

TECHNOLOGIES DE LA LANGUE EN EUROPE : MARCHE ET TENDANCES

SYNTHESE DE L'ETUDE

1. OBJECTIF

L'objectif de cette étude réalisée à la demande du Département « Technologie de l'Information et de la Communication » du Ministère délégué à la Recherche dans le cadre du programme Technolangue¹ est de qualifier et d'évaluer le marché européen des outils linguistiques traitant les données textuelles et vocales, et d'en identifier les axes de développement. Elle porte sur 13 pays européens (Allemagne, Autriche, Belgique, Danemark, Espagne, Finlande, France, Grèce, Irlande, Italie, Pays-Bas, Royaume-Uni, Suède), avec une attention particulière pour le marché français.

2. METHODOLOGIE

La méthodologie a été basée sur l'analyse de publications existantes, la réalisation d'une enquête par questionnaires auprès des offreurs de technologies, et des entretiens avec des utilisateurs et des experts pour spécifier et ajuster les résultats obtenus.

L'étude s'est déroulée, pour l'essentiel, sur l'année 2003 selon 3 phases :

- Phase 1 : caractérisation du marché de l'offre en Europe et en France, recensement de l'offre dans les principaux pays européens,
- Phase 2 : état de la demande et typologie des applications déployées (élaboration du panorama général des applications sollicitant des solutions et des produits informatiques utilisant des ressources linguistiques),
- Phase 3 : identification des tendances et des développements futurs (tendances et perspectives du marché en Europe et en France à l'horizon 2005-2007).

3. RESULTATS

Le marché des technologies de la langue est représenté par une multitude d'acteurs qui vendent des solutions logicielles déployant diverses fonctionnalités linguistiques. Certaines sociétés sont à la fois éditeurs de logiciels et offreurs de services. Un *overlap* de 60% a été retenu afin de ne pas surestimer la valeur de ce marché. S'agissant d'une étude de marché, la qualification des technologies a été réalisée selon une typologie basée sur les usages et les applications déployées. Cette typologie comprend neuf segments :

1. Gestion du contenu	4. E-business	7. Interfaces vocales, multimodales
2. Gestion des documents	5. Ressources linguistiques	8. E-learning
3. Traduction	6. Production et édition	9. Moteurs de recherche intelligents

3.1. IDENTIFICATION ET EVALUATION DE L'OFFRE

Facteurs structurels et conjoncturels

Le marché mondial des NTIC est évalué à 2292 milliards d'euros en 2001. Le marché des logiciels en représente 10% (230 milliards d'euros) et les services attachés 20% (480 milliards d'euros). La part de l'Europe est estimée à 29%. Le marché des logiciels en France est estimé à 20,6 milliards d'euros en 2002.

Les dépenses en NTIC dans la zone des 13 pays européens ont été de 493 milliards d'euros en 2002. Pour 2004, les prévisions s'élèvent à 6,9% du PIB contre 9,4% pour les USA.

¹ <http://www.technolangue.net>

La part des dépenses en R&D (investissements privés et publics) représente 157 milliards d'euros en 2002 soit 1,86% du PIB. L'Allemagne, la France, le Royaume-Uni et l'Italie représentent 120 milliards d'euros, soit 75% des dépenses totales en R&D. En Europe, la R&D orientée linguistique est estimée à 400 millions de dollars en 2002, et 800 millions de dollars en 2003.

La croissance de la numérisation, du stockage des données dans les réseaux influence fortement le marché avec la montée en puissance du multimédia (fichiers sons, images et vidéos) et la hausse de la production mondiale d'information (de 69% en volume entre 2000 et 2002).

L'élargissement de l'utilisation d'Internet est un des facteurs à ne pas négliger pour l'évolution des technologies de la langue. Parmi les nouveaux segments qui influencent ce marché, on peut citer le *e-content*, le *e-business*, le *e-learning*, le multimédia, le multimodal, l'Internet mobile, le web sémantique...

Le multilinguisme est un facteur important en Europe (l'Europe à 25 comprendra 450 millions de personnes, dont 55% parlant au moins une seconde langue).

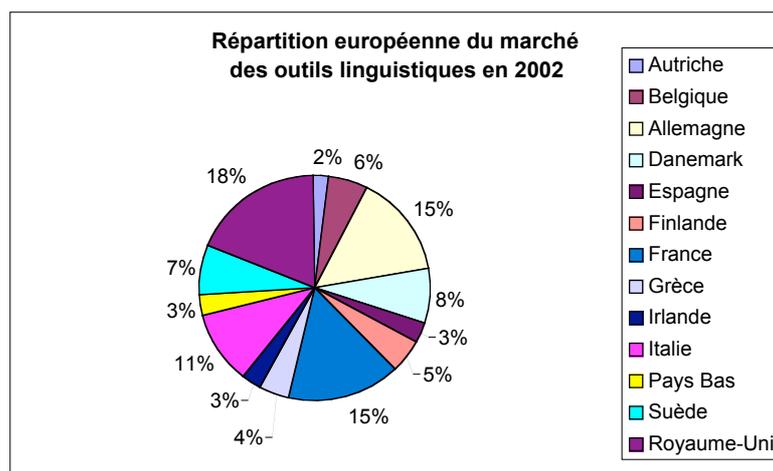
Evaluation du marché de l'offre en matière de technologies linguistiques :

En Europe, les créations de sociétés ont considérablement augmenté dans les années 90 et 2000, même si près de 30 sociétés européennes spécialisées dans le traitement du langage ont disparu (liquidation ou rachat) sur la période 2000-2003. En quinze ans, le nombre de sociétés créées a néanmoins été multiplié par plus de 10, passant d'une trentaine à plus de 400 sociétés en Europe en 2003.

377 offreurs européens sont présents sur au moins un des 9 segments d'application :

- L'offre en Europe se répartit entre le traitement du texte (80%) et le traitement de la voix (20%) ;
- Le marché européen est évalué à 510 millions d'euros en 2002 ;
- Le Royaume-Uni, la France, l'Allemagne, et l'Italie représentent 60% du marché européen.

Sur la base des chiffres d'affaires communiqués, le marché européen des outils linguistiques est estimé à environ **510 millions d'euros**, ainsi répartis:



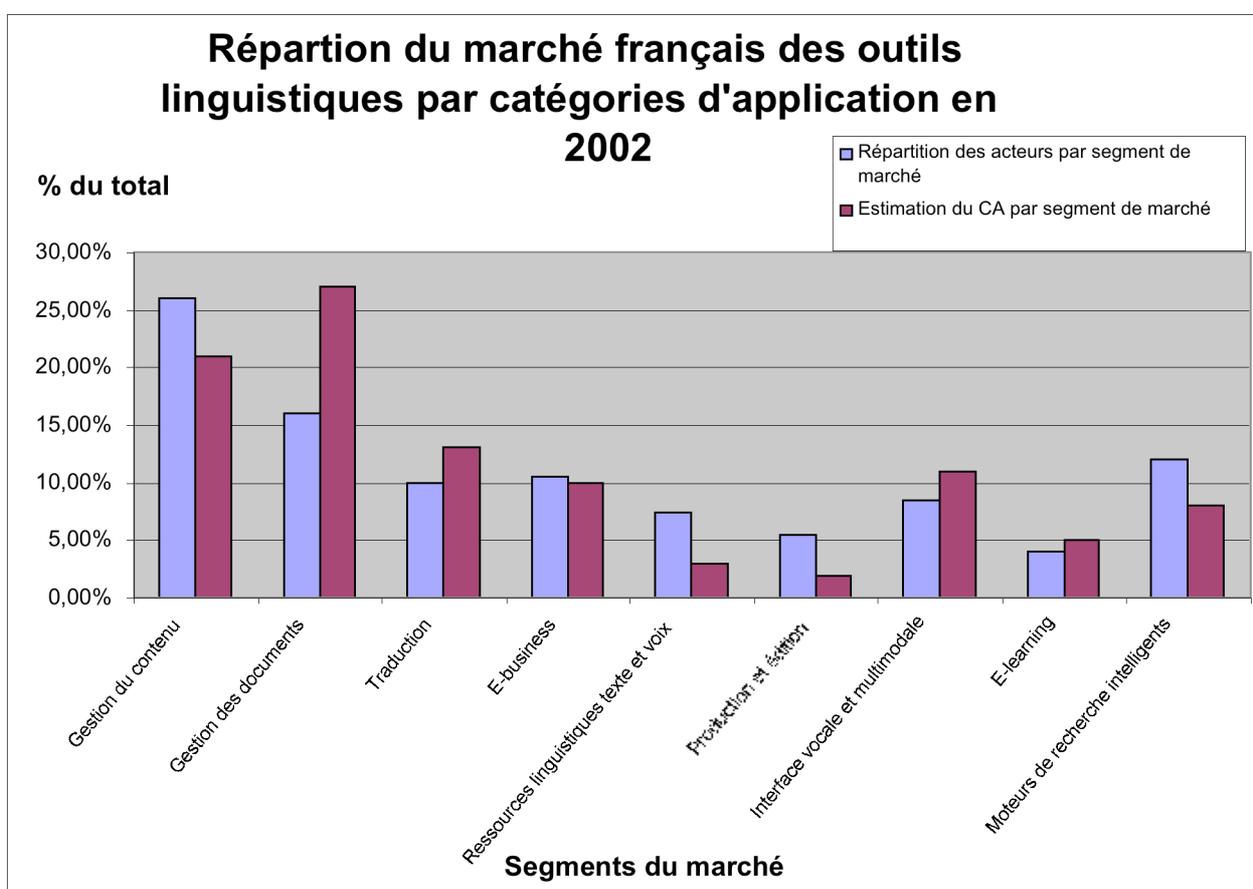
Les pays *leaders* du marché des outils linguistiques sont donc le Royaume-Uni, la France, l'Allemagne et l'Italie, représentant à eux quatre près de 60% du marché européen. La place de *leader* du Royaume-Uni peut s'expliquer par la place prédominante du traitement vocal dans ce pays.

Les 5 premiers acteurs (*Bowne Global Solutions* en Irlande, *Eckoh Technologies UK*, *Autonomy UK*, *Aculab UK*, et *Scansoft* en Belgique) représentent plus de 40% de la valeur dégagée par les 20 premières sociétés européennes du

marché des outils linguistiques. Parmi ces 5 *leaders*, 4 sont spécialisés sur le segment du traitement vocal : *Bowne Global Solutions*, *Eckoh Technologies*, *Aculab*, et *Scansoft*.

En France, 96 sociétés ont pu être identifiées dans le périmètre de l'ingénierie linguistique : le chiffre d'affaire dégagé représente environ 77 millions d'euros, soit environ 15% du marché européen et la France est le deuxième pays *leader* après le Royaume-Uni.

En se basant sur les *leaders* du marché en France, nous pouvons établir la prédominance du traitement du texte sur le marché français : près de 85% des sociétés *leaders* du marché français sont positionnées sur le texte, qui est encore aujourd'hui le moteur de la valeur du marché des outils linguistiques en France, mais les revenus générés par la voix sont en expansion importante.



Environ 40% des acteurs identifiés sont positionnés sur un seul segment d'application.

Les sociétés les plus présentes sur le marché des outils linguistiques sont celles traitant la gestion du contenu (26%), la gestion des documents (16%), les moteurs de recherche intelligents (12%), puis les spécialistes du e-business (10,5%), les sociétés de traduction par ordinateur (10%), et les sociétés traitant la voix (8,5%).

En terme de potentiel de CA, les sociétés de GED (27%) et de gestion de contenu (21%) représentent près de 50% du marché en 2002.

Le marché français reste concentré : les 23 premières sociétés du marché français des outils linguistiques représenteraient 65% du CA, les 38 suivantes seulement 20%. Le marché est donc concentré du fait, d'une part, de la position dominante de quelques *leaders* et, d'autre part, de la dispersion des acteurs les plus petits ou nouvellement arrivés.

3.2. ETAT DE LA DEMANDE ET TYPOLOGIE DES APPLICATIONS DEPLOYEES.

La demande a été évaluée selon trois approches :

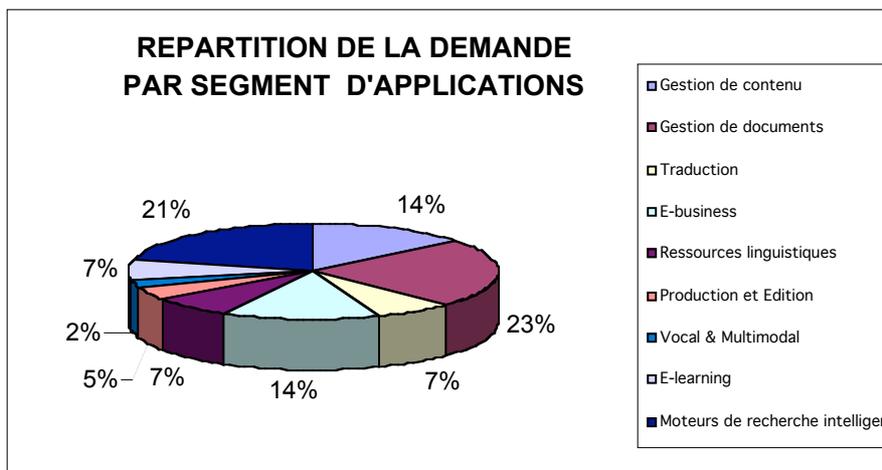
- Caractérisation des références fournies par l'offre ;
- Enquête ciblée auprès de 80 organismes ou entreprises représentatifs des segments de l'offre ;
- Etude détaillée de 20 organismes ou entreprises.

680 références de la demande ont été sélectionnées pour 96 acteurs de l'offre.

Les secteurs d'activités les plus représentatifs de la demande sont : l'Industrie y compris Pharmacie et Santé (36%), les Services / Banque Assurance Finances (20%), le Public (15 %) et les Transports/Tourisme (11%).

Certains secteurs sont clairement demandeurs d'applications avec des outils linguistiques sur des segments particuliers. Par exemple, le secteur Public est très concerné par les applications relatives au *e-learning* et, l'Industrie par les applications de Production et d'Edition, les ressources linguistiques et la traduction automatique.

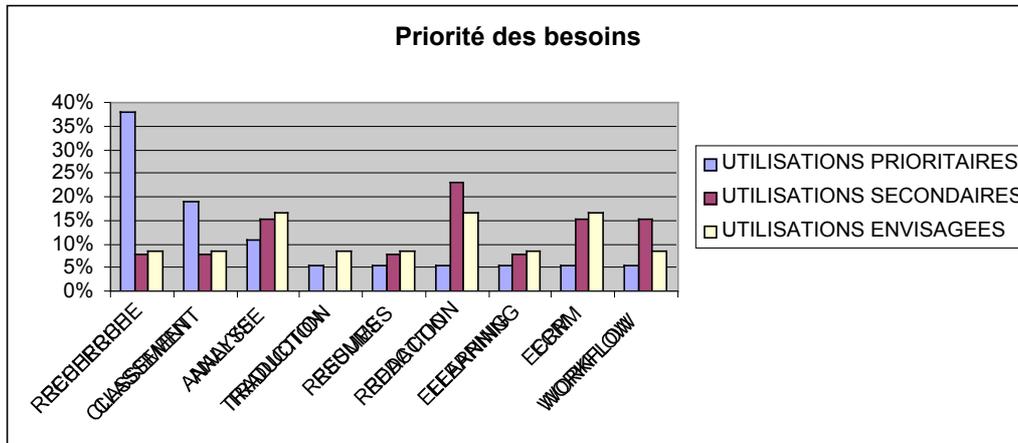
Les sociétés du CAC 40 et la plupart des administrations publiques représentent la majorité des entités demandeuses de solutions de traitement du langage. La répartition de la demande exprimée par segment d'applications se présente ainsi :



La répartition de la demande associée aux solutions mises en place est représentée par des applications dédiées à des groupes d'utilisateurs spécifiques au sein des sociétés : 52% des utilisations concernent une cible utilisateur de 100 à 1000 personnes, 12 % une utilisation pour plus de 1000 personnes. Les applications sont de plus en plus orientées réseau (Internet, Intranet, Extranet) avec des architectures client/serveur. Elles ciblent un groupe d'utilisateurs de plus en plus large, même si encore 18% des applications mises en place concerne un nombre restreint d'utilisateurs (nombre d'utilisateurs inférieurs à 100) lié à l'étroitesse et à la spécialisation de la cible.

Les applications mises en place en univers industriel (hors usage des particuliers) impliquent un processus décisionnel au sein de l'entreprise variant de 12 à 18 mois en moyenne. Plus de 60% des applications déployées datent de moins de 5 ans. Le marché de la demande est encore jeune dans la mesure où les outils déjà mis en place ne permettent pas de capitaliser intégralement la maîtrise de l'information dans l'entreprise.

En ce qui concerne les évolutions et les attentes de la demande, elles sont liées pour 82% à l'accroissement du volume des données à traiter et de la montée en puissance des réseaux. Ce qui implique que près de 35% de la demande est intéressé par les systèmes de résumés automatiques, et 41% par la veille automatisée sur Internet.



3.3. EVOLUTION DU MARCHÉ EUROPEEN DES TECHNOLOGIES DE LA LANGUE.

- **Les années 1990-2000 : de la logique de l'offre vers la logique de la demande**

Les années 1990 ont marqué l'avènement de la société de l'information où le traitement du texte est passé dans une phase opérationnelle, et où le traitement de l'image et de la voix apparaît : l'arrivée de nouveaux supports de communication, tels que les téléphones portables, les ordinateurs portables, les bornes interactives, les systèmes embarqués, participe au développement industriel des applications. Les connexions haut débit (connexion réseau câble et satellite) facilitent l'accès à l'information. L'offre devient de plus en plus concurrentielle avec une multitude de nouveaux acteurs : plusieurs *spin off* en provenance de centres de recherche publics émergent sur les 9 segments d'application définis. Les pôles de recherche et les programmes européens sont des soutiens à la R&D et à l'industrialisation du secteur.

Les années 2000 marquent l'avènement de la société de la communication avec un traitement du texte, de l'image, de la voix, de la vidéo en phase industrielle. La convergence des technologies est fortement marquée dans le domaine de la communication. L'informatique, les télécommunications et l'audiovisuel sont fédérés par la numérisation. Le contenu numérique concerne désormais la voix, le son et les images : l'ère du multimédia se concrétise avec le déploiement de kiosques multimédias, des téléphones mobiles de nouvelle génération, des consoles et des terminaux de divertissement, de la domotique. Les connexions aux réseaux à haut à débit se généralisent (ADSL, fibre optique, WIFI...).

Les applications concernent les intranets documentaires multimédia, la gestion de contenu multilingue, le *e-business*, le vocal/multimodal, le *e-learning*, la traduction automatique... En plus des sociétés du CAC 40, les clients types sont les FMN (Firmes multinationales), les PME orientées *e-services*, le secteur Public.

La dynamique du marché s'oriente désormais vers une logique de la demande : intégration des outils au sein d'architectures existantes, capitalisation de l'investissement, demande permanente de l'évolution des technologies avec la généralisation du multimédia et de l'Internet mobile, et le besoin de la communication multilingue.

- **L'intensité concurrentielle :**

Les perspectives d'évolution du marché des outils linguistiques dépendent avant tout de son intensité concurrentielle qui est de plus en plus forte : l'évolution de l'offre, de la demande et de l'environnement technologique engendre des mouvements de concentration du marché, une mise en concurrence systématique des technologies et des outils disponibles.

L'offre

Depuis ces dix dernières années, les fusions et acquisitions au sein du marché sont de plus en plus fréquentes et s'expliquent par :

- la frénésie du marché de l'Internet dans les années 1990 avec la course au premier entrant qui a permis de développer des partenariats et de valider des fusions entre acteurs concurrents ;
- le marasme économique du secteur des NTIC dans le début des années 2000 avec l'éclatement de la bulle Internet et la perte de valeur de plusieurs acteurs de l'offre ;
- le positionnement stratégique multinational des offreurs de technologie sur un marché des outils linguistiques qui devient global (interconnexion des acteurs européens et américains), aujourd'hui concrétisé par la pénétration des acteurs américains sur le marché européen ;
- la volonté de contrôle des technologies par la demande, enclin à racheter les sociétés qui ont développé des outils de gestion de connaissance dans un domaine hautement stratégique.

La demande

L'accès à l'information s'oriente vers des solutions en langage naturel qui autorisent l'interrogation multilingue dans des bases plurilingues. La diffusion de l'information avec les technologies associées a engendré l'éclatement de la fonction « Gestion de l'Information » : les systèmes de gestion sont orientés vers les utilisateurs finals et non plus ciblés vers les seuls experts de la documentation. L'éparpillement de la demande pose à l'offre le problème d'identification des besoins et des attentes d'une même communauté d'utilisateurs.

CONCLUSION

L'offre est diffuse et disparate : la taille des offreurs de technologie est très variée, et leur représentativité nationale n'est pas toujours facile à identifier.

De plus, d'une année sur l'autre, le périmètre est flottant voire turbulent : plusieurs sociétés disparaissent (liquidation, rachat ou fusion), d'autres se créent (*spin off* de sociétés grands comptes, de centres de recherche, *start up*). La comparaison d'une année à l'autre est donc délicate.

En France et en Europe, 95% des offreurs de technologie estiment que la croissance du marché est de 10 à 15% par an, segments texte et voix confondus. L'évolution des investissements en R&D orientés outils linguistiques est plus pessimiste : l'évolution moyenne des dépenses des acteurs européens recensés dans l'étude baisse sensiblement entre 2000 et 2002.

Le marché mondial des technologies de la langue devrait dépasser les 2 milliards d'euros en 2005 et dépasser les 3 milliards d'euros en 2007. Le marché européen devrait atteindre 1 milliard d'euros en 2007.

La progression de l'usage des NTIC laisse présager un élargissement du marché des outils linguistiques vers le grand public. Le besoin est ressenti d'engager des actions *marketing* pour optimiser l'adéquation de l'offre à la demande.

Bureau van Dijk Information Management
www.bvdim.com

Martine DEJEAN : mde@bvdim.com
Antoine RAULIN : ara@bvdim.com