

On commence à parler corpus. Un corpus c'est un ensemble cohérent de sources et en l'occurrence un corpus d'archives orales c'est un ensemble cohérent d'entretiens qui vont être réunis afin de constituer, de construire et aussi de conserver une mémoire collective. On passe là de la dimension individuelle de l'entretien à la dimension collective du corpus d'entretien et pour rendre les choses un peu plus vivantes, je vais vous parler d'un exemple d'un corpus numérique qui est en train d'être constitué, qui a été rassemblé en partie sur un site internet, qui s'appelle « Sciences : histoire orale », vous voyez ici une capture d'écran, qui est hébergé sur le site de l'école d'ingénieurs qui s'appelle ESPCI ParisTech, Physique Chimie Paris et qui a pour ambition d'associer plusieurs corpus. L'un de ces corpus, c'est le corpus de ma thèse sur l'histoire de la chimie du solide ; j'ai réalisé pendant un à deux ans, une trentaine d'entretiens ; donc cette trentaine d'entretiens forme un corpus. De même d'autres doctorants ont développé d'autres corpus, un camarade travaillait sur les nanosciences, par exemple, et il a formé un corpus d'une trentaine d'entretiens autour de la question des nanorobots et des nanosciences, un troisième travaillait sur la catalyse, lui aussi a constitué un corpus d'entretiens. Le but de ce site sciences expressions orales, c'est de former un corpus des corpus et de tous les rassembler au même endroit sur un site internet afin de pouvoir naviguer de l'un à l'autre. Qu'est ce qui rassemble tous ces corpus, c'est le thème de recherche qu'on appelle globalement la recherche sur les matériaux, matériaux au sens large. A partir de la deuxième moitié du XX<sup>e</sup> siècle, va se développer un vaste champ de recherches en lien avec la guerre froide qui s'appelle la recherche sur les matériaux dont la catalyse fait partie, dont les nanosciences font partie, dont la chimie fait partie et ça forme un champ extrêmement vaste et pour essayer de saisir ce champ là, on a voulu rassembler ces différents corpus d'archives orales sur un même site internet. Ca nécessite des compétences multiples, l'historien ne fait pas tout, d'abord il n'y a pas que des historiens, il y a des philosophes, des ingénieurs et informaticiens. Les ingénieurs et les informaticiens vont participer à la mise en ligne, à l'organisation des données, au traitement des données aussi. Se monte un projet de recherches inter-disciplinaires pour mettre en place cet objet sur le site de l' ESPCI ParisTech,

Ce site est auto-référent, c-à-d qu'il va aussi expliquer sa propre genèse, si vous circulez sur le site et c'est une des activités que je vous propose, vous allez comprendre comment a été constitué le site. Ces différents corpus de différents doctorants ont pu être rassemblés parce que tous ces doctorants étaient en thèse avec la même directrice de thèse, Bernadette B Vincent, qui est philosophe historienne de la chimie, qui lancé un certain nombre de doctorants sur des thèmes voisins qui s'incluent dans la recherche des matériaux et en même temps, elle a participé, au début des années 2000 aux Etats Unis, à un autre projet d'archives orales basé sur les Etats Unis. Ce projet d'archives orales a été financé au MA sur 5 ans et au bout de 5 ans, il n'y avait plus d'argent et donc le site a fermé Ce site internet a été rapatrié sur le site de CalTech, c'est l'institut technologique de Californie et une autre partie a été rapatriée en France. On a une partie du corpus, d'entretiens en anglais qui va être articulée avec différents entretiens en français, issus de nos thèses respectives. Donc un corpus plutôt nord américain d'un côté, un corpus plutôt européen de l'autre, anglais d'un côté, français de l'autre et notre but est de constituer un corpus homogène afin de conserver une mémoire collective sur la recherche matériaux.

Dans cette constitution, d'abord il faut constituer physiquement, c-à-d il faut avoir les entretiens, avoir les transcriptions des entretiens et ensuite il faut les mettre en ligne de manière où l'organisation du site internet permet de naviguer intelligemment au sein du corpus. C'est le 2<sup>ème</sup> temps et il y a un troisième temps c'est l'analyse du corpus. Comment est-ce qu'on analyse un corpus discursif d'archives orales et là on rentre dans les outils pour historiens. On peut faire de l'analyse quantitative, analyse linguistique, récurrence, occurrence des mêmes mots, le mot matériau, le mot chimie du solide, par exemple ondes de densité de charge, règle du tailleur et on voit la façon dont ces différents mots vont être liés les uns aux autres dans le corpus. Donc c'est une analyse linguistique, de type quantitative pour essayer d'établir une cartographie, essayer de représenter en 2 dimensions, 3 dimensions ces communautés savantes à travers leur discours de mémoire. Eventuellement, essayer de réaliser une cartographie dynamique, c-à-d essayer de faire une cartographie du champ en 1950, en 1970, en 1990 et aujourd'hui par exemple.

Comme activité, je vous propose de réfléchir à ce que vous pourriez faire comme analyse linguistique en fonction de vos centres d'intérêt pour essayer de voir quels mots clé vous proposeriez, quels types de relations vous proposeriez pour essayer de cartographier un objet et je vous donne un petit exemple : ici, sur le site du Caltech qui est accessible au public, vous pouvez vous connecter, aller sur ce site, on voit là une cartographie extrêmement basique qui associe, le projet américain il était centré sur les matériaux australiens, matériaux research et on voit qu'en 1960, matériaux research ça existe dans tout un tas de disciplines différentes : la physique, la chimie, la métallurgie, la biologie, la minéralogie, etc ... et on voit qu'il n'y a pas cohérence de ce champ de recherche.

Si on fait une cartographie en 1980, il doit y en avoir une sur le site du Caltech, il va y avoir une reconfiguration de ces différentes catégories au sein d'une nouvelle cartographie et l'on va voir l'évolution de l'organisation des sciences les unes par rapport aux autres.

Dans ce cas là, les outils numériques sont extrêmement intéressants parce qu'ils permettent de faire des calculs extrêmement rapides, extrêmement puissants, qui prennent en compte tout un tas de données que le cerveau humain ne peut pas saisir et néanmoins il faut toujours que le cerveau humain dise à la machine ce qu'elle doit faire. On arrive jamais à une machine qui travaille à votre place . Vous devez toujours la faire travailler.

Ce premier aspect, la constitution de cette mémoire active, conservation de la mémoire collective avec tout un tas de questions sur la pérennité, pérennisation de ce corpus, avec des bases de données qui doivent pouvoir être pérennes, vu la jeunesse du numérique, on peut se demander quelle est la pérennité des documents qui sont déposés sur internet sachant que les logiciels changent, que la lecture et le moissonnage des documents va poser problème au cours du temps, parce que les formats numériques se modifient, donc il y a toute une réflexion et les informaticiens autour de cette pérennisation,

Le deuxième aspect contenu dans ce site, c'est tout un aspect de mémoire, le site s'adresse tout à la fois aux historiens, aux philosophes mais aussi aux acteurs, un chimiste du solide, un catalyste, un matériau scintiste va pouvoir venir consulter le site et confronter la mémoire qui est disposée, la mémoire sous forme d'entretiens orales avec sa propre mémoire et notre but est que ce site soit un lieu d'échanges avec éventuellement un modérateur qui permet aux gens de discuter. Mais pour avoir un modérateur, il faut avoir de l'argent, payer quelqu'un pour faire ça. C'est peut être un moment où différentes mémoires vont être confrontées et où des aspects d'histoire, des aspects de mémoire vont être confrontés, donc c'est aussi un outil qui permet un rapprochement et une confrontation de discours et création de connaissances.

Puisque c'est un lieu ouvert au public, il faut respecter un certain nombre de règles de déontologie et en particulier il faut que chacun des entretiens, chacun des interviewés en ligne acceptent d'être interviewés, on a des formulaires d'accord, tout un tas de règles communes , est-ce qu'on met des supports sous forme orales ou des supports écrits, est-ce qu'on met des photographies ou pas, on a tout un tas de règles de déontologie pour ne pas qu'il y ait de problème.

Et enfin dernier élément, ce site mais comme d'autres archives orales constitue ce que l'UNESCO appelle, depuis les années 90, le patrimoine oral et immatériel de l'humanité, En même temps que du patrimoine physique : des bâtiments, des constructions, peuvent être sauvegardés par l'UNESCO, l'UNESCO a décidé, depuis les années 90 de sauvegarder aussi, ce qu'ils appellent des cultures, qui peuvent être des cultures orales, donc il y a un effort en faveur de la protection de ce patrimoine moral et matériel. Ces projets d'archives orales ont été suscités par deux événements importants : d'abord la Shoah au moment de la deuxième guerre mondiale, a suscité tout un tas de conservation de la parole, la parole de ceux qui sortaient des camps de concentration pour essayer de comprendre l'incompréhensible et en même temps dans les années 50 aux Etats Unis, il y a eu un courant de prise de paroles par les minorités indiennes, les minorités afro-américaines qui ont pris la parole, donc ces paroles là ont pu développer des programmes d'archives orales afin de garder une certaine tradition, par exemple les traditions orales des indiens des grandes plaines ont pu être ainsi, grâce à des anciens, qui connaissaient encore les mythes, les légendes, être conservées.

Donc ce site va apparaître à la fois comme un outil d'histoire, un outil d'historien au niveau de l'analyse du document mais aussi au niveau de la confrontation de la mémoire des acteurs les uns avec les autres en particulier, on peut très bien imaginer, que par rapport à des

passages d'entretiens qui sont cités sur la site sciences histoire orale d'autres témoins du domaine qui se souviennent, qui nous écrivent pour proposer une autre version du même épisode et là, on est dans un enrichissement de la mémoire collective par confrontation. Ça c'est au niveau des outils pour l'historien ou d'une manière plus générale pour les Sciences Humaines et sociales et d'un autre côté on a tout cet aspect de patrimonialisation de la mémoire . La mémoire devient un patrimoine commun, un patrimoine collectif de différentes cultures, et la Science, et les différents champs scientifiques, sont un type de culture parmi d'autres. Donc double aspect du corpus d'archives orales.

On va passer à la troisième partie, le corpus d'archives orales qui va être analysé, utilisé pour servir de base au récit historique.