

Spécificités du lexique dans un corpus de dialogue oral

Mariette Bessac* et Geneviève Caelen-Haumont**

Laboratoire de Communication Langagière et Interaction Personne-Système (CLIPS - IMAG)
Université Joseph Fourier
B.P. 53
38041 Grenoble cedex 9
FRANCE

* - Tel : (+33) (0)4.76.51.45.10 ** - Tel : (+33) (0)4.76.63.56.95

Fax : (+33) (0)4.76.44. 66.75

e-mail : Mariette.Bessac@imag.fr ; Genevieve.Caelen@imag.fr

Résumé

Le lexique utilisé en communication orale est bien différent de celui mis en œuvre à l'écrit, notamment au niveau de l'ordre des éléments, de leur nature et de leur fonction. Dans un contexte de compréhension ou de traitement automatique de la parole spontanée, il n'est donc pas possible d'utiliser des systèmes et/ou des procédures de l'écrit. Dans cette perspective, le lexique d'un corpus de dialogue a été étudié, avec le souci d'en privilégier l'aspect spontané.

Cette partie du travail s'est basée sur les transcriptions de 6 dialogues (12 locuteurs) durant lesquels deux locuteurs devaient échanger des informations pour accomplir des tâches à partir d'informations différentes et de buts divergents. Le contexte concerne l'orientation dans une ville et la planification d'activités.

Une analyse lexicale traditionnelle a rapidement montré qu'une grande partie des éléments qui relèvent de l'oral étaient difficilement descriptibles. De nouvelles classes lexicales ont donc été conçues, regroupant principalement les "appuis du discours". Ces derniers représentant 22 % du corpus, il est très important de les prendre en compte.

Du fait qu'à l'oral certains éléments ont une fonction spécifique, il est nécessaire de répertorier tous les items lexicaux dans un dictionnaire, et ce avant toute analyse lexicale automatique et, *a fortiori*, avant toute désambiguïsation. Un tel document (185 p.) a été réalisé ; chacune de ses entrées indique la nature et/ou les fonctions possibles de chaque élément dans ce contexte précis. Ce fascicule peut apporter une aide efficace à l'analyse lexicale de la langue orale.

1. Introduction

L'étude du dialogue naturel montre clairement à quel point la langue orale est différente de l'écrit et de la lecture. En effet, à tous les niveaux d'analyse, le discours oral spontané nécessite de nouveaux modèles et de nouvelles théories. Le contenu de cet article s'efforcera de montrer, en s'appuyant sur de nombreux exemples, les particularités lexicales de ce type de corpus.

L'analyse lexicale présentée ici a donné lieu à la rédaction d'un dictionnaire d'analyse des éléments lexicaux en contexte qui est le seul document du genre. Cependant, l'analyse lexicale n'est qu'une partie du travail de recherche effectué sur un corpus de dialogue naturel. Les recherches menées visent à mesurer les interactions entre lexique et prosodie, mais aussi entre prosodie et pragmatique. Dans une telle perspective, l'analyse lexicale est indispensable pour repérer les éléments significatifs qui sont alors analysés selon des modèles spécialement conçus puis étiquetés sur le signal.

Les parties 2 et 3 de cet article présenteront le cadre général du travail de recherche, le corpus d'étude et l'analyse lexicale ; les parties 4 et 5 s'attacheront à décrire les modifications apportées aux classes lexicales et le dictionnaire d'analyse des éléments lexicaux en contexte.

2. Corpus d'étude

2.1. Introduction

Le corpus d'étude n'a pas été conçu uniquement dans le but d'étudier le lexique de la langue orale spontanée (cf. [Bessac et Caelen-Haumont 1995]), mais il est cependant très représentatif du discours naturel. Il est de plus suffisamment important pour valider des investigations statistiques concernant par exemple la fréquence de certains éléments, comparativement aux données de l'écrit ou de la lecture.

2.2. Conception

Le corpus a été élaboré en fonction des besoins des analyses sémantiques et pragmatiques prévues pour le travail de recherche, le souci principal étant d'obtenir une grande variété d'actes de dialogue (cf. [Colineau et Moulin 1996]) et une qualité sonore optimale pour permettre de faire des mesures acoustiques fiables sur le signal (cf. [Caelen-Haumont 1991]).

Il a ainsi été décidé de mettre en scène deux locuteurs en situation d'appel téléphonique, l'un jouant le rôle d'un touriste et l'autre celui d'un employé de l'office du tourisme. Le dialogue est orienté par les deux tâches imposées : réactualiser le plan du touriste et prévoir un planning de visites (cf. [Gryl 1994]), sachant que les informations données aux deux locuteurs sont différentes et qu'ils ont des buts divergents (le touriste désire visiter les lieux sportifs et l'office a pour consigne d'orienter les visiteurs vers les lieux culturels). À ces consignes s'ajoutent des difficultés articulatoires, phonétiques et des degrés différents de surcharge de la mémoire de travail : noms peu évocateurs, rues piétonnes, voies en sens unique et travaux de voirie (que le touriste doit mémoriser), changements de nom de rues, construction de nouveaux bâtiments (que le touriste peut indiquer sur son plan). Le plan du touriste s'affiche et se modifie sur un écran, ce qui permet de contrôler les éléments à mémoriser ou à modifier.

2.3. Enregistrement

Les deux locuteurs sont placés dans une chambre sourde, se tournant le dos, et

communiquent par l'intermédiaire de casques et microphones. Ces derniers sont fixés au casque pour assurer une distance bouche-micro constante, et ce afin d'obtenir des mesures d'énergie cohérentes. L'enregistrement se fait simultanément sur bande D.A.T. (une piste par locuteur) et sur fichier P.C. (un fichier par locuteur). Les documents donnés à chacun sont fixés pour ne pas produire de bruit parasite. Les deux locuteurs sont familiarisés au préalable avec l'environnement de l'expérimentation et font un court essai avec l'expérimentateur. Cet essai sert également à optimiser le gain à accorder à leur voix.

D'autre part, les locuteurs répondent à des tests (cf. [Denis 1990]) destinés à connaître leur conception de l'espace, leur système de référence dans un espace abstrait et leur vitesse de traitement de l'information. Ces tests visent à déterminer les différences de conception de l'espace et de traitement de l'information entre les deux locuteurs mis en scène.

Les locuteurs sont au nombre de douze, trois couples d'hommes et trois couples de femmes. Ils sont tous étudiants. Les 6 dialogues durent entre 11 et 17 minutes chacun.

2.4. Transcription orthographique

Ces enregistrements ont ensuite fait l'objet d'une transcription, qui sert de base aux analyses lexicales, sémantiques et pragmatiques. Une transcription orthographique a été préférée à une transcription phonétique pour des raisons de lisibilité et de traitement lexical automatique. Les conventions de transcription sont celles de [Morel 1989], excepté ce qui concerne les recouvrements. En effet, ces derniers étant extrêmement nombreux, et certains n'interrompant pas les tours de parole, il est plus représentatif de transcrire les deux voix synchronisées. La difficulté est alors d'établir avec précision les limites des recouvrements des deux voix. Pour cela, la transcription est réalisée en visualisant les signaux synchronisés des deux locuteurs. La figure 1 présente un extrait de cette transcription orthographique.

Touriste : T1	
Office du tourisme : O1	

T1 -	(h) e l'office du tourisme° <hh> (h h) <h> oui
O1 -	<t> oui.: <u>bonjour mada.me</u>

T1 -	<u>bonjour°</u> (h) <u>je° suis°</u> - - à <u>l.a gare sncf là</u> (h:) et e: j'aurais
O1 -	<h h> <h> <t> (h) <u>oui:</u>

T1 -	voulu aller au centre commercial° (h:) <u>e donc</u> là je° vois
O1 -	<t> <u>oui: e:</u>

T1 -	qu' <u>j'ai un plan</u> <h> <u>mais: e:</u> <h> <u>oui</u> <h>
O1 -	<h:> (h) <u>oui vous:-êtes</u> à la gare sncf (h) <u>et vous</u>

T1 -	<h> (h) <u>et:</u> e je° suis--à vélo
O1 -	(v)oulez ou < l > centre°commercial° (h) <u>alors/</u> <<oui >>

N.B. :	(h) note une pause remplie par une inspiration audible
	<h> note une pause remplie par une expiration audible
	e note ce qui est généralement transcrit "euh" dans la graphie traditionnelle
	° indique une prononciation inattendue d'un son habituellement non prononcé
	: note l'allongement d'un son
	< > les caractères notés entre < > notent les bruits qui ne concernent pas la parole
	- note qu'une liaison habituelle n'a pas été faite
	souligné signale un recouvrement des deux voix

Figure 1 : Extrait de la transcription orthographique du premier corpus

On constate que l'enchaînement des énoncés n'est pas toujours chronologique, c'est-à-dire qu'une réponse peut ne pas arriver immédiatement à la suite de la question : elle peut intervenir avant que la personne ait terminé de poser sa question, ou débiter alors que la personne a commencé un autre énoncé. C'est pourquoi le dialogue n'est remis sous forme de tours de parole qu'après la transcription orthographique.

Lorsque les tours de parole sont reconstitués, l'analyse des énoncés en actes de dialogue est effectuée, suivie des analyses lexicale et sémantique, comme le montre le tableau 1 :

Analyse en actes de dialogue	Tours de parole		Analyse lexicale		Analyse de la connaissance partagée	
	Locuteurs		Lexique	Catégorie	Nouveauté	Analyse de l'information
	T1	O1				
demande de confirmation >	1		e	pphp		
			l'	dete		
			office du tourisme	lexn	nouveau	< connu à l'avance
confirmation >	2		oui	adva		
acte conventionnel >	3		bonjour madame	pphr		
acte conventionnel >	4		oui	pphp		
			bonjour	pphr		
information en initiative >	5		je	pper		
			suis	verb		
			à la	prel		
			gare sncf	lexn	nouveau	< connu à l'avance
			là	pphf		
acte phatique >	6		oui	pphm		
information d'un but >	7		et e	pphp		
			j'	pper		
			aurais	vaux		
			voulu	vern		
			aller	vern		
			au	prel		
			centre commercial	lexn	nouveau	< accessible par le plan
acte phatique >	8		oui	pphm		
			e	pare		

Tableau 1 : Extrait de l'analyse du corpus premier corpus

Chacun des 6 corpus contient entre 350 et 500 tours de parole.

L'analyse de la connaissance partagée et celle des énoncés en actes de dialogue ne sont pas présentées dans cet article. L'analyse lexicale est détaillée dans le paragraphe suivant.

3. Analyse lexicale

3.1. Introduction

L'analyse lexicale ne constituant qu'une partie du travail proposé, il a tout d'abord semblé suffisant de faire une analyse lexicale automatique classique, quitte à corriger quelques points manuellement par la suite. Le système PILAF [Courtin et al. 1992], adapté à l'analyse de la langue orale, a donc été choisi. Malheureusement, malgré ses qualités, l'analyse n'est pas satisfaisante. Les problèmes proviennent de la spécificité du corpus : la langue orale spontanée et le dialogue ne sont pas comparables en tous points à la lecture ou à l'oral finalisé. Les

différences posant problème sont de trois types :

- ◆ de nombreux éléments apparaissent presque uniquement dans la langue spontanée (prises de parole, phatiques de fin d'énoncé, etc.) ;
- ◆ d'autres éléments apparaissent exclusivement dans le dialogue (phatiques de maintien de contact, phatiques du rituel social, etc.) ;
- ◆ un grand nombre d'items ont ici une fonction qu'ils n'ont jamais à l'écrit ; inversement, certaines fonctions de l'écrit ne sont jamais utilisées dans la langue spontanée.

Chaque cas a donné lieu à des modifications spécifiques qui sont présentées dans les paragraphes suivants.

3.2. Nouveaux items

Les items dont l'analyse ne donne aucun résultat ont été répertoriés. Il s'agit principalement de particules phatiques qui peuvent avoir plusieurs fonctions selon leur place dans l'énoncé. Ce sont des items spécifiques de la langue orale et/ou du dialogue. Tous les nouveaux items sont recensés dans un dictionnaire d'analyse des éléments lexicaux en contexte, chaque entrée indiquant la liste des fonctions possibles, c'est-à-dire toutes les fonctions rencontrées dans le corpus d'étude (avec un exemple), mais également les fonctions absentes mais possibles et/ou probables.

Exemples :

“**ben**” peut être une particule phatique de prise de parole (“Ben autrement je serais peut-être bien allée au musée.”) ou une particule phatique simple (“Vous allez me dire par où vous passez et puis ben je vous dirai si c'est possible.”).

“**hein**” peut être une particule phatique de fin d'énoncé (“Pour faire du footing dans des conditions optimales, il y a le stade hein.”) ou une particule phatique du rituel social (“non attesté dans le corpus d'étude.”).

“**si on veut**” peut être une particule phatique du rituel social (“Si on veut, on peut prendre la visite guidée.”) ou une particule phatique de fin d'énoncé (“non attesté dans le corpus d'étude.”).

Il est tout à fait probable que de nouvelles fonctions apparaîtraient avec l'étude d'un autre corpus. La liste des nouveaux items n'est pas non plus exhaustive.

3.3. Nouvelles fonctions

L'autre type de problème rencontré provient du fait qu'une grande quantité d'items n'ont plus les mêmes fonctions qu'à l'écrit et/ou bénéficient de nouvelles fonctions. Par exemple, un “donc” est généralement décrit comme une conjonction introduisant une conséquence, exprimant une transition, l'incrédulité ou la surprise, ou encore renforçant le contenu d'un propos (cf. [Delatour et al. 1991], [Rey-Debove et Rey 1995] et [Wagner et Pinchon 1962]). A l'oral, et tout particulièrement en dialogue spontané, ce même “donc” peut prendre ces fonctions, mais il peut également avoir d'autres emplois. Ainsi, on le retrouve comme particule phatique simple ou comme particule phatique de prise de parole.

Cette profusion de nouvelles fonctions a donné lieu à une extension du dictionnaire. Comme pour les nouvelles entrées, les fonctions rencontrées dans le corpus d'étude (y

compris les nouvelles fonctions) sont assorties d'un exemple ; les fonctions non attestées mais possibles à l'oral sont indiquées sans exemple ; les fonctions réservées exclusivement à l'écrit ne sont pas mentionnées.

Exemple :

“**donc**” peut être une conjonction de coordination (“Elle ferme à 22h30, donc vous avez tout votre temps.”), un connecteur linguistique (“Donc c'est là le centre commercial.”), une particule phatique simple (“La rue neuve donc, que vous avez vue tout à l'heure.”), une particule phatique de prise de parole (“Donc, il est 2 heures 1/2”.) ou une expression d'une réaction (“non attesté dans le corpus d'étude.”).

On peut voir, dans les exemples donnés ci-dessus, que les classes lexicales traditionnelles ont dû subir quelques modifications en raison de la quantité de nouvelles fonctions et de nouveaux items. Ces modifications sont présentées dans le paragraphe suivant.

4. Restructuration des classes lexicales traditionnelles

4.1. Introduction

Les classes lexicales du système PILAF, déjà modifiées pour s'adapter aux particularités de la langue orale, ont subi une nouvelle refonte pour le dialogue spontané. La quantité des changements nécessaires s'est avérée trop importante pour modifier les classes et leur contenu au fur et à mesure de l'analyse. Une autre stratégie a donc été adoptée : les deux tiers du corpus sont analysés par l'expert, les données ainsi recueillies étant utilisées pour modifier PILAF tandis que le tiers restant sert de corpus de validation.

4.2. Nouvelles classes lexicales

À cause de la multitude de nouvelles fonctions relatives à la spécificité du corpus, certaines classes lexicales ont été regroupées et d'autres ont été scindées, et ce en fonction des besoins des analyses sémantique et pragmatique prévues sur le corpus. Par exemple, tous les adverbes relatifs au temps sont classés, non plus ensemble, mais dans des sous-groupes distincts.

De nouvelles classes lexicales ont également été créées. Il s'agit principalement d'erreurs, de bruits et d'appuis du discours. Dans chacun de ces groupes, les éléments ont été classés en différentes catégories :

- **Erreurs** : éléments indéterminables, erreurs de prononciation et items interrompus.
- **Bruits et silences** : bruits, inspirations, expirations, éléments à peine audibles, pauses de réflexion, pauses silencieuses.
- **Appuis du discours** : connecteurs linguistiques, expressions pragmatiques, expressions réactives, particules phatiques simples, particules phatiques de perte de contact, particules phatiques de fin d'énoncé, particules phatiques de maintien de contact, particules phatiques de prise de parole, particules phatiques du rituel social, présentatifs.

Les erreurs, bruits et silences ne constituent qu'une partie minime du lexique total (6,4 %). Par contre, les appuis du discours, composés d'éléments typiques du dialogue et de la langue orale, représentent 22 % du lexique. On voit donc à quel point il est indispensable de les intégrer dans les systèmes d'analyse automatique du dialogue spontané.

4.3. Modification du contenu de certaines macro-classes lexicales

Les classes lexicales du système PILAF correspondaient à la nature des éléments. Dans la nouvelle version, elles correspondent à la nature, mais aussi à la fonction des éléments. Cette différence de définition a pour conséquence de modifier les entrées et leur description. Par exemple, “à côté” est traditionnellement décrit comme une locution adverbiale. Le dictionnaire présenté ici décrit “à côté” comme un adverbe de lieu mais également comme une locution prépositionnelle de lieu, précisant que ce dernier emploi ne peut intervenir que dans le cas où la locution est suivie d'un pronom relatif composé. Dans cette description, la nature de l'item est l'adverbe et sa fonction peut être celle d'un adverbe ou celle d'une locution prépositionnelle.

L'item “à côté de” est traditionnellement analysé comme une locution prépositionnelle et elle est généralement suivie d'un déterminant. Pour l'analyse de PILAF, il était plus pratique de prendre en compte la forme “à côté de + déterminant” d'un seul tenant car le cas de “à côté du” ne pouvait pas être analysé correctement. En effet, dans cet exemple, le déterminant est inclus dans une partie de la locution prépositionnelle, empêchant tout découpage. L'analyse ne peut donc se faire que sur “à côté” puis “du” pour lesquels PILAF produit “adverbe de lieu + préposition”, ce qui ne représente pas la fonction de “à côté du”. Il a donc été décidé d'analyser toutes les prépositions et locutions prépositionnelles amalgamées avec le déterminant qui leur fait suite.

Le dictionnaire propose donc comme analyse de la famille “à côté” :

A COTE		
- adverbe de lieu : Par contre à côté il y a un 2 ^{ème} bâtiment.		advl
- préposition de lieu (+ pronom relatif composé) : non attesté dans le corpus.		prel
A COTE D' (+ substantif propre ou + déterminant autre qu'un article ou + pronom personnel tonique ou + adjectif indéfini)		
- préposition de lieu : non attesté dans le corpus.		prel
A COTE D'UN(E)		
- préposition de lieu : non attesté dans le corpus.		prel
A COTE DE (+ substantif propre ou + déterminant autre qu'un article ou + pronom personnel tonique ou + adjectif indéfini)		
- préposition de lieu : Je pensais pouvoir traîner mon vélo à côté de moi.		prel
A COTE DE L'		
- préposition de lieu : C'est de l'autre côté, à côté de l'office du tourisme.		prel
A COTE DE LA		
- préposition de lieu : Vous avez un complexe sportif à côté de la rue lakanal.		prel
A COTE DES		
- préposition de lieu : non attesté dans le corpus.		prel
A COTE DU		
- préposition de lieu : Il se trouve à côté du complexe sportif.		préposition de lieu
N.B. :	advl = adverbe ou locution adverbiale de lieu	
	prel = préposition de lieu	

Figure 2 : Extrait du dictionnaire d'analyse des éléments lexicaux en contexte

On remarque que la macro-classe “préposition” comprend à la fois des prépositions, des locutions prépositionnelles et des locutions prépositionnelles suivies de leur déterminant.

Une présentation plus globale du dictionnaire est proposée dans le paragraphe suivant.

5. Dictionnaire d'analyse des éléments lexicaux en usage

5.1. Introduction

Le document (185 p.) se compose des éléments présentés précédemment, à savoir les nouveaux items, spécifiques de la langue orale et/ou du dialogue, et les items qui peuvent avoir de nouvelles fonctions. À ces deux groupes se rajoute celui des homographes, très nombreux en dialogue spontané.

5.2. Homographes

Il s'agit d'items dont la transcription révèle une prononciation particulière à l'oral (par exemple “p't-êt(r)e” pour “peut-être”) mais également des éléments dont la prononciation est défectueuse mais qui n'entraînent pas de trouble dans la compréhension (par exemple “woilà” pour “voilà”). Lorsque l'erreur de prononciation (qu'elle ait des causes articulatoires ou phonétiques ou encore que ce soit un lapsus) entraîne un problème de compréhension de la part de l'allocutaire, l'item est répertorié dans les erreurs (par exemple “stalle” pour “salle”, qui provoque une interruption du dialogue et nécessite une répétition avec correction).

Les homographes rencontrés dans le corpus d'étude sont tous répertoriés dans le dictionnaire. Leur définition renvoie à l'entrée correctement orthographiée qui propose, comme pour les autres entrées, la nature et les fonctions possibles de l'item, comme on peut le voir dans l'exemple suivant.

P'T-ETRE	
- cf. entrée “peut-être”	
⋮	
PEUT-ETRE	
- adverbe de modalisation : J'ai peut-être quelques petites rectifications à faire.	advd
- connecteur linguistique : Ce que je propose c'est que peut-être je vous guide comment aller au complexe sportif.	cols
⋮	
VOILA	
- adverbe d'approbation : T2 - Ça se passe au stade ? O2 - Voilà .	adva
- préposition de temps : non attesté dans le corpus.	pret
- expression pragmatique : O5 - Vous me direz quand ceci sera fait. T5 - Voilà .	expp
- particule phatique : Vous voyez la rue Asterix, voilà ?	ppha
- particule phatique de fin d'énoncé : C'est tout le bloc qui descend jusqu'au sud là voilà .	pphf
- particule phatique de prise de parole : Voilà , je vous téléphone parce que j'ai un problème	pphp
- présentatif : Alors voilà ce qu'il y a eu aussi l'année dernière et cette année.	pres
- particule phatique de maintien de contact : Oui ... voilà ... oui.	pphm
⋮	
WOILA	
- cf. entrée “voilà”	

Figure 3 : Extrait du dictionnaire présentant deux exemples d'homographes

Le dictionnaire d'analyse des éléments lexicaux en contexte est écrit à partir de deux tiers du corpus d'étude. Il servira de document de base pour l'adaptation du système PILAF au

dialogue naturel. Le tiers restant sera utilisé pour valider la nouvelle version de PILAF.

6. Perspectives et conclusion

Dans le travail de recherche concernant le dialogue oral spontané, une analyse lexicale adaptée est indispensable pour repérer les items lexicaux à analyser selon d'autres critères (analyses sémantique et/ou pragmatique). Les systèmes d'analyse lexicale automatique actuellement disponibles sont adaptés aux types de corpus étudiés jusqu'à présent, c'est-à-dire des corpus de lecture et d'oral (généralement monologique). L'intérêt du corpus de dialogue spontané est donc d'une part de servir de base pour adapter ces systèmes aux spécificités de la langue orale spontanée et du dialogue et d'autre part de fournir un matériau suffisamment riche pour modifier à leur tour les modèles sémantiques et pragmatiques d'analyse du discours. Le dictionnaire d'analyse des éléments lexicaux en contexte est le document le plus adapté aux besoins de l'expert qui apportera les modifications nécessaires aux systèmes d'analyse lexicale.

7. Bibliographie

- Bessac, M. et Caelen-Haumont, G. Analyses pragmatique, prosodique et lexicale d'un corpus de dialogue oral homme-homme. In : *JADT 1995, Actes des IIIèmes journées internationales d'analyse de données textuelles, Vol. 1, Rome, 11-13 décembre, 1995*. Rome : Eurograf 2000, 1995. p. 363-370
- Caelen-Haumont, G. *Stratégies des locuteurs en réponse à des consignes de lecture d'un texte : analyse des interactions entre modèles syntaxiques, sémantiques et pragmatique et paramètres prosodiques*. Thèse de doctorat d'état : Lettres : Université de Provence Aix Marseille I, 1991. 2 Vol. 279 et 298 p.
- Colineau, N et Moulin, M. Un modèle connexionniste pour le reconnaissance d'actes de dialogue. In : *ILN'96, Actes du Colloque Informatique et langage naturel, Nantes, 9-10 octobre, 1996*. (à paraître)
- Courtin, J., Dujardin, D. ET Kowarski, I. PILAF : Software tools for lexicography and text research. In : *Complex'92, proceedings of the 2nd conference on computational lexicography and text research, Budapest, 1992*. p. 93-109
- Delatour, Y. et al. *Grammaire du français, cours de civilisation française de la Sorbonne*. Paris : Hachette, 1991. 336 p. Collection Français Langue Étrangère. ISBN 2-01-015601-3
- Denis, M. *Cognition : l'individuel et l'universel*. Paris : Presses Universitaires de France, 1990. Collection Psychologie d'aujourd'hui. ISBN 2-13-042707-3. Chapitre 3, approches différentielles de l'imagerie mentale, p. 91-120
- Gryl, A. Analyse cognitive des descriptions d'itinéraires. In : *Actes du premier colloque jeunes chercheurs en sciences cognitives, La Motte d'Aveillans - Isère, 23-25 mars, 1994*. Grenoble : CNRS, 1994. p. 211-220
- Morel, M.-A. (Resp.). *Analyse linguistique d'un corpus de dialogue homme/machine. Vol. 2. Deuxième corpus : Centre d'Information et d'Orientation de l'Université de Paris V*. Paris : Publications de la Sorbonne Nouvelle, 1989. 371 p. ISBN 2-903019-93-2
- Rey-Debove, J. et Rey, A. (Dir.). *Le nouveau petit Robert, dictionnaire de la langue française*. Paris : Dictionnaires Le Robert, 1995. 2555 p. ISBN 2-85036-390-1

- Wagner, R.L. et Pinchon, J. *Grammaire du français classique et moderne*. 2^{ème} édition. Paris : Hachette, 1962. 648 p. ISBN 2-01-001325-5