

Développement des métiers liés aux outils de la langue

Jacques Souillot

Les métiers traditionnels qui ont les langues pour objet sont ordinairement répertoriés, sinon définis, dans l'esprit du grand public. On pense immédiatement au traducteur, à l'interprète, au professeur de langue. Cependant, bien que l'on ait clairement conscience que ces professionnels utilisent, comme tout un chacun, des dictionnaires, on a souvent tendance à oublier de compter parmi les spécialistes des langues, les concepteurs-producteurs de dictionnaires que sont les lexicographes et les terminologues. Par ailleurs, on est généralement enclin à penser « langue » comme « langue étrangère » et en conséquence à oublier qu'un professionnel travaillant sur des choses énoncées dans sa propre langue appartient à la grande famille des langagiers. Les productions de notices techniques, de documents de vulgarisation, de comptes rendus analytiques, de synthèses de documents, ne sont en effet pas si éloignées que cela de l'activité du traducteur, qui, lui, fondamentalement, fait de la réécriture, ni même de l'activité du terminologue, toujours à batailler avec les imprécisions, les ambiguïtés des mots.

Les différents secteurs, leurs outils

Si le volume de la sphère des métiers et des industries des langues s'est accru sans discontinuer ces dernières années, sa densité, elle, s'est multipliée

LCN, volume 1, n° 3-2000, pages 97 à 110

à un rythme en perpétuelle accélération. Que l'on parle de l'aspect industrie de service ou de l'aspect industrie de produit, le paysage a subi des transformations si importantes qu'il n'est plus possible aujourd'hui de le décrire avec précision sans avoir recours aux jargons particuliers des sous-secteurs qui le composent, ou tout au moins à la langue de notre culture du numérique.

La traduction

On parle de TAO (Traduction assistée par ordinateur), de TA (Traduction automatique), de THAM (Traduction humaine assistée par la machine), d'environnement du poste du traducteur (avec ses outils matériels et logiciels, ses ressources propres et partagées, ses protocoles, ses interfaces).

Le traducteur, s'il reste le maître *es* langues qu'il a toujours été, a de moins en moins de rapport « charnel » avec le texte, les mots, le contenu, le sens même de ce qu'il traduit. En fait, on ne lui demande plus de traduire mais plutôt de produire une traduction. Dans ces circonstances, ses capacités d'adaptation à son nouveau rôle et ses compétences à gérer les multiples tâches que compte son poste prennent le pas sur son excellence linguistique. D'ailleurs, au sein d'une entreprise de traduction, la promotion des traducteurs se fait sur la base de leur efficacité à conduire un projet, les plus performants devenant *project managers*. Ce sont eux qui établissent les cahiers des charges, préparent les devis, distribuent les tâches, assurent la coordination, le contrôle qualité, la validation des différentes étapes, etc. Plutôt que d'être en relation permanente avec les arcanes des langues, ils sont investis de la logique et des impératifs des contrats, une réalité professionnelle que la plupart des jeunes linguistes sur les bancs des universités sont loin d'imaginer.

Le traducteur, quant à lui, qu'il appartienne à une entreprise de traduction ou à un grand service de traduction institutionnel, se voit soumis à des objectifs de productivité nettement affirmés. Dans les cas de figure les plus extrêmes, son rendement (en volume) est vingt fois supérieur à celui d'un artisan, ce facteur étant plus proche de quatre ou cinq en moyenne. Toutefois, avant que ne puisse être atteint ce régime de croisière si élevé, l'« environnement » du traducteur aura dû être aménagé de façon approfondie. Outre les logiciels de traduction automatique, il faut avoir mis en route des bases de données terminologiques exhaustives, éventuellement amorcé des mémoires de traduction (fichiers contenant un texte source et sa traduction, ce qui permet de récupérer, par comparaison, des segments déjà traduits), mis en ligne des dictionnaires bilingues spécialisés se rapportant aux domaines couverts par les textes.

La logistique au soutien du traducteur à « très grande vitesse », on le constate, requiert d'importants investissements. Les ressources compétences nécessaires aux travaux de préparation sont puisées très souvent chez les traducteurs eux-mêmes, puisqu'ils ont généralement reçu une formation poussée en terminologie et lexicographie. Pour les travaux de préparation destinés à des traductions pour des grands comptes (clients ayant des volumes astronomiques et des exigences serrées en matière de délais), l'acquisition des ressources nécessaires se fait à l'extérieur, soit chez des fournisseurs de ressources linguistiques, soit chez les clients eux-mêmes, qui, majoritairement, ont déjà produit lesdites ressources en interne, pour leurs besoins propres en communication et gestion documentaire, par exemple.

Il pourrait ressortir de ce que nous venons de décrire que la traduction « industrielle » a besoin et génère des grandes entreprises de traduction. C'est le cas, d'autant plus que le capital compétences et ressources linguistiques d'une entreprise est sans cesse augmenté et valorisé par le partage et les échanges internes. Une mémoire de traduction approvisionnée par plusieurs traducteurs, plusieurs projets, sera plus vite opérationnelle, des dictionnaires bénéficiant de nombreuses révisions et ajouts seront plus vite fiables. Il s'agit là du classique effet boule de neige. Les poids lourds du secteur seront donc plus compétitifs que les autres pour traiter des « quantités industrielles » de texte. Ils seront aussi davantage contraints, car les seuls textes qui peuvent subir de manière continue les processus de traduction auxquels ils ont recours sont ceux dont la langue est elle-même relativement contrainte (c'est-à-dire une langue à caractère technique, sans ambiguïtés lexicales ni syntaxiques).

On peut voir se dessiner, en fait, des acteurs et marchés divers : les grosses entreprises dominant sur le créneau grands comptes, avec leur tout automatique, les PME consacrant leur activité aux clients dont la demande en volume est moins importante, ou dont les besoins ne trouvent des réponses qu'avec des technologies plus dépendantes du traducteur (TAO, THAM), les traducteurs indépendants conservant les domaines non accessibles à l'industrie.

Terminologie et lexicographie

Les activités des terminologues et des lexicographes sont directement liées aux activités de traduction. Bien que pour une part elle soit souvent partagée avec les traducteurs, la responsabilité des constructions de terminologies et de lexiques ou dictionnaires bilingues est de plus en plus confiée aux spécialistes de ces questions. Les rangs les plus fournis de ces

métiers sont bien sûr ceux des linguistes ; ils sont largement complétés par ceux des documentalistes, qui, pour leurs besoins en classification, indexation et recherche d'information, se forgent les ressources linguistiques adéquates. Autant que faire ce peut, toutefois, on adjoint des terminologues professionnels aux équipes de documentalistes afin d'assurer la cohérence linguistique des ensembles à produire, ainsi que leur portabilité vers les systèmes qui y ont recours.

Les grands projets linguistiques, de nos jours, quel que soit leur objectif final, passent régulièrement par l'étape « terminologie ». Les efforts pour automatiser cette étape sont importants. Car elle nécessite le traitement de masses considérables de données et est dévoreuse de temps et de compétences. Les technologies avancées de la terminologie (la terminotique) reposent aujourd'hui sur le traitement de vastes corpus linguistiques (collections de textes représentatifs d'un domaine répondant à un certain nombre de critères, construites suivant des règles et des formats précis¹).

Après que le corpus à exploiter ait été soumis à nombre d'opérations préparant le terrain pour le terminologue, les fichiers enrichis deviennent la pâture d'un logiciel d'extraction terminologique [BOU 98]. Qu'il soit entièrement ou semi-automatique, il requiert le pilotage du terminologue qui seul est en mesure d'apprécier la validité linguistique des sélections effectuées par le logiciel. Quant à la correspondance des termes sélectionnés avec la réalité du domaine contenu dans les textes, elle va demander l'intervention d'un expert dudit domaine qui devra s'assurer de la propriété des termes et, en collaboration avec le terminologue, valider la pertinence des termes retenus.

Vu l'investissement massif en temps et expertise nécessité par la production de terminologies, on vise très souvent à rendre les produits réutilisables, et le principe de réutilisabilité est inscrit dans les projets dès la conception, dès l'établissement du cahier des charges. Un produit terminologique peut d'ailleurs devenir une ressource linguistique stratégique et être distribué vers des utilisateurs extérieurs. Il va de soi que ceux-ci devront pouvoir l'exploiter et l'intégrer dans leurs applications rapidement et efficacement.

Avec l'explosion du multilinguisme dans tous les médias, les besoins en dictionnaires bi- et plurilingues ne cessent de s'accroître. La constitution de ces dictionnaires, si l'on veut assurer leur parfaite fiabilité, implique que

1. Domaine de la « linguistique de corpus », l'une des bases du traitement automatique du langage naturel (TALN ou encore TAL) [HAB 97].

soient établies au préalable des terminologies multilingues. Or, on l'a vu plus haut, créer une terminologie sur un domaine particulier, pour une seule langue, est déjà très exigeant que ce soit au niveau temps, compétences ou ressources linguistiques. Aussi, de tels projets terminologiques sont rares et ne sont guère menés que par des équipes universitaires travaillant sur des programmes bénéficiant de subventions institutionnelles. Mais des efforts colossaux sont déployés par ces « bénévoles », et, à l'heure actuelle, de nombreuses équipes en Europe sont engagées dans des programmes de collaboration, tel le projet EuroWordnet² dont les objectifs sont de couvrir toutes les langues du continent.

Néanmoins, les lexicographes spécialistes des dictionnaires bilingues ne restent pas inactifs en attendant, et produisent eux aussi les ressources linguistiques dont sont si friands les logiciels de traduction notamment. La maintenance et l'enrichissement de ces dictionnaires sont eux-mêmes des activités technologisées pour lesquelles les traitements de corpus sont fondamentaux, et dans le cadre desquelles s'est développé le concept de « veille linguistique ». Pour ce qui concerne les dictionnaires papier du monde de l'édition, ceux-ci n'ont pas encore vraiment bénéficié des acquis de ce que l'on appelle la lexicomatique. Même dans leur version numérique (cédérom), leur conception n'a pas été sensiblement modifiée. Une nouvelle génération de dictionnaires est attendue, et avec elle un plus grand nombre de combinaisons de langues et une prépondérance moindre de l'anglais dans les couples linguistiques offerts³.

Localisation

La prise en compte du multilinguisme, dont le principe revêt essentiellement une dimension humaniste, est un élément central du développement des industries des langues ; l'internationalisation des marchés est aussi le catalyseur de ce développement. Tout le monde sait que vendre, c'est parler la langue de son client : plus on veut atteindre un nombre élevé de clients, plus on doit ajuster sa communication à différents profils. Les producteurs des industries de loisir, en particulier le cinéma, ont toujours tendu à élargir leur distribution à l'international. Avec les TIC, cela leur est devenu plus facile. Par conséquent, pouvant atteindre un public plus vaste, ils sont à même d'assurer des retours sur investissement encore plus

2. <http://www.hum.uva.nl/~ewn/> ; projet reprenant les spécifications de Wordnet, qui est d'origine américaine.

3. Actuellement, les dictionnaires bilingues dédiés au japonais, par exemple, n'offrent pratiquement pas d'autre combinaison que celle de l'anglais !

favorables. Le cas des films distribués sur support DVD est caractéristique des avantages procurés par l'association des TIC et du multilinguisme ; en effet, sur un même DVD, il y a un film dont la bande son et/ou les sous-titrages sont disponibles en plusieurs langues. Le poids de la logistique pour le producteur/distributeur est considérablement réduit.

Les secteurs du doublage (*dubbing*) ou du sous-titrage de film sont donc en pleine expansion. L'art des linguistes y joue un rôle fondamental, mais il s'appuie désormais sur des technologies permettant une production plus rapide et des environnements de travail plus conviviaux.

Ces activités ne sont pas nouvelles, mais elles sont les précurseurs de ce que l'on a vu émerger ces dernières années sous le nom de « localisation ». Ce nouveau secteur prend en charge la conversion d'un logiciel dans une autre langue, une sorte de « naturalisation ». C'est plus qu'une traduction. Tout d'abord, la connaissance approfondie de l'utilisation, voire du fonctionnement du logiciel est impérative (on ne traduit que ce que l'on comprend !) : sans elle, pas d'assise terminologique sérieuse, pas de reconnaissance des nouvelles acceptions qu'a pu prendre un mot ou une expression dans ce contexte nouveau [PET 96]. Les besoins en créations de néologismes doivent être analysés, pesés (ce qui requiert une solide expérience dans le domaine), et des solutions lexicales proposées. La responsabilité du linguiste à cet égard ne doit pas être sous-estimée : c'est en partie de lui que va dépendre la non-prolifération dans notre langue de termes barbares tels que *backup* (pour « copie de secours »), ou *downloader* (pour « télécharger ») !

L'un des points cruciaux du travail de localisation est celui des menus du logiciel. Des menus dont les éléments traduits sont devenus flous ou totalement inadéquats vont rendre le logiciel inutilisable, ou tout au moins entraîner pour l'utilisateur la mise en œuvre de mécanismes de corrections mentales qui nuiront gravement à son efficacité. En fait un logiciel mal localisé risque d'être un échec commercial. L'échec est garanti si par ailleurs les modules d'aides ne sont pas en cohérence avec les menus, si les index ont perdu leur précision, et si, là-dessus, le logiciel est accompagné d'un manuel au contenu tout aussi approximatif. D'autres éléments entrent également en ligne de compte dans la localisation des logiciels : couleurs, icônes, et toute particularité entrant en conflit avec les spécificités culturelles des clients visés.

Les produits demandant le plus de précision dans l'adaptation sont les logiciels multimédias. Ce sont aussi les produits qui, plus que tout autre, requièrent une gamme étendue de localisations, car restreindre leur diffusion à un marché « natif » (en une seule langue, sauf pour ce qui est de

l'anglais) ne permet guère de couvrir les frais de production engagés : il n'existe de marché viable, en matière de multimédia, qu'international. Le créneau le plus porteur dans ce secteur d'activité est celui de la localisation des logiciels de jeux, ce qui n'est pas surprenant quand on sait que l'industrie du multimédia ludique représente un volume d'activité et une puissance financière nettement supérieurs à ceux de l'industrie du cinéma.

Les linguistes et l'internet

La toile absorbe aujourd'hui un nombre grandissant de nouveaux experts linguistes. En premier lieu, les besoins en localisation y sont très aigus. Communiquer avec le monde entier ou simplement affirmer sa présence sur la toile conduit à offrir aux internautes des pages qu'ils pourront lire et comprendre, des sites sur lesquels ils pourront interagir et opérer des échanges (intellectuels, commerciaux ou autres⁴).

La diversification des langues sur les sites, ressentie comme nécessaire, demande donc un renfort important en linguistes. Outre leurs fonctions de localisateurs, ces derniers se voient confier l'animation du site dans la version langue étrangère dont ils sont spécialistes (ce devrait être leur langue maternelle). Ce rôle de médiateur linguistique et culturel ne représente cependant qu'une partie de ce qui est du ressort des linguistes dans les technologies de l'internet. Ceux d'entre eux qui ont acquis une formation davantage orientée vers les techniques documentaires se voient confier les travaux de recherche d'information, de fouille, de veille (technologique, commerciale...), de filtrage.

Là encore, leur expertise en terminotique est un atout majeur, et leur appréhension fine des phénomènes linguistiques, la condition de l'aboutissement des tâches qu'ils conduisent. Repérer des octets parmi les océans de pages HTML est une chose que les moteurs d'indexation accomplissent automatiquement ; les moteurs de recherche trouvent une information particulière avec succès (et plus ou moins de précision, plus ou moins de silence et de bruit) ; mais extraire du sens (c'est-à-dire exploiter les données éclatées sur toute la toile pour obtenir des informations structurées) et mettre en forme des connaissances demande, quel que soit le degré de technologisation des diverses opérations entreprises, l'intervention d'une expertise humaine de très haut niveau. Les linguistes exerçant dans ces champs d'application se métamorphosent de plus en plus en ingénieurs de

4. Or, bien que la moitié des internautes à l'heure actuelle soit américaine, c'est encore plus de 85 % de la toile qui reste en langue anglaise.

la connaissance, les méthodes et techniques qu'ils emploient relevant directement de l'intelligence artificielle et des sciences cognitives.

Les divers nouveaux profils de linguiste apparaissant avec le développement des autoroutes de l'information et l'orientation « société cognitive » qui y est associée mettent clairement en relief le fait que les qualifications d'un expert linguiste dépassent les savoirs et savoir-faire relatifs aux langues et aux sciences du langage, qui autrefois constituaient l'essentiel de sa spécialité. Aujourd'hui, le linguiste, informaticien ou non, est appelé à maîtriser les outils des TIC et à les utiliser dans des projets relevant de domaines connexes. Bien entendu, son expertise de linguiste demeure dominante, mais l'enrichissement de ses qualifications s'avère indispensable, tant pour la conduite de son travail au niveau personnel que pour son intégration aux équipes pluridisciplinaires avec lesquelles il « mutualise » ses connaissances.

Il en découle qu'il est devenu plurispécialiste, qu'il est tout aussi bien expert en documentation qu'en analyse du discours, sémiotique, multimédia, sciences de l'information et de la communication...

Multimédia et didactique des langues

Le portrait du nouveau linguiste protéiforme qui vient d'être esquissé n'est qu'un des extrêmes vers lesquels on tend dans les métiers de la langue, des langues. On n'oubliera pas en effet que l'enseignement des langues s'est depuis longtemps embarqué dans la technologisation de ses outils avec l'Enseignement assisté par ordinateur (EAO). Il est plus juste d'y faire référence de nos jours en tant qu'apprentissage des langues assisté par les TIC⁵. Si les bases techniques, méthodologiques et pédagogiques de l'EAO de langues sont désormais plus assurées et le public toujours aussi fasciné par les produits multimédias qu'il rencontre, il demeure que par maints aspects, l'EAO de langue n'a pas encore atteint une maturité satisfaisante (en ce qui concerne ses buts, en particulier). On peut d'ailleurs se poser la question de savoir si cette variété d'enseignement verra jamais un quelconque équilibre, puisque, les technologies évoluant sans arrêt, de nouvelles possibilités d'exploitation et d'orientation voient le jour au même rythme. On a ainsi pour les langues de grands chantiers d'enseignement à distance qui viennent superposer leur problématique didactique à celles des environnements communautaires interactifs, ou celles des environnements axés sur l'apprentissage autonome (les centres de ressources de langues).

5. Voir la revue électronique ALSIC (*Apprentissage des langues et systèmes d'information et de communication*) : <http://alsic.org>.

Ce cadre mouvant est celui dans lequel les linguistes ingénieurs de la connaissance mettent en œuvre les méthodologies et techniques rodées de leur métier, tout en intégrant les dimensions pédagogiques et didactiques nécessaires. Les enseignants de langue sont très souvent leurs doubles ou homologues, nombre d'entre eux ayant acquis les savoirs et savoir-faire linguistiques et technologiques requis.

L'industrie du multimédia de langue n'apparaît cependant pas entravée par les difficultés de définition de ce que sont, pourraient ou devraient être les applications des TIC à l'apprentissage des langues. D'ailleurs, les cédéroms de langues viennent largement en tête des ventes dans la catégorie multimédias culturels et pédagogiques.

Les technologies de la parole

Lorsque l'on parle des industries de la langue, ne devrait-on pas mentionner les technologies de la parole? Synthèse et reconnaissance vocales ne sont généralement pas reconnues comme des apports des linguistes. Depuis leur début, les études dans ce domaine font appel à des moyens technologiques peu accessibles du public, ainsi qu'à des traitements mathématiques qui l'effraient. Là-dessus on replace le tout dans le cadre scientifique de l'acoustique, on renforce son lien avec l'électronique, et on laisse cela aux mains (et aux cerveaux) des industries des télécommunications. La phonologie et la phonétique ont pourtant leur mot à dire dans ce contexte, même si l'on a inventé le téléphone avant qu'elles ne puissent se prononcer. Et, de même que l'on n'a pas eu besoin des linguistes pour encoder les chaînes de caractères sur les premiers ordinateurs, on a pu se passer de leur service pour soumettre les sons d'une langue au format électronique.

Mais phonéticiens et linguistes se sont retrouvés et font converger leurs technologies⁶. Il est évident qu'une machine capable de reconnaître des mots ou des voix doit encore évoluer pour comprendre le sens de ce qui lui est dit; il est tout aussi évident qu'une machine capable de porter des mots à nos oreilles a également besoin d'être perfectionnée avant de pouvoir engager un dialogue « naturel » avec l'homme.

Bien que l'on n'ait pas atteint à ce jour les objectifs évoqués ci-dessus, les progrès réalisés dans ce domaine ces dernières années ont bouleversé les

6. ELSNET (European network in language and speech) joue d'ailleurs un rôle moteur dans la promotion de synergies entre les acteurs des technologies de la langue et ceux des technologies de la parole (<http://www.elsnet.org>).

données économiques du monde des télécommunications. La miniaturisation des composants des machines, les gigantesques capacités de mémoire et de calcul, les extraordinaires vitesses de transport de l'information (que ce soit par satellite ou réseau) sont bien sûr à la base de cette révolution. En fait, on pourrait presque dire que phonéticiens et linguistes ont seulement aujourd'hui les moyens d'associer leurs recherches, recherches dont les résultats permettront aux industries de produire les outils de communication dont nous semblons rêver.

L'interaction vocale (partie cruciale du dialogue homme-machine), malgré ses limites actuelles, présente déjà une convivialité appréciable, ses capacités à automatiser certaines tâches ont prouvé sa rentabilité économique et son apport a pu se révéler indispensable à la mise en œuvre de certaines technologies. On en a pour preuve les applications destinées aux services d'information, les logiciels de dictée, d'aide vocale aux malvoyants, les dispositifs de commande vocale (dont un exemple est le pilotage de caméra exploratoire en chirurgie). On notera également qu'on peut être séduit par les messages vocaux des machines, car ceux-ci induisent moins de tension que des signaux sonores basiques, souvent ressentis comme agressifs.

Les ressources linguistiques

Que ce soit pour les technologies de la parole ou pour celles du langage, on a besoin dans toute circonstance (création d'outils, test, exploitation) de ressources linguistiques importantes. En fait plus ces ressources seront volumineuses plus elles auront de chance d'être utilisables. Il leur faudra également être parfaitement représentatives de la langue que l'on doit traiter. En toute logique on n'utilisera pas un corpus de textes relatifs à la biologie végétale pour mettre au point des applications destinées au domaine de l'aéronautique. Cela implique que chaque domaine, sous-domaine et spécialité ait à sa disposition les corpus nécessaires à la mise au point de ses outils linguistiques. C'est dire l'ampleur du travail dans ce secteur des industries de la langue. Elle est encore augmentée par le nombre de langues à traiter. Bien sûr, les pays technologisés construiront eux-mêmes les ressources concernant leur(s) propre(s) langue(s), mais qu'en est-il des langues des régions défavorisées, des « petites » langues, des langues minoritaires⁷ ?

7. A ce sujet, on prévoit que la survie des langues de faible diffusion va dépendre de leurs ressources linguistiques numérisées et donc de leur potentiel technologique. Se pose là un problème de politique linguistique à l'échelle mondiale qui pourrait risquer d'être trop vite réglé par les acteurs économiques les plus puissants.

Même si les pays à haute technologie ne donnent pas dans l'humanisme et l'humanitaire à tout crin pour préserver les langues en voie d'extinction, il leur reste à établir, à côté de leurs propres corpus écrits, les corpus oraux, tout aussi variés que les précédents.

Le TAL. L'ingénierie linguistique

La production des outils et méthodes à destination des divers métiers des industries de la langue peut elle-même être considérée comme pourvoyeur de ressources linguistiques spécifiques : les logiciels de traitement linguistique. C'est le sous-secteur lourd de ces industries qui crée et met au point, au sein de laboratoires de recherche et de développement privés ou de laboratoires universitaires, les logiciels auxquels nous avons fait allusion tout au long de cet article, sans en faire une liste exhaustive (il faudrait y ajouter les outils d'aide à la rédaction, les logiciels de génération ou de résumé automatique de texte, les correcteurs orthographiques, les logiciels de reconnaissance de caractères...).

C'est aussi vers le TAL que l'on se tourne pour l'établissement des ressources linguistiques. Dès la numérisation des textes [HAB98], l'approche des linguistes informaticiens est strictement scientifique et toute de rigueur systématique. Leurs protocoles de contrôle et de validation sont présents à toutes les étapes des traitements. Encodages, formats, normes sont leur premier souci : on sait très bien que si l'on n'a rien avec rien, on obtient aussi n'importe quoi en utilisant n'importe quoi (*garbage in, garbage out*). Les opérations auxquelles ils soumettent ensuite les données textuelles constituent le cœur de leur métier : analyse (morphologique, syntaxique, sémantique), étiquetage, indexation, traitements statistiques, éventuellement annotations spécifiques, création de concordanciers, alignement de corpus de langues différentes.

Toutefois, tous ces traitements qui reposent sur l'exploitation des capacités des ordinateurs à effectuer des calculs à des vitesses phénoménales ne nous renseignent guère sur les fondements de la discipline TAL. En effet, TAL (ou linguistique informatique) ne signifie pas simplement linguistique + informatique. Les propositions de la linguistique théorique n'ont que très rarement des applications directes en TAL : la modélisation, la formalisation nécessaires à toute recherche appliquée sont généralement étrangères aux préoccupations des théoriciens. D'autre part, les approches informatiques pures et dures, animées par des algorithmes à forte coloration mathématique, si elles rendent de grands services en temps et lieux où elles sont requises, se révèlent inadéquates lorsqu'elles sont utilisées comme stratégies de débroussaillage de la langue. C'est dès la conception des

logiciels et à chaque étape de leur développement que les représentations linguistiques doivent être présentes. Si les pages de texte sont solubles dans l'informatique, les objets linguistiques, eux, ne le sont pas, et leurs liens restent tout autant complexes et invisibles.

Conclusion

Comme nous venons de le voir, les développements des métiers liés aux outils de la langue sont soumis à un bouleversement important dû à l'implantation des TIC. Ceci entraîne entre autres l'exigence d'une adaptation des préparations à ces métiers.

La demande en ingénieurs linguistes ne cesse de croître et ne manque pas de susciter des inquiétudes tant chez les employeurs, pour qui les viviers de recrutement sont trop rares, que chez les citoyens informés, qui voient tout un secteur économique en perte d'équilibre du fait de sa pauvreté en ressources et compétences humaines. Cet aspect négatif de la situation est en outre aggravé par la pénurie constatée en informatique : des jeunes gens formés en linguistique informatique révisant leurs choix professionnels et optant pour l'informatique de gestion par exemple. Les influences économiques sur les orientations des jeunes sont également marquées par les phénomènes de concurrence internationale, avec le trop traditionnel phénomène de « fuite des cerveaux ».

Parallèlement, les formations à caractère traditionnel, les plus courantes en lettres et sciences humaines, ne semblent pas réagir devant les statistiques d'embauche de leurs diplômés, qui sont comparativement très défavorables. Les systèmes universitaires en général n'ont guère appréhendé le sens des évolutions de nos nouvelles sociétés axées sur les technologies, l'information, la mondialisation. En conséquence l'adaptation radicale de certains de leurs cursus n'a pas pu être préparée, ni même envisagée. En linguistique informatique de rares poches de formation sont présentes⁸, mais celles-ci n'offrent pas toujours des cursus complets : un cursus de 2^e cycle ici, un DESS là ; et pour ce qui est des écoles doctorales on constate un déficit en thésards symptomatique, aggravé par le fait que les étudiants choisissent plus facilement un emploi que 4 ou 5 années d'études supplémentaires.

8. L'université Paris VII, *via* Atala (Association pour le traitement automatique des langues), tient un répertoire des formations « industries de la langue » sur son site internet : <http://www.biomath.jussieu.fr/ATALA/formation/>.

C'est avec effroi que l'on peut considérer le futur si les professeurs de demain ne sont pas formés maintenant. L'immobilisme risque de se payer très cher dans l'avenir, que ce soit en termes de crédibilité pour l'université (qui n'aura pas la capacité d'assurer les formations nécessaires et certainement pas en nombre suffisant) ou en termes de compétitivité économique.

L'importance économique du secteur des industries de la langue ne devrait plus avoir à être démontrée, sans compter que son potentiel de développement⁹ est à la mesure de ce qui fermente et bouillonne dans toutes les nouvelles technologies. Par ailleurs l'équilibre géopolitique des industries de la langue se calque sur celui des divers secteurs de pointe, et la stabilité de leurs entreprises est soumise aux mêmes tangage et roulis que ceux-ci : mouvements de capitaux, fusions, concentrations.

Il semblerait donc vital que de sérieux efforts soient faits pour éviter que ne s'installent ou pérennisent des monopoles ou quasi-monopoles, porteurs de risque d'asphyxie, stérilisation, déresponsabilisation. Finalement, pour reprendre le thème de « La guerre des langues et les politiques linguistiques » de J.-L. Calvet [CAL 99], si guerre il y a effectivement, c'est aujourd'hui sur ce front des industries de la langue qu'elle est la plus active. Mais ce n'est pas une guerre de tranchées...

Bibliographie

[BLO 99] BLOOTHOOFT G., van DOMMELEN W., FELLBAUM K., HAZAN V., HUCKVALE M., LEATHY M., WIGFORSS E. (eds), *The Landscape of Future Education in Speech Communication Sciences*, vol. 1-3, Utrecht, OTS Publications, 1997-1999.

[BOU 98] BOURIGAULT D., HABERT B., « Evaluation of Terminology Extractors: Principles and Experiments », *Actes de la première conférence internationale sur les ressources linguistiques et l'évaluation*, vol. 1, p. 299-305, Grenade, Espagne, 1998.

[CAL 99] CALVET L.-J., *La guerre des langues et les politiques linguistiques*, Hachette Littératures, 1999.

[HAB 97] HABERT B., NAZARENKO A., SALEM A., *Les linguistiques de corpus*, Armand Colin/Masson, 1997.

[HAB 98] HABERT B., FABRE C., ISSAC F., *De l'écrit au numérique : constituer, normaliser et exploiter les corpus électroniques*, Masson, 1998.

9. Un aperçu de la situation est donné sur le site du Riofil (Réseau international des observatoires francophones de l'infrastructure et du traitement automatique des langues) : <http://www.riofil.org/>.

[PET 96] PETRALLI A., *Neologismi e nuovi media : Verso la globalizzazione multimediale della comunicazione ?*, CLUEB, Bologna, 1996.

[SME 99] SMEDT K. (de), GARDINER H., ORE E., ORLANDI T., SHORT H., SOUILLOT J., VAUGHAN W. (eds), *Computing in Humanities Education: a European perspective*, University of Bergen, 1999.