

**Peter Stockinger**

*Équipe Sémiotique cognitive et nouveaux médias (Escom), FMSH, Paris*

## **DES ARCHIVES AUDIOVISUELLES MONOLINGUES SUR UN SITE MULTILINGUE**

Notre propos ici est de discuter, d'une manière succincte mais pratique, certains enjeux d'Internet en tant que « marché multilingue et multiculturel de connaissances » et d'exposer ce qu'il est possible de faire pour ouvrir des œuvres intellectuelles intrinsèquement monolingues à un tel marché.

Nous avons commencé à rencontrer la problématique de la traduction dans un contexte très concret qui est celui de la diffusion de documents audiovisuels sur Internet. En 2001, nous avons mis en place, avec l'aide de la Fondation Maison des Sciences de l'Homme à Paris (FMSH), un projet de recherche visant la constitution progressive et l'« exploitation » d'un patrimoine scientifique en sciences humaines et sociales à l'aide de l'audiovisuel numérique. En réalisant des entretiens – souvent de plusieurs heures – avec des chercheurs travaillant dans différentes disciplines des sciences de l'homme et de la société (SHS), ainsi que des tournages de séminaires de recherche et de manifestations collectives telles que colloques, tables rondes ou autres journées d'étude, ce projet de recherche a donné naissance aux Archives audiovisuelles de la recherche (AAR)<sup>1</sup> qui proposent actuellement, sur leur site-portail officiel environ 3 700 heures de vidéos consultables en ligne. Un « événement » scientifique (tel qu'un séminaire, un entretien, etc.) est publié sur le portail des AAR sous forme d'un site dynamique (généralisé automatiquement à partir des données collectées au préalable par les réalisateurs auprès des auteurs et à partir des métadonnées issues du travail de segmentation des vidéos, de la classification des segments vidéo ou encore de leur indexation et annotation). La consultation elle-même se fait sous forme d'un « feuilletage » de différents « chapitres » (segments audiovisuels) qui composent le « vidéo-livre interactif » de l'événement filmé dans le cadre des AAR.

Comme la plupart des autres programmes ou projets visant la constitution d'un patrimoine culturel, les AAR se trouvent confrontées, aujourd'hui, aux trois grands problèmes suivants :

– La qualité et la richesse de leur contenu sont en quelque sorte écrasées par la quantité, le volume d’heures offertes à la communauté intéressée (composée de chercheurs, enseignants ou étudiants, mais aussi de journalistes et professionnels travaillant dans des ONG, des associations ou des agences gouvernementales).

– Le contenu scientifique est presque entièrement monolingue (soit uniquement en français – ce qui est le cas majoritairement –, soit uniquement en une des douze autres langues actuellement représentées dans les AAR).

– Le contenu (le ou les propos) véhiculé par les vidéos des AAR possède son identité propre, liée à l’habitus aussi bien *individuel* de l’auteur (chercheur, expert, personne interviewée, enseignant-chercheur responsable d’un séminaire, etc.) que *collectif* des usages et traditions disciplinaires et institutionnels (comme on le sait bien, deux linguistes, pour prendre cet exemple, travaillant sur la grammaire d’une langue donnée peuvent en faire un objet de connaissance fort différent).

Le premier problème concerne avant tout la question de la « bonne » description, classification et indexation des corpus audiovisuels ou autres. Le deuxième problème est celui qui est effectivement associé à la traduction au sens habituel du terme, c’est-à-dire au fait de remplacer un document réalisé en une langue 1 par un autre document réalisé en une langue 2, tout en prenant comme « norme de qualité de traduction » – si rien d’autre n’est stipulé – une équivalence aussi proche que possible de sens entre les deux documents. Le troisième problème concerne plus particulièrement l’explicitation du contenu d’un document en tenant compte à la fois de son identité spécifique et du « profil » de son destinataire (projeté ou réel). Comme on le voit, tout en étant distinct des deux premiers, ce troisième problème partage néanmoins certains aspects : avec le premier, par exemple, le souci de rendre accessible à un destinataire un certain type de contenu ; avec le second de « traduire » (*lato sensu*) un contenu donné pour un destinataire possédant son propre profil (non pas ou pas seulement linguistique mais, plus généralement, cognitif, voire idéologique).

Restreignons-nous au deuxième et troisième problème. Ils constituent, pour nous, un défi crucial. En effet, en consultant les statistiques de fréquentation des AAR, produites par une société externe à la FMSH, on se rend évidemment compte que la langue dans laquelle les vidéos des AAR sont réalisées (dans notre cas, majoritairement le français) constitue une barrière dans la diffusion de celles-ci. En moyenne, des visiteurs de 40 à 55 pays du monde, explorent tous les jours les AAR mais, bien entendu, ce ne sont que les internautes d’environ 10 à 15 pays qui en constituent des « grands consommateurs ». Tous ces pays parlent soit français, soit une autre langue romane (seules exceptions : États-Unis, Allemagne et Grande-Bretagne). Les mêmes statistiques (complétées par des enquêtes ponctuelles) montrent également des préférences pour certaines thématiques, certains événements filmés, voire certains passages (« chapitres ») vidéo dans un « vidéo-livre » donné. Enfin, malgré leur intérêt intrinsèque, certains documents audiovisuels sont difficilement exploitables en dehors d’un cercle de spécialistes à cause de leur profil cognitif très particulier exigeant des connaissances préalables importantes pour leur assimilation.

Toutes ces raisons nous ont obligés, d'une manière tout à fait pratique, à réfléchir aussi bien à la question de la *traduction (linguistique)* que sur celle de l'*adaptation (cognitive)* des ressources audiovisuelles en ligne relevant du domaine du patrimoine scientifique ou culturel. C'est le hasard qui a voulu que la question de l'ouverture multilingue, voire multiculturelle des ressources numériques soit devenue, plus ou moins en même temps, un des axes stratégiques de la recherche appliquée en Europe dans le cadre de ses grands programmes de R&D.

## La republication d'un corpus audiovisuel numérique

Le problème de la « traduction » (les guillemets sont intentionnels) de ressources audiovisuelles constitue, pour nos activités dans le cadre des AAR, un cas spécifique du cas plus général de la *republication de corpus audiovisuels* (et, plus généralement, multimédia) *numériques* dans un contexte multilingue ou multiculturel (le processus de republication lui-même pouvant être compris, ainsi que nous le verrons, comme une activité centralisée ou distribuée).

La « republication » (en anglais *re-authoring* ou *re-purposing*) recouvre un ensemble d'activités visant à réutiliser un corpus de documents numériques (textuels, audiovisuels, visuels, etc.) pour des usages spécifiques auxquels les documents sources, dans leur forme initiale, ne peuvent que partiellement répondre. Pour les AAR, les usages spécifiques les plus récurrents se rencontrent dans l'enseignement formel et informel, dans la recherche elle-même, dans le journalisme scientifique ou encore dans la documentation professionnelle.

La republication est donc un processus, parfois très complexe, d'adaptation d'un document ou d'un corpus de documents sources à des usages spécifiques. Ce processus d'adaptation peut concerner tous les plans constitutifs d'un document (Stockinger, 1981, 1999 et 2003), c'est-à-dire aussi bien le plan du contenu que celui de l'expression. Il s'accomplit à travers un ensemble d'activités intellectuelles et de gestes techniques et en référence à des modèles ou genres de publications qui intègrent les contraintes typiques des contextes et des communautés d'usage auxquels un document ou un corpus de documents republiés est destiné.

La notion du *genre de publication* (à distinguer d'autres notions plus spécialisées telles que celle du *genre de discours* ou du *format de publication* !) est en effet importante car c'est elle qui oriente le processus de la republication dans son ensemble. Dans le cadre des AAR, les genres de republications actuellement en cours sont le *vidéo-livre hypermédia*, l'*encyclopédie hypermédia*, le *vidéo-lexique*, le *trailer* (la *bande-annonce*) ou encore la *version bilingue*. Ces modèles ou genres de publications nous servent à donner une certaine forme soit à un document source soit à un corpus de documents sources en vue de les préparer pour des usages plus ciblés.

Les activités composant le processus de la republication recouvrent notamment la constitution d'un corpus de travail pertinent, l'adaptation d'une « ontologie de description » (vocabulaire général de

description utilisé pour indexer un fonds audiovisuel), la segmentation des données audiovisuelles (sélection et « extraction logique » de segments considérés pertinents), la description textuelle (description des caractéristiques d'un segment audiovisuel), la description thématique et l'indexation de chaque segment, la description rhétorique et discursive proprement dite (identification et présentation des actes et genres de discours utilisés pour développer une thématique), la description dite « péritextuelle » d'un segment qui consiste en son « enrichissement » sous forme de liens hypertextuels ou de production de mini-analyses, d'opinions, d'interprétations, de commentaires, etc.

## La traduction comme activité de republication

La traduction constitue une activité spécifique de republication d'un document audiovisuel ou d'un corpus de documents audiovisuels. Mais elle peut prendre des formes très variées. Une première distinction peut être faite entre d'une part la traduction au sens d'une aide à la compréhension (linguistique) du contenu de tel ou tel document audiovisuel, et d'autre part la traduction proprement dite en tant que version dans une langue cible d'un document audio-visuel réalisé dans une langue source.

Comme exemple du premier cas, citons un entretien avec César Itier, spécialiste de la famille linguistique quechua à l'Inalco (Institut national des langues et civilisations orientales) à Paris, réalisé par Élisabeth de Pablo pour les AAR en janvier 2007. Cet entretien en français, publié sous forme d'un vidéo-livre interactif également en français, est découpé et monté sous forme de segments thématiquement homogènes (caractérisés par une isotopie thématique centrale). Chacun de ces segments est, cependant, accompagné d'un petit texte proposant une version « rapprochée » du contenu pour des lecteurs/auditeurs anglophones. Le but est ici de permettre aux visiteurs maîtrisant trop peu le français d'avoir une « idée » relativement exacte de cet entretien. Le point important à souligner ici est le suivant : il ne s'agit pas d'une traduction au sens habituel ou « standard » du terme (cf. la définition ci-dessus) mais plutôt d'une *aide linguistique* à la compréhension d'un contenu exprimé dans une langue pour laquelle le lecteur/auditeur possède des compétences trop réduites. Une telle aide doit être suffisamment bien articulée pour que le lecteur/auditeur ait au moins une vision globale du contenu en question. Il est vrai qu'il s'agit ici d'un critère très intuitif de qualité qui exige une discussion impossible à mener ici.

Toujours est-il que nous distinguons, dans le cadre de nos activités de republication de corpus audiovisuels, entre différents types de traduction au sens d'aide à la compréhension linguistique de tel ou tel segment d'un document audiovisuel :

- *Traduction semi-automatique* des thèmes appartenant à l'ontologie de la description, ainsi que des mots-clés partiellement réalisés sur la base de vocabulaires normalisés (c'est le cas des vocabulaires des langues, des pays, des métiers, des disciplines etc.).
- *Traduction sous forme de sous-titrage* de certains segments occupant une place centrale dans le développement d'un propos (segments appelés « highlight » ou « point culminant »).

- Traduction « rapprochée » sous forme d'une synthèse du contenu d'un segment choisi.
- Version plus ou moins libre, adaptée au contexte d'usage visé (il s'agit ici surtout des usages pour l'apprentissage ou l'enseignement informel de certains segments réalisés dans le cadre de notre programme AAR).

Considérons maintenant le deuxième style de traduction cité ci-dessus, à savoir celui de la production d'une seconde version dans une langue cible d'une publication audio-visuelle réalisée dans une langue source<sup>2</sup>. Cette activité s'appuie sur un genre, un modèle de publication qui peut être aussi bien une « copie » de la publication originale adaptée à un destinataire donné qu'un genre original (« hybride »). Ainsi une version en langue cible d'un événement filmé et publié en langue source peut prendre la forme d'un dossier thématique, d'un dossier pédagogique, d'un *trailer* (bande-annonce), d'un vidéolexique, etc. Elle peut aussi prendre la forme d'un genre assez original combinant, par exemple le *trailer* avec le vidéolexique afin de permettre à un public non francophone de connaître en quelques minutes les « grands titres » d'un corpus audiovisuel réalisé en langue française pour s'informer ensuite d'une manière bien plus ciblée et approfondie sur telle ou telle notion, tel ou tel terme spécialisé, etc.

« Version en langue cible » veut dire ici toute forme de traduction/adaptation, plus ou moins fidèle ou libre, d'une sélection de segments audiovisuels pour un public ne possédant pas les compétences linguistiques nécessaires pour comprendre le contenu développé dans les segments en question (voir à ce sujet les différentes formes de traduction ouverte ou « couverte » discutées par J. House, 1973). Cela n'inclut pas seulement la traduction d'une langue source en une langue cible (lorsque les deux langues en question sont deux langues naturelles différentes comme, par exemple, le français et l'allemand), mais aussi la traduction dans un *registre différent* à l'intérieur d'une même langue. Le cas le plus typique est le remplacement d'une terminologie scientifique par des expressions de la langue courante ou encore toute forme d'explicitation « pédagogique » tentant de « faire passer » une connaissance à un destinataire possédant un « profil » culturel et cognitif spécifique<sup>3</sup>.

D'un point de vue très pratique, on voit bien que traduction « linguistique » et adaptation cognitive font partie d'une même problématique plus générale qui est celle de la *valeur ajoutée* (de la pertinence, de l'intérêt, de l'utilité, ...) d'un document ou d'un fonds de documents au sens d'une ressource (cognitive) potentielle pour un acteur et un contexte d'usage donné.

## **Perspectives et enjeux**

Actuellement, deux projets de « R&D » (recherche et développement) – Saphir<sup>4</sup> et Logos<sup>5</sup> – nous permettent d'expérimenter des approches, des méthodologies et des solutions techniques visant à mettre en place des scénarios de republication et notamment de traduction *lato sensu* de corpus audiovisuels. Comme nous l'avons dit plus haut, l'un des défis majeurs auxquels se heurtent les AAR, comme les autres projets de même type, consiste à trouver comment faire circuler des documents monolingues à profils

fortement marqués (généralement des documents scientifiques) dans un « marché » intrinsèquement multilingue et multiculturel tout en devant supporter la concurrence de documents réalisés en langue anglaise. Ce défi est conceptuel et technique, mais aussi social et économique.

Nous avons esquissé ci-dessus rapidement sous le terme de « republication » les principales solutions conceptuelles que nous privilégions dans notre travail de R&D : des solutions conceptuelles supportées et encadrées par des outils informatiques développés principalement dans le cadre des deux projets que nous venons de citer<sup>6</sup>.

Mais, comme on le sait, il est parfaitement impossible, d'un point de vue économique, de songer à la traduction – ne serait-ce que partielle – d'un fonds audiovisuel composé de plusieurs milliers d'heures, dans l'ensemble des langues dites officielles de la Communauté Européenne (sans parler de la multitude des autres langues composant le contexte de diffusion actuel des AAR sur Internet).

Outre le fait donc d'« inventer » des formes de traduction/adaptation plus ou moins légères, plus ou moins informelles sous l'étiquette d'« aide linguistique » à la compréhension, ce qui est en jeu ici, c'est le fait qu'un tel travail d'ouverture ne peut se concevoir que d'une manière *coopérative et distribuée*, c'est-à-dire non pas sous forme d'un travail de service fourni par une « institution » de traduction ou de quelque chose de ce genre, mais plutôt comme une activité qui fait partie inhérente de l'activité de publication et de communication soit des « producteurs de contenu » eux-mêmes (dans le cas des AAR, des chercheurs ou des enseignants) soit d'un réseau d'acteurs qui fonctionnerait un peu comme une coopérative d'entraide visant la diffusion maximale d'une œuvre numérique. Les « membres » d'un tel réseau pourraient être, par exemple, certaines catégories de visiteurs/lecteurs d'un fonds tel que celui des AAR, des collègues, des « producteurs de contenu », des jeunes étudiants-chercheurs ou des professionnels de la traduction souhaitant contribuer à la diffusion dans une langue cible visée d'un contenu produit dans une langue source donnée. Il va de soi que la mise en place et le fonctionnement d'un tel réseau seront confrontés à un ensemble de problèmes qui ne peuvent être résolus par un coup de baguette magique. Certains concernent plus particulièrement le statut intellectuel et aussi la propriété du contenu produit par une personne, d'autres concernent davantage les modes de coopération et de gestion d'un tel réseau.

Il nous semble néanmoins évident qu'un tel réseau coopératif et distribué d'acteurs serait la solution la plus appropriée pour imaginer une gestion efficace d'un multilinguisme se basant sur de grandes *langues relais* (au détriment d'un monolinguisme toujours limité et trop asséchant) et sur la *circulation* d'informations et de connaissances provenant de multiples langues sources dans un monde qui reste intrinsèquement multilingue.

## NOTES

1. Site officiel des AAR : <<http://semioweb.msh-paris.fr/aar/fr/>>.
2. Pour des exemples concrets, on peut consulter la rubrique « version bilingue » sur le site officiel des AAR (<<http://semioweb.msh-paris.fr/aar/fr/versions.asp>>), par exemple : Olga ANISSIMOVA, *A Russian and English Version of the Conference with Faleh A. Jabar (Institut irakien d'études stratégiques, Irak) on « Ideological Roots of Suicide Bombers - The Salafi Case in Iraq »*, colloque « Martyr(e) et suicide dans l'Islam contemporain », Paris, FMSH, mars 2006.
3. Un bon exemple de cette forme de traduction/adaptation intra-langue est la republication en français d'un entretien fort spécialisé réalisé pour les AAR par Anne Gotman avec l'anthropologue Jean Claude Galey (EHESS) sur l'anthropologie sociale du monde indien en avril 2004. Les deux auteurs de cette republication, Muriel Chemouny et Natalia Semenova, proposent, sous le titre « L'Inde, un univers de sens particulier », une adaptation de cet entretien visant à sensibiliser les jeunes étudiants aux problématiques interculturelles. Cette republication ainsi que plusieurs autres du même type peuvent être consultées sur le site Web des AAR ainsi que sur celui de l'Escom : <<http://www.semionet.com>>.
4. Saphir (« Système d'aide à la republication de corpus audiovisuels sous forme de genres spécifiques de publications hypermédia ») est un projet français coordonné par l'Institut national de l'audiovisuel (INA) et financé dans le cadre du programme Riam. Il vise le développement d'outils de republication de corpus audiovisuels sous forme de styles particuliers de publication. Pour plus d'information, voir le site de l'Escom : <<http://www.semionet.com>>.
5. Logos (« Knowledge on Demand for Ubiquitous Learning ») est un projet européen coordonné par la télévision hongroise (Antenna Hungarica) et financé dans le cadre du 6<sup>e</sup> PCRD. Il vise la republication-traduction de corpus audiovisuels pour l'enseignement informel. Pour plus d'information, voir le site du projet : <<http://www.logosproject.com>>.
6. Outre des outils internes à l'Escom pour la publication et la gestion d'archives audiovisuelles, il s'agit notamment de graphes conceptuels (développés notamment par le LIRMM de l'Université de Montpellier II et l'Université d'Angers), d'un outil de segmentation, d'analyse et d'indexation (développé par l'Escom et l'INA), ainsi que de divers outils de publication (développés à la fois par l'Escom et l'INA). Dans le cadre du projet européen Logos, s'ajoutent à cet environnement des outils permettant plus particulièrement l'exploitation de corpus audiovisuels pour l'enseignement formel et surtout informel.

## RÉFÉRENCES BIBLIOGRAPHIQUES

- HALLIDAY, M. A. K., *Explorations in the Functions of Language*, Victoria, Edward Arnold, 1973.
- HOUSE, J., *A Model for Translation Quality Assessment*, Tübingen, Gunter, 1977.
- STOCKINGER, P., *Semiotik. Beitrag zu einer Theorie der Bedeutung*, Stuttgart, Hans Dieter Heinz, 1983.
- STOCKINGER, P., *Les Nouveaux Produits d'information. Conception et sémiotique du document*, Paris, Hermès Science Publications, 1999.
- STOCKINGER, P., *Le Document audiovisuel*, Paris, Hermès Science Publications, 2003.