Le "Poème du Cid" étudié à l'ordinateur : le système prépositionnel

Autor(en): **Pellen, René**

Objekttyp: **Article**

Zeitschrift: **Revue de linguistique romane**

Band (Jahr): **40 (1976)**

Heft 157-158

PDF erstellt am: **27.04.2024**

Persistenter Link: https://doi.org/10.5169/seals-399613

Nutzungsbedingungen

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.
Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.
Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

Haftungsausschluss

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

Ein Dienst der *ETH-Bibliothek*
ETH Zürich, Rämistrasse 101, 8092 Zürich, Schweiz, www.library.ethz.ch

http://www.e-periodica.ch

LE « POÈME DU CID » ÉTUDIÉ A L'ORDINATEUR. LE SYSTÈME PRÉPOSITIONNEL

1. Orientation.

Vouloir décrire le système des prépositions dans le *Cid* implique une succession de choix théoriques qu'il est nécessaire de définir avant d'entrer dans le vif du sujet. Précisons d'abord que ce travail repose sur un dépouillement automatique du Poème et une série d'analyses qui ont débouché sur la constitution de nombreux fichiers, dont le Vocabulaire exhaustif du texte. Chaque vocable a été affecté à une catégorie grammaticale donnée, pour qu'il devienne possible d'étudier dans leur dynamisme fonctionnel tous les « mots du texte ». Le système linguistique codé s'articule donc en un certain nombre de structures, dont chacune est fixée par convention, mais nullement figée au nom de critères dogmatiques qui se prétendraient définitifs ; au contraire : la recherche est une entreprise vivante, qui à tout instant peut modifier ses perspectives si le besoin s'en fait sentir. En particulier, toute classification demeure, quoi qu'en pensent de nombreux théoriciens trop pressés, ou trop exclusifs, une construction mentale de type heuristique ou, si l'on préfère, une structuration opérationnelle des connaissances en fonction d'une théorie, d'une actualité, d'un projet. La catégorie des prépositions s'inscrit dans un cadre général où les éléments de la langue et du discours se distribuent en classes. Mais l'analyse en classes ne peut se dissocier d'une théorie linguistique sous-jacente (généralement implicite), ni du moment où elle s'effectue (actualité émanant des recherches déjà réalisées, des résultats acquis, des hypothèses élaborées), ni du programme limité que le chercheur retient à ce moment-là.

Pour toutes ces raisons, les éléments qui entreront ici dans la catégorie des prépositions ont été triés en fonction de plusieurs critères, qui font qu'en d'autres circonstances ils appartiendraient aussi bien à d'autres ensembles. En effet, une catégorie linguistique n'est pas univoque. Qui s'aventurerait à verser *allende* dans une seule classe ? Ce terme peut être adverbe ou pré-

position, mais encore démonstratif si on le compare à *cerca*, lui-même adverbe ou préposition. En outre, il existe plusieurs niveaux possibles d'analyse linguistique ; pour n'en citer que quelques-uns : le mot, le syntagme, la phrase. Ne seront étudiés dans cet article que les « mots prépositionnels [1] », c'est-à-dire une catégorie du Vocabulaire fonctionnel du *Cid*, que l'on abordera dans son ensemble et dans ses éléments. Les deux approches fournissent des renseignements complémentaires, également précieux, sur la nature et le fonctionnement du système. En un premier temps, on examinera donc le système globalement, pour mettre en évidence ses caractères statistiques et distributionnels, tant il paraît certain, comme l'écrit I. I. Revzin, que « les catégories d'une langue manifestent, d'une façon ou d'une autre, leur existence par des critères statistiques [2] » ; puis l'attention se portera sur les divers éléments du système, chacun ayant son histoire propre, son comportement particulier... L'étude des syntagmes prépositionnels ou prépositions complexes sera laissée pour une autre occasion [3].

Quels critères secondaires permettent d'identifier les prépositions simples ? Ce sont, essentiellement, des critères grammaticaux : à la différence des « adverbes », ces éléments fonctionnels seront transitifs (*cerca* a été classé adverbe quand il entre dans le syntagme *cerca de* : 560 *çerca del agua*, ou quand il est employé absolument : 76, 212, etc. ; préposition lorsqu'il est suivi immédiatement de son régime : 3316 *çerca Valençia la grand*) ; ces opérateurs syntaxiques, d'autre part, ne fonctionnent qu'en association avec des éléments appartenant aux catégories dites lexicales, substantifs, verbes principalement, par rapport auxquels ils interviennent, pour le castillan médiéval, en position proclitique. Voilà qui suffit à en dresser l'inventaire. Mais il convient d'ajouter que la perspective de cette enquête est résolument diachronique. Il importe de dater le système observé afin d'en extraire le plus possible d'information sur la date probable du document dont il dérive et, par là même, d'apprécier de façon exacte la valeur historique de ce document. Diverses études catégorielles seront ainsi réalisées [4] ; si leurs conclusions concordent, il sera plus facile de dater le système

1. Sur la définition du mot dans les traitements informatiques, cf. « Le Poème du Cid étudié à l'ordinateur. Vocabulaire des Noms Propres »..., à paraître dans la revue *Études de Linguistique Hispanique Médiévale*, Paris.

2. *Les Modèles linguistiques*, Dunod, 1968, p. 10.

3. Elle relève d'une analyse syntaxique générale du texte, dont la réalisation est en cours.

4. Sur le système du possessif, cf. « *Poema de Mio Cid*. Étude du possessif », à paraître dans *Hommage à Ch. Camproux*, Montpellier.

linguistique du *Cid* dans son ensemble et d'intégrer les témoignages qu'il renferme dans une histoire plus rigoureuse de la langue castillane.

2. L'ensemble prépositionnel du Cid. Description.

2.1. *Tableau général.*

Cid		1	2	3
1133	*a*	300	373	460
6	*allende*	1	3	2
21	*ante*	6	7	8
2	*cabo*	1		1
1	*cerca*			1
291	*con*	71	107	113
5	*contra*	2	1	2
1485	*de*	394	470	621
7	*delante*	2	1	4
1	*des*		1	
2	*desde*		2	
607	*en*	188	194	225
20	*entre*	8	6	6
1	*hacia*			1
32	*hasta*	8	11	13
5	*par*			5
66	*para*	17	25	24
326	*por*	98	84	144
45	*sin*	20	12	13
2	*so*			2
36	*sobre*	16	7	13
8	*tras*	3		5
4102	22 vocables	1135	1304	1663

(Dans la colonne de gauche figurent les fréquences totales des vocables, sommes des fréquences par chants présentées sous 1, 2, 3.)

2.2. *Examen de ce tableau. Caractères statistiques.*

Sur un total de 22 vocables, 6 seulement sont devenus archaïques dans l'espagnol moderne : *allende, cabo, cerca, des, par, so*. Le corpus, considéré globalement, offre donc un aspect relativement actuel, si l'on pense qu'on y trouve 17 vocables apparaissant dans la liste des 20 prépositions « simples usuelles » que cite María Moliner (manquent dans le *Cid* : *bajo, cabe, según* ; María Moliner, d'autre part, précise que *cabe* et *par* ne sont plus d'usage courant). En revanche, on ne rencontre ni *pues, iestra, cada, latus* (liste de García de Diego, *Gram. hist. esp.*, éd. 1970, p. 256), ni les variantes compo-

sées de *contra escuentra, escontra* ou *salvo* (Pidal, *Gram. hist.*, p. 336-337), auxquelles on pourrait ajouter les emplois prépositionnels de *aprés, cara, dentro, depós, detrás, empós, entro, fuera,* etc. Par ailleurs, les 22 vocables du *Cid* ne se situent pas dans les mêmes zones de fréquence. On enregistre dans le Poème 125 classes de fréquence. Cet ensemble peut être divisé en quatre sous-ensembles, où les prépositions se distribuent comme suit :

1–31	32–69	70–143	> 143
13	4	0	5
D	C	B	A

Il est aisé de reconnnaître des familles distinctes. Les prépositions de haute fréquence sont *a, con, de, en, por* ; un grand vide (ensemble B) sépare ces vocables du groupe de moyenne fréquence inférieure 32-69 : *hasta, para, sin, sobre* ; les autres vocables peuvent être dits de très basse fréquence, puisqu'en dehors de *ante* (22) et de *entre* (20) ils ont une fréquence au plus égale à 8 (*tras*). Or toutes les prépositions devenues archaïques appartiennent à cet ensemble, avec des fréquences telles (*allende* 6, *cabo* 2, *cerca* 1, *des* 1, *par* 5, *so* 2) qu'elles paraissent témoigner déjà d'une existence épisodique, voire résiduelle. Une simple addition rendra plus nette encore cette différence de nature entre les 22 vocables du corpus : l'ensemble A réunit 3842 occurrences sur 4102 (reste : 247), C 179 (reste : 81, effectif de D). Cette discrimination en niveaux au regard de la fréquence est capitale quand on veut interpréter correctement l'économie du système et, sur un autre plan, l'apprécier en diachronie.

En effet, la fréquence moyenne de l'ensemble A ($\bar{f} = 768{,}4$) est typique des éléments que la tradition lexicographique dénomme « fonctionnels ». La fréquence de C ($\bar{f} = 44{,}75$) reste encore beaucoup plus élevée que la fréquence moyenne des vocables dits « lexicaux » ($\bar{f} = 10{,}91$ pour le *Cid*) ; tandis que les vocables de D ($\bar{f} = 6{,}23$) ont un rendement moindre que les verbes, les substantifs et les adjectifs (respectivement 15 ; 9,5 ; 6,3) ! La distribution traduit à sa manière ce caractère statistique : seuls 5 des vocables de D reviennent dans les trois chants (*allende, ante, entre, delante, contra*) ; des 8 autres, seuls 2 (*cabo, tras*) figurent dans deux chants différents. En diachronie, un tel contraste soulève trois problèmes complémentaires : 1) ces vocables étaient-ils déjà usés à l'époque du *Cid* ? 2) étaient-ils au contraire

trop récents pour se répéter souvent ? 3) étaient-ils et ont-ils toujours été des vocables peu fréquents ? Une indication est immédiatement fournie par la comparaison avec l'espagnol actuel. Certains de ces vocables sont bien vivants (*ante, contra, desde, entre, hacia, tras* = 6) ; d'autres, presque ou complètement oubliés, en tout cas comme prépositions (*allende, cabo, cerca, delante, des, par, so* = 7). Sans doute faut-il donc analyser séparément le cas des deux sous-groupes. Pour les vocables du premier groupe, la question devient la suivante : quand se sont-ils imposés dans la langue ? Pour les vocables du deuxième groupe on peut se demander quelles traces de leur vieillissement sont attestées par l'histoire de la langue.

2.3. *Vocables peu fréquents toujours usuels.*

Ante, issu directement du latin, semble avoir toujours existé, mais avec un emploi concurrentiel, au Moyen Age, d'adverbe et de préposition. *Contra, entre, tras*, eux aussi, sont directement issus du latin. Si leurs valeurs n'ont pas toujours coïncidé avec leurs valeurs actuelles, ils n'ont jamais cessé d'exister en tant que vocables, à la différence de *desde* et *hacia*, qui sont des créations romanes. *Desde*, à en croire Cor-*dce* [1], étant plus récent (XII^e-XIII^e) que *hacia* (X^e, ou avant même, mais avec une résurgence tardive). D'où vient leur faible fréquence ? Pour *desde* on pourrait invoquer la concurrence de *de, des*, l'époque tardive à laquelle il s'est affirmé. Pour les autres vocables ces raisons ne valent pas. Or, d'après la liste de Julliand, les vocables de ce groupe occupent dans la hiérarchie des 5025 termes qu'elle retient un rang très élevé : *entre* 54, *desde* 78, *ante* 184, *hacia* 204, *tras* 1240, qui les rapproche des prépositions appartenant dans le *Cid*, à l'ensemble C : *para* 19, *sin* 36, *sobre* 46, *hasta* 56. Le texte du *Cid* est trop court pour qu'on puisse tirer des fréquences rapportées ci-dessus des conclusions valables pour la langue médiévale en général. Tout au plus parlera-t-on de tendance : les moins fréquentes des prépositions du *Cid* usitées encore aujourd'hui semblent avoir accru leur fréquence depuis le Moyen Age. A l'époque du *Cid*, elles se trouvaient donc peut-être à la phase initiale de leur carrière en castillan, bien que la plupart aient été héritées du latin.

Tout autre est la situation des vocables du second groupe : *allende, cerca, delante* étaient des adverbes fonctionnant parfois comme prépositions, mais qui se voyaient concurrencés soit par des vocables voisins (*cerca/cabo*,

1. On renverra sous ce sigle à Corominas, *Diccionario crítico etimológico de la lengua castellana*. Cor-*Br* désignera le *Breve diccionario etimológico de la lengua castellana*.

delante/ante) soit par des constructions complexes dans lesquelles eux-mêmes fonctionnaient comme adverbes (*cerca/[a]cerca de, delante/delante de* ; à noter que *allen-de* est une construction de ce type très anciennement lexicalisée, ce qui fait qu'on a employé plus tard *allende de* [1]. *Cabo, des* avaient également des rivaux ; *par* ne servait qu'à introduire des formules de serment ; *so* ne tarderait pas à s'effacer devant *bajo* (entre le XIII^e^ et le XVI^e^, d'après COR-*dce* et *Br.*). L'impression qui domine ici est donc celle d'une grande fragilité ; les éléments du deuxième groupe sont des éléments mal fixés, syntaxiquement ambigus ou faibles. Ce groupe était presque voué, linguistiquement parlant, à la disparition, comme tous les sous-ensembles en déséquilibre ou insuffisamment pertinents d'un système. Il ne faut pas oublier non plus que la création d'une langue s'opère en fonction de paradigmes qui tendent, s'ils présentent un haut degré de pertinence et de simplicité, à élargir leur champ d'application. Dans le cas des oppositions adverbe/préposition, l'un de ces paradigmes était constitué par l'opposition « élément simple-adverbe » / « élément composé-préposition » (*cerca/cerca de* ; cf. actuellement l'extension de ce modèle à *tras* → *tras de*, alors que l'adverbe est *detrás*). Le système prépositionnel du *Cid*, dans ses vocables de basse fréquence, manifeste ainsi les symptômes d'une langue en devenir. L'évolution du castillan vers la fin du Moyen Age ou, selon les cas, à l'époque classique, a rejeté progressivement de l'usage *allende, cerca*... Néanmoins, leur faible fréquence dans le *Cid* n'est sans doute pas à imputer à un seul ordre de faits : sans qu'on puisse écarter l'hypothèse d'une fréquence moyenne constamment faible en synchronie (XII^e^-XIII^e^ siècles) pour ces vocables, il n'est pas exclu que tel d'entre eux (par exemple *so*, assez courant dans d'autres textes) doive sa fréquence dans le Poème à des causes internes : mode de représentation (cf. en face de *so* 2, *sobre* 36), choix stylistique (*ante* 22/*delante* 7), traditions expressives, notamment formulaires, liées à la transmission orale (*por/par*)... Seule une étude comparative du *Cid* et d'autres œuvres pourrait éclairer cet aspect de la question, qui relèverait moins de l'état que de l'utilisation du système linguistique.

2.4. *Morphologie et diachronie.*

En ce qui concerne l'évaluation diachronique du système, diachronie impliquant chronologie et dates, il est indispensable, pour mieux appréhen-

1. Cf. Hanssen, *Gramática histórica*, p. 315. Cf. également *Lazarillo*, éd. Clás. Cast., I 80 : 1, VII 234 : 12, VII 238 : 3.

der l'état de langue du texte, de dresser un inventaire exhaustif des formes sous lesquelles les vocables sont attestés [1] :

1133 A (*a* 299/371/458 ; *ha* 1/2/2)
6 ALLENDE (*alen* 0/0/1 ; *alende* 1/0/0 ; *alent* 0/3/1)
21 ANTE (*ant* 0/1/0 ; *ante* 6/6/8)
2 CABO (*cab* 0/0/1 ; *cabo* 1/0/0)
1 CERCA (*çerca* 0/0/1)
291 CON (*co* 0/1/1 ; *con* 71/106/112)
5 CONTRA (*contra* 2/1/2)
1485 DE (*d* [2] 27/53/77 ; *de* 367/417/544)
7 DELANTE (*delant* 1/0/3 ; *delante* 1/1/1)
1 DES (*des* 0/1/0)
2 DESDE (*desd* 0/1/0 ; *desde* 0/1/0)
607 EN (*e* 1/0/2 ; *en* 187/194/223)
20 ENTRE (*entr* 2/0/1 ; *entre* 6/6/5)
1 HACIA (*faza* 0/0/1)
32 HASTA (*fasta* 3/4/5 ; *fata* 5/6/8 ; *ffata* 0/1/0)
5 PAR (*par* 0/0/5)
66 PARA (*para* 1/0/0 ; *pora* 16/25/24)
326 POR (*por* 98/84/144)
45 SIN (*si* 1/0/0 ; *sin* 16/9/13 ; *sines* 1/2/0 ; *sjn* 2/1/0)
2 SO (*so* 0/0/2)
36 SOBRE (*sobr* 3/3/4 ; *sobre* 13/4/9)
8 TRAS (*tras* 3/0/5).

Si l'on néglige les cas de variation purement graphique (apocope, oubli de signe diacritique pour *-n*, etc.), on constate peu de variété morphologique dans ces mots courts. On ne rencontre pas *cuentra*, qui n'est pas rare dans les textes du XIII[e] [3], ni *fazia*, antérieur à *faza* [4], ni *sen* ni *sien*, attestés par ailleurs [4]. L'ensemble morphologique présente une certaine modernité, lui

1. Les vocables sont écrits en capitales, les formes en minuscules et en italiques ; chaque forme est accompagnée de ses fréquences dans les trois chants, séparées par une barre oblique. Dans le cours de l'article, lorsqu'il ne sera plus indispensable de distinguer formes et vocables, les vocables seront transcrits comme ils l'étaient jusqu'ici.

2. Ont été traités comme « mot » *d* les *d* isolés dans le texte ou le résultat d'analyse des graphèmes *daqui, dalent*, etc. ; *del* a été analysé *de* + *l*, comme *al* = *a* + *l*, *contral* = *contra* + *l*.

3. Cf. Pidal, *Cid*, p. 146 : 10.

4. COR-*dce*.

aussi, sur laquelle tranchent seulement *(f)fata* et *sines*. Encore faut-il vite corriger cette apparente opposition : *sines*, qui a dû être créé par extension du paradigme adverbial en *-s* [1], est attesté au moins jusqu'au XIV^e^ [2] ; il est fréquent au XIII^e^. Quant à *fata*, qui alterne avec *fasta*, on rappellera à son sujet que « los manuscritos antiguos de Berceo vacilan entre *fata* y *fasta* [3] », que « en castellano antiguo *fata* es tan frecuente como *fasta* [4] ». Est-ce un lapsus ? COR-*Br.* date HASTA du XIII^e^. Voilà qui rejoint les approximations ou hésitations relevées dans une étude sur le système possessif du *Cid* [5].

Or, si l'on tient compte et des signes positifs (éléments ou aspects représentés) et des signes négatifs (éléments ou aspects non-représentés) dans le système prépositionnel, force est de reconnaître que rien ne permet de parler d'archaïsme ; tout, au contraire, inciterait, comme on va le voir, à identifier le système du *Cid* avec un système moyen se laissant dater du XIII^e^ siècle. Il serait trop simple de supposer que le scribe du XIV^e^ a entièrement modernisé la morphologie : il a bien laissé *cuemo, pora, fata...* Pourquoi n'aurait-il pas respecté de même **cuentra, *fazia, *sien, *ata*, etc., si la version du manuscrit qu'il recopiait avait véhiculé ces formes ? Le système prépositionnel du *Cid* constitue un ensemble déjà normalisé du point de vue phono-morphologique. L'emploi quasi général de *pora* (à une exception près) n'est pas une objection à retenir puisque, comme le souligne COR-*dce*, « también en el XIII se halla normalmente *pora* (...). En la *Primera Crónica General* la parte escrita en vida del Rey Sabio tiene *pora* ». La présence de *par* ne l'est pas davantage : il faut attendre le XV^e^ ou le XVI^e^ pour qu'à *par* soit substitué définitivement *por* ou *para* dans les serments ou formules [6].

1. Pidal, *Cid*, p. 296 : 24.

2. *Ibid.*, p. 391 : 5. On pourrait ajouter aux exemples cités par Pidal les exemples du *Fuero de Alarcón* (seconde moitié du XIII^e^), 108 *b* « sines danno », 387 *c* « sines mandamiento », *in* J. Roudil, *Les Fueros d'Alcaraz et d'Alarcón*, Klincksieck, 1966, p. 144 et p. 302. COR-*dce* note que « hasta muy modernamente en Aragón, sobre todo en la combinación *sines de* » on rencontre cette forme. Il est remarquable que tous les documents cités en référence à propos de *sines* renvoient à l'Est de la Castille ou à l'Aragon : Berceo, *Libro de Apolonio, Sta. Ma. Egipciaca, Treys Reys d'Orient, Auto de los Reyes Magos*, les fueros plus haut mentionnés...

3. COR-*dce*.

4. COR-*Br.* Il faut préciser que *fata* n'est pas propre au castillan archaïque : dans 1335-Juan Manuel, *El Conde Lucanor*, on lit « et perdiera toda su fama que *fata* entonçe guardara meior que omne del mundo » (Clásicos Castalia, p. 229 : 5-7). Pour d'autres exemples, cf. Gifford et Hodcroft, *Textos lingüísticos del Medievo español* ; p. 239 *a*, s. v. HATA.

5. Voir p. 9, note 4.

6. Cf. *para* dans *Quijote* I, V, Clás. Cast. vol. 1, p. 144 : 14 ; I, XXI, vol. 2,

Cabo, d'après Pidal, n'est relégué qu'au XVIII-XIX^e siècle à la poésie un peu affectée [1]. *Des* constitue un cas un peu particulier : il est utilisé dans le *Cid* avec un adverbe (v. 1710 : « desa qui uos sean mandadas »), formant un syntagme dont le modèle se retrouve dans *desi*, très courant jusqu'au Siècle d'Or, et proche de la construction *desque*, tôt lexicalisée. Il semble que de bonne heure *des* n'ait plus été employé qu'en combinaison et que l'exemple du *Cid* soit un phénomène résiduel, comme *lo so*, *los sos*, par exemple, dans la *General Estoria* [2], témoin d'un état de langue dépassé et ne survivant que dans des expressions toutes faites.

Bref, tant le corpus des vocables que l'ensemble des formes tendraient à affirmer plus la modernité que l'antiquité du Poème en ce qui concerne les prépositions. A un moment où les travaux de nombreux érudits, historiens ou linguistes, mettent en question la date et l'origine du *Cid* [3], peut-être est-il fécond pour la compréhension des problèmes linguistiques que pose le texte d'oublier les conclusions un peu prématurées de Menéndez Pidal et de se laisser porter par l'information telle qu'elle se livre à nous dans sa neutralité documentaire.

3. Confrontations diachroniques.

3.1. *Le système prépositionnel du XII^e.*

Car enfin, que renferment les documents de la première moitié du XII^e ? Souvent une langue mi-latine mi-romane. Il suffit de parcourir la *Crestomatía* même de Pidal pour s'en rendre compte : p. 28-29, dans un document de 1100, on rencontre les prépositions *ad* (l. 2, 4, 5...), *in* (l. 10), *per* (l. 13) ; dans le document suivant, de 1109 (Logroño), *ad*, *ennos*, *enna*, formes qui n'apparaissent jamais dans le *Cid*. P. 30-31, un document de 1111 (Huesca) présente de même *in*, *per*, *ad*, *usque ad*. Tous ces documents, écrits par des gens de loi ou par des personnes habituées à s'exprimer en latin, sont moins du castillan que du latin romanisé. Il faut attendre les textes déjà littéraires de la deuxième moitié du XII^e pour lire une langue authentiquement romane :

p. 172 : 2 ; I, XXX, vol. 3, p. 116 : 8 ; II, XXXII, vol. 6, p. 263 : 21 ; II, XLV, vol. 7, p. 163 : 16 ; mais encore « *Par* Dios » I, XXIX, vol. 3, p. 94 : 12.

1. *Cid*, p. 389 : 5-6.

2. Respectivement 11 *b* : 29 et 327 *b* : 33. De *desque* Cor-*dce* dit que « subsistió bien viva hasta el s. XVI, aunque Juan de Valdés ya censura su uso con el valor de *cuando*, pero todavía la usa Mariana y siguen empleándola en verso muchos autores hasta el s. XIX ».

3. Cf. notamment A. Ubieto Arteta, J. Horrent, L. Chalon.

Auto de los Reyes Magos, *Disputa del alma y el cuerpo*. Mais le système de ces poèmes reste plus archaïsant que celui du *Cid*. Comme prépositions le premier emploie *a*, *de*, *con*, *en*, *(h)ata*, *par*, *por*, *sines*, *so*, *sobre*, mais aussi *ad*, *in* (simples graphies ? ; d'autres latinismes sont visibles dans certaines formes verbales : *uinet* 19, *percibistis* 97, ou non-verbales : *cumo* 67, 88, *december* 16, *gramatgos* 122, *clamado* 139...). Le second, peu étendu, n'atteste que *a*, *de*, *con*, *en*, *par*, *por*, *so*. Globalement — ce qui correspond à la probabilité théorique d'occurrence — c'est le corpus du sous-ensemble le plus fréquent qui se répète dans ces documents, ce qui n'est pas d'un grand secours pour dater le système du *Cid*, puisque les prépositions rencontrées, hormis *par*, *so*, *sines*, demeurent à travers toute l'histoire de la langue castillane.

Les documents non littéraires, heureusement, viennent en aide au linguiste. Rien de plus instructif qu'une lecture attentive des textes cités par la *Crestomatía*, p. 62-71 :

— *Fazañas de Palenzuela* (c. 1150-1160) : *a* 18 / *ad* 2 ; *con* 3 / *cum* 5 ; *de* 37 ; *en* 6 / *in* 5 ; *entre* 1 ; *epos* 1 ; *por* 3 / *per* 1 / *pro* 1 / *para* 1 ; *sobre* 2, *sobrie* 1 ; *tras* 1. Cet échantillon, on le constate, est extrêmement riche en prépositions. Plusieurs remarques s'imposent : 1) les formes latines sont encore nombreuses (*ad*, *cum*, *in*, *per*, *pro*), les formes romanes encore mal fixées (cf. en plus des formes précédentes, *sobrie*) ; 2) dans un texte court peuvent apparaître des vocables que le *Cid* n'emploie pas (EMPOS) : ils existaient donc bien dans le lexique des XII^e^-XIII^e^ siècles, mais ne s'actualisent pas nécessairement, même dans un texte relativement long ; chaque discours naît d'une sélection du lexique ; 3) sauf erreur de transmission ou d'interprétation, la forme *para* est bien attestée antérieurement au XIII^e^ siècle, alors que COR-*dce* la date de c. 1250 ; rien n'oblige donc, dans le *Cid*, à y voir une modernisation tardive ; 4) l'ensemble opérant dans ce fragment est sans contexte moins homogène que l'ensemble relevé dans le *Cid*, surtout si l'on tient compte de la longueur respective des documents (45 lignes/ 3730 vers).

— *Fuero de Avilés* (1155, « copia poco posterior ») : *a* 23/ *ad* 4 ; *con* 2 / *cum* 7 ; *contra* 1 ; *de* 40 ; *en* 9 (*e* 2, *em* 1, *en* 6) / *in* 4 ; *par* 1 ; *por* 3 / *per* 21 ; *sin* 1, *sen* 1 ; *sobre* 2 / *subr-* 1. Les remarques précédentes vaudraient pour ce texte asturien, dont la morphologie est marquée à la fois par le latin (*ad*, *cum*, *in*, *per*, *subr-*) et par une romanisation affirmée (*par*, *sen*). Mais on pourrait objecter que l'asturien n'est pas le castillan moyen, quoique ce fuero ait été accordé par Alphonse VII, roi de Castille et de Léon. Les autres échantillons s'avèreront-ils pourtant si différents ?

— *Fuero de Valfermoso de las Monjas* (Guadalajara, 1189) : *a* 7 / *ad* 15 ; *ata* 1 ; **con* 0 / *cum* 9 ; *d*[*e*] 60 ; *e*[*n*] 3 / *in* 17 ; *por* 24 / *per* 8 ; **sin* 0 / *sine* 2 ; **sobre* 0/ *super* 2 ; *usque* 2 (une fois *usque in*). Ce texte, bien que postérieur, est encore plus marqué par le latin que le précédent puisque, sauf pour *ata* qui n'est pas d'origine latine et *de* qui a survécu sans changement, à chaque forme romane correspond une forme latine, souvent plus fréquente, et que pour certains vocables seule la forme latine est attestée. Or, d'après la datation traditionnelle, le système du *Cid* serait antérieur au système de ce document...

— *Fuero de Madrid* (c. 1170-1202 ; ms. 1202) : *a* 40 / *ad* 26 ; *con* 10 / *cum* 29 ; *de* 75 ; *desde* 1 ; *en* 3 / *in* 23 ; **por* 0 / *per* 27 / *pora* 1 ; *si*[*n*] 1 / *sine* 3 ; **sobre* 0 / *super* 1 ; *usque* (*ad*) 1 — notons au passage 3 occurrences de *per a*[*d*].

Les nomenclatures de Madrid et de Guadalajara ne s'écartent pas notablement des deux premières. Il ne semble donc pas téméraire, devant la grande ressemblance des tableaux que livrent les documents de la seconde moitié du XII^e^, de considérer qu'il existe un sensible décalage entre les systèmes utilisés un peu partout dans l'aire castillane du XII^e^ siècle et le système du *Cid*. Ce dernier ne présente aucune trace latinisante ; ses variations morphologiques restent limitées et s'alignent sur les variations le plus couramment enregistrées au XIII^e^ [1]. De sorte qu'une conclusion prudente de cet examen rapide pourrait être la suivante : la comparaison des documents du XII^e^ siècle et du *Cid* dans le domaine prépositionnel ne dégage aucun argument de nature à prouver le caractère archaïque du Poème. Son système prépositionnel actualise une partie du lexique prépositionnel médiéval commun. Sur le plan morphologique il est même très stable, 4 seulement des 22 vocables attestés (*hacia, hasta, para, sin*) n'étant pas encore fixés — pouvaient-ils l'être ? On est donc amené, si l'on ne veut pas forcer la réalité, à préférer une date récente (XIII^e^ siècle) à une date plus ancienne (XII^e^) comme référence diachronique du système linguistique dont ce sous-système du *Cid* n'est qu'une actualisation partielle.

3.2. *Le système prépositionnel du XIII^e^ siècle.*

En effet, si en un deuxième temps l'on confronte le système prépositionnel du *Cid* avec le système illustré par un document de la seconde moitié

1. Lorsqu'un grand nombre de documents auront été dépouillés dans leur intégralité, cette appréciation deviendra mesurable. Il sera en outre plus facile d'évaluer les fluctuations dialectales.

du XIIIe siècle comme la *General Estoria* (date probable, selon Solalinde : entre 1272 et 1284 [1]), la réalité diachronique du Poème s'affirme encore davantage. Examinons de la *General Estoria* les pages 3-46, qui équivalent à peu près à un texte de la longueur du *Cid* (environ 31 300 mots). Pour des raisons de commodité (en attendant un dépouillement automatique de l'œuvre), abstraction sera faite de l'ensemble A (*a, con, de, en, por*). L'ensemble C présente les fréquences suivantes :

	Cid	*G. E.*
hasta	32	51
para	66	71
sin	45	6
sobre	36	98

Si l'on excepte *sin* (différence due à la manière d'envisager les sujets traités ?), les chiffres des deux textes reflètent la même tendance. Pour interpréter correctement les fréquences élevées de *hasta* et *sobre* il faut garder présente à l'esprit l'importance dans la *General Estoria* des considérations géographiques et chronologiques ; plus particulièrement, pour *sobre*, de la référence constante aux sources utilisées (« sobreste logar »..., « sobrello »...). *Sobre*, dans ces conditions d'emploi, devient une marque stylistique renvoyant à l'attitude du locuteur face à son récit et à son public, face à une tradition, à un savoir que l'on rappelle (*auctoritates*), à une diversité d'opinions exigeant souvent un choix, etc. [2]. Ce sont toutes ces raisons qui le font entrer dans l'ensemble B du *Cid* (fréquence comprise entre 70 et 143 : ensemble vide dans le Poème), *para* se situant juste au seuil de cette classe. Notons que si un seul *para* est attesté dans le *Cid*, cette forme ne revient que deux fois dans le fragment analysé de la *General Estoria* (4 *a* : 29 et 11 *b* : 1), ce qui confirme l'indication de Corominas concernant la *Primera Crónica*

1. Cf. édition de Madrid, Junta para Ampliación de Estudios e Investigaciones Científicas, 1930 ; introduction de A. G. Solalinde, en particulier p. XXI *sqq*.

2. La répétition de *sobre*, surtout en fin d'exposé ou de chapitre, tourne au tic de langage. On comparera, dans le *Cid*, pour *sin*, avec des répétitions de tournures telles que « *sin* falla », « *sin* art », qui reviennent comme des chevilles. Mais la répétition même a valeur expressive : *sin* est un opérateur contrastif puissant qui, en écartant ce qui amoindrirait la réalité de l'imaginaire, accentue le relief du signifié : « *sin* color », « *sin* todo pavor ». Dans la *General Estoria*, *sobre* est un indicateur de provenance, signe d'érudition et de tradition. On le voit, on ne peut réduire la fonction linguistique de ces prépositions à celle de connexion logique.

General citée plus haut : la forme *pora* est la forme normale au XIII^e^, en tout cas dans les deux premiers tiers du XIII^e^, et ne constitue en rien un archaïsme dans le *Cid*. S'il est vrai que *para* se rencontre épisodiquement dès le XII^e^, la distribution *pora/para* du *Cid* serait tout à fait normale pour un texte du XIII^e^.

Maintenant, est-ce que l'ensemble des prépositions les moins fréquentes dans le *Cid* se retrouve dans la *General Estoria* ? A ce niveau, 8 vocables seulement sur 13 sont représentés : *allende* 1 (*Cid* 6), *ante* 16 (21), *cerca* 8 (1), *contra* 26 (5), *delante* 1 (7), *desde* 1 (2), *entre* 23 (20), *so* 15 (2). Les cinq non représentés sont *cabo, des* (on ne lit que *desi* et *desque*), *hacia, par, tras* (qui n'existe que comme préfixe). En revanche, la *General Estoria* possède des prépositions qui ne figurent pas dans le *Cid* : *aderredor* 1, *empos* 15, *segun* 25 (employé 48 fois comme conjonction ; il est à rapprocher, dans sa fonction d'indicateur de sources, de *sobre*). Les prépositions communes aux deux documents ont un comportement statistique voisin puisqu'aucune n'atteint le seuil supérieur de la classe, 31. Si *contra* augmente considérablement sa fréquence, ce phénomène s'explique par l'emploi qui est fait de *contra* pour situer géographiquement, en alternance avec *de parte de* [1] ; dans la *General Estoria* il ne subit pas encore la concurrence de *hacia*. Quant à *so*, il revient presque rituellement dans certains contextes (« so el cielo ») et le thème du déluge, abondamment développé dans ces premières pages, n'est pas étranger à l'accroissement de sa fréquence. Dans l'ensemble, donc, on ne peut guère tirer de conclusion très affirmative dans le sens d'une évolution clairement dessinée du système. Si l'on tient compte des thèmes traités dans les deux œuvres et des particularités du style historique comme du style épique, il semble qu'on puisse imputer au hasard les différences de fréquence.

Car l'absence même de tel élément du lexique dans l'un ou l'autre document ne peut être acceptée comme le signe d'un décalage diachronique. La faible longueur des textes étudiés pourrait suffire à expliquer des « trous » dans la représentation du lexique. On rappellera que sur les 5 vocables du *Cid* absents de la *General Estoria* 3 (*cabo, des, hacia*) n'ont qu'une fréquence 1, deux (*par, tras*) une fréquence supérieure déjà appréciable (5 et 8). Ce sont par conséquent des vocables rares, qui peuvent ne pas apparaître dans un texte de longueur égale à celui auquel ils appartiennent [2]. Il n'est pas

1. Cf. 350 *a* : 14, 448 *b* : 46-47, 48, etc. ; on trouve même *contra part de* II, 16 *b* : 31 et 83 *b* : 10.

2. A l'inverse, cf. *supra*, on peut les rencontrer dans des textes beaucoup plus courts ; ex. *par, tras*.

exclu qu'on les voie ressurgir à l'occasion de dépouillements plus importants, comme c'est le cas pour *tras*, moins rare, à la suite d'une simple consultation manuelle (cf. 326 *a* : 16, 32, 33 ; 364 *b* : 49 ; 433 *b* : 36). D'autres textes, de toute manière, confirment leur existence au XIII^e^ (cf. *cabo*, *Sta. Ma. Egipciaca* 657 ; *des*, utilisé comme dans le *Cid* : « a los que des aqui uernan morar », *Fuero de Alcaraz*, deuxième moitié du XIII^e^ [1], « desaqui en adelante », *Fuero de Alarcón*, même époque [2] ; *hacia* : *faza*, *Primera Crónica General* 6 *b* : 17, exemples de Pidal, *Cid*, p. 389-390 ; *par* : Pidal, *Cid*, p. 387-388). Et c'est là l'essentiel ; ces vocables s'identifient comme des éléments du lexique du XIII^e^ (voire du XIV^e^) ; ils ne commandent pas une datation antérieure. Pour les raisons déjà invoquées, on ne peut pas s'étonner de l'absence, dans le *Cid*, de telle préposition représentée dans la *General Estoria* : certains outils logiques (*empos*, *segun*) sont plus nécessaires à l'historien qu'au poète. Il sera aisé de montrer que dans le domaine des noms ou des verbes, par exemple, où l'effectif des vocables est plus élevé, les convergences sont très grandes entre le corpus du *Cid* et le corpus de la *General Estoria* ou celui de la *Primera Crónica General*. Pour en rester au domaine des prépositions, la *General Estoria*, comme le *Cid*, varie peu sur le plan morphologique (*pora*/*para*). La tentation est forte d'assigner les deux documents à des époques voisines ; des études complémentaires, notamment d'histoire générale, devraient préciser l'importance de l'écart.

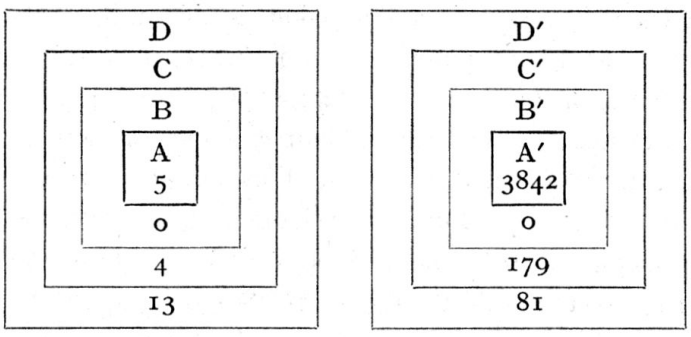
4. Distribution et fonctionnement du système.

Mais la description serait incomplète si les éléments du système n'étaient pas appréciés d'une manière plus approfondie en fonction de leur distribution. Car la fréquence à elle seule ne fournit qu'une présomption sur le rendement et la permanence d'une préposition. La stabilité ou l'évanescence d'un élément s'exprimeront plus nettement dans sa distribution.

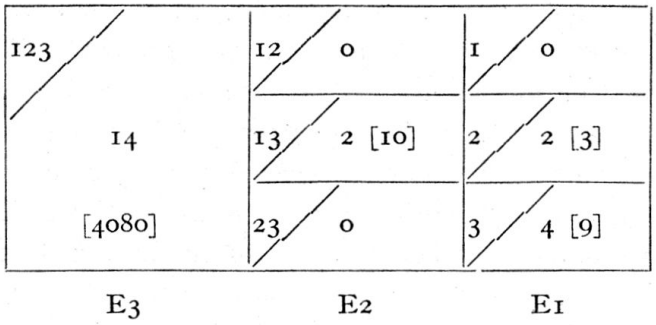
4.1. *Ensembles distributionnels et noyau du système.*

On pourrait résumer en trois graphiques les données distributionnelles concernant les prépositions : les ensembles I et II reprendront les renseignements commentés au paragraphe 2 ; l'ensemble III montrera comment se répartissent entre les chants les vocables de I et les occurrences de II (chiffres entre crochets ; à l'angle supérieur gauche, les numéros des chants) :

1. Éd. cit. p. 15, n. 2, I 1 *a*, p. 83.
2. *Ibid.* 354 *d*, p. 282.



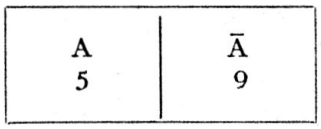
I. *Vocables* (22) II. *Occurrences* (4102)



III. *Distribution générale des vocables et occurrences.*

Il y a une disproportion écrasante entre la population de E3 et la population de E2 ou de E1. Dans E2, 12 et 23 sont des ensembles vides. Seul 13 a une existence linguistique sous l'espèce de *cabo* (vv. 56/3682) et *tras* (vv. 969, 990, 998/2290, 3333, 3365, 3676, 3679). Dans E1, l'ensemble 1 est vide. En 2 on rencontre *des* (v. 1710) et *desde* (vv. 1730, 2035) ; 3 est plus riche : *cerca* (v. 3316), *hacia* (v. 3060), *par* (vv. 2832, 3028, 3140, 3186, 3509) et *so* (vv. 2287, 3077) totalisent 9 occurrences [1]. Sur les 22 vocables de I, 8 sont donc intermittents, à des degrés divers. Ils possèdent vraisemblablement une moins grande stabilité que les 14 vocables de E3. On retiendra cependant que A, de I, ne comprend que 5 éléments, alors que E3 en comporte 14 : c'est dire que la permanence et la fréquence sont deux caractères distincts. Aussi n'est-il pas sans intérêt de comparer A et E3 :

1. Les références dans le texte pour les autres éléments de D∩E3 sont : *allende* : 911/1156, 1620, 1639/2409, 2425 ; *ante* : 169, 264, 269, 432, 502, 696/1318, 1548, 1747, 1843, 1873, 2093, 2128/2299, 2548, 2593, 2624, 2673, 2788, 3239, 3344 ; *contra* : 558, 558/1090/2910, 3471 ; *delante* : 327, 715/1577/2934, 3324, 3324, 3615 ; *entre* : 191, 547, 595, 603, 797, 842, 968, 1066/1236, 1549, 1737, 1774, 2087, 2254/2348, 2810, 2959, 2972, 3058, 3295.



E3

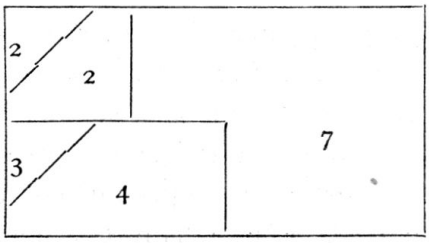
Ā comprend *allende, ante, contra, delante, entre, hasta, para, sin, sobre,* ou, par ordre de fréquence décroissante :

— *para, sin, sobre, hasta* ∈ C ;

— *ante, entre, delante, allende, contra* ∈ D.

Distribution qui fait apparaître que tous les éléments de C (fréquence moyenne inférieure) appartiennent au stock commun des prépositions, alors que 5 seulement de D (fréquence basse) en font partie ; en sont exclus ceux qui ont été cités plus haut. La stabilité de certains éléments n'est donc pas en relation directe avec leur fréquence — leur rendement. Faut-il supposer qu'elle traduit une réalité linguistique essentielle, que certains éléments de basse fréquence exprimeraient des représentations fondamentales, ce qui leur confèrerait un caractère de constante disponibilité et à la fois de nécessité ? Peut-être que la distribution enregistrée, combinée avec la répartition des fréquences, permettrait de distinguer plusieurs types de vocables : 1) les vocables sans cesse disponibles et nécessaires ; 2) les vocables sans cesse disponibles mais non nécessaires, en tout cas pas au même point que ceux du premier groupe (*tras,* figurant dans E2 et dans D) ; 3) les vocables éventuellement disponibles, mais dont la probabilité d'actualisation est faible ; ils constituent la frange menacée du système et deux possibilités seulement s'offrent à eux : ou bien ils s'effacent, pour finalement disparaître (*cabo, des, par*), ou bien ils se renforcent à la suite de modifications dans l'équilibre du système et tendent à rejoindre C (*desde, hacia*).

Ces considérations aideront à définir le « noyau » le plus solide du système, si l'on tient compte d'un autre aspect de la distribution, les relations entre D et E1 :



D = 13

Dans l'ensemble des vocables de fréquence inférieure à 32, 6 sur 13 ne figurent que dans un chant, soit près de la moitié ; leur fréquence est basse : $\bar{f} = \frac{12}{6} = 2$. Les vocables de $\overline{E1}$ (*allende, ante, cabo, contra, delante, entre, tras*) ont une fréquence bien supérieure : $\bar{f} = \frac{69}{7} = 9{,}85$. Statistiquement, on a tout lieu de penser qu'on a affaire à des populations de nature différente, même si tel élément fait figure d'exception (*cabo*) ; effectivement, 5 des 7 vocables de $\overline{E1}$ appartiennent à l'ensemble commun E3, les deux autres à E2, ce qui amène à retenir comme critères de sélection pour définir le corpus prépositionnel stable du *Cid* 1) l'appartenance à A (fréquence élevée), 2) l'appartenance à E3 ou E2 (constance dans l'actualisation). Si l'on fixe le seuil d'actualisation à 3 ($f = 1$ pour chaque chant) — seuil arbitraire qui pourrait être relevé — le noyau du système s'établit comme le montre le tableau suivant (les vocables apparaissant entre parenthèses simples sont classés par ordre de fréquence décroissante) :

$$\sum \begin{cases} A = \{a, con, de, en, por\}\ (de, a, en, por, con) \\ E3 \cap C = \{hasta, para, sobre\}\ (para, sobre, hasta) \\ E3 \cap D = \{allende, ante, contra, entre, delante\}\ (ante, entre, \\ \qquad delante, allende, contra) \\ E2 \cap C = \{sin, tras\}\ (sin, tras). \end{cases}$$

Soit 15 vocables, sur 22. Sont exclus : *cabo* (E2, mais $f = 2$), *par* ($f = 5$, mais E1), *des, desde, cerca, hacia, so* ($f < 3$ et E1), qui forment, par conséquent, le sous-ensemble le plus instable du système ($\overline{\Sigma}$). On constate la coïncidence presque absolue, qui était probable, entre $\overline{\Sigma}$ et le sous-ensemble de D regroupant les éléments de fréquence basse. Il ne faudrait pas pourtant en déduire qu'*a priori* tout élément de E1 ne saurait être que marginal. Les critères retenus ici s'appliquent à des éléments linguistiques dont la répétition dans le discours est une caractéristique première. Ils ne vaudraient pas nécessairement pour d'autres catégories, en particulier pour les catégories lexicales, où la survenance — et même la fréquence — d'un vocable peuvent être liées à un thème (cf. chant 1, *airar* 6 ; chant 2, *boda* 6), ou à une situation (noms propres comme *Alcocer* 16, chant 1 ; *Toledo* 8, chant 3), tandis que les thèmes traités ou la manière de les traiter repoussent et maintiennent en état de disponibilité des éléments du lexique qui, dans d'autres circonstances, occuperaient le devant de la scène (chant 1, *asmar* 3, *peón* 5, *virtud* 5 ; chant 3, *juicio* 4 ; etc.). A cette réserve près, il semble que la définition de Σ

soit non seulement justifiée, mais vérifiable : lorsque de nombreux textes auront été dépouillés exhaustivement, il sera loisible de comparer leur noyau prépositionnel avec le noyau du *Cid*, en prenant en considération l'étendue respective des textes et d'éventuelles particularités qu'on ne peut prévoir dans l'abstrait. Des constantes systématiques larges se manifesteront, qui permettront d'évaluer avec plus de finesse l'originalité de chaque document. Dans l'état actuel de la recherche, force est de se limiter à une analyse interne du système étudié.

4.2. *Évolution du système dans le Cid.*

Cette limite, réelle, ne réduit en rien la valeur informative des renseignements recueillis et n'implique aucune servitude dans l'approfondissement de notre connaissance du système cidien. Le Poème est un texte assez long pour qu'on puisse aborder le problème du fonctionnement statistique des prépositions. En outre, l'articulation en trois chants suggère après l'examen statique de l'ensemble E ($\Sigma \cup \bar{\Sigma}$), une enquête sur son comportement dynamique à travers l'œuvre. Des contrastes surgissent, en effet, entre les trois *cantares*, qui signaleront peut-être aux chercheurs qui s'intéressent au *Cid* d'utiles points de départ pour des études plus spécialisées.

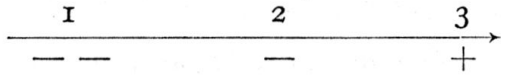
Si l'on se réfère au texte intégral, la longueur de discours fournie par le système prépositionnel est de 13,09 % pour l'ensemble, cette moyenne variant légèrement d'un chant à l'autre dans le sens d'un plus grand pourcentage :

12,5 12,8 13,6

Faut-il attribuer à une telle variation une signification spéciale ? (On pourrait relever aussi que si les chants 1 et 2 utilisent 16 vocables prépositionnels, le chant 3 en utilise 20 ; mais ce n'est pas l'accroissement du nombre des vocables qui influe sur la moyenne puisque les 4 vocables propres au chant 3 — *cerca, hacia, par, so* — n'apportent globalement au texte que 9 occurrences !) Examinée en probabilité, la distribution des prépositions (ensemble E) se présente de la façon suivante :

	1	2	3	*Cid*
r	1135	1304	1663	4102
t	1192,10	1311,98	1597,91	4101,99
é	— 57,10	— 7,98	+ 65,08	0
χ^2	2,73	0,04	2,65	5,42

Pour 2 d.d.l., à cette valeur du χ^2 correspond une probabilité comprise entre 0,05 et 0,10. Bien qu'on ne puisse en déduire avec une entière certitude une évolution significative de l'emploi des prépositions dans le *Cid* — c'est-à-dire, sur ce plan, une évolution caractéristique de la syntaxe —, les chances sont assez fortes pour que la modification orientée qui se lit dans les écarts ne soit pas due uniquement au hasard, ce qui mérite attention. On pourrait matérialiser cette orientation (les déséquilibres se situent aux deux extrêmes) par le schéma suivant :



Le déficit du chant 1 en prépositions et l'excédent complémentaire du chant 3 exprimeraient-ils une transformation de la syntaxe et du style impliquant d'autres changements dans des catégories grammaticales avec lesquelles sont compatibles les prépositions ? C'est une hypothèse de travail intéressante, qui ne peut être étudiée dans le cadre de cet article, mais qui devra faire l'objet d'une exploration minutieuse lorsqu'il s'agira de définir les caractéristiques générales du Vocabulaire du *Cid* [1]. Nous nous contenterons pour l'instant de retenir que le chant 2, dont la fréquence réelle tend à coïncider avec la fréquence théorique, apparaît comme un axe d'équilibre, comme un échantillon linguistiquement neutre du point de vue de l'emploi prépositionnel, et, d'autre part, que le discours cidien ne cesse d'augmenter ses articulations [2].

Mais nous avons distingué dans le système des prépositions plusieurs sous-ensembles, ou niveaux d'analyse. La question qui se pose, au-delà de l'évolution globale du système, est de savoir si A′, C′, D′ se comportent statistiquement comme E. On interrogera successivement les distributions de ces ensembles, en commençant par D′ :

$$D' = \{23, 21, 37\} = 81$$

1. Ce travail est en préparation. Il paraîtra dans la revue *Études de Linguistique Hispanique Médiévale*. L'évolution de la catégorie des prépositions s'inscrit dans un cadre beaucoup plus large d'évolution où le système verbal — cf. *infra* — manifeste une originalité déroutante.

2. Une analyse syntaxique globale, distincte du travail précédent, est également en préparation, pour déterminer à quel niveau se renforcent ces articulations : de verbe à verbe (faible probabilité), de verbe à nom, de nom à nom, etc.

Si l'on rapporte ces effectifs à la longueur des chants, on obtient (avec un grossissement de 10 000 pour rendre les valeurs plus lisibles) les indices suivants, mesurant la richesse relative des chants en occurrences prépositionnelles à l'intérieur de D' :

212 / 176 / 254.

Au chant 2 correspond un creux, bien que le nombre de vocables de D soit le même (7) que dans le premier chant. C'est que chaque élément intervient moins souvent (3 ne sortent qu'une fois : *contra, delante, des*) ; ce sont des vocables rares au même titre que les vocables lexicaux de fréquence basse, sortes de buttes témoins du système et en tant que tels très riches d'information. Se situant au seuil de l'absence, ils se chargent d'un coefficient d'actualité d'autant plus grand qu'ils émergent à peine du virtuel (c'est-à-dire des secteurs en pure disponibilité du système). Mais, corrélativement, leur fréquence interdit de dépasser une interprétation au niveau de l'individu — en pratique, de l'occurrence : chaque occurrence se convertit en un cas d'espèce, intéressant, certes, sur le plan philologique, mais de peu d'utilité pour une appréciation statistique du système, les sous-ensembles en opposition devenant par trop disparates (*contra* 1/ *de* 470 dans le chant 2).

En tant qu'ensemble, D' marque un net excédent pour le chant 3. Ceci est dû sans doute au fait que les vocables sont plus nombreux (11 au lieu de 7), car 3 d'entre eux, comme dans le chant 2, n'ont que la fréquence 1 (*cabo, cerca, hacia*). Il convient de souligner deux phénomènes remarquables. Des 5 prépositions tombées en désuétude (*allende, cabo, des, par, so*) 4 (toutes sauf *des*) figurent dans le chant 3 (deux : *par* et *so*, seulement dans ce chant), alors que le chant 1 et le chant 2 n'en possèdent chacun que 2 (*allende, cabo / allende, des*). Si l'on projetait sur les trois séquences du *Cid* une vision diachronique hâtive, on pourrait être tenté d'en conclure que c'est ce dernier chant qui présente le plus de traces d'archaïsmes. En réalité, cette conclusion serait incorrecte, pour deux raisons : 1) à l'époque du *Cid*, les vocables concernés étaient toujours en usage, même si leur emploi était rare (*des* seul serait déjà figé dans une survie syntaxiquement limitée ; au demeurant, *des* appartient au chant 2) ; 2) le nombre des éléments en jeu est trop faible, les textes sont trop courts pour qu'on puisse en tirer des conclusions générales ; à cet égard, des comparaisons avec d'autres corpus importants du Moyen Age apporteraient de précieux éclaircissements sur les aspects diachroniques de phénomènes qu'une information trop partielle

permet difficilement de cerner et de dater avec précision. Du reste, une évaluation en probabilité de la distribution 23/21/37 fait ressortir, pour un χ^2 de 1,86 avec 2 d.d.l., une probabilité comprise entre 0,30 et 0,50, ce qui montre combien il serait aventuré de s'appuyer sur cette distribution pour défendre quelque hypothèse que ce soit en matière de diachronie. Le nombre des vocables, quant à lui, n'autorise aucune appréciation statistique.

L'ensemble C′ est déjà plus étoffé (179 occurrences), bien qu'il repose sur un nombre plus réduit d'éléments (4 vocables). La distribution par chants est la suivante :

hasta	8	11	13
para	17	25	24
sin	20	12	13
sobre	16	7	13
	61	55	63

Au premier abord, on constate une variation globale de même nature que pour D′. Mais les valeurs absolues sont un écran derrière lequel se dissimule une baisse constante de la représentation : réduites, comme ci-dessus, à un coefficient relatif, elles deviennent 562 / 461 / 433. Toutefois, il serait douteux qu'un écart aussi faible entre le chant 2 et le chant 3 ne relève pas du hasard seul. Effectivement, si l'on calcule le χ^2 correspondant à cette distribution, la valeur obtenue (2,26) donne une probabilité comprise entre 0,30 et 0,50. Les ensembles D′ et C′ considérés globalement se démarqueraient donc de l'évolution générale manifestée par E. Dans ces conditions, il est probable que c'est le comportement de A′ qui détermine cette évolution. Cette déduction est-elle vérifiée par l'expérience ?

L'ensemble A′ (5 vocables) comporte 3842 individus :

a	300	373	460
con	71	107	113
de	394	470	621
en	188	194	225
por	98	84	144
	1051	1228	1563

La valeur du χ^2 (6,78) s'avère ici très significative puisqu'elle assigne à la distribution une probabilité comprise entre 0,02 et 0,05. C'est dire qu'on enregistre entre le premier et le troisième chants du *Cid* une évolution très

nette de la classe des prépositions les plus fréquentes. Tout aussi remarquable est la quasi identité de l'écart (en sens opposé : — 65,54/+66,38) entre chants 1 et 2 d'une part, chants 3 et 2 de l'autre. Le chant 2 est conforme au modèle d'une distribution parfaitement homogène des prépositions ; les chants 1 et 3, avec leur passif ou leur excédent, présentent les signes d'une modification importante de l'articulation syntaxique des phrases. Affinant le premier sondage, on peut tester chacun des éléments en cause, qui constitue lui-même un sous-ensemble dont la composition est variable et la distribution autonome. Le classement de ces prépositions par ordre de χ^2 décroissant donne le tableau que voici :

	Cid	1	2	3
de.	6,44	3,26	0,05	3,12
por.	6,33	0,11	3,93	2,28
con.	4,26	2,17	2,08	0
a.	3,70	2,60	0,31	0,78
en.	1,31	0,76	0	0,55

C'est ce tableau qui livre les points de départ les plus suggestifs pour une recherche détaillée sur la syntaxe et le style du Poème. Tout d'abord il fait apparaître que seules deux prépositions, *de* et *por*, ont un χ^2 permettant d'écarter valablement l'hypothèse nulle. On accordera néanmoins une certaine attention au fait que si *en* présente une distribution très homogène, *con* et *a* accusent dans le chant 1 — *con* également dans le chant 2 — des valeurs qui ne sauraient laisser le linguiste indifférent : elles expriment des écarts assez sensibles et dans les deux cas de sens négatif pour le chant 1 ; la valeur du χ^2 de *con* dans le chant 2 traduit la surreprésentation de cette préposition dans le deuxième *cantar* pour équilibrer exactement le déficit du chant 1, alors que la compensation pour *a* s'étale sur les chants 2 et 3. L'orientation est donc bien d'un déficit à un excédent, même pour les éléments dont la distribution globale n'est pas absolument significative. En ce

qui concerne *de* et *por*, le décalage entre les chants d'écart faible, voire presque nul, est frappant : la distribution de *de* s'organise autour de l'axe du chant 2 ; elle est orientée (selon le schéma — — / — / +) ; la distribution de *por* met en opposition les chants 2 et 3 ; elle n'est pas orientée (schéma + / — / + +), ne coïncide pas avec le schéma général de distribution des prépositions mis en évidence ci-dessus. Dans une interprétation qui voudrait rendre compte des faits concrets observés au niveau des réalisations de discours, il faudrait garder constamment à l'esprit ces variations individuelles dans le cadre plus large d'une évolution d'ensemble orientée, des jeux de substitution ou de compensation entre prépositions étant probables en fonction des thèmes abordés.

Cette analyse de détail ne peut être entreprise ici. On se contentera de deux exemples pour essayer de montrer l'intérêt des perspectives ouvertes. 1) Sur le plan général, le chant 2 se distingue par sa neutralité statistique ; or voilà que *por* et *con*, sur ce fond de grisaille syntaxique, se signalent par une fréquence non espérée ; qu'est-ce qui dans le chant 2 conduit donc le poète à privilégier ces représentations ? 2) Lorsqu'on analyse le Vocabulaire du *Cid* dans son entier, on s'aperçoit que la catégorie dont la distribution s'écarte le plus du modèle théorique est la catégorie verbale. Elle est suivie par la catégorie prépositionnelle. La tentation est grande de mettre en rapport ces deux phénomènes et de se demander si une évolution cohérente ne se dessinerait pas à travers l'œuvre modifiant l'emploi des modes verbaux en fonction de leur affinité avec les relateurs prépositionnels. Mais il se trouve que le schéma de distribution, pour le verbe (du plus au moins) est juste l'inverse du schéma exprimant la distribution prépositionnelle (du moins au plus). Autrement dit, c'est dans le chant le plus riche en occurrences verbales que les prépositions manquent, et vice-versa. Il est donc peu probable que l'évolution de la catégorie verbale soit en rapport direct avec l'évolution de la catégorie prépositionnelle et vraisemblable que l'évolution enregistrée dans l'emploi des prépositions les plus fréquentes affectera plus particulièrement d'autres catégories du vocabulaire, notamment celle du substantif. Toutes ces orientations seront reprises et contrôlées dans des travaux ultérieurs, le nombre d'occurrences, ne serait-ce que pour un vocable comme *de* (1485 unités), exigeant un recours à l'informatique. Mais d'ores et déjà il est clair que si la distribution des relateurs se modifie, c'est la relation même entre les éléments des autres catégories grammaticales qui se voit modifiée, et ce, à l'intérieur des structures syntaxiques de base, phrases et propositions. Autant dire que le profil de ces dernières évolue dans le

Poème, dans des conditions qui ne manqueront pas de surprendre si l'on restitue à ce phénomène le contexte qui est le sien : celui d'une prodigieuse stabilité prosodique et syntaxique. La longueur du vers, en effet, ne varie que dans les limites que fixe le jeu du hasard autour de la valeur 8,39 (8,34/8,50/8,35) ; quant à la longueur de la proposition [1], elle reste voisine de 6,27 mots (6,17/6,30/6,34), et toujours inférieure à la longueur du vers. Ces données permettent d'esquisser à grands traits les caractères syntaxiques du discours cidien. Si la fréquence d'occurrence des modalités conjuguées du verbe est identique dans les trois chants, alors que le chant 1 manifeste un excédent de verbes, et le chant 3 un net déficit, c'est que les modalités non conjuguées du verbe (appelons-les paranominales) ne sont pas distribuées d'une façon aussi régulière que les modalités conjuguées. Mais comme les prépositions sont sous-représentées dans le chant 1, il n'existe pas un lien direct entre la diminution des modalités paranominales et l'emploi des prépositions. Tout au contraire, c'est dans une conjoncture d'effacement des modalités paranominales du verbe que la catégorie prépositionnelle renforce sa représentation. On peut donc penser que la nature même des verbes employés change et que les compléments introduits ou déterminés par relateurs augmentent, un substantif précédé d'une préposition pouvant, par exemple, « occuper la place » d'une modalité paranominale non précédée d'une prépositon. Ce processus de conversion peut également concerner la catégorie très variée des pronoms.

5. Hypothèses et perspectives.

Ainsi l'analyse des données premières du système fonctionnel du *Cid*, abordé par le biais des prépositions, met au jour quelques aspects fondamentaux du texte et contribue à clarifier les problèmes qu'il pose encore aux érudits. Si l'on s'en tient aux faits principaux, il apparait que vocables et formes prépositionnels renvoient obstinément à un état de langue qui coïnciderait avec celui que d'autres documents attestent pour la première moitié du XIII[e] siècle. Sur ce plan, le *Cid* présente une indubitable homogénéité lexicale, formelle et statistique. Il va de soi, néanmoins, que des confrontations avec d'autres Vocabulaires complets, établis suivant des normes comparables à celles qui ont été retenues pour le Poème, seraient les bienvenues ; elles aideraient à préciser la constance des rapports entre classes de fréquences, l'impact du genre, du milieu, de la tradition, les influences

1. Au sens de ' construction reposant sur un verbe conjugué '.

dialectales (s'il y en a)... En particulier, des comparaisons devraient être effectuées entre genres proches (épopée en tant que poésie / autres poésies ; épopée en tant que genre et ensemble thématique, stylistique, etc. / prosifications de l'épopée), entre textes suffisamment longs de toutes époques (*Cid* / XII^e^ ; *Cid* / première moitié du XIII^e^ ; *Cid* / seconde moitié du XIII^e^ ; *Cid* / premier tiers du XIII^e^ ; *Cid* / deuxième tiers du XIII^e^ [1]), entre textes castillans — ou jugés tels — et textes non castillans [2].

Provisoirement, étant donné que le système possessif et le vocabulaire en général renvoient eux aussi à l'état du castillan dans la première moitié du XIII^e^ siècle, nous adopterons cette référence diachronique comme hypothèse de travail. Cela revient à poser une nouvelle synchronie, avec toutes ses implications, linguistiques, historiques, littéraires. Les études comparatives proposées fourniront peut-être bientôt les moyens de restreindre cette synchronie. Mais d'ores et déjà il est possible d'envisager une autre démarche, s'appuyant sur une radiographie informatique multipolaire du texte. Elle se fonderait sur un certain nombre de constatations convergentes (dont le détail sera exposé ailleurs [3]) et à la fois contradictoires : la distribution des catégories grammaticales, les caractères de la syntaxe contrastent avec l'homogénéité que présentent d'autres aspects du texte (cf. *supra*). Le lexique, non encore étudié, manifeste, dans sa pauvreté, une grande hétérogénéité d'un chant à l'autre. On en arrive alors à s'interroger sur la nature de l'œuvre en relation avec le temps. La synchronie plus haut évoquée, d'une part, les contraintes du genre épique, de l'autre, suffisent-elles à expli-

1. Les universités de Poitiers et de Villetaneuse, au moins, vont procéder, dans les années qui viennent, à l'analyse informatique de la *General Estoria*, de la *Primera Crónica General* et d'autres documents du Moyen Age, afin de constituer des thesaurus de référence.

2. Malgré tout ce qui a pu être affirmé sur la nature profondément castillane du *Cid*, des chercheurs non prévenus découvrent de plus en plus d'éléments non castillans dans ce texte. Ce fait ne prouve évidemment pas que le document ne soit pas castillan. Il rappelle simplement que les langues péninsulaires médiévales n'étaient pas aussi cloisonnées, ni fixées, qu'on affecte trop souvent de le croire aujourd'hui, pour des raisons étrangères à la recherche linguistique. La connaissance du document ne peut qu'avancer par une identification toujours plus poussée des éléments marginaux ou clairement étrangers du Poème (emprunts, contamination de systèmes en contact, etc.). Pour le linguiste, un élément aragonais et un élément castillan ne s'inscrivent pas dans deux systèmes de valeurs différents (et qui à la limite s'excluraient), mais dans deux registres, deux réalités qui se sont sans doute mutuellement influencés.

3. Cf. article sur les caractères statistiques du Vocabulaire du *Cid* mentionné note 1, page 26.

quer les traits d'homogénéité ? Peut-on, inversement, assigner à une même synchronie soit historique (cadre temporel) soit psychologique et mentale des actualisations aussi divergentes d'un système linguistique donné dans les limites d'un document de 3730 vers ? On se heurte, en apparence, à deux ordres d'observations qui normalement tendraient à s'exclure.

Cette contradiction pourrait se résoudre (théoriquement) en imaginant, à l'intérieur de la synchronie large de la première moitié du XIII^e, une diachronie de moments distincts, et non seulement d'instants (durée nécessaire à tout acte linguistique, à la composition d'une œuvre, etc., relativement brève pour un texte court) ou d'époques (décalages entre synchronies historiquement opposables). Pour que cette hypothèse soit utile, le chercheur se doit, bien sûr, de renoncer à toute une tradition et à des préférences intellectuelles ou sentimentales. Mais d'un autre côté, il peut se laisser séduire par les diverses interprétations qui découlent de cette structuration du temps à plusieurs degrés. Admettre, par exemple, que les chants ne constituent pas nécessairement les chapitres d'une œuvre conçue initialement comme une œuvre unique et homogène. Ou bien supposer que l'œuvre initialement conçue par un auteur unique comme une œuvre homogène a subi des retouches, voire de véritables refontes — plusieurs générations de poètes y participant, chacun en fonction de ses goûts, de ses idées, mais dans le cadre d'un système linguistique, d'une idéologie, d'une tradition littéraires donnés. Ou bien encore considérer le document que nous connaissons comme la surface émergée de l'iceberg ; ne serions-nous pas victimes d'une illusion historique lorsque nous cherchons à interpréter comme l'ensemble d'une œuvre ce qui n'en constitue qu'une partie ?... Chacune de ces approches possibles peut s'étayer d'arguments solides. En tout cas, une chose paraît certaine : devant l'accumulation des faits qui ne se laissent pas dater de la première moitié du XII^e siècle, nul ne devrait plus attribuer automatiquement comme par le passé, la date de c. 1140 à tout élément linguistique du *Cid*. Une révision de la chronologie espagnole médiévale s'impose ; avec elle sans doute, un bouleversement d'habitudes et de conceptions. Tout renouvellement est un peu une rupture ; mais si nos connaissances de la langue et de la culture médiévales acquièrent par là plus de cohérence, si le déséquilibre ainsi provoqué oblige à d'autres remises en question, on a tout intérêt à explorer les directions indiquées ci-dessus, quitte à les abandonner si l'expérience les condamne ou à les infléchir en fonction des obstacles rencontrés. Les faits demeureront, qu'il faudra expliquer. Les résultats nouveaux à leur tour susciteront des hypothèses dont le mérite ne sera

pas seulement d'expliquer d'une façon plus adéquate la réalité observée, mais probablement d'inspirer des recherches jusque-là négligées ou insoupçonnées, tant il est vrai que la recherche engendre sans cesse de nouveaux objets de recherche — avec eux, des problèmes nouveaux — et que notre connaissance des langues ibériques du Moyen Age n'a guère dépassé le stade embryonnaire.

Université de Poitiers. René PELLEN.