

**Claire BLANCHE-BENVENISTE**

Université de Provence  
EPHE Paris

## **Aspect lexical de la confrontation entre langues romanes Existe-t-il un lexique européen ?**

On peut dire qu'il existe actuellement, dans les textes écrits d'Europe occidentale, une sorte de *lexique européen*. Le plus frappant est sans doute cette part de lexique commun hérité du latin, tel qu'on le retrouve dans les langues romanes d'aujourd'hui comme le portugais, l'espagnol, le français ou l'italien et qui a été longuement étudié par les romanistes. Mais le partage va plus loin ; il est évident, par exemple, que les journaux et les magazines européens utilisent les mêmes mots techniques et scientifiques pris au fonds gréco-latin, les mêmes emprunts à l'anglais et, au-delà du lexique lui-même, beaucoup de tournures grammaticales, de procédés rhétoriques et d'argumentations très semblables. Des spécialistes de la typologie des langues ont même forgé récemment le terme de *Standard Average European* pour désigner ce phénomène linguistique d'*européen standard moyen* (Decsy 2000, Faarlund 1995, Haspelmath 1998, 2001, Ramat 2000, Ramat & Stolz 2002).

La part importante de lexique commun devrait faciliter la compréhension mutuelle des langues en Europe, au moins sous leur version écrite et au moins à l'intérieur d'un même groupe de langues. On pourrait imaginer que ce lexique commun permettrait à tout le monde de comprendre une bonne partie des textes qui circulent, au moins dans des langues apparentées à la langue maternelle. Il n'en est rien. Seuls des « initiés » y parviennent : par exemple des personnes qui ont étudié sérieusement le latin, ou d'autres qui, sans être spécialistes, connaissent un peu la linguistique des langues romanes, germaniques ou scandinaves ; ou des personnes qui, maîtrisant déjà deux ou trois de ces langues, trouvent très banal d'en absorber une nouvelle. « Très

facile », disent ces initiés. « Insurmontable », disent les profanes. Il y a là un curieux paradoxe.

Ce paradoxe nous amène à poser quelques questions fondamentales : Quel est le rôle du lexique dans la compréhension d'un texte ? Que veut dire *comprendre* quand il est question de comprendre un mot ou de comprendre un texte ? Que veut dire *comprendre* quand il s'agit de lire différentes sortes de textes, dans sa propre langue ou dans une langue étrangère ? Quel rôle joue le lexique dans toutes ces opérations ? Et quel rôle joue plus précisément ce *lexique européen*, qui serait disponible pour les initiés, et qui n'est pourtant pas accessible aux profanes ? En observant expérimentalement ces diverses questions, on voit qu'il faut distinguer entre plusieurs sortes de connaissances du lexique : connaissance active et connaissance passive ; connaissance par écrit ou par oral ; connaissance exacte ou approximative ; connaissance purement déclarative (celle que transmettrait un enseignant qui expliquerait la parenté des langues) et connaissance procédurale qui permet de s'approprier ce lexique.

Je tenterai d'apporter quelques éléments de réponse à ces questions, en m'appuyant sur une expérience d'intercompréhension des langues romanes que nous avons menée avec des adultes, dans quatre pays européens, dans les années 1990 (Blanche-Benveniste et alii 1995, Blanche-Benveniste et Valli, 1997). J'utiliserai également des observations accumulées en étudiant la production et le maniement du lexique dans la langue maternelle (plus précisément en français parlé).

Certaines des notions que j'utiliserai ont été mentionnées dans ce colloque par plusieurs collègues. Le professeur De Mauro a insisté sur la nécessité qu'ont toutes les langues de parler de l'inconnu et sur l'utilité de la notion de compréhension « vague » pour y parvenir. Rosaria Solareno a décrit, à partir de ses propres observations, la démarche qui consiste à « deviner » ce qu'on doit comprendre. Le professeur Banfi et ses collaborateurs ont montré quelles sont les stratégies utilisées pour palier le manque de mots. Grazia Basile a utilisé la notion de « script » qui implique une « connaissance partagée » par les usagers de la langue. Nous sommes donc en terrain connu.

Je parlerai en premier lieu de l'opposition entre démarche d'initiés et démarche de profanes dans la compréhension du lexique des langues romanes et ensuite de phénomènes qui montrent différents degrés de « compréhension du lexique » dans la langue maternelle parlée ; je mentionnerai quelques usages actuels du terme d'*intercompréhension* ; enfin je proposerai un petit bilan sur la relation entre le rôle du lexique et le rôle des structures linguistiques dans la construction du sens d'un texte.

## **1. Le lexique des langues romanes**

Tous ceux qui ont suivi des études classiques de romanistique ont appris qu'il existe, entre les lexiques des langues romanes actuelles, des relations systématiques permettant de passer de l'une à l'autre, à partir d'un modèle latin commun. Cette ancienne démonstration a été constamment revivifiée au cours des siècles et elle a été récemment modernisée et présentée dans la perspective de différentes théories linguistiques (AGARD 1984, SALA 1987, HOLTUS, METZELTIN & SCHMITT 1988, REINHEIMER et TASMOWSKI 1997, RENZI & ANDREOSE 2003, TEYSSIER 2004).

### 1.1. Le lexique des langues romanes vu par les romanistes

Comme tous les spécialistes de la comparaison des langues, les romanistes ont fondé leur démarche sur les ressemblances formelles entre les mots des différentes langues étudiées. Ils ont présenté des tableaux montrant les correspondances régulières entre les phonèmes et les groupes de phonèmes de plusieurs langues romanes. Par exemple, il est bien connu que, aux mots latins commençant par les groupes de consonnes pl-, cl-, fl-<sup>1</sup>, le français répond systématiquement par pl-,cl-,fl-, l'italien par pi-,chi-,fi-, le portugais par ch- et l'espagnol par ll-

latin	P	E	I	F
plenum	cheio	lleno	pieno	plein
clavem	chave	llave	chiave	clé
flamma	chama	llama	fiamma	flamme

Ces ressemblances formelles permettent à quiconque a étudié la comparaison des langues romanes de reconnaître facilement un même radical latin *plen-*, *clav-* ou *flamm-*, sous les mots portugais *cheio*, *chave*, *chama* et donc d'en connaître immédiatement le sens. On peut ainsi « déchiffrer » toute une partie du vocabulaire des langues romanes, même si l'on n'est pas capable d'en produire activement les mots.

Les ressemblances formelles sont plus fortes entre les langues qui appartiennent à certains sous-groupes, le sous-groupe des langues ibériques ou le sous-groupe français-italien. D'un sous-groupe à un autre, le passage est parfois moins régulier. Pour un Portugais qui possède le mot *cerveja* (« bière »), l'espagnol *cerveza* est absolument transparent, alors que les équivalents italiens et français, *birra* et *bière*, sont pour lui opaques. De la

<sup>1</sup> Pour des raisons de commodité, et pour mieux faire le lien avec l'expérience d'*Eurom4*, je cite ici tous les mots sous leur forme graphique usuelle.

même façon, à partir du français *manger*, il est facile d'identifier l'italien *mangiare*, alors que les équivalents du groupe ibérique, *comer* en portugais et en espagnol, ne sont pas transparents, sauf si l'on sait faire le détour par le dérivé plus « savant », *comestible* ; mais, pour savoir faire ce type de détour, il faut déjà être un bon initié !

Autre détour à maîtriser, celui des mots de la langue maternelle qui existent dans des usages plus marginaux que les équivalents des autres langues; c'est le cas par exemple des mots français *péril*, *clos*, qui correspondent exactement dans leur forme à l'italien *pericolo* et *chiuso*, mais à un niveau de bien moins grande disponibilité :

P	E	I	F
cerveja	cerveza	birra	bière
achar	hallar	trovare	trouver
comer	comer	mangiare	manger
mesa	mesa	tavola	table
perigo	peligro	pericolo	danger (péril)
fechado	cerrado	chiuso	fermé (clos)

Celui qui maîtrise des relations entre mots de base et dérivés dits « savants », sait d'emblée diviser le lexique en couches différentes, avec des « transparences » différentes. Le dérivé *oculiste* est absolument transparent pour toutes les langues romanes (et bien au-delà des langues romanes) , alors que les mots « de base », *olho*, *ojo*, *occhio*, *œil*, le sont beaucoup moins :

port	esp	ital	fran
olho	ojo	occhio	oeil
oculista	oculista	oculista	oculiste

Il existe un domaine important du vocabulaire des langues romanes vivantes pour lequel on pourrait dire qu'il y a une totale transparence : termes formés par décalque des mots latins ou grecs, sans presque aucune trace des évolutions historiques particulières qui ont par ailleurs affecté chacune des langues, comme *parlamento/parlement* :

parlamento	parlamento	parlamento	parlement
memória	memoria	memoria	mémoire
catálogo	catálogo	catalogo	catalogue

Comme les pays d'Europe ont pour la plupart quantité d'institutions et de caractéristiques culturelles en commun, ce vocabulaire occupe actuellement une très grande place. En consultant de grands dictionnaires, on s'aperçoit que, pour les lexiques techniques, les langues d'Europe ont en partage plus de 5.000 mots ( Hagège 1992). Les radicaux sont généralement strictement les mêmes. Ce qui, par écrit, rend ces mots différents les uns des autres, ce sont seulement les finales, certains accents graphiques, quelques particularités d'orthographe et parfois la forme de certains préfixes et suffixes :

personagem	personaje	personaggio	personnage
desagradável	desagradable	sgradevole	désagréable

Ces petites différences ne gênent pas les initiés, qui n'y voient que des phénomènes secondaires et qui estiment que ce vocabulaire est absolument transparent. Il ne l'est pourtant pas pour les profanes, qui risquent d'être troublés par des différences purement orthographiques. Nous avons vu des Français hésiter à reconnaître le mot italien *psicologia*, auquel ils trouvaient que manquaient les lettres grecques du français *psychologie*, et des Italiens gênés par la forme graphique du français *théâtre*, qui, pour eux avait vraiment trop de diacritiques, par rapport au simple *teatro* auquel ils étaient habitués. La connaissance du lexique, tel qu'il est écrit, suppose que l'on sait traiter ces différences.

Pour disposer de ces diverses sortes de connaissance du lexique, il n'est même pas nécessaire d'avoir étudié la source historique qu'est le latin ; celui qui connaît trois ou quatre langues romanes actuelles peut aisément trouver le système de correspondances qui lui donnera accès à la cinquième ou sixième qu'il rencontrera. Avant même d'aborder une langue nouvelle, les initiés – qu'ils se soient formés par la discipline historique du latin ou par la pratique actuelle des langues vivantes – disposent donc d'une sorte de réserve de lexique sur laquelle ils peuvent s'appuyer pour aller plus avant dans leurs connaissances.

## **1. 2. Le lexique roman vu par les non-initiés : l'expérience d'Eurom4**

### 1. Les conditions de l'expérience

Nous avons lancé l'expérience *EuRom4*, dans les années 1990, avec l'idée qu'il devait être possible de faire partager à des profanes certaines des connaissances qui permettent de passer d'une langue romane à une autre. Nous nous sommes adressés à des adultes (et non à des élèves captifs d'un système d'enseignement) qui se déclaraient volontaires pour tenter une expérience qui devait les habiliter à lire des journaux dans trois autres langues romanes que la leur. Il était bien précisé qu'il s'agissait seulement de

comprendre et non de produire (parler ou écrire une langue nouvelle est une aventure bien différente), qu'il s'agissait de comprendre seulement les langues écrites et non les langues parlées (percevoir oralement des langues étrangères est une tout autre aventure) et que l'expérience serait restreinte à la lecture de textes faisant appel à un savoir commun largement partagé en Europe : sujets de société, politique étrangère, mais rien qui aurait exigé une connaissance préalable de contenus culturels nouveaux.

Nous avons présenté aux participants des articles authentiques, auxquels nous avons seulement ajouté des titres plus explicites que dans les originaux (afin que le contenu soit plus facile à prévoir). La longueur des articles a progressé au cours des séances, passant d'une dizaine de lignes pour les premiers à une centaine de lignes pour les derniers. Les articles présentés par écrit étaient accompagnés d'une lecture à haute voix par un natif de la langue<sup>2</sup>. Les participants abordaient les trois langues dans la même séance, à raison de vingt minutes environ pour chacune. La consigne était de « traduire » successivement chacune des phrases, en remplaçant chaque mot inconnu par un « fantôme », *machin*, *truc* ou *chose*. En fait, il n'agissait pas vraiment de traduire, mais de fournir une interprétation avec un équivalent dans sa langue maternelle, avec éventuellement différents degrés de compréhension. La même expérience a été menée dans quatre pays (Portugal, Espagne, Italie, France).

## 2. Les hypothèses

Nous faisons l'hypothèse que les participants pouvaient devenir des *initiés* dans une période assez courte. Comme ils étaient tous de bons lecteurs dans leur langue maternelle, nous pensions qu'ils pouvaient transférer à une langue étrangère les procédés de lecture qu'ils pratiquaient déjà sans le savoir dans leur langue maternelle, en particulier celui qui consiste à « sauter » par-dessus les mots inconnus pour faire du sens à partir de l'ensemble d'un texte, en utilisant des séries d'« inférences » (Hosenfeld 1984). Nous faisons également l'hypothèse qu'ils pouvaient déclencher rapidement des procédés intuitifs de comparaison entre les langues romanes (en y intégrant leur propre langue), qu'ils pouvaient fabriquer des systèmes d'équivalences et construire ainsi un socle de connaissances lexicales à partir duquel ils pouvaient s'émanciper. Nous ne savions pas comment ils allaient mémoriser ces acquisitions, ni comment ils allaient traiter les relations entre le lexique et la grammaire des langues qu'ils exploraient.

---

<sup>2</sup> Le volume de 800 pages édité par Nuova Italia est accompagné d'un CD-Rom qui contient, entre autres choses, les enregistrements de lectures de tous les articles.

### 1.2.3. Les étapes de l'apprentissage

Les participants, dans leur ensemble, sont passés par trois grandes étapes.

La première étape a duré environ six à dix séances. Au début, les participants déclaraient qu'ils ne comprenaient rien aux textes et qu'ils n'y comprendraient jamais rien. Certains ont mentionné les difficultés qu'ils avaient toujours ressenties dans leurs apprentissages scolaires (« J'ai passé l'âge de me faire punir parce que je ne connais pas les verbes irréguliers italiens »). Ils ne s'intéressaient qu'au vocabulaire, sans aucune préoccupation pour la grammaire ni pour le contenu. Ils semblaient angoissés de n'avoir aucune réserve lexicale pour aborder les textes comme nous les y invitons. Ils ne s'intéressaient absolument pas aux renseignements que nous pouvions leur donner sur les ressemblances formelles entre les mots : la leçon des romanistes n'était, à cette étape, d'aucune utilité. Beaucoup se méfiaient des mots les plus *transparents* et craignaient d'utiliser tout procédé consistant à *deviner* sans être sûr du résultat. Ils auraient voulu consulter un dictionnaire ou, à défaut, ils auraient souhaité que nous leur donnions le sens des mots inconnus, ce que nous avons toujours refusé de faire.

La deuxième étape a pris place entre la dixième et la vingtième séance environ. Les participants avaient mémorisé à peu près cent à deux cents mots fréquents. Dès que ce stock a été constitué, ils ont commencé à s'intéresser à d'autres aspects que le vocabulaire, par exemple un peu à la forme des mots, davantage à certains détails de la grammaire et de plus en plus aux ressemblances et différences entre les langues.

La troisième étape a occupé le reste des séances (de vingt à vingt-cinq ou trente). Les participants se sont intéressés aux relations entre les mots et à la grammaire. Ils ont discuté sur ce qu'ils étaient en train de faire et sur les différentes formes de compréhension qu'ils découvraient, des plus approximatives aux plus exactes. Plusieurs se sont émerveillés de constater que, d'une façon ou d'une autre, ils commençaient à « tout comprendre », comme ils le disaient. Ils ont alors commencé à admettre qu'on pouvait comprendre sans pouvoir pour autant fournir de « belles traductions ». A la fin de l'expérience, ils pouvaient *comprendre* des articles de presse assez longs, pour lesquels il leur manquait seulement cinq ou six mots.

### 1.2.4. Le recours systématique à l'inférence

Lorsque nous observons ces participants qui commencent à lire dans une langue romane étrangère, qui avaient reçu la consigne de deviner les mots non transparents, différents mécanismes de compréhension lexicale nous apparaissent. Placé devant une phrase tirée d'un article en portugais, le non-

initié prend appui sur tous les mots transparents qu'il rencontre : *interrupção* est très proche du français *interruption*, *voluntária* de *volontaire* et *legalizada* de *légalisée*. Seul le mot *gravidez* est ici opaque. D'où cette première interprétation :

« A interrupção voluntária da <i>gravidez</i> foi legalizada em 1973 », « L'interruption volontaire de <i>machin</i> fut légalisée en 1973 »
---

Dans une deuxième phase, deux mécanismes se mettent en route. Le premier consiste à activer une association faite de plusieurs mots, *interruption volontaire de grossesse*, bien connue en français. Le mot *gravidez*, qui ne pouvait pas être compris à l'état isolé, est un candidat très plausible dans ce contexte (dans une étape ultérieure, les participants seront en état de faire le rapprochement paradigmatique avec des termes techniques comme *gravide*, s'ils les connaissent). Il s'agit de reconstituer une expression qui fonctionne comme un stéréotype. Le deuxième mécanisme est celui du *cadre* (au sens de *frame* que propose Fillmore 1985). Tout Européen peut faire l'hypothèse suivante : une disposition législative, concernant un processus d'interruption volontaire, appartient certainement aux lois adoptées dans les années 1970 par certains pays européens pour l'avortement ; ce qui confirme que *gravidez* veut dire grossesse. Dans d'autres cas, lorsque les participants ne connaissent pas encore le cadre culturel dans lequel se place le texte (mouvements racistes en Allemagne, par exemple), le seul recours au stéréotype peut suffire pour interpréter *superato il limite* en tant que *dépassé la limite* :

- |   |
|---|
| 1. La violenza racista <i>ha superato</i> in Germania il limite<br>2. la violence raciste a [machiné] en Allemagne la limite<br>3. La violence raciste a dépassé la limite en Allemagne |
|---|

On voit que le procédé respecte le schéma grammatical, sujet + verbe + complément, mais que la tête syntaxique de la construction, le verbe, peut rester lexicalement opaque pendant la première phase d'interprétation, sans que cela empêche une compréhension ultérieure.

Lorsque l'inférence doit se faire seulement sur la base du cadre de connaissances partagées, l'opération est plus délicate. Dans l'exemple suivant, le mot espagnol *seres* restait opaque pour les participants français ; aucun stéréotype ne pouvait les aider, parce que le contexte, *trafic illégal de... humains*, pouvait être complété par plusieurs éléments (drogues ? organes ? armements ?). Il a fallu laisser ce mot opaque et avancer pour voir que le texte traitait du trafic d'émigrants entre le Maroc et l'Espagne et qu'il pouvait s'agir de *trafic illégal d'êtres humains* ; d'où l'interprétation rétroactive : *seres* = *êtres*. Ce n'est que plus tard encore, à l'occasion d'un autre texte, qu'il est apparu que *seres* était le pluriel de *ser*.

1. Una de las principales organizaciones que se dedican a este tráfico ilegal de *seres* humanos opera en la plaza de Feddan
2. une des principales organisation qui se livrent à ce trafic illégal de *machins* humains opère en la place de Feddan
3. Une des principales organisations qui se livrent à ce trafic illégal d'*êtres* humains opère *sur* la place de Feddan

Même effet rétroactif par le cadre général du texte pour interpréter l'italien *pazzo* (remarquons aussi que, entre la troisième et la quatrième interprétation, la qualité de la traduction s'améliore grâce à l'adoption d'un stéréotype, *passer pour fou*):

1. Cristoforo Colombo fu preso par *pazzo* 500 anni fa
2. Christophe Comb fut pris pour *machin* il y a 500 ans
3. Christophe Colomb fut pris pour *fou* il y a 500 ans
4. Christophe Colomb *passa pour fou* il y a 500 ans

Reconstruire le cadre sémantique général est facile lorsqu'il s'agit de connaissances communes partagées et de *scripts* prévisibles, comme il y en a dans les rubriques journalistiques traitant de thèmes internationaux. L'opération est moins facile pour les rubriques de politique intérieure de chaque pays. Voici trois exemples tirés de la rubrique *Internacional* du *Diário de Notícias* de Lisboa, dans lesquels les mots opaques pour un Français débutant, *apoio*, *nega*, *ficar*, peuvent aisément être interprétés par inférence :

1. Europa reforça apoio aos críticos de Lukachenko.
2. Berlusconi nega manobra para dividir patronato italiano.
3. Tropas americanas podem ficar no Iraque até 2009.

En revanche, dans la rubrique *Interior* du même journal, il peut paraître plus difficile de comprendre que *poupar* veut dire *épargner* :

Veja como poupar no Seguro Automóvel

Comme il n'est pas possible de faire des inférences très rentables pour les connecteurs, ce sont presque toujours, dans toutes les langues, des mots particulièrement opaques et difficilement mémorisables :

*De sorte que, ainsi que, afin que*

### 1.2.5. Relation entre transparence, fréquence et mémorisation

Le déchiffrement des mots par inférence a une vertu : il facilite la mémorisation. Les participants se souviennent des mots qu'ils ont atteints au terme d'une recherche personnelle bien mieux que si la solution leur est donnée par quelqu'un d'autre. Tout se passe comme si le chemin fait par atteindre le mot l'ancrait dans la mémoire.

La fréquence d'occurrence facilite aussi la mémorisation. Les mots qui reviennent à plusieurs reprises dans un même texte ou dans des textes différents sont stockés plus facilement (à l'exception des connecteurs). Il est bien connu que les 100 à 200 mots les plus fréquents fournissent un stock qui permet d'absorber une bonne partie des textes. Voici les statistiques que rappelait P. Teyssier (2004 :57) :

Mots les plus fréquents	pourcentage des mots dans un texte
Les 100 plus fréquents	60%
Les 1.000	85%
Les 4.000	97,5 %

C'est ce qui explique que, à l'issue de la première étape de l'expérience *Eurom4* pendant laquelle ils ont accumulé un petit stock lexical, les participants se soient sentis à ce point rassurés dans leur démarche.

Les mots les plus fréquents sont en partie transparents d'une langue romane à une autre, en particulier les verbes. Voici une comparaison entre les vingt verbes les plus fréquents en italien et en français, par écrit et par oral (Cresti-Moneglia 2005 : 105-126):

	Italien écrit	Italien parlé	Français écrit	Français parlé
1	Fare	fare	Etre	être
2	Dire	Dire	Avoir	Avoir
3	Venire	Andare	Pouvoir	Faire
4	Andare	Vedere	faire	aller
5	Vedere	Sapere	Dire	dire
6	Stare	Stare	Devoir	pouvoir
7	Sembrare	Venire	Donner	voir
8	Mettere	Capire	Voir	savoir
9	Chiedere	Mettere	Prendre	vouloir
10	Parlare	Dare	Aller	falloir

11	Trovare	Partire	Mettre	passer
12	Prendere	Sentire	Savoir	prendre
13	Diventare	Arrivare	Permettre	mettre
14	Entrare	Prendere	Vouloir	venir
15	Continuare	Pensare	Falloir	parler
16	Prevedere	Parlare	Venir	arriver
17	tenere	Guardare	Trouver	Penser
18	Decidere	Trovare	Tenir	trouver
19	Spiegare	portare	Concerner	devoir
20	Pubblicare	Crede	Agir	donner

Mais les noms, les adjectifs et les adverbes les plus fréquents sont en revanche souvent opaques d'une langue à une autre et ils demandent un effort spécifique de mémorisation. Pour un Français qui n'a pas étudié le latin, les mots qui désignent la « femme » en espagnol et en italien, *mujer*, *moglie*, ou la « jeune fille » en portugais et en italien, *rapariga*, *ragazza*, ne sont pas transparents et ne peuvent généralement pas être décryptés par inférence. Pour les usagers des autres langues romanes, les mots français *vite*, *deuxième*, le verbe *il faut*, qui se rencontrent si souvent dans les textes, ne sont aucunement transparents.

Au total, l'expérience *Eurom4* nous a montré qu'il existe diverses approches du lexique d'une langue étrangère et que, faute d'observations *in vivo*, les formes d'enseignement habituelles en tiennent rarement compte. C'est pourquoi, sans doute, les professeurs de langues vivantes n'ont pas beaucoup apprécié notre méthode. Ils ont été surtout choqués par le fait que nous acceptions la notion de « connaissances approximatives ». Les participants avançaient en effet dans les textes à coups d'approximations et, même lorsqu'ils étaient très heureux d'avoir apparemment *tout compris*, il restait encore une bonne part d'approximation dans la phase finale de leurs interprétations, qui n'étaient pas de *belles traductions*. Comme ce point me paraît capital, je voudrais l'illustrer avec quelques réflexions sur l'approximation dans la langue maternelle et dans le phénomène d'*intercompréhension* entre langues (Blanche-Benveniste 2005, 2006).

## **2. Les connaissances approximatives dans la langue maternelle**

### **2.1. Les inférences du « bon lecteur »**

La notion de *bon lecteur* a été étudiée surtout par les pédagogues et par les auteurs intéressés aux calculs de lisibilité (BOYER 1992, FAYOL et al., 1992, GELINAS & PREFONATINE 1999, FUKKINK and al. 2001) . Un bon lecteur dans sa langue maternelle sait *comprendre* un texte sans se laisser arrêter par des mots qu'il ne connaît pas. Il ne procède pas « mot par mot » mais par grands ensembles. Il pratique des inférences à divers niveaux avant de parvenir à *deviner* une partie des mots inconnus. C'est du reste ainsi que, nous autres, bons lecteurs, nous apprenons du vocabulaire nouveau. Le bon lecteur tolère donc, pendant un temps plus ou moins long, d'avoir des zones de compréhension *approximative*. Voici quelques exemples.

Dans un article de vulgarisation scientifique, trois termes techniques peuvent arrêter un bon lecteur, s'il n'est pas spécialiste du domaine : *enzyme, silice, lipases*. Est-ce à dire qu'il ne comprend pas du tout le texte ?

L'activité de l'enzyme peut être optimisée en jouant sur la nature chimique de la silice comme le montre l'exemple des lipases

Une forme de compréhension approximative est disponible, si l'on fait abstraction des mots opaques :

L'activité de l'..... peut être optimisée en jouant sur la nature chimique de la....., comme le montre l'exemple de .....

On peut en donner une version lexicalement « complète », en dissociant lexique et syntaxe, si l'on remplace les vides lexicaux par des pronoms (Quine 1960):

L'activité de celui-ci peut être optimisée en jouant sur la nature chimique de quelque chose, comme le montre l'exemple des autres.

La désignation des référents a disparu, mais il reste dans la phrase une forme incontestable de compréhension, qui peut se préciser soit dans la suite du texte, soit en questionnant un spécialiste. De la même façon, il est possible de « réduire » la difficulté de compréhension de *ubiquitaires* et *eukaryote* dans l'exemple suivant en passant provisoirement de la version (1) à la version (2) :

- (1) Les moteurs linéaires sont *ubiquitaires* dans la vie *eukaryote* .
- (2) Les moteurs sont *comme ça* dans *ce type de vie* .

Le *bon lecteur* dispose donc, pour le même texte, de plusieurs types de compréhension, les unes très approximatives et les autres plus précises.

## 2.2. *Approximations dans la production du lexique*

C'est une caractéristique bien connue de la langue parlée : au moment de les produire, nous cherchons nos mots et il arrive fréquemment que nous n'y arrivions pas du premier coup, parce que la production du mot adéquat demande du temps. Parmi les nombreuses solutions disponibles, il en est une très fréquemment utilisée en français parlé : celui qui parle produit une amorce du syntagme où doit figurer le mot, puis une hésitation, puis, dans un deuxième temps, le syntagme complet :

Elle voulait un...un poulain

Pour bien représenter cette succession, je dispose les deux réalisations l'une sous l'autre, de façon à faire ressortir le vide lexical de la première étape :

elle voulait un.... un poulain
-----------------------------------

Très souvent, l'opération qui consiste à chercher le mot est avouée en tant que telle, avec une série de commentaires métalinguistiques : *comment dire, je ne trouve pas le mot, je ne sais pas si c'est le bon mot* :

Il avait aussi      une <i>comment dire</i> une aide
---

Parfois, même s'il a produit quelque chose qui se rapproche du mot recherché, celui qui parle fait des commentaires sur l'adéquation, bonne ou mauvaise, de ce qu'il propose :

c'était une maison très huppée ( <i>huppée je ne sais pas si c'est bien le mot qu'on doit dire, mais enfin...</i> )
--

Il introduit parfois d'assez longues parenthèses pour justifier l'utilisation d'un mot:

Il fallait vraiment se battre se se débrouiller ( <i>comme on dit à ce moment-là</i> ) pour avoir quelque chose
---

---

Tous ceux qui ont observé de près les productions de langue parlée ont rencontré ces phénomènes. Ces paraphrases et ces commentaires soulignent à quel point celui qui parle estime qu'il utilise des expressions approximatives, pour un sens qui est en train de se construire, pendant le moment même de sa parole. Cette observation est importante : elle montre que l'on ne peut pas présenter l'échange linguistique, comme on le fait souvent, sous la forme d'un *message*, qui coderait un contenu sémantique déjà tout prêt, et qui l'enverrait à son interlocuteur. Le contenu sémantique se met au point, progressivement, à travers plusieurs tâtonnements, et reste souvent approximatif par rapport aux désirs exprimés par le locuteur :

« Corpus semantics therefore rejects the view that the speaker encodes a meaning (the message) into the language and that the hearer decodes the message from the utterance” (Teubert 2000).

Malheureusement, si on ne les considère que d'un point de vue normatif, ces approximations paraissent maladroitement, de sorte que certains observateurs n'y ont vu que des fautes (*dysfluencies*), accompagnées de tentatives de réparation (*repairs*). Les termes mêmes de *dysfluences* et de *réparations* sont gênants, dans la mesure où ils introduisent une perspective normative qui empêche de comprendre les mécanismes mis en œuvre. D'autres observateurs y ont vu, au contraire, des indices d'opérations cognitives importantes :

« Repair is more than a method by which participants may “correct an error” [...] Repair provides a method by which to expand the lexico-semantic resources available in designing talk” (Fox and Jaspersen 1995: 116).

Il est beaucoup plus difficile d'observer ces opérations dans la phase d'acquisition du lexique, où l'on peut supposer, cependant, que les valeurs de sens approximatives (les sens *vagues*), jouent un rôle tout aussi important.

### **2.3. Le procédé de la comparaison approximative**

Le procédé de la comparaison approximative est intéressant. La comparaison est déjà en elle-même un procédé d'approximation, puisqu'elle suggère une ressemblance et non une identité de désignation. Dire d'une maison que *c'est comme un château*, c'est faire passer le sens par une approximation. Mais il y a plus : certaines comparaisons sont accompagnées d'un ensemble de précautions visant à expliquer que la comparaison ne peut pas être vraiment prise au pied de la lettre, *presque comme un château mais pas tout à fait ; presque comme cela mais pas tout à fait ; pas comme cela mais presque ; pas comme cela , mais pas loin* . La comparaison est utilisée en

quelque sorte comme une *approximation d'approximation*. Le procédé est si fréquemment utilisé qu'il semble remplir une tâche sémantique de première importance. En voici quelques exemples :

il est encore euh pas tout à fait apprenti mais pas loin (Apos82,9)

il est un petit peu rouge mais pas trop (Beaut 17,12)

L1 -non j'irais pas jusque là mais bon

L2 - mais pas loin quoi (Cirill 1, 7-13)

ça m'a pas fait ni chaud ni froid mais pas loin (Lesaum 1,16-8)

Tout se passe comme si la désignation lexicale était une opération très délicate, qui ne peut pas être garantie intégralement. Voici un exemple très explicite de cette prudence dans la désignation, à propos d'un vêtement, qui pourrait être désigné approximativement comme *blue jeans*:

Elle portait des.... ce qui m'a semblé être des jeans d'un bleu très passé (Chalne 13,14 )

#### **2.4. Limites du droit à l'approximation**

Les désignations approximatives ne conviennent pas à tous les domaines. Deux grandes exceptions : les noms propres et les termes techniques. Dans les productions de langue parlée, celui qui cherche un nom propre le cherche par différents tâtonnements, mais il ne se satisfait généralement pas d'une approximation. Voici un locuteur qui cherche et trouve le nom d'une accordéoniste très connue :

Ah, comment est-ce qu'on l'appelle cette accordéoniste ?... Ah, ça me reviendra peut-être, son nom [...] - - Ça y est, mon nom, il me revient : Yvette Horner (Vie par 263,3)<sup>3</sup>

Les termes techniques sont conçus comme des désignations qui ne supportent pas l'approximation. Par exemple, l'expression française *brevet de secouriste*, terme technique défini par la législation, doit aboutir à une forme lexicale codifiée, et non à une désignation vague, ni à une paraphrase comme pourrait l'être l'expression *diplôme de secouriste*, que le locuteur rejette :

---

<sup>3</sup> Comme les exemples de français parlé sont donnés ici avec peu de contexte, ils sont pourvus d'une ponctuation qui en facilite l'interprétation.

Il fallait avoir le euh ... Ah zut j'arrive plus à trouver le nom.... Le brevet, ... le...., le diplôme là, ....de secouriste : le brevet de secouriste (Buttaf 91,3).

De la même façon, un maçon récuse explicitement un terme générique qui lui vient en premier lieu à l'idée, *un tuyau*, pour produire le terme technique, *une buse*, que son interlocuteur ne connaît sans doute même pas :

Il faudrait...bon ici on met un caniveau – pas un caniveau un tuy- ... une buse .....normalement il faudrait faire un puisard (maçon 232,11).

Ces limites données à l'approximation sont révélatrices. La métaphore du message, du code et de la transmission, qui paraît totalement inadéquate pour le *langage ordinaire* est en revanche tout à fait adéquate dans ces domaines ; Il s'agit d'opérations lexicales totalement différentes.

Les exemples d'approximations lexicales ont été pris dans le maniement de la langue maternelle, soit dans la lecture de textes difficiles, soit dans la langue parlée. Les opérations que nous y observons ont beaucoup d'analogies avec ce que nous avons observé dans l'expérience *Eurom4* pour l'acquisition du lexique de langues étrangères. Des phénomènes similaires sont notés dans ce qu'il est convenu d'appeler l'*intercompréhension*.

### ***3. Le phénomène de l'intercompréhension***

Le terme anglais de *mutual intelligibility*, traduit en français par *intercompréhension*, a servi et sert encore de mesure pratique dans certaines situations où il s'agit de décider du degré d'apparementement entre langues et dialectes. L'intercompréhension est le phénomène par lequel des locuteurs estiment qu'ils se comprennent, au moins partiellement, dans une chaîne de parlars mutuellement intelligibles.

Cette notion reçoit beaucoup d'applications pratiques. C'est le cas lorsque par exemple, dans une zone très fortement dialectalisée, des institutions politiques décident d'établir des repères standard pour l'enseignement primaire ; elles se fondent souvent sur le critère de l'intercompréhension (Mexique, zone Maya). Des mesures pratiques ont été établies, en dehors même de toute évaluation donnée par des linguistes professionnels : une intercompréhension d'environ 80% signifierait qu'il s'agit de dialectes d'une même langue, entre lesquels une politique linguistique peut établir un standard commun. Au-dessous de 80%, il s'agirait de langues différentes (Projet CASAS, The Center for Advanced Studies of African Society, revue *Notes and Records Series*)<sup>4</sup>. Ce qu'on mesure, dans ces cas-là, ce sont les

---

<sup>4</sup> Ces mesures d'intercompréhension utilisent la plupart du temps des listes lexicales, ce qui permet d'établir des pourcentages.

possibilités de compréhension approximative : à 80%, il manque un mot sur cinq, mais cela n'empêche pas certaines formes de *compréhension* et d'interactions sociales. Dans les zones fortement dialectalisées, la notion même de compréhension exacte et rigoureuse serait totalement inadéquate.

Les différentes formes d'intercompréhension que l'on peut observer dans le monde sont étroitement dépendantes des habitudes culturelles. Haugen, qui parlait en ce cas de *semi-communication*, notait que, en Norvège, la culture nationale est tout à fait compatible avec l'idée de variations linguistiques et de changements réguliers (Haugen 1966). Ces changements sont du reste intégrés dans les habitudes scolaires, ce qui prépare les jeunes enfants à l'idée d'intercompréhension.

Les phénomènes d'intercompréhension sont graduels : ils s'échelonnent selon les régions traversées et peuvent, de proche en proche, se diffuser sur de grandes étendues. Ils peuvent au contraire être bloqués très vite par des raisons culturelles et des sentiments d'appartenance à des communautés distinctes.

Ces phénomènes mettent en jeu des relations non-symétriques : les habitants d'une zone A peuvent dire qu'ils comprennent à 80% au moins les habitants d'une zone B mitoyenne, sans que l'inverse soit vrai ; les habitants de la zone B peuvent fort bien soutenir qu'ils ne comprennent pas ceux de la zone A. C'est ce qui passe, semble-t-il, sur de nombreuses zones frontalières, par exemple entre Polonais et Russes (les Polonais disant qu'ils comprennent les autres et les Russes qu'ils ne les comprennent pas), ou entre Espagnols et Portugais (les Portugais conviennent généralement qu'ils comprennent les voisins, alors que les Espagnols n'en conviennent pas).

Trudgill pense que l'intercompréhension était autrefois un phénomène beaucoup plus étendu qu'aujourd'hui, avant que s'impose l'idée qu'une langue coïncide avec une identité nationale :

« In pre-koiné linguistic pool, most varieties are mutually intelligible » [...] Speakers accommodate to the speech of their interlocutors in order to promote a sense of common identity » (Trudgill 2005).

C'est sans doute pourquoi nous ne pouvons pas comparer ce qu'était autrefois l'intercompréhension entre les langues romanes et ce qu'elle est aujourd'hui (R. WRIGHT 1982, 1994). Les locuteurs d'autrefois estimaient probablement qu'ils avaient, d'une langue romane à une autre, un stock lexical commun suffisamment grand pour se comprendre approximativement. Le reste était affaire d'accommodation :

« Communication was not necessarily perfect then any more than it is now, but at least it happened regularly enough for the system to work » (R. Wright 1994: 4).

Au XXème siècle, de nombreuses générations d'ouvriers migrants (Italie, France, Espagne, Portugal) avaient, vaille que vaille, mis cette idée en pratique. La notion d'appartenance commune n'est sans doute plus aussi active aujourd'hui.

#### 4. Conclusion

Cette incursion dans ce que pourrait être aujourd'hui un *lexique européen* nous oblige à considérer plusieurs problèmes.

Le premier et le plus intrigant est sans doute la différence d'attitude entre ceux que j'ai appelés des *initiés* et les *non-initiés*. Seuls les initiés tirent parti de la communauté de lexiques qui existe actuellement en Europe. On pourrait penser que les professeurs de langues vivantes sont précisément là pour agir sur les non-initiés et pour les faire accéder à ce thesaurus. Mais nous rencontrons là le deuxième problème : une bonne partie des professeurs de langues vivantes envisagent de transmettre des connaissances qui vont bien au-delà de la maîtrise d'un stock lexical commun. Les formes de connaissances passives leur paraissent peu sérieuses. Il semble pourtant qu'elles aient joué un rôle considérable dans tous les changements linguistiques de l'histoire de l'Europe :

“[...] people can understand all sorts of uses that they would rarely actually say themselves” (R. Wright 1994: 18).

Le deuxième problème est celui de la diversité des connaissances et des compétences. J'ai présenté une sorte de plaidoyer pour des compétences partielles et même pour des formes approximatives de connaissances lexicales. C'est en effet ce qui nous avait beaucoup frappés dans l'expérience *Eurom4* : il est possible de transmettre, dans un temps assez court, des compétences qui permettent aux participants de s'émanciper et de continuer seuls leurs acquisitions. Il serait faux de penser que, à l'issue de l'expérience, ils « savent » trois langues romanes de plus. Ils n'en ont qu'une connaissance passive très partielle. Mais l'expérience a changé leurs attitudes : ils se comportent, à partir de là, comme des initiés capables de conquérir le lexique de ces langues. Ils sont devenus ce que Firth désignait drôlement comme des êtres linguistiques *solvables* :

« To be linguistically solvent, you must be able to exchange your terms somewhere and sometimes for gold of intrinsic social value” (Firth 1964: 176).

Le troisième problème est celui de l'unité de calcul utilisée pour traiter du lexique. Toutes les expériences montrent que les acquisitions ne se font pas mot par mot mais qu'elles passent par la maîtrise de l'expérience ou par la

maîtrise des textes. Il s'agit d'unités tout à fait différentes. Or la linguistique ne nous a pas habitués à traiter ce niveau d'unités contextuelles :

« En somme, centrée sur le problème du signe, la linguistique peine à concevoir le rapport entre signe et texte, et plus généralement le rapport du local au global, médiatisé par plusieurs paliers d'organisation de complexité croissante (le syntagme, la période, etc.) » (Rastier 2005 : 36)..

Pour traiter de ces unités plus complexes que le mot et de leurs différentes formes d'acquisition, il serait utile de faire converger plusieurs formes de connaissances : celles des linguistes, celles des spécialistes de l'acquisition, celles des traducteurs et interprètes professionnels, celles des ingénieurs informaticiens intéressés à la construction du sens à partir des données attestées et celles des spécialistes de la politique linguistique. L'accès à ce lexique européen dont il a été question oblige en effet à poser des problèmes sémantiques à plusieurs facettes.

## Bibliographie

- BLANCHE-BENVENISTE, Claire, (dir. 1995), *Méthode d'apprentissage simultanée de quatre langues romanes : portugais, espagnol, italien, français*. Firenze : Nuova Italia.
- BLANCHE-BENVENISTE, Claire (2000), *Approches de la langue parlée en français*. Paris : Ophrys.
- BLANCHE-BENVENISTE, Claire (2005) « L'étude grammaticale des corpus de langue parlée en français », in Geoffrey WILLIAMS (dir.), *La Linguistique de corpus*. Rennes : Presses Universitaires de Rennes : 47-66.
- BLANCHE-BENVENISTE, Claire (2005) « Les aspects dynamiques de la composition sémantique à l'oral », in Anne CONDAMINES (dir.), *Sémantique et Corpus*. Paris : Lavoisier, 39-70.
- BLANCHE-BENVENISTE, Claire (2005), « Accepter l'approximation dans l'apprentissage », *Le Français dans le Monde* N° 340, 25-26.
- BLANCHE-BENVENISTE, Claire, (à paraître), « Formes de compréhension approximative », in E. CASTAGNE, actes de la rencontre de Gap en 2005 sur l'apprentissage des langues.
- BLANCHE-BENVENISTE, Claire et André VALLI (éds.) (1997), *L'Intercompréhension: le cas des langues romanes*. Numéro spécial de *Le Français dans le Monde* (Recherches et Applications). Paris : Hachette.
- BLOOM, Paul (2000), *How children learn the meaning of words*. Cambridge: MIT Press.
- BOYER, Y.(1992), "La lisibilité", *Revue Française de pédagogie*, avril-mai-juin.
- CLARK, Eve (1993), *The Lexicon Acquisition*. Cambridge: Cambridge University Press.
- CLARK, Eve (1998), "Word Choice and Conceptual Perspective". Amsterdam: Koninklijke Akademie.
- FAUCONNIER, Gilles & TURNER, Mark, (2002), *The way we think. Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books.

- FAYOL, M., GOMBERT, J.E., SPENGER-CHQROLLES, L. & ZAGAR, D.(1992), *Pédagogie cognitive de la lecture*. Paris : PUF.
- FILLMORE, Ch. (1985), Frames and the Semantics of understanding », *Quaderni di Semantica* 6-2, 222-254.
- FIRTH, J.R. (1964), *The Tongues of Men and Speech*. Cambridge: Cambridge University Press.
- FOX, Barbara & JASPERSON, R. (1995), « A syntactic exploration of repair in English conversation », in DAVIS (ed.), *Alternative linguistics: Descriptive and Theoretical Modes*. Amsterdam: Benjamins: 77-134.
- FUKKINK, Ruben, BLOK, Henk & DE GLOPPER, Kees (2001), “Deriving word meaning from Written Context: a multi-componential skill”, *Language Learning* 51-3: 477-96.
- GELINAS, C & PREFONATAINE, C. (1999), *Lisibilité. Intelligibilité des documents d'information*. Département de Linguistique, Université du Québec à Montréal.
- HAGEGE, Claude (1992), *Le souffle de la langue. Voies et destins des parlers d'Europe*. Paris : O. Jacob.
- HALLIDAY, Michael, TEUBERT, Wolfgang, YALLOP, Colin & ČERMÁKOVÁ, Anna (2004), *Lexicology and Corpus Linguistics*. London/ New York: Continuum.
- HAUGEN, Einar (1966), *Language Conflict and Language Planning. The Case of Modern Norwegian*. Cambridge, Mass: Harvard University Press.
- HOLTUS, G., G. METZELTIN & C. SCHMITT (1988), *Lexicon der Romanistischen Linguistik* (8 ol.).Tübingen: Niemeyer.
- HOSENFELD, Carol (1984), “Case Studies of Ninth Grade Readers”, in ALDERSON, J.Ch. & URQUHART, A.H., *Reading in a foreign Language*. London/New York: Longman, , 231-244
- MICHNICK-GOLINKOFF, Roberta & HIRSH-PASEK, Kathryn (eds.2000), *Becoming a word-learner: debate on lexical acquisition*. Oxford: Oxford University Press.
- QUINE, W. V. O. (1960), *Word and Object*. Cambridge: MIT Press.
- RAMAT, Paolo, “Standard Average European“ (2000), “The notion of Standard Average European”, *Le Lingue d'Europa*. Bologna ottobre 2000.
- RAMAT, Paolo & Th. STOLZ (2002), *Mediterranean Languages. Papers from the MEDTYP Workshop*, Terrenia 2000. Bochum: Brockmeyer.
- RASTIER, François (2005), “Enjeux étymologiques de la linguistique de corpus”, in Geoffrey WILLIAMS (dir.), *La Linguistique de corpus*. Rennes : Presses Universitaires de Rennes 31-46.
- RENZI, Lorenzo & ANDREOSE, Alvise (2003), *Manuale di Linguistica e filologia romanza*. Bologna: Il Mulino
- SALA, Marius (1987), *Vocabularul reprezentativ al limbilor romanice*. Bucuresti: Editura Academiei.
- TEUBERT, Wolfgang (2000), « Corpus Semantics », *International Journal of Corpus Linguistics* (5-1).
- TEYSSIER, Paul (2004), *Comprendre les langues romanes. Du français à l'espagnol, au portugais, à l'italien & au roumain. Méthode d'intercompréhension*. Paris : Chandeigne
- TOMASELLO, Michael & FARRAR, M.J.(1984), « Cognitive bases for lexical development : object permanence and relational words », *Child Language* n°11: 477-93.
- TRUDGILL, Peter (2005), compte rendu de TUTEN, *Koineization in medieval Spanish*, *Language* 751-2.
- WRIGHT, Roger (1982), *Late Latin and Early Romance in Spain and Carolingian France*. Liverpool: Francis Cairn.
- WRIGHT, Roger (1994), *Early Ibero-Romance*. Newark: Juan de la Cuesta.