



HAL
open science

Le droit de la parole

Olivier Baude

► **To cite this version:**

Olivier Baude. Le droit de la parole. Données orales : les enjeux de la transcription, 2005, Perpignan, France. halshs-01162543

HAL Id: halshs-01162543

<https://shs.hal.science/halshs-01162543>

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.

1.3 Le droit de la parole^{1 2}

Les aspects juridiques de la constitution et de l'exploitation des corpus oraux ont très longtemps été négligés même si une préoccupation éthique a toujours été au cœur des travaux des chercheurs de terrain - notamment chez les ethnolinguistes et les sociolinguistes. La situation change avec les nouveaux usages et l'engouement actuel pour la diffusion des corpus oraux. Ainsi, la numérisation des données ouvre des possibilités de stockage, de diffusion et d'utilisation seconde – ressources pour des outils didactiques et pour l'ingénierie – tout en suscitant de nombreuses questions juridiques : Comment recueillir le consentement des personnes enregistrées ? Faut-il anonymiser les données ? A qui appartiennent les enregistrements, les transcriptions, les annotations ? La transcription est-elle couverte par le droit d'auteur? Etc.

Les questions sont complexes et les réponses ne vont pas de soi. Trois grands domaines juridiques sont concernés : le droit d'auteur et la propriété intellectuelle, les données personnelles et le respect de la vie privée et les responsabilités des "exploitants" et diffuseurs. Il convient donc dans un premier temps de définir le statut juridique d'un corpus oral afin de procéder à la gestion contractuelle des droits des personnes repérés et enfin de définir les responsabilités de tous ceux qui interviendront lors de l'exploitation du corpus (de la collecte à la diffusion).

1.3.1 Quel est le statut juridique d'un corpus et de ses transcriptions?

Définir le statut juridique d'un corpus nécessite de décrire avec précision son contenu puis les conditions d'élaboration et d'exploitation de celui-ci. La forme d'un corpus oral est très variable et son contenu diffère selon les méthodologies, les théories et les usages de ses concepteurs. Dans la majorité des cas les corpus oraux sont constitués d'enregistrements analogiques ou numériques – et qui dans le cas de supports analogiques ont une durée de vie très courte avec une perte de qualité, lors des migrations – de productions de locuteurs ; de données contextuelles sur les locuteurs et la situation d'enquête qui peuvent être en partie des données personnelles (nom propre, profession, adresse, lieu,...) ; de transcriptions "annotations-primaires" phonétiques, orthographiques, multilinéaires, etc. (sous la forme de fichiers indépendants ou offrant une synchronisation avec le signal) ; d'annotations secondaires (informations sur les conditions de production des énoncés, précisions sur les phénomènes sonores tels que les rires et les bruits) ; d'annotations enrichies (étiquetage morphologique, syntaxique, annotations prosodiques pragmatiques).

¹ Les problèmes juridiques liés à la diffusion des corpus oraux ont été l'occasion d'une démarche originale adoptée par une communauté scientifique ouverte à un travail pluridisciplinaire mené dans le cadre du programme *Corpus de la Parole* de la Délégation générale à la langue française et aux langues de France – Ministère de la Culture et de la Communication. Cette démarche a comporté plusieurs étapes. Une lecture croisée des textes juridiques par les linguistes et les juristes a permis de repérer les problèmes. Les chercheurs ont ensuite accepté d'explicitier leurs pratiques au regard de la législation. Cette étape fondée sur la réflexivité a permis d'élaborer des propositions pour de bonnes pratiques partagées par la communauté scientifique et de repérer des aspects juridiques qui posent des difficultés dans l'état actuel du droit.

² Ce texte présente principalement les résultats du groupe de travail publiés sous la forme d'un guide : *Corpus oraux, guide des bonnes pratiques 2006* rédigé par O. Baude coordinateur (Université d'Orléans/DGLFLF), C. Blanche-Benveniste (EPHE/Université de Provence), M-F. Calas (DMF), P. Cappeau (Université de Poitiers), P. Cordereix (BnF), L. Goury (CNRS), M. Jacobson (CNRS), I. de Lamberterie (CNRS) C. Marchello-Nizia (ILF/ENS-LSH-Lyon) et Lorenza Mondada, (Université Lyon 2). La paternité de la grande majorité du contenu de ce texte en revient donc à l'ensemble des auteurs à qui le rédacteur exprime toute sa gratitude.

Le statut juridique d'un corpus se définit également par les conditions d'élaboration et d'exploitation de celui-ci. En effet, l'exploitation par les chercheurs des enregistrements oraux et multimodaux modifient considérablement l'objet : les opérations d'annotations et notamment la transcription intègrent de nombreux aspects théoriques et interprétatifs mais aussi juridiques et éthiques. De fait, "*dans le passage de l'oral à l'écrit graphico-visuel, de nombreuses opérations de catégorisation sont effectuées, soit quant aux formes linguistiques, segmentées visuellement en unités (Blanche-Benveniste & Jeanjean, 1987 ; Mondada, 2000), soit quant à l'identité des locuteurs eux-mêmes (Mondada, 2003). Du point de vue de la protection de l'image et de l'identité des personnes enquêtées et enregistrées, il convient d'apprécier ces effets pour éviter la surinterprétation, la stéréotypisation (Jefferson 1996) et la stigmatisation des locuteurs et de leurs façons de parler*"³.

Les opérations d'annotations ne sont donc jamais neutres en termes d'effet de la recherche sur des données produites par les locuteurs. La description rigoureuse des traitements permet d'évaluer les effets de ceux-ci sur des productions dont on doit respecter les droits de leurs auteurs. De plus, seule une description rigoureuse fournit la possibilité d'attribuer la paternité et la responsabilité de niveaux de traitement à un ou plusieurs auteurs.

Cette étape de description du corpus est donc fondamentale pour différencier les composantes susceptibles d'être protégées par le droit d'auteur ainsi que les composantes contenant éventuellement des données personnelles. C'est à l'issue de cette étape qu'il conviendra de procéder à la gestion contractuelle des droits repérés et aux traitements nécessaires à ceux-ci comme le recueil de consentement ou l'anonymisation.

1.3.2 Droit d'auteur et propriété intellectuelle

Les transcriptions d'un corpus, tout comme les autres éléments de celui-ci, sont elles protégées par le droit d'auteur ? D'un point de vue juridique il convient de classer les composantes d'un corpus comme relevant du domaine public – dans ce cas son exploitation est totalement libre –, ou comme étant protégées par le droit d'auteur .

Pour qu'un corpus ou qu'une composante d'un corpus soit considéré comme protégé par le droit d'auteur il faut qu'il remplisse trois conditions : qu'il corresponde à l'exigence d'une *activité créatrice*, qu'il ait une *forme définie* et que cette forme soit *originale*⁴. Si tel est le cas l'auteur, qui est en principe la (ou les) personne(s) physique(s) sous le nom de laquelle (ou desquelles) l'œuvre est divulguée⁵, accède à des droits patrimoniaux et aux prérogatives du droit moral. "*Les droits patrimoniaux se résument en un droit exclusif au profit de l'auteur (ou des titulaires) ou des ayants droits (bénéficiaires d'une cession, héritiers...) d'autoriser ou interdire la reproduction ou la communication au public de l'œuvre protégée. Quant aux prérogatives du droit moral toujours attachées à la personne physique créatrice de l'œuvre protégée, elles sont au nombre de quatre : le droit de divulgation, le droit de repentir et de retrait, le droit à la paternité et le droit au respect de l'œuvre*"⁶.

Ces précisions juridiques ne permettent pas une réponse tranchée mais précisent les questions juridiques : le contenu d'une langue, son expression phonique de même que son expression graphique font-ils ou non partie du domaine public ? Le travail scientifique de collecte comme celui de transcription et d'annotation doivent-ils être considérés comme des activités

³ *Corpus oraux, Guide des bonnes pratiques, 2006, p73.*

⁴ *Corpus oraux, Guide des bonnes pratiques, 2006, p39.*

⁵ Le travail scientifique suppose l'intervention de nombreux acteurs dont bon nombre sont susceptibles de revendiquer la qualité d'auteur sur les résultats de la recherche. Certains corpus oraux, comme les autres produits de la recherche, peuvent rester l'œuvre d'un auteur unique, alors que d'autres peuvent être l'œuvre de plusieurs auteurs.

⁶ *Corpus oraux, Guide des bonnes pratiques, 2006, p 40.*

créatrices d'œuvres à la forme définie et originale ? Cette dernière question est d'autant plus complexe que le statut d'auteur-fonctionnaire n'est pas clairement défini actuellement.

Seule une réflexion sur les pratiques des chercheurs, menée par l'ensemble de la communauté scientifique permettra de construire des réponses. Il est alors d'autant plus important de signaler une solution intermédiaire où les corpus protégés par le droit d'auteur peuvent être mis en libre accès dans le cadre d'une licence accordée par les titulaires de droits autorisant l'utilisation et l'exploitation des résultats⁷.

1.3.3 Le respect de la vie privée

Le second grand volet des aspects juridiques concerne la gestion d'éventuelles données personnelles relevant du respect de la vie privée. La présence de données personnelles dans un corpus implique obligatoirement de se conformer à la *loi informatique et liberté (licéité et loyauté, information préalable, obtention du consentement)*, ou de procéder à l'anonymisation *irréversible* de celles-ci. Avant de présenter les modalités du recueil de consentement et les techniques d'anonymisation il est nécessaire de définir plus précisément ce que sont "*les données personnelles*".

La loi informatique et liberté ne restreint pas le respect de la vie privée aux précautions à prendre pour gérer des données "nominatives". La question légale est celle, plus vaste, de l'impossibilité d'identifier des personnes. En effet, l'objectif est de protéger la vie privée des personnes enregistrées en dépersonnalisant les données, ainsi, les textes législatifs abandonnent les termes de « données nominatives » au profit de « données personnelles » :

...pour déterminer si une personne est identifiable, il convient de considérer l'ensemble des moyens susceptibles d'être raisonnablement mis en œuvre soit par le responsable du traitement, soit par une autre personne, pour identifier ladite personne (directive 95/46/CE).

Une attention toute particulière doit donc être apportée aux données personnelles contenues dans un corpus qui permettraient d'identifier directement un témoin (formes nominatives, données personnelles, profession, statut, titres, activités sociales, parenté, réseaux, référence à des lieux, référence à des caractéristiques de la personne, caractéristiques physiques, etc.) mais aussi à tout ce qui peut permettre indirectement une identification (notamment les possibilités de recoupement d'informations).

1.3.4 Techniques d'anonymisation des données primaires et des transcriptions

L'anonymisation n'est pas une opération obligatoire. Elle est toutefois indispensable dans les cas où le corpus contient des données personnelles sans que le consentement des personnes concernées ait été recueilli. Ainsi, le nom propre n'étant pas le seul élément devant donner lieu à un traitement spécifique, il serait plus juste de parler de "*dépersonnalisation*" des données.

La première phase consiste donc à repérer les données permettant l'identification directe et indirecte mais aussi celles qui pourraient porter préjudice.

Dans une seconde phase les données primaires sont traitées au moyen d'opérations techniques (bippage, effacement, déformation, etc.).

Enfin la troisième phase concerne la "dépersonnalisation" des transcriptions. Le principe général est celui de la substitution. L'information est remplacée par un segment vide, un hyperonyme ou une abréviation (NN, NPersonne), par des caractères spéciaux (****), par un pseudonyme (Pierre à la place de Paul).

Une éventuelle quatrième phase concerne l'anonymisation des métadonnées par cryptage ou structuration de bases de données séparées.

⁷ Corpus oraux, *Guide des bonnes pratiques*, 2006, p 39.

L'objectif de l'impossibilité d'identification est totalement irréaliste dans le cadre des corpus contenant des enregistrements (il faudrait systématiquement déformer la voix des locuteurs), il est donc fondamental de prendre en compte la notion de "*moyens susceptibles d'être raisonnablement mis en œuvre*" (Cf. supra) ; et d'avoir une démarche éthique et scientifique suffisamment rigoureuse pour permettre une exploitation sereine de corpus oraux. Enfin la communauté scientifique se doit d'inventer des procédés prenant en compte ses usages d'exploitation et de diffusion. Ainsi la loi québécoise « *concernant le cadre juridique des technologies de l'information* » propose de protéger l'anonymat non pas en modifiant les données, mais en limitant les possibilités de recherche, voire en les adaptant à la personne qui consulte la base selon des critères bien précis (sa profession, une autorisation, sa présence dans le fichier, etc.). Cette dernière perspective offre pour la constitution et l'exploitation de corpus oraux la possibilité de faire coïncider les obligations légales avec les nécessités du travail de recherche. Toute donnée étant potentiellement sensible, une anonymisation systématique s'avère de plus en plus complexe ; elle peut même mettre en danger l'intérêt de certaines recherches. En effet, des détails concernant les personnes comme par exemple le nom, ou le lieu d'habitation peuvent constituer un élément important du corpus, et des résultats qui peuvent en être obtenus. C'est pourquoi la possibilité de ménager des niveaux d'accès selon des critères stricts (ex : chercheur ou non, présence d'autorisation, but de la consultation, etc.) semble une alternative efficace. Il existe d'autres procédés à inventer⁸.

1.3.5 Consentement

Le recueil du consentement des personnes enregistrées reste la meilleure solution éthique et juridique, pour autant que celle-ci ne soit pas réduite à une vague demande d'autorisation. Or sans informations préalables précises la demande d'autorisation n'a pas d'objet ni de sens. Pour que cette autorisation soit pertinente il faut établir un consentement "éclairé" qui démontre que le signataire est informé des finalités de la recherche et des conséquences à son égard d'une participation au projet.

Dans le cadre du recueil de données et d'exploitation de corpus oraux, le consentement devrait tenir compte de l'adéquation au destinataire - les informations fournies, pour être comprises doivent être adaptées aux compétences de compréhension du destinataire -, et de l'explicitation des finalités de l'enquête - qui toutefois ne doivent pas renforcer le paradoxe de l'observateur en pointant l'objet de l'observation -. Le consentement devra préciser l'objet de la demande : les actions effectuées par les chercheurs dans le cadre du projet, les formats et les conditions de l'enregistrement, les conditions de diffusion des données et des résultats, les contextes de diffusion des données et des résultats⁹.

De plus, les explications sur le projet scientifique, doivent être complétées par des informations précises notamment sur la façon dont les données seront anonymisées (le cas échéant) et sur la forme que prendront les énoncés transcrits. En effet, la transcription est empreinte de lourds enjeux éthiques. "*La représentation écrite de la langue surprend souvent les locuteurs, et peut même leur déplaire considérablement. Il arrive qu'ils refusent l'image de leur langue transmise par la transcription, qu'ils désavouent le chercheur et qu'ils refusent son travail. Ainsi le recours à l'API et à l'orthographe adaptée peut produire des effets de stigmatisation et d'asymétrie à l'encontre de locuteurs* »¹⁰.

⁸ Corpus oraux, Guide des bonnes pratiques, 2006, p 109

⁹ Il est à noter que les formes de l'autorisation ne sont pas imposées par le législateur et qu'une demande orale enregistrée peut être valide et même parfois indispensable.

¹⁰ Corpus oraux, Guide des bonnes pratiques, 2006, p 76

Il n'existe pas de réponses juridiques toutes faites aux questions posées par le droit de la parole. Les solutions ne consistent pas à appliquer simplement une législation existante, mais passent par l'élaboration de bonnes pratiques dans le respect d'une démarche éthique. Cela implique que le chercheur *sache ce qu'il fait* et qu'il soit donc capable d'explicitement sa démarche dans un véritable travail réflexif. Cette explicitation est indispensable à la définition même d'un corpus constitué d'enregistrements, de transcriptions et d'annotations et à la connaissance de ses conditions d'élaboration et d'exploitation mais aussi à l'évaluation des enjeux empirico-théoriques. Ainsi, l'élaboration de bonnes pratiques, loin d'être restreintes à l'application de contraintes juridiques est aussi l'opportunité pour une communauté scientifique de s'acquitter de la dette que le chercheur contracte envers ceux qui produisent les données, en prenant en charge, dans toutes leurs dimensions – y compris juridiques et éthiques - les objets scientifiques qu'elle construit.

Bibliographie

- Baude, O., coord. (2006). *Corpus oraux, Guide des bonnes pratiques 2006*, CNRS éditions et PUO, Paris et Orléans.
- Baude, O. (2004). Les corpus oraux entre science et patrimoine. L'expérience de l'observatoire des pratiques linguistiques. In *Actes du Colloque international du GRESEC « La publicisation de la science »* (Grenoble), 7-11.
- Beaud, S. & Weber, F. (1997). *Guide de l'enquête de terrain : produire et analyser des données ethnographiques*. Paris, La Découverte.
- Becker, H.S. & Geer, B. (1960). Participant observation : the analysis of qualitative field data. In Adams & Preiss (eds.), 267-289.
- Bergounioux, G. (ed.) (1992). Enquêtes, Corpus et Témoins. *Langue Française* 93.
- Biber, D. (1985). *Variations across spoken and written language*. Cambridge, CUP.
- Biber, D. (1999) *Longman Grammar of Spoken and Written English*. Londres, Longman.
- Bilger, M. (ed.) (2000). Linguistique sur corpus, études et réflexions. *Cahiers de l'université de Perpignan*, Perpignan, Presses Universitaires de Perpignan.
- Bilger, M. (ed.) (2000). *Corpus, Méthodologie et applications linguistiques*. Paris, Champion.
- Blanche-Benveniste, C. & Jeanjean, C. (1987). *Le français parlé : transcription et édition*. Paris, Didier-Erudition.
- Blanche-Benveniste, C. (1997). Transcription et technologie. *Recherches sur le Français Parlé* 14, 87-100.
- Blanche-Benveniste, C., Bilger, M., Rouget, C. & van den Eynde, K. (1999). *Le Français Parlé : Études grammaticales*. Paris, CNRS-Éditions.
- Blanche-Benveniste, C., Rouget, C. & Sabio, F. (2001). *Choix de textes de français parlé : trente-six extraits*. Paris, Champion.
- Bourdieu, P. (1982). *Ce que parler veut dire. L'économie des échanges linguistiques*. Paris, Fayard.
- Bourdieu, P. (1993). *La misère du monde*. Paris, Le Seuil.
- Calas, M-F. & Fontaine, J-M (1996). *La conservation des documents sonores*. Paris, CNRS Editions.
- Callu, A. & Lemoine, H. (2004). *Patrimoine sonore et audiovisuel français : entre archive et témoignage : guide de recherche en sciences sociales*. Paris, Belin, 7 vol., 1 CD-Rom, 1 DVD-Rom.
- Cameron, D., Frazer, E., Harvey, P., Rampton, M. & Richardson, K. (1991). *Researching Language : Issues of Power and Method*. London, Routledge.
- Condamines, A. (ed.) (2006). *Sémantique et corpus*. Paris, Hermes.
- Cresti, E. & Moneglia, M. (eds.) (2005). *C-ORAL-ROM, Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam, Benjamins.
- Cribier, F. & Feller, E. (2003). *Projet de conservation des données qualitatives des sciences sociales recueillies en France auprès de la « société civile »*. Rapport au Ministre délégué à la Recherche et aux nouvelles technologies, dactylogr., 2 vol.
<http://www.iresco.fr/labos/lasmas/rapport/Rapdonneesqualita.pdf>
- Encreve, P., & Fornel (de) M. (1983). Le sens en pratique. *ARSS* 46, L'usage de la parole.
- Gadet, F. (2003). *La variation sociale en français*. Paris, Ophrys.
- Guilhaumou, J., Mesini, B. & Pelen, J.N. (1997) Récits de vie. Dynamiques et autonomies des récits de vie dans le champ de l'exclusion. *Cahiers de littérature orale*, 41, 91-126.

- Gumperz, J.J., & Hymes, D. (eds.) (1972). *Directions in Sociolinguistics : The Ethnography of Communication*. New-York, Hold, Rinehart & Winston.
- Habert, B., Nazarenko, A. & Salem, A. (1997). *Les linguistiques de corpus*. Paris, A. Colin.
- Jacobson, M. (2004). Corpus oraux en linguistique de terrain. *Traitement Automatique des Langues*, 45/2, 63-88.
- Jacobson, M. (2004). Les archives sonores au LACITO. *Bulletin de liaison de l'AFAS* 26 (<http://afas.mmsh.univ-aix.fr/bulletin/Bulletin AFAS 26.pdf>).
- Joutard, P. (1979). Historiens, à vos micros. Le document oral, une nouvelle source pour l'histoire. *L'Histoire*, 12, 106-113.
- Kennedy, G. (1998). *An introduction to Corpus Linguistics*. Londres, Longman.
- Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphie, University of Pennsylvania Press.
- LAMY, Droit de l'informatique et des réseaux (S. Marcellin, L. Costes & al. eds., Paris, 2004).
- Leech, G. (1992). The state of the art in corpus linguistics. In Aijmer & Altenberg (eds.), 8-29
- Mondada, L. (1998). Technologies et interactions sur le terrain du linguiste. Le travail du chercheur sur le terrain. Questionner les pratiques, les méthodes, les techniques de l'enquête. Actes du Colloque de Lausanne (13-14.12.1998), *Cahiers de l'ILSL* 10, 39-68.
- Mondada, L. (2006). Video recording as the reflexive preservation-configuration of phenomenal features for analysis. In Knoblauch, H., Raab, J., Soeffner, H.G., Schnettler, B. (eds.)
- Mondada, L. (à paraître) « La demande d'autorisation comme moment structurant pour l'enregistrement et l'analyse des pratiques bilingues », *Tranel*, Université de Neuchâtel.
- Quééré, L. & al. ed. (1984) *Arguments ethnométhodologiques*, Paris, Centre d'Étude des Mouvements Sociaux, EHESS.
- Recherches sur le Français Parlé* 5 (1984). Pourquoi le français parlé est-il si peu étudié ?.
- Revue Française de Linguistique Appliquée* (1996) I-2, (1999) IV-1.
- Sacks, H. (1984). Notes on methodology. In J.M. Atkinson & J. Heritage (eds.), 21-27.
- Sankoff, D., Sankoff, G., Laberge, S. & Topham, M. (1976). Méthodes d'échantillonnage et utilisation de l'ordinateur dans l'étude de la variation grammaticale. *Cahiers de Linguistique* 6, 85-125.
- Shaffir, W.B. & Stebbins, R.A. (eds.) (1991). *Experiencing Fieldwork : An inside View of Qualitative Research*. Londres, Sage.
- Silverman, D. (ed.) (1997). *Qualitative Research. Theory Method and Practice*. Londres, Sage.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Londres, OUP.
- Sinclair, J. (1996). *Preliminary recommendations on corpus Typology*. Technical Report, Eagles.
- Sinclair, J. & Coulthard, R.M. (1975). *Towards an Analysis of Discourse*. Londres, OUP.
- Speech Communication* (2001) Speech Annotation and Corpus Tools. Vol. 33, 1-2, S. Bird & J. Harrington (eds.).
- Welland, T. & Pugsley, L. (eds.) (2002). *Ethical Dilemmas in Qualitative Research*. Aldershot, Ashgate.