine culturel". Une grande partie des questions posées par la transcription se fondent alors dans les enjeux de la conservation (que conserver? sous quelle forme figée, etc.)

Le Guide des bonnes pratiques

C'est dans ce contexte que la Délégation générale à la langue française (direction du ministère de la culture) et le Cnrs ont constitué un groupe de travail pluridisciplinaire pour favoriser la collecte et l'exploitation de corpus oraux en résolvant notamment les problèmes juridiques.

Ce groupe de travail comporte donc des linguistes experts et de chercheurs de "terrain" porteurs de projets actuels, des représentants des fédérations de laboratoire du CNRS, des juristes, des représentants des grands organismes de conservation sous la tutelle du Ministère de la Culture et de juristes de ces institutions. L'objectif premier était de permettre le travail en commun sur un objet scientifique (l'oral et sa transcription), et de favoriser sa conservation et surtout sa diffusion (diffusion auprès de différentes équipes de recherche mais aussi auprès d'un public plus large). Or il est très vite apparu que les aspects juridiques étaient les premiers obstacles à la diffusion de l'oral transcrits (qui est propriétaire de quoi? Qui est responsable de la diffusion? Quel sont les autorisations à recueillir? Qu'en est-il du droit d'auteur?, etc.). Enfin, ce travail sur les aspects juridiques a très vite été lié à une réflexion sur l'éthique du chercheur et l'occasion d'une démarche réflexive sur ses méthodes.

Dans un premier temps, le groupe de travail s'est orienté vers l'élaboration par la communauté scientifique "de bonnes pratiques" avec les contraintes suivantes: premièrement il n'existe pas de réponses juridiques simples à l'exploitation de l'oral et à la transcription des données et deuxièmement les solutions passent systématiquement par un travail réflexif sur la démarche du chercheur, seul moyen pour qualifier le statut des enregistrements et les objets exploités et diffusés. Les "bonnes pratiques " consistent donc à clarifier les questions juridiques, mais aussi (et c'est là un point fondamental) à porter une réflexion sur le travail scientifique des linguistes dans le respect d'une éthique validée par la communauté scientifique.

Clarifier les points juridiques

Les problèmes juridiques se concentrent autour de deux grands domaines : protection de la vie privée et propriété intellectuelle. Le premier point ne concerne pas vraiment la transcription de séminaires, il s'agit principalement du respect de l'anonymat et plus généralement de la sensibilité des données personnelles contenues dans les enregistrements et leurs transcriptions. Notons toutefois que la réflexion autour de l'anonymisation d'enregistrement (quand il s'agit d'inconnus) s'est très vite déplacée au profit d'un travail sur l'acte du recueil de consentements. Ne pas rendre anonyme l'auteur , implique de reconsidérer la question du consentement éclairé. Ainsi, parmi les bonnes pratiques, une part importante est consacrée au recueil de consentement et d'autorisation. Les propositions contenues dans le Guide refusent de considérer que la question de la diffusion de paroles (quelles soient transcrites ou non) puisse se réduire à une demande d'autorisation souvent très allusive. La situation est le plus souvent particulièrement complexe et les changements de finalités multiples surtout avec l'évolution des médias et des moyens de diffusion. En ce qui concerne la transcription, le guide propose d'éclairer le consentement en fournissant notamment au témoin, des exemples de la forme que prendrait cette dernière. L'idée fondant cette proposition est simple : elle repose sur la méconnaissance supposée du locuteur de ce qu'implique un travail de transcription notamment en terme d'enjeux de représentation.

Le deuxième domaine concerné par les aspects juridiques est celui de la propriété intellectuelle et du droit d'auteur. Là également un travail pluridisciplinaire permet de révéler toute la complexité de la situation et les nombreux problèmes non résolus. La difficulté principale reste de qualifier le statut juridique des corpus oraux.

Les bonnes pratiques et les techniques de transcription

Le guide des bonnes pratiques ne se résume pas à des aspects juridiques. Il aborde également les "bonnes pratiques" de transcription. De fait, derrière les questions juridiques de protection des données personnelles et de droit d'auteur se profile la question du respect de la personne enregistrée et de l'éthique du chercheur. C'est bien cette même question que l'on retrouve dans l'acte de transcription.

Il convient tout d'abord de constater l'hétérogénéité des pratiques que recouvre le terme de transcription. On parle de transcription, qu'il s'agisse de transpositions ou adaptations utilisées pour ne transmettre que le contenu en changeant librement la forme orale ; d'une mise en norme orthographique et syntaxique consistant à transposer l'ensemble du contenu dans une forme respectant les règles de l'écrit (ponctuation, suppression des répétitions, restitution des formes contractées par la prononciation, inversion du sujet dans les formes interrogatives etc.) qu'elle soit rendue obligatoire par les règles de diffusion (publication d'un livre) ou pour des questions de traitement automatique (requête et moteur de recherche demandent une régularité dans la transcription) ; de transcriptions orthographiques respectant les particularités des productions orales (répétitions, hésitations lapsus, amorces de mots) ; de transcription avec information sur la prosodie, les gestes, les regards, les phénomènes d'interaction (chevauchements, phatiques,...) ; de transcription en alphabet phonétique ; etc.

Cette hétérogénéité ne provoque pas simplement des choix de granularité d'analyse, mais également des enjeux théoriques de première importance tant il est vrai que toute transcription engage nécessairement une théorie sous-jacente. Ainsi, entre une transcription en API et une transposition de l'orale en écriture il existe dorénavant des transcriptions graphiques redonnant toute leur dimension aux productions orales tout en restant lisible par tous. Les linguistes peuvent fournir des transcriptions sans ponctuation, avec des élisions, des assimilations, des répétitions, des troncations, des chevauchements, des reformulations, des phénomènes prosodiques, des indications sur la prononciation, sur les bruits de gorge, les rires, les silences, les pauses, mais aussi les gestes les mimiques, les regards, etc. Ce travail sur la transcription n'est pas aisé mais il apporte des éléments permettant la mise en écrit d'une *parole*.

Les linguistes ont trouvé, pour leur domaine, une première solution au problème du parti pris qu'impose une transcription combiné à la volonté de rendre les corpus oraux disponibles à l'ensemble de la communauté scientifique avec différents choix et variations de transcription. Il s'agit des outils offrant des transcriptions multiples toutes alignées sur le même signal sonore. Ainsi, plusieurs outils informatiques actuels permettent d'associer une transcription à un enregistrement avec un ancrage temporel. La voix redevient alors intimement liée à sa représentation graphique. De plus ces outils juxtaposent différentes transcriptions ou différents niveaux de transcription. Une même portion sonore peut ainsi être écoutée tout en lisant comme sur une partition avec différentes portées, une transcription phonétique et en dessous une transcription semi-orthographique, puis une transcription transposée sur une troisième ligne.

La variation d'interprétation est enfin possible. Cette possibilité (révolutionnaire pour le travail des linguistes) rend compte de la complexité du travail de transcription et des variations d'interprétation.

Cependant le respect des variations ne doit pas laisser penser qu'on puisse faire l'économie d'une transcription "figée". Parmi toutes les possibilités il faut, notamment si on veut diffuser ou publier une transcription, élaborer une forme stabilisée. Le choix de cette dernière est bien difficile. Il s'agit, là aussi, de respecter l'oral, mais surtout le locuteur. La forme de ce respect a donné lieu à différentes controverses comme la polémique développée autour de la publication des témoignages transcrits par Pierre Bourdieu dans son livre *la Misère du Monde*. Pierre Bourdieu a choisi d'adapter les transcriptions arguant qu'il s'agissait de l'image des locuteurs perçus à travers une catégorisation socioculturelle non exempte de stéréotypisation et de catégorisation péjorative. Ce choix lui a été reproché, cette intervention du porte-parole sur la parole des témoins étant présentée comme un exemple de la violence symbolique pourtant dénoncée par l'auteur.

Cet exemple résume la complexité de la question. La transcription est un travail lourd et complexe qui implique que les linguistes continuent ce qu'ils ont commencé bien tardivement : ne rien céder sur cette question théorique tout autant qu'éthique. C'est une condition essentielle pour que le travail scientifique que constitue l'acte de transcrire soit reconnu dans toute sa richesse.