



Recherche d'un lexique pour la denomination de commandes en messagerie electronique

D. Tasset

► To cite this version:

D. Tasset. Recherche d'un lexique pour la denomination de commandes en messagerie electronique. RT-0032, INRIA. 1984, pp.89. inria-00070125

HAL Id: inria-00070125

<https://inria.hal.science/inria-00070125>

Submitted on 19 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



CENTRE DE ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
BP 105
78153 Le Chesnay Cedex
France
Tél. (3) 954 90 20

Rapports Techniques

N° 32

RECHERCHE D'UN LEXIQUE POUR LA DÉNOMINATION DE COMMANDES EN MESSAGERIE ÉLECTRONIQUE

Daniel TASSET

Janvier 1984

Groupe de Psychologie Ergonomique de l'INRIA

**RECHERCHE D'UN LEXIQUE POUR LA DENOMINATION
DE COMMANDES EN MESSAGERIE ELECTRONIQUE**

Daniel TASSET

Janvier 1984



PAPIER RECUPERE ET RECYCLE

SOMMAIRE

INTRODUCTION	7
I - METHODOLOGIE	9
1. Les fonctions	9
2. Principe de la méthode	10
3. Le dictionnaire	10
4. Les termes employés	11
5. Les synonymes	11
II - RESULTATS	14
1. Les matrices booléennes	14
2. Les graphes	14
3. Cheminement dans les graphes	16
4. Caractéristiques des graphes	17
4.1 Graphes fortement connexes	18
4.2 Eléments particuliers	18
a) élément minimal	18
b) premier élément	18
c) élément maximal	20
4.3 Cas particulier de graphe	21
5. Configurations graphiques	22
6. Réunion de tous les graphes	22
7. Recherche de groupes sémantiques disjoints	24
8. Niveau de généralité des termes	27
CONCLUSION	30
BIBLIOGRAPHIE	33
ANNEXES	
1. Liste des 25 fonctions	35
2. Données quantitatives pour chacune des 25 matrices booléennes ..	41
3. Représentations graphiques des 25 graphes	45
4. Liste pour chaque graphe des éléments minimaux et maximaux	73
5. Intersections lexicales entre matrices deux à deux	77
6. Représentation graphique du T-graphe	81

Résumé

La conception des langages de commande d'ordinateur se heurte à un certain nombre de difficultés, notamment en ce qui concerne la sélection des termes parmi les candidats possibles à la dénomination. Notre objectif est de fournir au concepteur un outil sous la forme d'une base de données lexicales qui lui permette de faire des choix de dénomination plus systématiques. Dans cette perspective, nous avons utilisé une méthode de recherche lexicographique consistant de manière itérative et systématique, pour une fonction donnée, à rechercher des synonymes dans un dictionnaire. Nous avons obtenu un lexique que nous avons étudié et décrit, et dont nous avons regroupé les termes en fonction de leurs relations au sein des graphes de synonymie. Le problème de la validité psychologique de ces résultats se trouve posé en termes de niveaux de généralité des éléments du lexique, de degrés de proximité sémantique et de prototypie, d'ambiguïtés et de regroupements en catégories sémantiques disjointes.

Mots-clefs : Langages de commande, lexique, vocabulaire, synonymie.

Abstract

When designing computer command languages many difficulties arise, particularly with regard to the selection of a word among other possible ones. Our aim is to provide the designer with a tool consisting of a lexical data base allowing a more systematic selection for the naming. For that purpose we used a lexicographical search method, iterative and systematic, to obtain, for a given function, all synonyms found in a dictionary. The obtained lexicon was studied and described, the words of the lexicon have been grouped according to their relations within graphs of synonymy. The problem of the psychological validity of these results is stated in terms of levels of generality of the lexical elements, degrees of semantic proximity and prototypicality, ambiguities and clusterings in semantics categories without overlapping.

Key words : command languages, lexicon, vocabulary, synonymy.

INTRODUCTION

Lors de la conception des langages de commande en langue naturelle restreinte, se pose le problème de la dénomination des objets et des fonctions.

Quelles appellations faut-il donner à ces objets et ces fonctions pour que leur signification paraisse claire et non ambiguë aux opérateurs, et qu'elles soient bien rappelées?

Si les chercheurs s'intéressent aux problèmes psychologiques que pose l'utilisation d'un langage de commande en langue naturelle restreinte (Barnard et al. 1982; Carroll, 1980; Scapin, 1982), peu de travaux concernent la constitution d'un vocabulaire pour une application informatique particulière; Landauer et al. (1983), cités par Branscomb et Thomas (1983), montrent que, pour une commande donnée, la probabilité de "deviner spontanément" le nom attribué à cette commande est faible; ils suggèrent pour remédier à cela de constituer un répertoire de synonymes. Quant à Rosenberg (1982), il propose pour le concepteur une méthode simple d'élaboration d'un vocabulaire adapté aux utilisateurs. Cette méthode consiste, en premier lieu, à évaluer le "caractère évocateur" (suggestiveness) d'un ensemble de termes de commande; ceci à l'aide d'une technique de jugement sémantique avec présentation des mots par triades; les justifications données par les sujets permettent de représenter les mots par un ensemble de "traits" (features), à partir desquels est estimée, par l'application d'un modèle de traitement mathématique élaboré par Tversky (1977), la similitude sémantique entre deux mots; une seconde expérience a pour objectif d'évaluer la valeur prédictive du concept de "caractère évocateur", ceci à l'aide d'une technique d'appariement mettant en correspondance les verbes et les actions.

Il nous a donc paru essentiel d'analyser les problèmes d'ordre lexicaux et sémantiques liés à la conception de vocabulaires spécifiques à une utilisation particulière, et de nous donner les moyens de constituer un vocabulaire à partir du lexique général du langage naturel (1).

(1) En linguistique, on oppose couramment "lexique" à "vocabulaire"; le lexique concerne l'ensemble des unités de la langue, c'est-à-dire ce qui est commun à tous les locuteurs de cette langue; le vocabulaire concerne les unités de la parole, c'est-à-dire ce qui est effectivement utilisé par un ou plusieurs locuteurs. Le terme "vocabulaire" est notamment employé dans le cas d'études portant sur des corpus spécialisés (e.g. vocabulaire politique, vocabulaire de l'informatique...).

Ces travaux nous amèneront à l'élaboration d'un document de synthèse pouvant prendre la forme d'une base de données lexicales, ceci afin :

- d'offrir au concepteur de langages de commande une base de travail pour la dénomination : ce document lui permettra de choisir un ou des termes du langage naturel non plus seulement à partir de sa propre intuition de la langue, ou encore à partir d'un vocabulaire spécifique à son activité professionnelle, mais en fonction des relations sémantiques qui, pour les utilisateurs, existent entre les termes.
- de constituer les bases de connaissances d'un système qui offrirait à l'utilisateur un dialogue plus flexible avec la machine : plusieurs termes de commandes peuvent être utilisés pour désigner une même fonction, ce qui évite d'imposer à l'utilisateur la mémorisation d'un terme unique pour désigner une fonction du système.

Les travaux présentés dans ce rapport constituent une première étape dans la réalisation de notre objectif : celle de recherche lexicographique.

I - METHODOLOGIE

Nous avons restreint notre étude à un domaine particulier d'application des systèmes informatiques, celui de la bureautique. Plus spécifiquement, nous avons choisi des systèmes de messagerie électronique français. Cela nous a amenés à considérer les cinq messageries suivantes :

- AGORA-KAYAK : Réalisée dans le cadre du projet pilote Kayak à l'INRIA, projet de construction d'un buroviseur comprenant l'automatisation d'un ensemble de tâches de bureau, cette messagerie est la plus complète des cinq, celle qui offre le plus de possibilités, notamment la constitution de téléconférences.
- AGORA-TELETEL : Il s'agit du service messagerie dont le maître d'oeuvre est TELESYSTEMES, inclus dans l'ensemble des services proposés au grand public dans l'expérience de télématique pour la ville de VELIZY.
- AGORA-SIRIUS : Système conçu dans le cadre d'un groupe de recherche de l'INRIA sur le thème des bases de données réparties, le projet pilote SIRIUS.
- MISSIVE : Système mis en service par les P.T.T. à titre expérimental fin 1981; les termes du langage de commande de cette messagerie sont traduits de l'anglais.
- PAPYRUS : Expérience de messagerie répartie, conçue par des chercheurs de l'INRIA, à l'Université de Grenoble (Cf. Richy, Quint et Sasyan, 1981).

1 - Les fonctions

Nous avons relevé l'ensemble des fonctions permises par les différentes messageries pré-citées, puis sélectionné les fonctions à partir desquelles sera effectué le recensement lexicographique. Nous n'avons pas retenu certaines fonctions qui, selon nous, étaient soit trop spécifiques d'un système, soit trop étrangères à l'activité de messagerie (e.g. la fonction Idées concernant l'envoi de suggestions d'amélioration du système); en outre, nous avons écarté tout ce qui concerne les procédures de connection et d'arrêt (log in/out).

De plus nous avons regroupé certaines fonctions qui, selon nous, étaient similaires sous des termes plus généraux : ainsi, "demander la liste des abonnés

dont le nom commence par une lettre donnée", et "demander la liste des abonnés dont le nom est DUPONT", sont regroupées dans une même fonction plus générale consistant en une recherche sélective d'un ou plusieurs objets dans un ensemble. La liste des 25 fonctions de départ pour la recherche lexicographique est donnée en annexe 1.

2 - Principe de la méthode

L'objectif de notre travail est de recenser l'ensemble le plus complet possible de termes du langage naturel pour dénommer une fonction donnée, et ceci pour chacune des fonctions de départ sélectionnées. Pour remplir notre objectif, nous avons utilisé une démarche proposée par Bisseret et Scapin (1980) lors d'une étude ergonomique du langage de commande de la messagerie AGORA-TELETEL. Cette technique permet le recensement du plus grand nombre possible de termes ayant le même sens (synonymes), ou un sens voisin, ceci à l'aide d'un dictionnaire général de la langue française.

La démarche comporte plusieurs étapes :

- 1) A chacune des fonctions recensées, on fait correspondre un terme de commande; ce terme est le mot de départ, pour la recherche lexicographique.
- 2) On cherche dans le dictionnaire l'ensemble des mots synonymes, ou de signification proche, de ce mot de départ.
- 3) Pour chacun des synonymes relevés à l'étape précédente, on applique de nouveau la même procédure qu'en 2 et ceci de manière itérative, jusqu'à l'arrêt de la recherche lexicographique.

A partir d'un mot, on obtient donc un ou plusieurs synonymes puis le(s) synonyme(s) de ce(s) synonyme(s), etc.

3 - Le dictionnaire

Nous avons préféré un dictionnaire général de la langue française à un dictionnaire de synonymes pour deux raisons :

- le premier a l'avantage sur le second de donner la définition du terme, ainsi que des nuances de sens, emplois et exemples d'emplois différents.
- le second a pour inconvénient, par rapport à notre méthode, d'opérer une sélection des entrées lexicales.

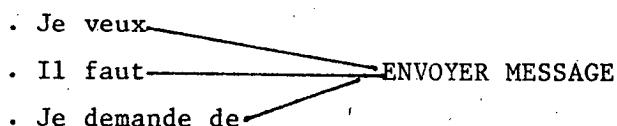
Le dictionnaire que nous avons choisi est le "Petit Robert", ou "Dictionnaire alphabétique et analogique de la langue française" (1978). Il propose pour chaque entrée lexicale "un éventail de mots synonymes dans un ou plusieurs emplois, ou de sens voisin" (Cf. préface de P. Rey, in Petit Robert 1978, page XVI), privilégiant ainsi les relations entre les mots.

4 - Les termes employés

Les termes retenus sont de deux types :

- un verbe à l'infinitif, sous la forme non-pronominale (ex. : ENVOYER) ou pronominale (ex. : SE JOINDRE).
- une locution verbale, c'est-à-dire un groupe de mots formant une unité significative (ex. : METTRE A PART), également à l'infinitif.

La forme infinitive est la plus courante pour désigner un verbe en français (c'est aussi celle utilisée par le dictionnaire); de plus, cette forme grammaticale exprime bien une situation de commande du type :



Pour chaque fonction, nous choisissons un terme de départ, de préférence dans les langages de commande existant dans les messageries étudiées; ces mots sont soit utilisés par une ou plusieurs, voire la totalité des messageries, soit des mots relevés dans les définitions données dans les manuels d'utilisation.

Le mot de départ de la recherche lexicographique doit décrire suffisamment bien la fonction qu'il désigne, mais cela ne permet pas d'accorder à ce terme de départ un statut privilégié par rapport aux mots relevés dans le dictionnaire. Nous verrons plus tard que chacun des termes retenus pour une fonction donnée peut constituer le départ d'une procédure de recherche des autres mots obtenus pour la même fonction, dans le graphe qui relie ces mots entre eux.

5 - Les synonymes

Au sens large, la synonymie est définie de la manière suivante : "sont synonymes des mots de même sens, ou approximativement de même sens, et de formes différentes" (Dictionnaire de linguistique, Dubois et al., 1973).

Si l'on consulte le Petit Robert (1978), nous trouvons une définition du même type que la précédente : l'adjectif "synonyme se dit de mots ou d'expressions qui ont le même sens ou une signification très voisine".

La synonymie est une relation sémantique très dépendante du contexte; ce contexte est, dans notre travail, l'utilisation d'un langage de commande pour dialoguer avec un système informatisé. Nous appliquerons le concept de synonymie à des mots ou locutions verbales qui ont le même sens ou une signification voisine, dans les limites de ce contexte.

Comment décider qu'un mot fera partie de la liste des synonymes, et pas un autre? A partir de quand considérer qu'il s'agit d'un sens différent?

La recherche lexicographique s'arrête lorsque l'on se trouve dans l'un des trois cas suivants :

- ou bien il n'y a pas de synonymes;
- ou bien le mot a déjà été obtenu antérieurement;
- ou bien le mot est à rejeter.

En ce qui concerne ce dernier cas, nous nous sommes donné plusieurs critères de rejet :

- tout d'abord, le rejet des verbes que nous jugeons peu familiers, d'usage peu fréquent ou ancien, dans notre contexte; l'exemple que nous citons est celui du verbe RECOLER, qui signifie "passer en revue, vérifier et pointer sur inventaire...";
- ensuite, le rejet des verbes ou expressions argotiques; par exemple, parmi les synonymes du verbe PARTIR, nous rejetons SE TAILLER, METTRE LES VOILES, FOUTRE LE CAMP;
- de plus, au cours de la recherche lexicographique dans le dictionnaire, nous prenons en compte deux entrées lexicales, celle du verbe dans sa forme infinitive (par exemple : COMPTER) et celle du substantif verbal, s'il y en a un, correspondant à ce verbe (COMPTE); le critère de rejet spécifique à ce dernier type d'entrée lexicale consiste à écarter les substantifs que l'on ne peut pas transformer en verbe : ainsi, en ce qui concerne l'entrée lexicale COMPTE, nous écartons des mots tels que MONTANT, STATISTIQUE, BILAN.

Lorsque les critères que nous venons de citer ne s'appliquaient pas, dans certains cas, la décision de rejet était facile à prendre; ainsi l'exemple le plus caricatural concerne un synonyme de EMETTRE; en effet, si l'on suit la méthode que nous avons décrite, nous partons de ENVOYER, puis ADRESSER, puis

EMETTRE, et parmi les synonymes de ce dernier, nous trouvons le verbe EJACULER; dans les limites du contexte que nous nous sommes donné, ce verbe n'est pas synonyme des précédents. Dans d'autres cas, ce fut plus difficile. Il nous fallait trouver un moyen pour placer chaque mot-candidat dans le contexte des langages de commande. Le moyen le plus simple nous a paru de placer le verbe-candidat dans une phrase du type :

SUJET - VERBE - OBJET

Exemple : Je veux - Envoyer - Message

L'objet est une entité manipulable par une fonction donnée. Quand plusieurs objets sont manipulables par une même fonction donnée, le verbe testé est appliqué à ces plusieurs objets différents. Cela revient à se demander si tel verbe peut être applicable à tel objet.

Dans tous les cas, c'est nous qui avons pris la décision finale concernant la possibilité d'appliquer un verbe à un objet, la similitude des actions décrites, l'acceptabilité des phrases-tests. Notre propre "intuition" (ou connaissance) de la langue française a pris une certaine part dans certains jugements d'acceptabilité ou de rejet.

II - RESULTATS

Un programme implémenté sur un microordinateur HP 9845B et écrit par D.L. Scapin nous a permis d'effectuer un certain nombre de manipulations sur les données.

1. Les matrices booléennes

Pour chacune des 25 fonctions de départ la relation binaire : "le mot X a pour synonyme le mot Y" est représentée sous la forme d'un tableau à double entrée. Nous avons donc 25 matrices booléennes.

Nous obtenons pour chacun des mots d'une matrice donnée :

- le nombre de synonymes de ce mot, (c'est-à-dire ses "descendants");
- le nombre de fois que ce mot a été obtenu comme synonyme à partir d'autres mots de la matrice (c'est-à-dire le nombre de ses "ascendants").

Le nombre des descendants est variable d'un mot à l'autre, certains mots n'en ayant aucun, d'autres en ayant une dizaine voire plus; le nombre d'ascendants est également très variable d'un mot à l'autre. Le nombre de mots par matrice est également très variable, de 7 à 30; si on additionne les nombres totaux de mots des 25 matrices, on obtient un total général de 427, soit en moyenne 17 par matrice.

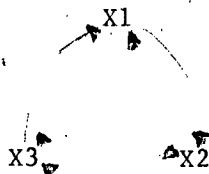
On donne en annexe 2 les résultats matrice par matrice; les matrices sont classées par ordre alphabétique du mot de départ et elles sont numérotées de 1 à 25.

2. Les graphes

Nous avons utilisé des concepts de la théorie des graphes (Berge, 1967; Roy, 1969) pour présenter les données; la représentation graphique facilite la lecture des données lorsque celles-ci sont organisées suivant un modèle de type relationnel. Afin de nous aider dans la construction des graphes, le programme nous communiquait, pour chaque matrice :

- l'ensemble des "cliques à 3 sommets" : une clique est un pré-ordre complet et symétrique, un pré-ordre étant un graphe transitif et réflexif (on

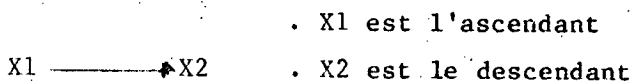
admet la réflexivité de la relation de synonymie (1), mais nous ne dessinerons pas les boucles sur les graphes, ceci pour en faciliter la lecture)



- l'ensemble des "cliques à 2 sommets", ou arcs symétriques, sauf ceux qui étaient déjà cités parmi les cliques à 3 sommets.



- l'ensemble des arcs orientés du graphe



Nous donnons en annexe 3 les représentations graphiques des 25 graphes. De la même façon que pour les matrices, nous désignerons les graphes par leur numéro d'ordre dans le classement alphabétique. Ainsi, la matrice 13 est représentée par le graphe G 13.

Nous avons tenté de faire en sorte que la disposition spatiale des sommets et des arcs assure aux graphes la meilleure lisibilité possible; le nombre parfois élevé de sommets et d'arcs de certains graphes nous a rendu la tâche difficile. Pour faciliter la lecture des graphes, nous avons représenté de manière différente les arcs non-symétriques et les arcs symétriques. Alors que les premiers sont représentés par une flèche reliant un sommet initial à un sommet terminal, nous n'avons pas représenté les seconds par une double flèche, afin de ne pas alourdir le dessin; cependant il ne faut pas lire le trait unique et non fléché reliant 2 sommets comme un arc non orienté mais comme un arc symétrique. Tous les arcs des graphes de synonymie restent des arcs orientés.

(1) Nous incluons dans la définition que nous donnons plus haut de la synonymie l'identité formelle des items lexicaux.

Prenons pour exemple le cas d'une matrice simple, la matrice 14 qui correspond au mot de départ ETIQUETER.

	1	2	3	4	5	6	7	8
	ETIQU	DENOM	DESIG	NOMME	APPEL	QUALI	TITRE	INTIT
1 ETIQUET	*							
2 DENOMME	1	*	1	1	1			
3 DESIGNE	1	1	*	1	1	1		
4 NOMMER		1	1	*	1	1		
5 APPELER		1	1	1	*	1		
6 QUALIFI			1	1	1	*	1	
7 TITRER						1	*	1
8 INTITUL							1	*

A chacun des 8 mots de la matrice nous associons un numéro. Si nous décidons, tout à fait arbitrairement, de partir du mot NOMMER, nous définissons une application θ_4 qui consiste à reproduire tous les 1 existant en ligne 4 sur chacune des lignes h qui possèdent un 1 en colonne 4; c'est-à-dire que tous les synonymes du mot NOMMER seront cités comme synonymes par les mêmes mots qui ont cité comme synonyme NOMMER.

Ou encore, dans le langage de la théorie des graphes, tout descendant d'un mot aura les mêmes ascendants que ce mot.

Dans notre exemple, l'application θ_4 amène à inscrire un 1 en ligne 2 et colonne 6, et un autre 1 en ligne 6 et colonne 2. On choisit ensuite une autre application jusqu'à avoir effectué les 8 applications. La matrice finale, $\mathcal{Z}(G)$ ne renferme que des 1, sauf sur la ligne 1 qui ne comporte que des 0.

4. Caractéristiques des graphes

La méthode décrite permet de regrouper les graphes en fonction des caractéristiques suivantes :

4.1 Graphe fortement connexe

Un graphe est fortement connexe si la matrice $\mathcal{Z}(G)$ renferme uniquement des 1; les huit graphes vérifiant la propriété de connexité forte correspondent aux fonctions suivantes :

- G2 : ajouter un élément à un ensemble
- G3 : annuler une procédure prévue, ou une décision déjà établie
- G5 : changer une entité par une autre, "changer de ..."
- G6 : ajouter un commentaire sur un document que l'on a reçu
- G7 : consulter un document, une liste, etc.
- G11 : détruire un document
- G18 : quitter une réunion, un groupe
- G21 : répondre à une interrogation, une demande...

Cette propriété ne dépend pas du nombre de mots par graphe : ainsi par exemple, G2 est composé de 30 mots alors que G6 n'en comporte que 8.

Une hypothèse que l'on peut faire est que la propriété de connexité forte d'un graphe, sur un plan sémantique, est l'expression d'une plus grande similitude de sens entre l'ensemble des mots du graphe, qu'entre les mots constituant un graphe non fortement connexe.

4.2 Élément particulier

Parmi les autres graphes, 15 d'entre eux comportent un ou plusieurs éléments particuliers :

a) Élément minimal : c'est un élément qui n'admet aucun précédent dans le graphe. En d'autres termes, l'élément minimal d'un graphe ne comporte aucun ascendant dans ce graphe. Dans la matrice $\mathcal{Z}(G)$, la ligne correspondant au mot "élément minimal" ne comporte aucun 1 (si on exclut la diagonale).

Trois graphes comportent un élément minimal; ils correspondent aux fonctions suivantes :

- G1 : prendre un abonnement
- G4 : archivage d'un document
- G25 : présenter un document, l'afficher sur un écran

Les éléments minimaux sont respectivement S'ABONNER, ARCHIVER et VISUALISER.

b) Premier élément du graphe : si de plus cet élément minimal est unique et précède n'importe quel autre élément du graphe, on l'appelle "premier élément",

ou "élément minimum". C'est-à-dire qu'à partir du premier élément d'un graphe il existe au moins un chemin qui mène à l'un quelconque des autres éléments du graphe : d'où, si l'on exclut de ce graphe le premier élément, on obtient un sous-graphe fortement connexe (ce n'est pas le cas en ce qui concerne les graphes comportant un élément minimal). La matrice $\mathcal{Z}(G)$ renferme des 1 partout sauf dans la ligne correspondant au premier élément.

Les deux graphes comportant un premier élément sont les suivants :

- G14 : donner un nom à une entité, un document
- G19 : faire passer un document reçu à plusieurs destinataires

Les premiers éléments sont respectivement ETIQUETER et REDISTRIBUER.

On peut noter que les éléments minimaux et premiers éléments sont toujours des mots de départ de la recherche lexicographique.

Aussi bien les éléments minimaux que les premiers éléments sont des mots peu reliés aux autres mots du graphe auquel ils appartiennent; non seulement ils n'ont aucun ascendant, mais encore ils ont peu de descendants : VISUALISER a trois descendants, ETIQUETER en a deux, les trois autres éléments particuliers n'ayant qu'un seul descendant. Cela signifie-t-il qu'il est possible, dans un contexte donné, d'utiliser des mots tels que DENOMMER, DESIGNER, APPELER, etc. (pour prendre l'exemple du graphe G14), à la place du mot ETIQUETER, alors que l'inverse ne l'est pas? Certains linguistes distinguent à propos de la synonymie, une substitution "symétrique" ou équivalence entre les items lexicaux et une substitution "orientée", dans laquelle les items ne sont pas équivalents (Cf. Pêcheux et Fuchs, 1975, pp. 61, 66 et 75). C'est à partir de la notion de spécificité des termes que l'on pourrait interpréter la présence d'items lexicaux non équivalents, les éléments minimaux et premiers éléments, dans les graphes. En effet, si nous prenons l'exemple du graphe G4, ARCHIVER un ou plusieurs objets consiste effectivement à les CLASSER; les RANGER, pour les CONSERVER, mais pas n'importe où : en effet le verbe ARCHIVER désigne une action de classement "dans les archives" et pas ailleurs. L'action d'archiver est une action particulière de classement; de ce fait le verbe ARCHIVER n'est pas considéré par le lexicographe comme synonyme du verbe CLASSER, plus général. Si le contexte a pour effet d'opérer une sélection parmi les composants sémantiques des termes lexicaux, on peut légitimement penser que dans un contexte mettant l'accent sur l'objet "Archives" (par exemple un contexte de tâche administrative de bureau), ARCHIVER soit considéré par des sujets comme synonyme de mots tels

que CONSERVER, CLASSER, etc. Dans ce cas, le verbe ARCHIVER ne serait plus un élément minimal du graphe des relations de similitude sémantique entre les mots.

c) Elément maximal : C'est un élément n'ayant aucun suivant dans le graphe. Cet élément correspond au cas d'un sommet n'ayant aucun descendant dans le graphe. La présence d'un élément maximal dans la matrice $\mathcal{Z}(G)$ se traduit par une colonne totalement vide, ne comportant aucun 1 (si on exclut la diagonale). Un même graphe peut comporter plusieurs éléments maximaux.

Les treize graphes comportant un ou plusieurs éléments maximaux correspondent aux fonctions suivantes :

- G1 : prendre un abonnement
- G4 : archivage d'un document
- G9 : créer
- G10 : calculer le nombre d'éléments d'un ensemble
- G12 : écrire le texte d'un document
- G13 : envoyer un document à un destinataire
- G15 : imprimer, garder une trace écrite
- G16 : lire un document
- G17 : modifier le texte d'un document
- G20 : se joindre à un groupe, réunion...
- G23 : consultation sélective, partielle, d'un document, d'une liste
- G24 : s'assurer de la destination de documents envoyés
- G25 : présenter un document, l'afficher sur l'écran

Le nombre d'éléments maximaux par graphe varie de 1 à 6. Par exemple, le graphe G12 comporte 4 éléments maximaux qui sont : COPIER, ENREGISTRER, AGENCER, FORMULER. On trouvera en annexe 4, la liste, pour chacun des treize graphes cités, des éléments maximaux qu'ils renferment.

Parmi les 26 éléments maximaux recensés, on trouve 11 locutions verbales dont 6 sont classées parmi les éléments maximaux non pas parce qu'elles ne possèdent pas de synonymes mais parce que l'on ne trouve pas dans le dictionnaire d'entrée lexicale correspondant à ces locutions.

Ces 6 locutions sont les suivantes :

- dans le graphe G15 : LAISSER TRACE, LAISSER EMPREINTE
- dans le graphe G20 : SE FAIRE ADMETTRE
- dans le graphe G25 : FAIRE LIRE, FAIRE APPARAÎTRE, RENDRE VISIBLE

Mis à part le cas des 6 locutions verbales que nous venons de citer, on peut penser qu'un mot est élément maximal du graphe soit parce que sa signification est trop distante de la signification de l'ensemble des autres mots du graphe (exemple : SAVOIR dans le graphe G24), soit parce que ce mot est spécifique, ou bien peu fréquent dans la langue (par exemple AFFINER en G17, ou QUANTIFIER en G10). Dans un cas le mot peut comporter un nombre élevé de synonymes dans la langue, mais aucun ne "convient" pour désigner le concept étudié; dans l'autre cas le nombre des synonymes fournis par le dictionnaire serait plus faible. En situation expérimentale, les sujets auront peut-être tendance à éliminer le premier type de mot mais pas le second si la fréquence d'usage du mot est suffisante.

En ce qui concerne les ascendants des éléments maximaux, leur nombre est généralement de 1, parfois 2; un cas est à part, celui du verbe ORGANISER (graphe G9), qui est précédé par 6 ascendants, dans le graphe.

4.3 Cas particulier de graphe

L'application de la procédure de fermeture transitive nous a permis de mettre en évidence une structure particulière de graphe, que nous pouvons décrire ainsi : à l'intérieur du même graphe, deux groupes de sommets sont reliés par un ou plusieurs arcs, tous orientés et "incidents vers l'intérieur" du même groupe.

Deux graphes sont dans ce cas :

- G8 : corriger les erreurs dans un texte
- G22 : retirer un ou plusieurs objets d'un ensemble d'objets

Considérons par exemple le graphe G8 : sur les 18 mots de ce graphe, un premier groupe A comporte 11 mots tels que CORRIGER, AMELIORER, MODIFIER, etc.; un second groupe B comporte 7 mots tels que RETIRER, OTER, ENLEVER, etc. Deux arcs orientés relient le groupe A vers le groupe B.

Ces deux arcs sont les suivants :

- . CORRIGER —————> SUPPRIMER
- . RECTIFIER —————> FAIRE DISPARAITRE

Aucun arc n'a pour extrémité initiale un des sommets du groupe B et pour extrémité terminale un des sommets du groupe A. Cette structure particulière pourrait correspondre à une partition du graphe en deux groupements sémantiques distincts.

Si l'on regarde le graphe G22, le premier groupe contient 21 mots tels que RETIRER, OTER, EXCLURE, etc., et le second 6 mots centrés autour de DIMINUER; ici l'action décrite paraît semblable dans les deux groupes, il s'agit toujours de retirer un objet d'une liste; mais alors que le premier groupe de termes concerne l'action elle-même appliquée à un ou plusieurs objets de la liste, le second groupe concerne la conséquence de cette action : en effet, ENLEVER ou RETIRER des objets d'un ensemble a pour conséquence de RESTREINDRE, DIMINUER cet ensemble; les deux groupes se différencient sémantiquement par l'objet auquel s'applique l'action évoquée par les mots.

5 - Configurations graphiques

Si l'on considère les graphes du point de vue de leur forme, nous pouvons distinguer trois types de configurations graphiques :

- Type 1 : une configuration de type "rayonnante"; un même mot est en relation avec un grand nombre des mots du graphe; c'est le cas par exemple des graphes G10, G21, G25.
- Type 2 : une autre configuration intéressante est la "figure", formée d'un ensemble de plusieurs mots très reliés entre eux; c'est notamment le cas des graphes possédant des cliques à quatre ou à cinq sommets; les graphes les plus caractéristiques de ce type de configuration sont par exemple les graphes G9, G12, G14, G19.
- Type 3 : enfin dans un troisième cas, nous pouvons dire que le graphe est "sans relief", c'est-à-dire sans configuration particulière : les arcs sont répartis de façon quasiment uniforme sur l'ensemble des mots du graphe.

Cette distinction permet de formuler des hypothèses intéressantes en ce qui concerne le niveau de généralité des mots d'un graphe. Nous verrons plus loin quelles sont ces hypothèses.

6 - Réunion de tous les graphes

Nous souhaitons présenter une vue d'ensemble de tous les graphes, car certains mots appartiennent simultanément à plusieurs matrices.

Nous avons donc commencé par répertorier toutes les intersections lexicales entre matrices deux à deux (cf. annexe 5). La plupart des intersections n'ont

qu'un ou deux mots en commun mais le nombre le plus élevé de mots en commun entre deux matrices est égal à 17.

Un graphe a été élaboré pour représenter les 25 matrices liées par la relation "X a des éléments communs avec Y". Cette relation est bien sûr symétrique.

Ce graphe est composé d'une juxtaposition de 5 sommets isolés, un sous-graphe à 4 sommets et 4 arcs et un autre sous-graphe rassemblant les autres sommets et arcs; il permet d'aboutir à la représentation des mots et des relations entre ces mots telle que chacun des mots n'apparaisse qu'une fois sur le dessin de l'ensemble des graphes, même si ce mot appartient à deux ou plusieurs graphes différents.

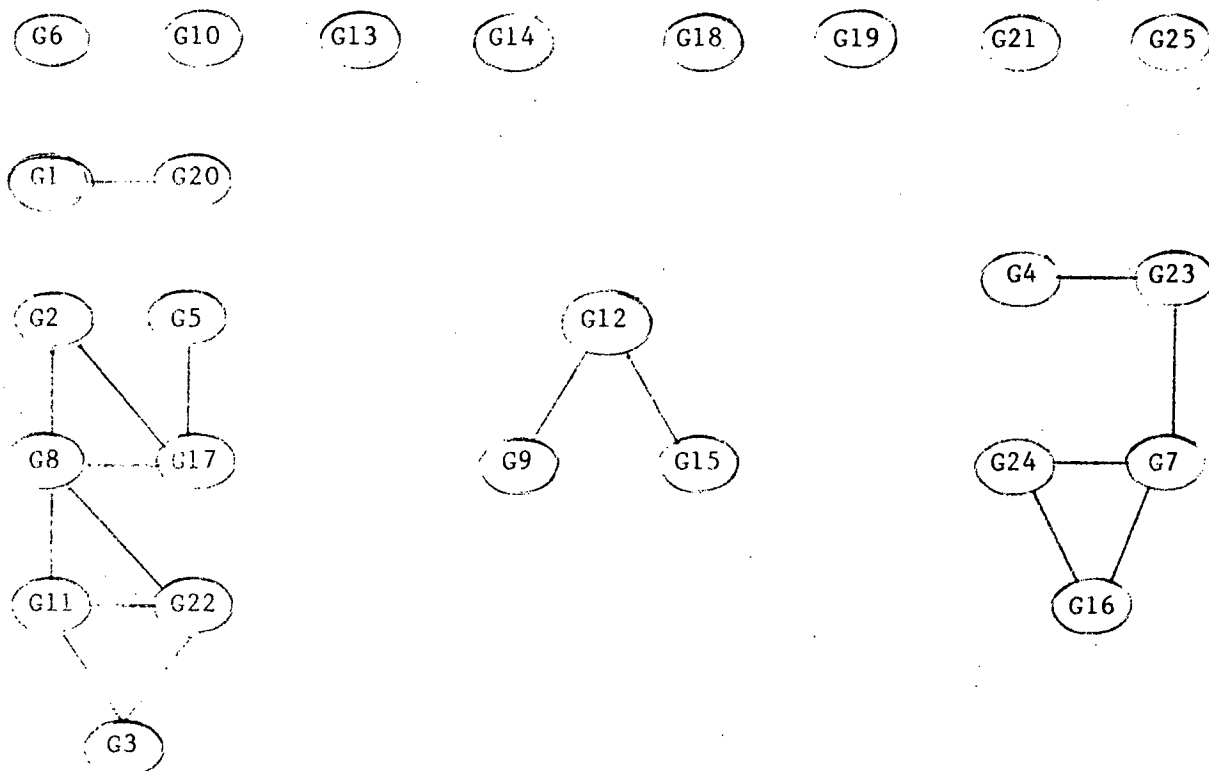
Nous avons cependant remarqué que dans certains cas, les items communs à deux graphes différents n'ont pas le même sens selon qu'ils sont dans un graphe ou dans un autre. C'est le cas pour les items lexicaux suivants :

- . 1'item INSPECTER, entre les graphes 7 et 24
- . 1'item ADDITIONNER, entre les graphes 2 et 10
- . 1'item JOINDRE, entre les graphes 2 et 20
- . 1'item SUPPRIMER, entre les graphes 3 et 8 ainsi qu'entre les graphes 3 et 11
- . 1'item PRODUIRE, entre les graphes 9 et 25, ainsi qu'entre les graphes 12 et 25
- . les items APPELER et DESIGNER entre les graphes 14 et 23
- . les items DETACHER et SEPARER entre les graphes 22 et 23

Par exemple, le verbe JOINDRE a d'après le dictionnaire le sens de "Joindre à, mettre avec..." dans le graphe G2 (Ajouter un objet à un ensemble d'objets) et le sens de ATTEINDRE, REJOINDRE... dans le graphe G20 (Se joindre à une réunion, un groupe, etc.).

Ou encore, le verbe APPELER a le sens de "Donner un nom" dans le graphe G14 et le sens de "Demander quelque chose" dans le graphe G23. Cela nous amène, pour les cas cités ci-dessus, à faire apparaître sur le dessin du graphe autant de fois le même mot qu'il a de sens différents. Ceci change la structure du graphe des interactions, certaines des interactions n'existant plus : 7 arcs sont supprimés, ce qui amène à 8 (au lieu de 5) le nombre de graphes isolés et à 4 (au lieu de 2) le nombre de groupements de graphes.

Ce graphe des interactions entre les matrices est le suivant :



La représentation graphique de l'ensemble des 25 graphes reliés entre eux par les recouvrements lexicaux est donnée en annexe 6.

Nous désignerons cette représentation par T-Graphe (pour Tous les Graphes).

Ce T-Graphe ne comprend que les relations symétriques, ceci afin de ne pas le rendre illisible; cependant, nous avons représenté les arcs non-symétriques chaque fois qu'un sommet se trouvait isolé, ou qu'aucune relation symétrique ne liait deux groupes de mots.

7 - Recherche de groupes sémantiques disjoints

A partir de cette représentation, nous cherchons à obtenir si possible un ensemble de groupes disjoints de mots, chacun des groupes représentant un concept différent.

Les observations que nous avons menées sur les graphes et le T-graphe nous ont amenés à distinguer deux manières d'obtenir des groupes disjoints.

- Tout d'abord; nous avons présenté précédemment (cf. paragraphe 4.3) le cas où, dans un graphe, deux groupes de mots sont reliés par un ou plusieurs arcs tous orientés et incidents vers l'intérieur du même groupe. Nous pouvons alors faire l'hypothèse selon laquelle tout graphe possédant ce type de structure sera constitué d'autant de groupes sémantiquement disjoints qu'il y aura de groupes de mots qui, dans le graphe, sont reliés de la façon précédemment décrite.

- Ensuite, en ce qui concerne le T-graphe, nous constatons que dans certains cas, le fait d'"enlever" un mot permet d'obtenir deux groupements disjoints.

Si on considère, par exemple, le groupement composé des 3 graphes G9, G12 et G15, nous constatons qu'en supprimant le verbe COMPOSER nous obtenons deux groupes distincts, l'un concernant l'action de Créer, Produire, Préparer, etc. et l'autre celle d'Ecrire, de Rédiger. De même, si nous enlevons le mot REPRODUIRE, deux groupes se trouvent disjoints, l'un concernant l'action d'Imprimer, et l'autre celle d'Ecrire.

Nous avons de cette manière dénombré 10 verbes susceptibles d'être enlevés de façon à former des groupes disjoints d'items; parmi eux 6 verbes forment le lien entre plusieurs graphes :

. ENTRER	distinction : G1 / G20
. APPELER	" : G23 / G16, G7, G24
. METTRE A PART	" : G23 / G4
. EXAMINER	" : G24 / G16, G7
. COMPOSER	" : G9 / G12
. REPRODUIRE	" : G12 : G15

En ce qui concerne les 4 autres verbes, leur suppression amène à distinguer plusieurs groupes au sein d'un même graphe :

. COMPTER (G10)	distinction : Calcul / Enumération
. ADDITIONNER (G2)	" : Ajouter / Augmenter
. INTERCALER (G2)	" : Ajouter / Insérer
. TRANSMETTRE (G13)	distinction : Envoyer (avec intermédiaire pour effectuer le transport) / Livrer, Donner (pas nécessairement d'intermédiaire pour le transport)

Dans un certain nombre de cas, la méthode de suppression d'un seul mot ne s'applique pas aussi aisément :

- soit qu'il faille enlever plusieurs mots pour obtenir des groupes dis-joints : c'est par exemple le cas du graphe G18, ou de regroupements tels que celui des graphes G3, G11, G22.
- soit que la structure du graphe est telle que si on enlève un mot "central" dans le graphe, la plupart des mots apparaissent isolés (Cf. par exemple G21).

Il est également possible d'appliquer la méthode de suppression non plus aux sommets mais aux arcs du graphe. En effet, si on prend l'exemple du graphe G2 sur le T-Graphe, il est possible qu'on enlève tout par le mot *INSERER* mais l'arc symétrique reliant les mots *JOINDRE* et *INTERCALER*.

Ces interprétations ne sont présentées ici qu'à titre d'hypothèse; nous pourrions proposer à des sujets un exemple de situation concrète pour chacun des groupes hypothétiques que nous avons distingué, et leur présenter une liste de mots en leur demandant quels mots employer pour décrire cette situation; par exemple, en ce qui concerne le groupe correspondant à l'opération d'insertion (graphe G2), on proposerait une situation consistant à introduire un nouveau client sur le fichier contenant la liste par ordre alphabétique des clients de l'entreprise. Dans ce cas, nous pouvons faire l'hypothèse que les sujets emploieront de préférence les mots correspondant à "Insérer un objet dans une liste" que ceux correspondant à "Ajouter un objet à un ensemble d'objets".

Une question intéressante concerne le nombre des mots et/ou des arcs qu'il faudrait enlever pour obtenir des graphes sémantiques disjoints. Ce nombre est vraisemblablement un bon indicateur du degré de proximité sémantique entre plusieurs catégories sémantiques.

Un autre centre d'intérêt concerne les caractéristiques mêmes des mots à enlever : l'appartenance de certains de ces mots à plusieurs groupes sémantiques dépendrait-elle de leur niveau de généralité? ou bien ces mots sont-ils tous des mots ambigus, c'est-à-dire, pour un même niveau de généralité, qu'ils peuvent désigner plusieurs opérations différentes, d'où un risque de confusion entre ces opérations?

La question est importante, le vocabulaire que nous voulons obtenir sera organisé de façon hiérarchique, un terme général permettant d'accéder à plusieurs groupes sémantiques distincts de mots plus spécifiques; il est donc essentiel d'éliminer les mots ambigus, les mots de même niveau de généralité pour lesquels on ne peut pas décider de leur appartenance à un seul et même groupe sémantique; par contre on conserve les mots généraux non ambigus entre eux.

Le nombre des descendants d'un même mot pourrait constituer un bon critère pour distinguer, sur le T-graphe, les mots spécifiques ambigus et les mots généraux.

Si l'on considère par exemple le groupement formé par les graphes G9, G12 et G15, le sommet COMPOSER possède 8 descendants, REPRODUIRE n'en possédant que 3, ce qui amène à penser que le premier est général alors que le second est plutôt spécifique, donc ambigu. Nous explorerons ce point plus en détail dans le paragraphe suivant consacré à l'étude du niveau de généralité des termes.

Enfin, il est possible d'obtenir des groupes sémantiques disjoints en utilisant une technique différente de celles que nous venons de présenter : il s'agit, non plus de formuler des hypothèses à partir d'observations menées sur le T-graphe, mais de partir de l'hypothèse selon laquelle il existe une relation entre la longueur du chemin qui relie deux mots entre eux sur le graphe, et la distance sémantique existant entre ces deux mots; nous chercherions, pour tout couple de mots du graphe, le plus court chemin reliant ces deux mots entre eux. A l'aide de techniques d'analyse combinatoire telles que celles utilisées en analyse des données, il serait alors possible de grouper les mots du graphe en fonction de la longueur du chemin existant entre ces mots; en attribuant à chacun des groupes de mots ainsi individualisés une structure de graphe, ceci en prélevant dans le T-graphe toutes les relations existant entre les mots composant ce groupe, nous nous donnerions une méthode permettant d'étudier ces graphes du point de vue de leur connexité : si un graphe est fortement connexe, nous pouvons penser que l'ensemble des mots du graphe forme un groupe sémantique.

8 - Niveau de Généralité des termes

Il serait possible de penser que les sommets comportant un nombre élevé d'ascendants et de descendants, ceux-ci n'appartenant éventuellement pas tous au même groupe sémantique, sont des termes généraux du lexique. Ainsi des mots tels que : CHANGER, MODIFIER, SUPPRIMER, pourraient être considérés comme généraux. Or du point de vue psychologique, le mot spécifique semble contenir plus de traits sémantiques que le mot général, une idée intéressante étant que le premier apporte plus d'informations que le second (Le Ny, 1976). Dans ce cas, on pourrait considérer, par exemple, le verbe ARCHIVER comme plus spécifique par rapport aux mots GARDER, CONSERVER (Cf. graphe G4); en effet, à l'action de conserver un document, classer ce document, ou le mettre à sa place, le verbe

ARCHIVER ajoute une information supplémentaire, la localisation : en l'occurrence, le lieu où l'on garde, conserve des documents.

Dans un contexte d'utilisation de système informatique, les sujets auront-ils tendance à choisir des termes généraux ou des termes spécifiques? On peut penser que cela dépendra en partie du niveau de précision du contexte.

Nous avons relevé dans le T-Grappe les mots dont le nombre des descendants est supérieur à 7.

Ces mots sont les suivants :

Nombre de descendants	Mots
10	COMPTER, DISTINGUER, FAIRE, REPANDRE
9	AUGMENTER, CHANGER, CORRIGER, EXAMINER
8	COMPOSER, CREER, GARDER, MONTRER, REPONDRE
7	ABANDONNER, MODIFIER, PARTICIPER, SUPPRIMER

Nous nous sommes demandés si ces 17 mots que nous venions de répertorier étaient des mots plus fréquemment employés dans la langue française.

Pour y répondre nous avons consulté un livre de statistique lexicale, celui de Gougenheim et al. 1964. Malheureusement les auteurs ne fournissent qu'un corpus limité à 1063 mots, recueilli à partir d'enregistrements oraux.

Cependant le fait de trouver le mot cherché dans ce corpus constitue déjà un bon indicateur de fréquence d'emploi du mot.

Sur l'ensemble des 17 mots répertoriés, nous en avons trouvé 7.

Mots	N° d'ordre dans le classement des 1063 mots (par ordre de fréquence décroissante)
FAIRE	19ème
CHANGER	295ème
COMPTER	323ème
REPONDRE	396ème
GARDER	431ème
MONTRER	534ème
CREER	957ème

Certains mots tels que COMPOSER, MODIFIER, etc. n'apparaissent pas. Par contre nous trouvons des mots situés en position très périphérique sur le graphe : par exemple, dans le groupe sémantique centré sur les mots FAIRE, COMPOSER, CREER, etc., alors qu'on ne trouve pas le mot COMPOSER, on trouve ORGANISER avec un numéro de classement de 836; alors qu'on ne trouve pas EXAMINER, on trouve dans la table lexicale les verbes VOIR (43ème), CHERCHER (231ème), mots pourtant plus périphériques que le mot EXAMINER. Il ne faut cependant pas oublier que les mots n'ont généralement pas la même fréquence d'usage dans la langue parlée et dans la langue écrite; le corpus analysé par les auteurs étant d'origine orale, les résultats pourraient être différents pour un corpus d'origine écrite.

CONCLUSION

Il s'agissait lors de cette première étape dans notre travail, de recenser tous les termes de la langue française "susceptibles" de convenir pour la dénomination d'un certain nombre de fonctions de systèmes automatisés en bureautique. Nous avons relevé ces termes du lexique dans un dictionnaire général de la langue française, ce qui nous a permis d'aboutir à un sous-ensemble du lexique dans lequel les mots sont reliés par une relation de synonymie. Puis nous avons donné une représentation graphique de ce lexique, représentation que nous avons appelé T-graphe.

Les questions que nous avons soulevées à partir de ce T-graphe ne concernent pas seulement les relations de synonymie entre les éléments du lexique, mais également l'organisation sémantique de ce lexique, les relations de généralité entre les mots. Ces questions nous permettent de formuler un ensemble d'hypothèses en vue d'études ultérieures à caractère expérimental.

Ainsi, nous explorerons le problème des niveaux de généralité des verbes du lexique recensé dans le dictionnaire. Ce qui caractérise un verbe général est la possibilité de dénommer, à partir du contenu sémantique de ce mot, un ensemble de plusieurs groupes sémantiques de mots, chacun de ces groupes correspondant à une action ou opération particulière. Selon que les synonymes d'un même mot appartiennent ou non à un même groupe sémantique, les hypothèses concernant ce mot sont différentes.

Un premier ensemble d'hypothèses concerne le niveau de généralité des verbes dans un lexique tel que celui que nous avons recensé. Si les synonymes d'un mot donné réfèrent à des actions différentes, alors ce verbe sera, ou bien "général" si le nombre de ses synonymes est élevé, ou bien plutôt "ambigu" s'il y a peu de synonymes (c'est-à-dire un mot plutôt spécifique).

Dans le cas où plusieurs verbes réfèrent à une même action ou opération constituant un groupe sémantique isolé, les hypothèses seront différentes selon le type de configuration graphique (cf. paragraphe II-5).

- configuration rayonnante (type 1) : le mot central est probablement le prototype (1), le meilleur représentant de la classe des verbes désignant une même action.

(1) Parmi les divers exemplaires d'une catégorie, certains sont plus "typiques", plus représentatifs de la catégorie; certains auteurs (cf. Rosch, 1978) désignent ces exemplaires par "prototypes". Ainsi l'exemplaire "moineau" sera pour la plupart des individus plus représentatif que l'exemplaire "autruche" de la catégorie des "oiseaux" (Pour une revue d'ensemble concernant la notion de prototypie, cf. Cordier et Dubois, 1981).

- configuration avec figure (type 2) : nous pouvons penser que les mots composant la figure centrale du graphe constituent ce que Rosch et Mervis (1975) appellent une catégorie sémantique de base, représentant un niveau optimal d'abstraction; par exemple, des termes tels que "voiture", "automobile", de même "bagnole" appartiennent à une catégorie de base; "véhicule" est un terme trop général, et "Renault 5" un terme trop spécifique. Ce type de configuration ne permet pas de formuler d'hypothèse en ce qui concerne la prototypie. Quant aux mots plus périphériques du graphe, ce sont des mots soit plus spécifiques, soit dont le sens est éloigné du concept central.
- dans le cas d'une configuration (type 3) sans relief particulier, tous les mots se situeraient à un même niveau de généralité, sauf peut-être les mots reliés seulement par des arcs orientés avec les autres mots du graphe (i.e., éléments minimaux, premiers éléments, éléments maximaux).

D'un point de vue ergonomique, il paraît essentiel de disposer pour une fonction donnée du mot le plus représentatif pour les utilisateurs, ainsi que, pour chaque groupe sémantique du vocabulaire, d'une évaluation du degré de proximité sémantique de chacun des mots par rapport au concept central. Le fait de relever des synonymes dans un dictionnaire ne permet pas d'affirmer qu'ils seront considérés comme synonymes au même degré par un ensemble de sujets; c'est ce que montrent, par exemple, Whitten et al. (1979) dans une étude où ils soumettent un ensemble de mots de la langue anglaise, de même fréquence, et prélevés dans un dictionnaire de synonymes, à une épreuve de jugement de similitude sémantique des mots présentés par paires : le degré de similitude observé pour chacun des couples de mots varie, sur une échelle à 7 points, de 2.24 à 6.79.

D'autres questions peuvent être soulevées en ce qui concerne l'existence d'une relation entre les propriétés des graphes telles que nous les avons décrites plus haut et l'organisation sémantique d'un ensemble de termes appliqués à un contexte particulier. Pour tout groupe sémantique isolé, c'est-à-dire tel que les verbes de ce groupe renvoient tous à la même action :

- le graphe associé à ce groupe est-il toujours fortement connexe?
- existe-t-il une relation entre la longueur du chemin entre deux mots du graphe, et le degré de proximité sémantique tel qu'on peut l'évaluer à l'aide de méthodes expérimentales en psycholinguistique : notamment, la

longueur du chemin entre deux éléments du graphe mesure-t-elle la distance sémantique entre deux items lexicaux? Peut-on identifier des frontières sémantiques entre concepts, par exemple en définissant sur le graphe une longueur maximale du chemin le plus court existant entre les deux sommets les plus éloignés d'un groupe sémantique donné?

Enfin, en ce qui concerne l'ensemble du lexique, notre objectif étant d'obtenir un ensemble de groupes sémantiques disjoints, chaque groupe correspondant à une action particulière, nous soumettrons à l'épreuve de la méthode expérimentale, l'organisation sémantique dégagée à partir de l'analyse du T-graphe.

Un certain nombre d'études à caractère expérimental permettront d'explorer les hypothèses auxquelles nous venons d'aboutir, l'objectif général étant, à partir d'un lexique, ou ensemble d'items lexicaux prélevés dans un dictionnaire général de la langue française, d'obtenir un vocabulaire spécifique et adapté aux utilisateurs pour des applications informatiques de type bureautique.

BIBLIOGRAPHIE

- BARNARD, P., HAMMOND, N., McLEAN, A., and MORTON, J. (1982). Learning and remembering interactive commands. Proceedings of Human Factors in Computer Systems, Gaithersburg, 15-17 March.
- BERGE, C. (1967). Théorie des graphes et ses applications. Paris Dunod.
- BISSERET, A. et SCAPIN, D.L. (1980). Ergonomie du service Agora-Teletel. II - Etude du langage de commande de la messagerie. Rapport technique INRIA BUR R06.
- BRANSCOMB, L.M. and THOMAS, J.C. (1983). Ease of use : a system design challenge. Proceedings of IFIP 83 Paris, North Holland, 431-438.
- CARROLL, J.M. (1980). Learning, using and designing command paradigms. IBM Research Report RC 8141.
- CORDIER, F. et DUBOIS, D. (1981). Typicalité et représentation cognitive. Cahiers de Psychologie Cognitive, 1,3, 299-333.
- DUBOIS, J., GIACOMO, M., GUESPIN, L. MARCELLESI, C., MEVEL, J.P., (1973). Dictionnaire de linguistique. Paris Larousse.
- GOUGGENTHEIM, G.K., MICHEA, R., FRIVENC, P., SAUVAGEOT, A., (1964). L'élaboration du français fondamental (1er degré). Etude sur l'établissement d'un vocabulaire et d'une grammaire de base. Paris Didier.
- LE NY, J.F. (1976). Sèmes ou mêmes? in S. Ehrlich et E. Tulving (Eds.). La mémoire sémantique. Bulletin de psychologie, numéro spécial, 46-54.
- PECHEUX, M., et FUCHS, C., (1975). Mises au point et perspectives à propos de l'analyse automatique du discours, Langages, 37, 7-80.

- RICHY, H., QUINT, V., SASYAN, S. (1981). Le service de messagerie PAPYRUS. Rapport INRIA MSG 2.526.
- ROBERT, P. (1978). Le Petit Robert, Dictionnaire alphabétique et analogique de la langue française.
- ROSCH, E. (1978). Principles of categorization, in E. Rosch and B. Lloyd (Eds.): Cognition and categorization, Hillsdale, N.J., L. Erlbaum, 27-48.
- ROSCH, E., and MERVIS, C.B. (1975). Family resemblances : Studies in the internal structure of categories. Cognitive Psychology, 7, 573-605.
- ROSENBERG, J. (1982). Evaluating the suggestiveness of command names. Proceedings of the Conference on Human Factors in Computer Systems. Gaithersburg, Maryland, March 1982, 12-16.
- ROY, B. (1969). Algèbre moderne et théorie des graphes, tome 1, Dunod.
- SCAPIN, D.L. (1982). Ergonomics of computer commands in restricted natural language. Proceedings of the 8th Congress of the International Ergonomics Association. Tokyo, September 1982, 546-547.
- TVERSKY, A. (1977). Features of similarity. Psychological Review, 84,4, 327-352.
- WHITTEN, W.B., NEWTON-SUTER, H.W., FRANCK, M.L. (1979). Bidirectional synonym ratings of 464 noun pairs. Journal of Verbal Learning and Verbal Behavior, 18, 109-127.

A N N E X E 1

Liste des 25 fonctions avec pour chacune d'entre elles :

- le mot de départ
- une définition explicitant à l'utilisateur ce que la fonction lui permet d'effectuer
- un ou plusieurs exemples d'objets auxquels peut s'appliquer la fonction

Mots de départ	Définitions	Exemples d'objets
1. S'ABONNER	- L'accès d'un système étant réservé aux personnes connues de ce système, cette commande permet à un utilisateur potentiel de se faire connaître auprès du système avant de l'utiliser pour la première fois	- système de messagerie électronique
2. AJOUTER	- Permet d'ajouter un ou plusieurs objets sur une liste déjà constituée	- liste de diffusion, conférence, etc.
3. ANNULER	- Permet de supprimer une procédure soit prévue, soit en cours	- téléconférence, liste de diffusion
4. ARCHIVER	- Permet de stocker des documents que l'utilisateur désire conserver dans une mémoire prévue à cet effet (Archives)	- message, texte écrit
5. CHANGER	- Permet de substituer une entité à une autre	- mot de passe, nom d'un texte
6. COMMENTER	- L'utilisateur ayant reçu un message, cette commande lui permet de préparer un commentaire à ajouter à ce message avant de le transmettre à un autre utilisateur	- message déjà reçu
7. CONSULTER	- Permet de regarder quelque chose pour y chercher des renseignements	- liste d'abonnés, de messages,...
8. CORRIGER	- Permet de supprimer les fautes, erreurs de forme, dans un texte	- message, texte...
9. CREER	- L'utilisateur veut créer dans le système un nouvel élément qui restera à sa disposition	- liste, fichier...

Mots de départ	Définitions	Exemples d'objets
10. DENOMBRER	- Permet d'obtenir le nombre d'objets contenus dans un fichier, une liste, ou une partie de cette liste	- messages dans boîte à lettres
11. DETRUIRE	- Permet d'effacer dans la mémoire du système certains des éléments d'un ensemble tel que boîte à lettres, liste	- messages reçus, liste de diffusion
12. ECRIRE	- Permet à l'utilisateur de préparer un texte écrit pour l'"entrer" dans le système	- messages, notes, textes
13. ENVOYER	- Après avoir écrit un message, cette commande permet de faire parvenir le message à un ou plusieurs destinataires	- messages
14. ETIQUETER	- Permet de donner un nom à un texte, à un message, à une liste, ou à un groupe d'utilisateurs, non encore étiquetés	- messages, textes, listes
15. IMPRIMER	- Permet à l'utilisateur de garder une trace écrite sur papier de tout document apparu sur l'écran	- messages, listes, textes
16. LIRE	- Permet à l'utilisateur de prendre connaissance du contenu de messages en les faisant apparaître sur l'écran	- messages
17. MODIFIER	- Permet de changer le contenu d'un texte déjà écrit et destiné à être envoyé, ou d'une liste	- messages, textes, listes
18. QUITTER	- Permet à un utilisateur membre d'une conférence ou de tout autre groupe, de quitter ce groupe	- conférences, groupes

Mots de départ	Définitions	Exemples d'objets
19. REDISTRIBUER	- Après avoir reçu un message, permet de diffuser ce message à d'autres utilisateurs	- messages
20. REJOINDRE	- Permet à un abonné de se joindre à une conférence ou tout autre groupe d'utilisateurs	- conférences, groupes
21. REPONDRE	- Après avoir reçu un message, permet d'envoyer à l'émetteur du message une réponse	
22. RETIRER	- Permet de retirer un ou plusieurs objets d'une liste d'objets déjà constituée	- listes de diffusion, conférences
23. SELECTIONNER	- Permet d'effectuer une consultation sélective de documents, c'est-à-dire de ne consulter qu'une partie ou certains éléments d'une liste ou d'un fichier	- messages, abonnés, etc
24. VERIFIER	- Permet de s'assurer que les messages ont été lus par leur destinataire	- messages
25. VISUALISER	- Permet de demander que quelque chose de déjà créé et existant dans le système apparaisse sur l'écran	- listes, fichiers, documents, messages, statistiques

A N N E X E 2

Données quantitatives pour chacune des 25 matrices :

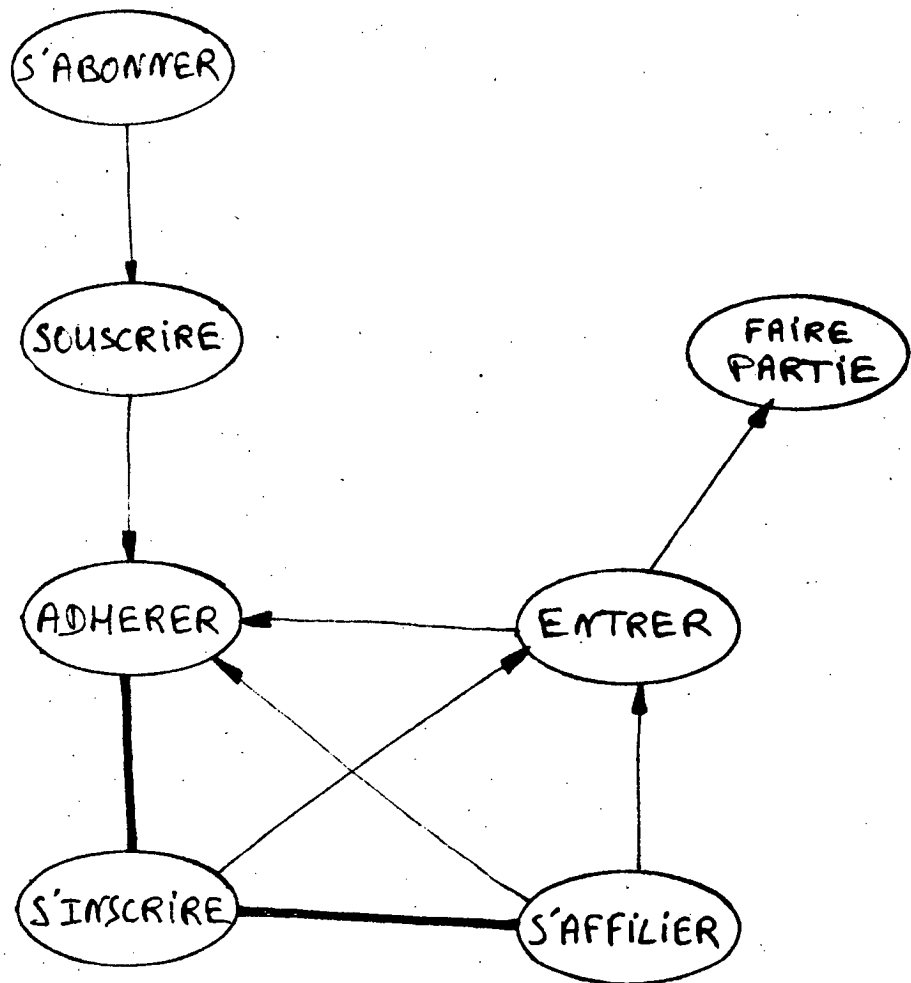
- nombre de mots par matrice
- nombre total de liaisons entre les mots par matrice
- parmi ces liaisons, nombre des liaisons symétriques par matrice.

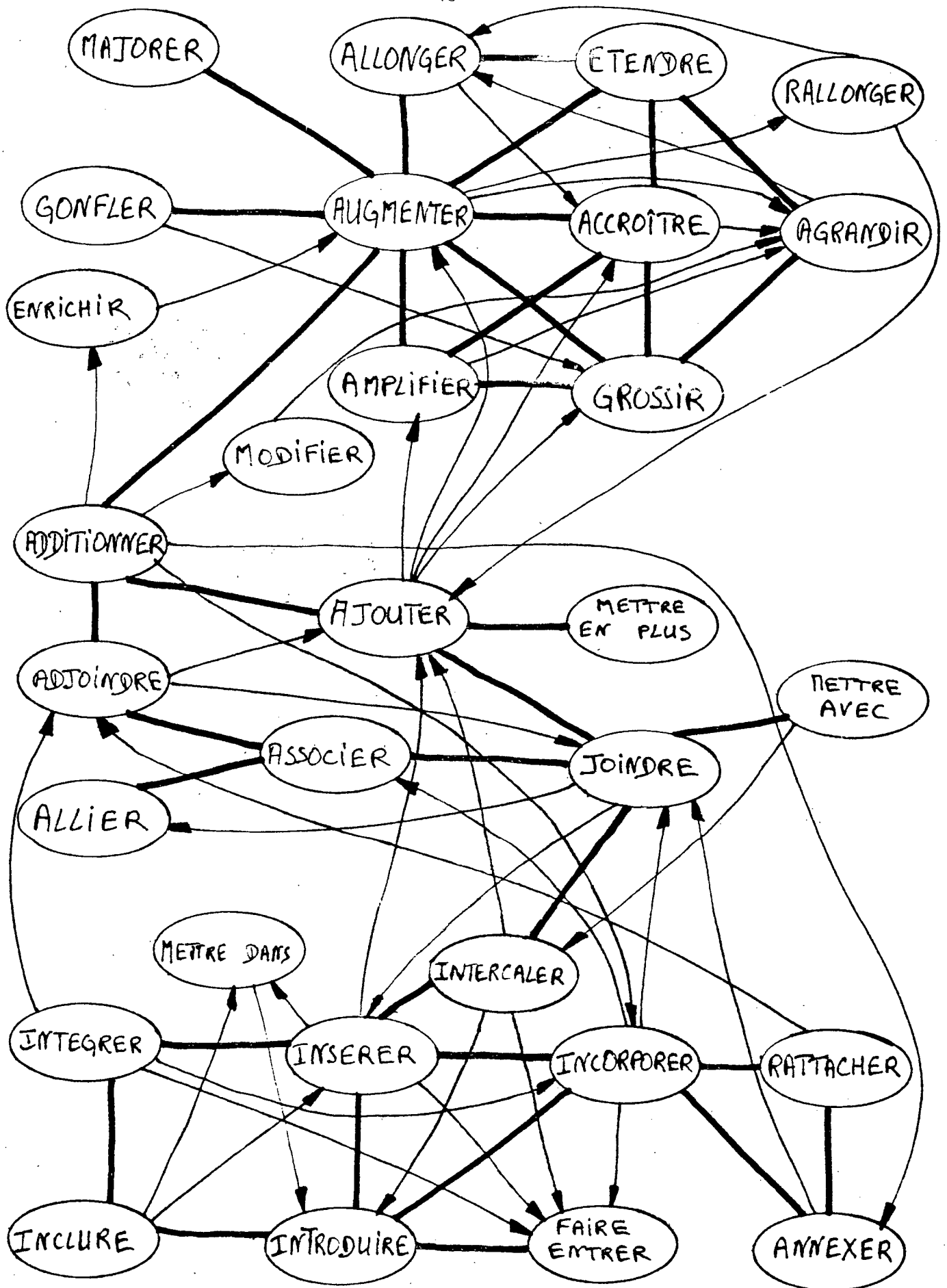
Remarque : nous dirons qu'il y a liaison entre deux mots différents chaque fois qu'un des deux mots est relié à l'autre par la relation binaire : "le mot X a pour synonyme le mot Y". Parmi ces liaisons, certaines sont appelées liaisons symétriques : ce sont les liaisons telles que le mot X est relié au mot Y et le mot Y est relié au mot X.

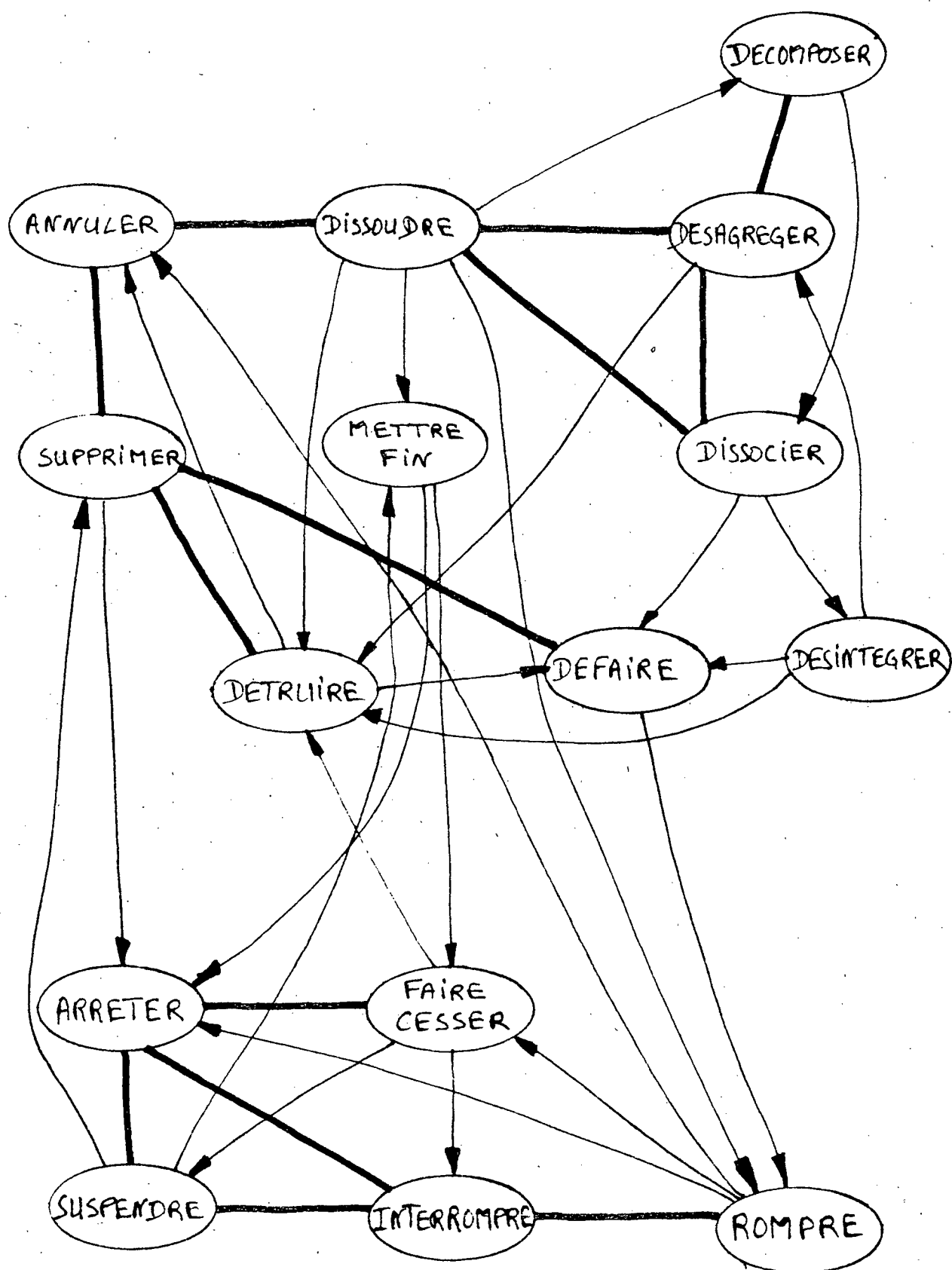
N°	Mots de départ	Nb de mots par matrice	Nb liaisons par matrice	Nb liaisons symétriques par matrice
1	. ABONNER	7	11	2
2	. AJOUTER	30	111	35
3	. ANNULER	15	51	13
4	. ARCHIVER	22	62	18
5	. CHANGER	9	25	11
6	. COMMENTER	8	21	8
7	. CONSULTER	16	45	16
8	. CORRIGER	18	71	26
9	. CREER	22	104	32
10	. DENOMBRER	18	72	26
11	. DETUIRE	19	59	20
12	. ECRIRE	28	110	36
13	. ENVOYER	18	53	19
14	. ETIQUETER	8	24	11
15	. IMPRIMER	10	21	8
16	. LIRE	12	39	14
17	. MODIFIER	21	65	21
18	. QUITTER	19	75	23
19	. REDISTRIBUER	20	76	30
20	. REJOINDRE	17	46	15
21	. REPONDRE	11	27	11
22	. RETIRER	27	98	36
23	. SELECTIONNER	26	59	18
24	. VERIFIER	10	27	10
25	. VISUALISER	17	36	10

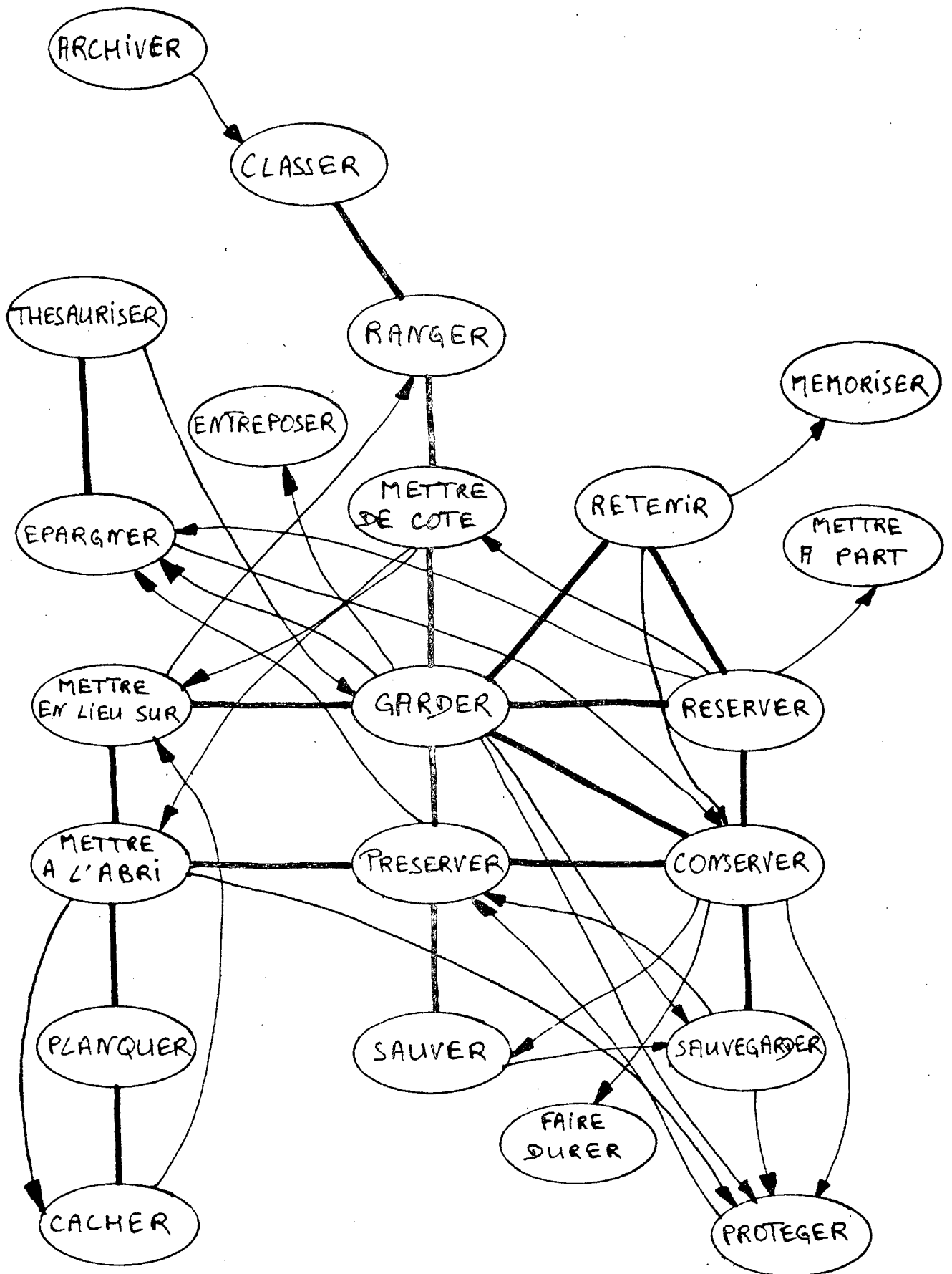
A N N E X E 3

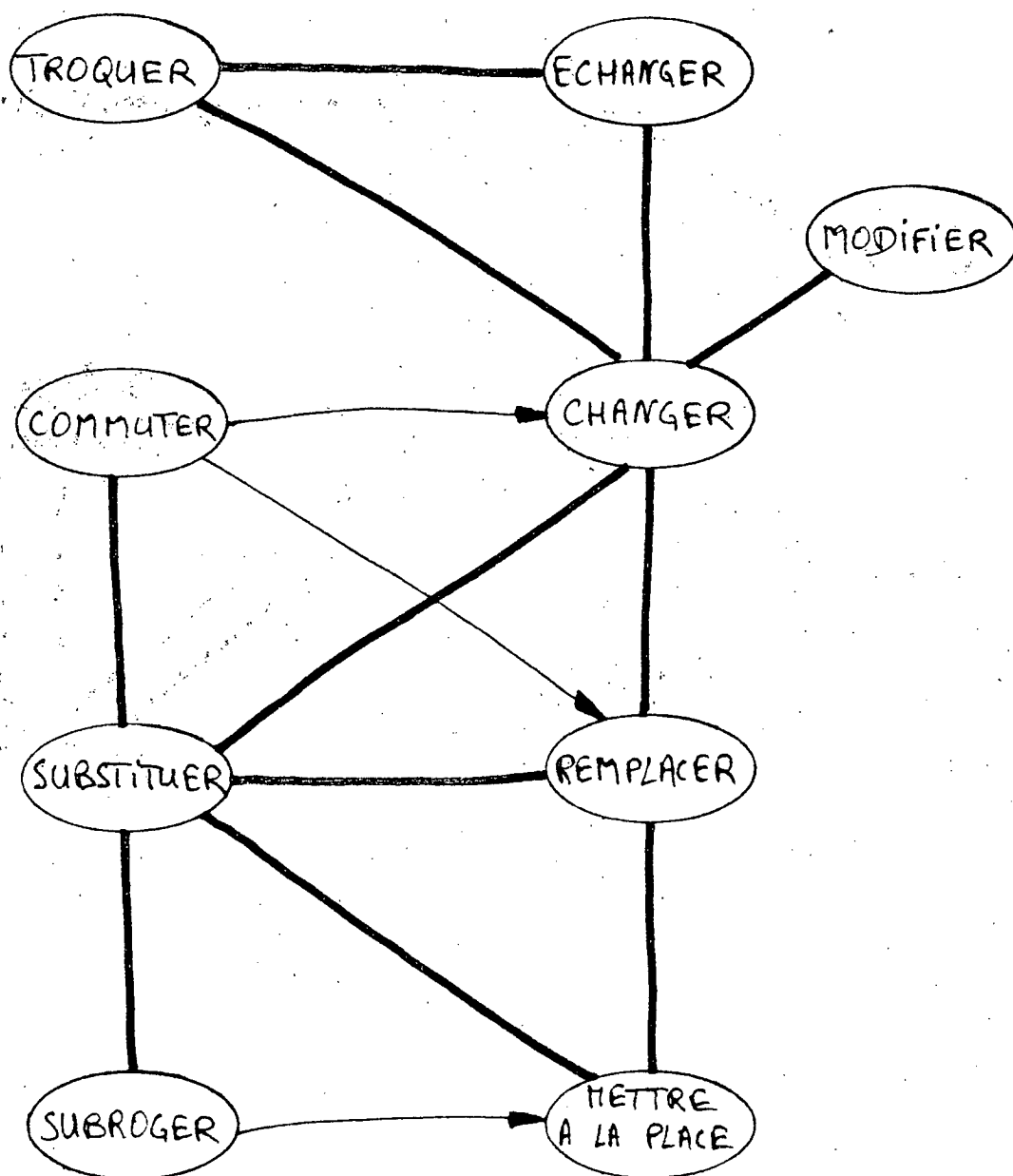
Représentation graphique des 25 graphes

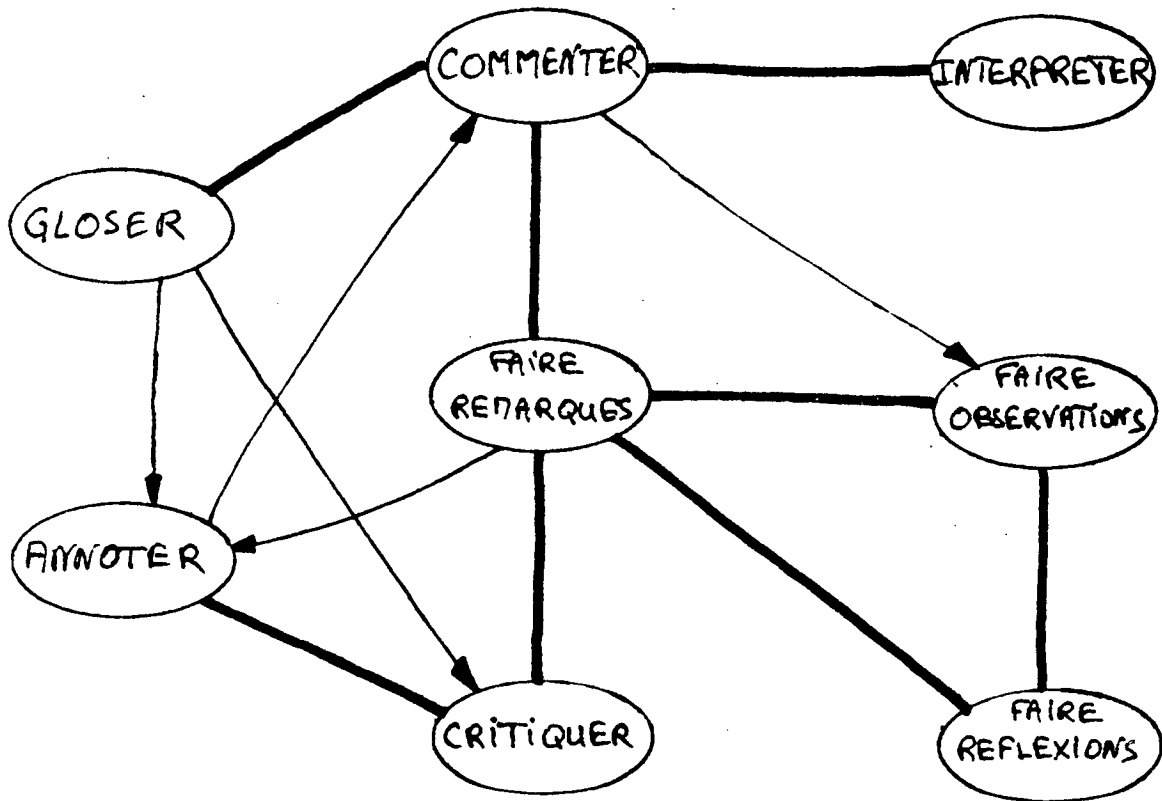


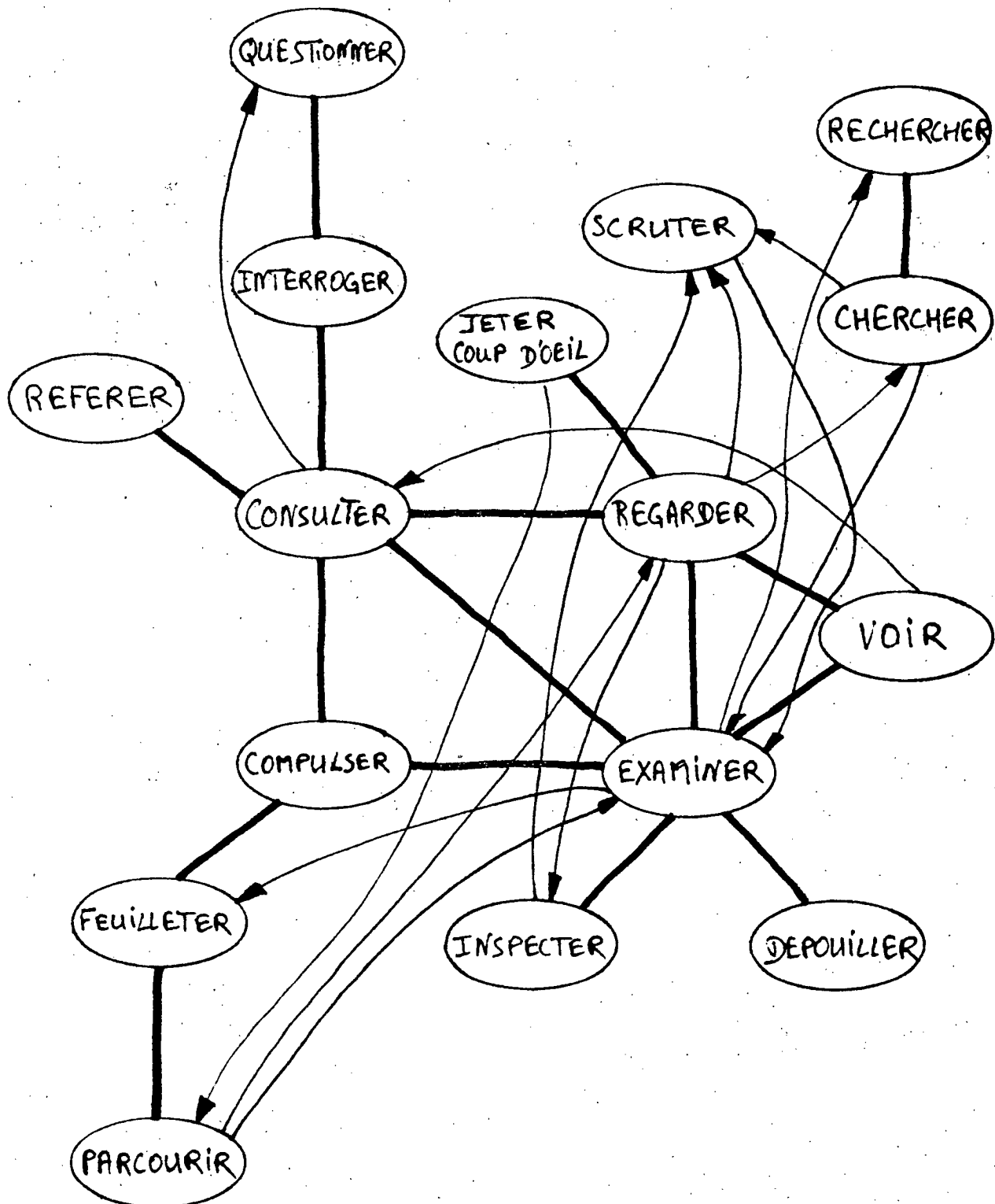


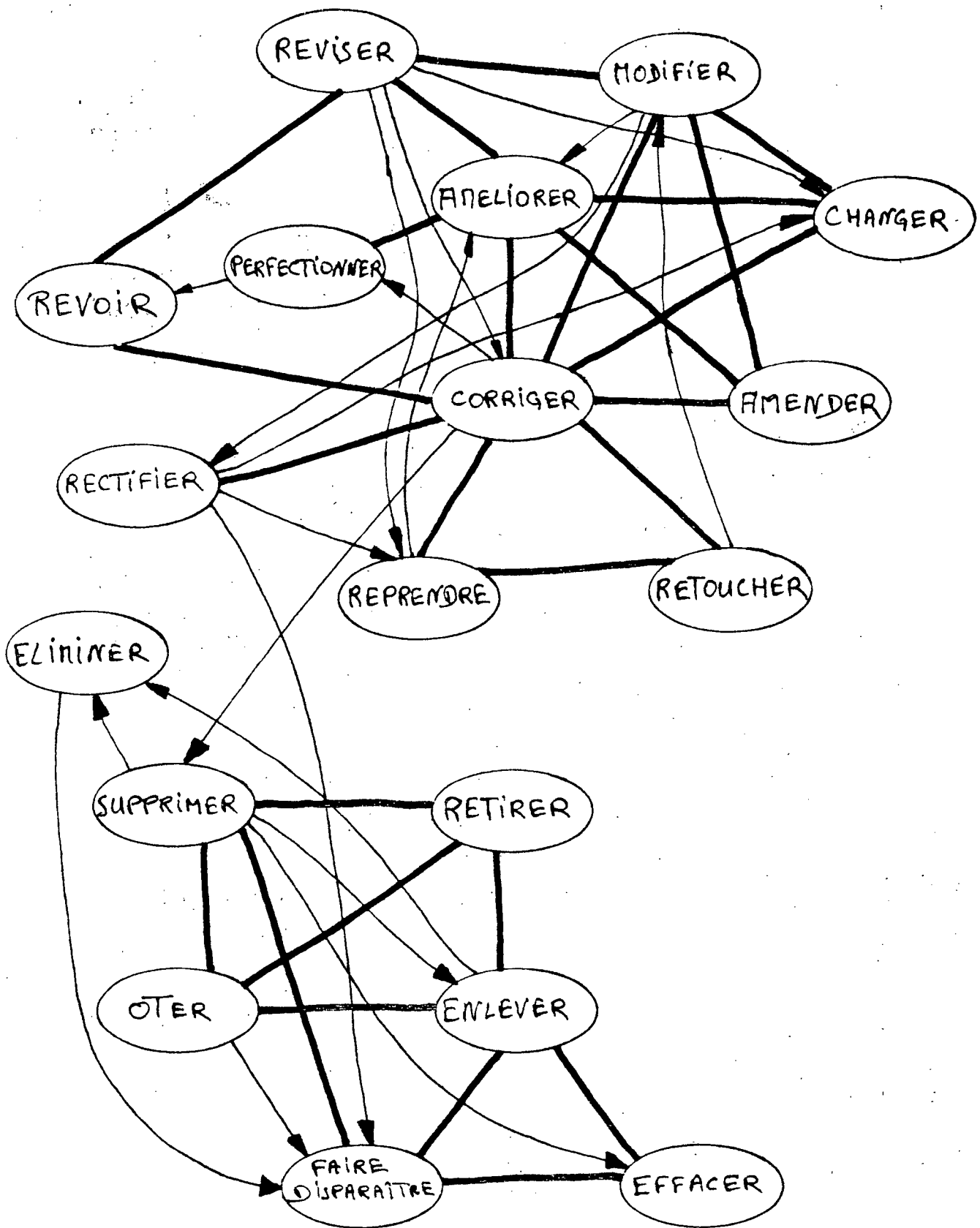


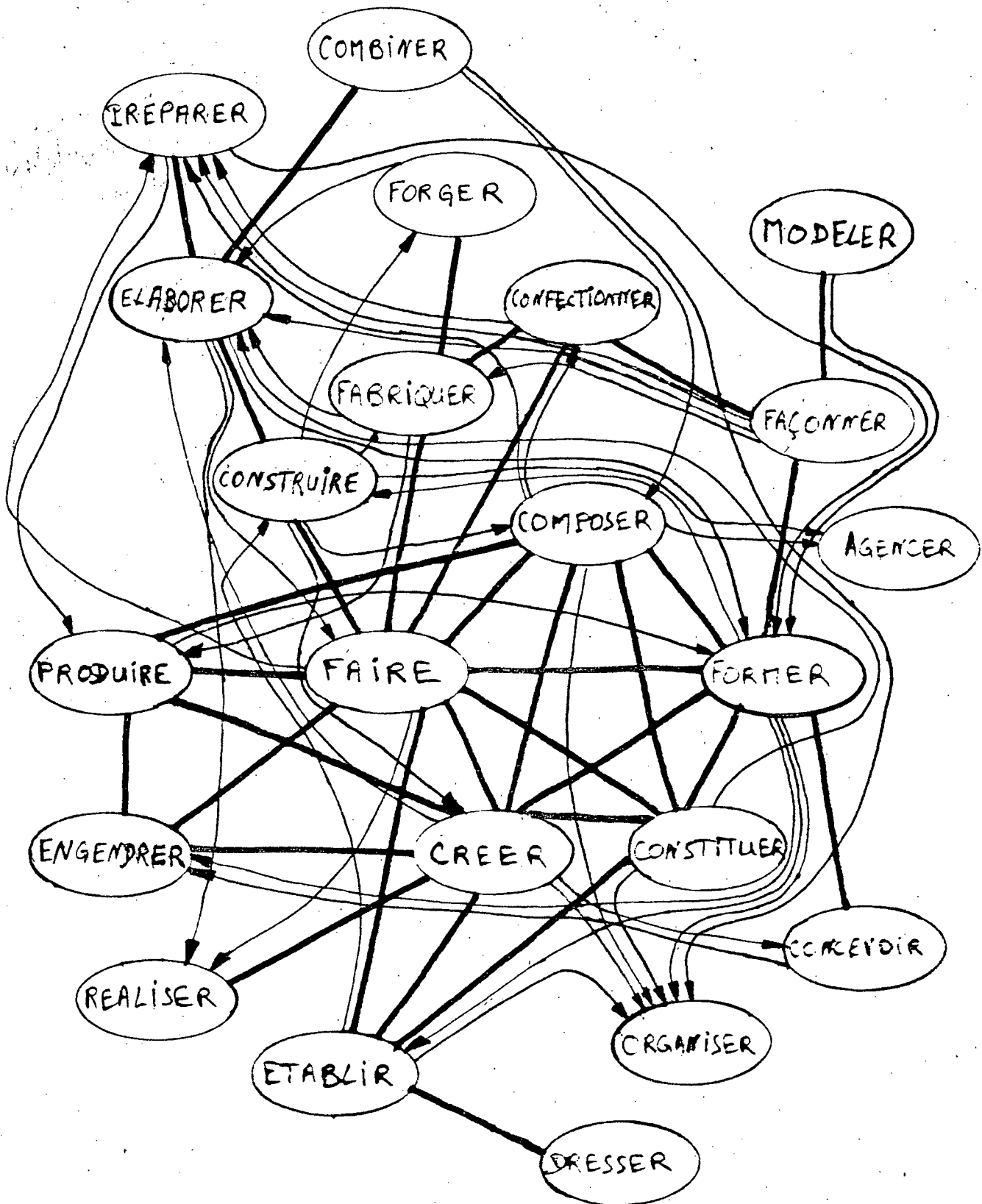


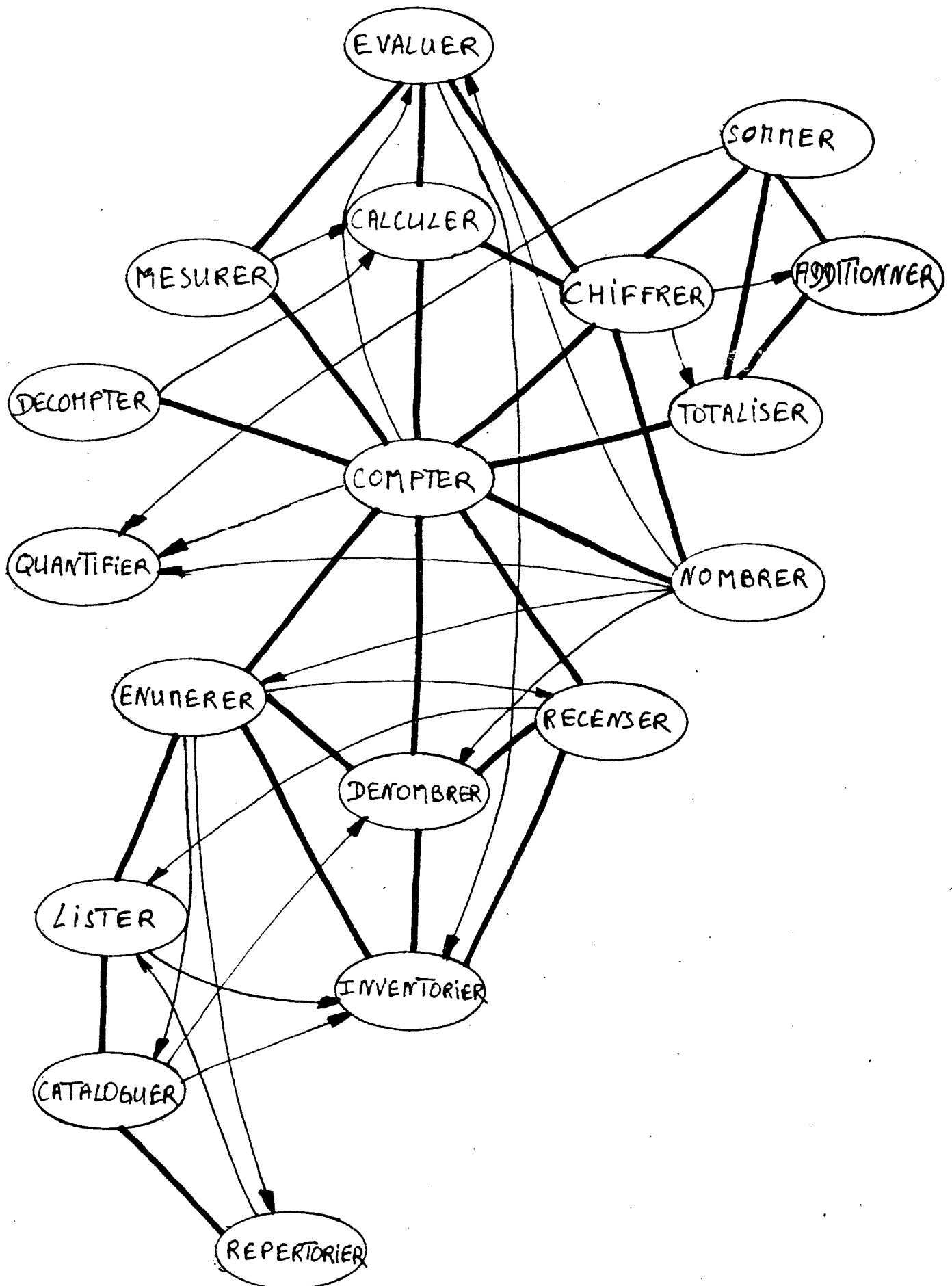


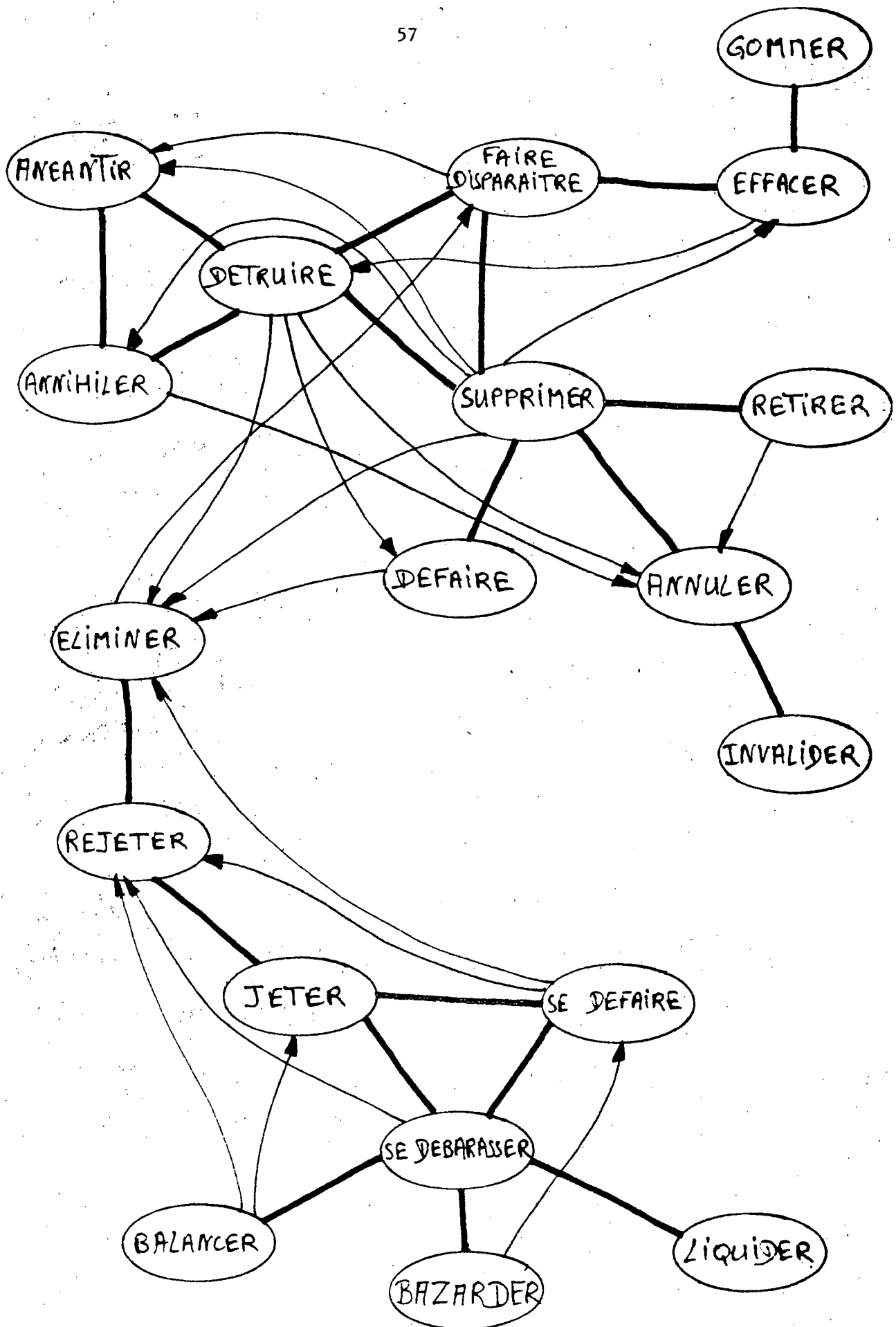


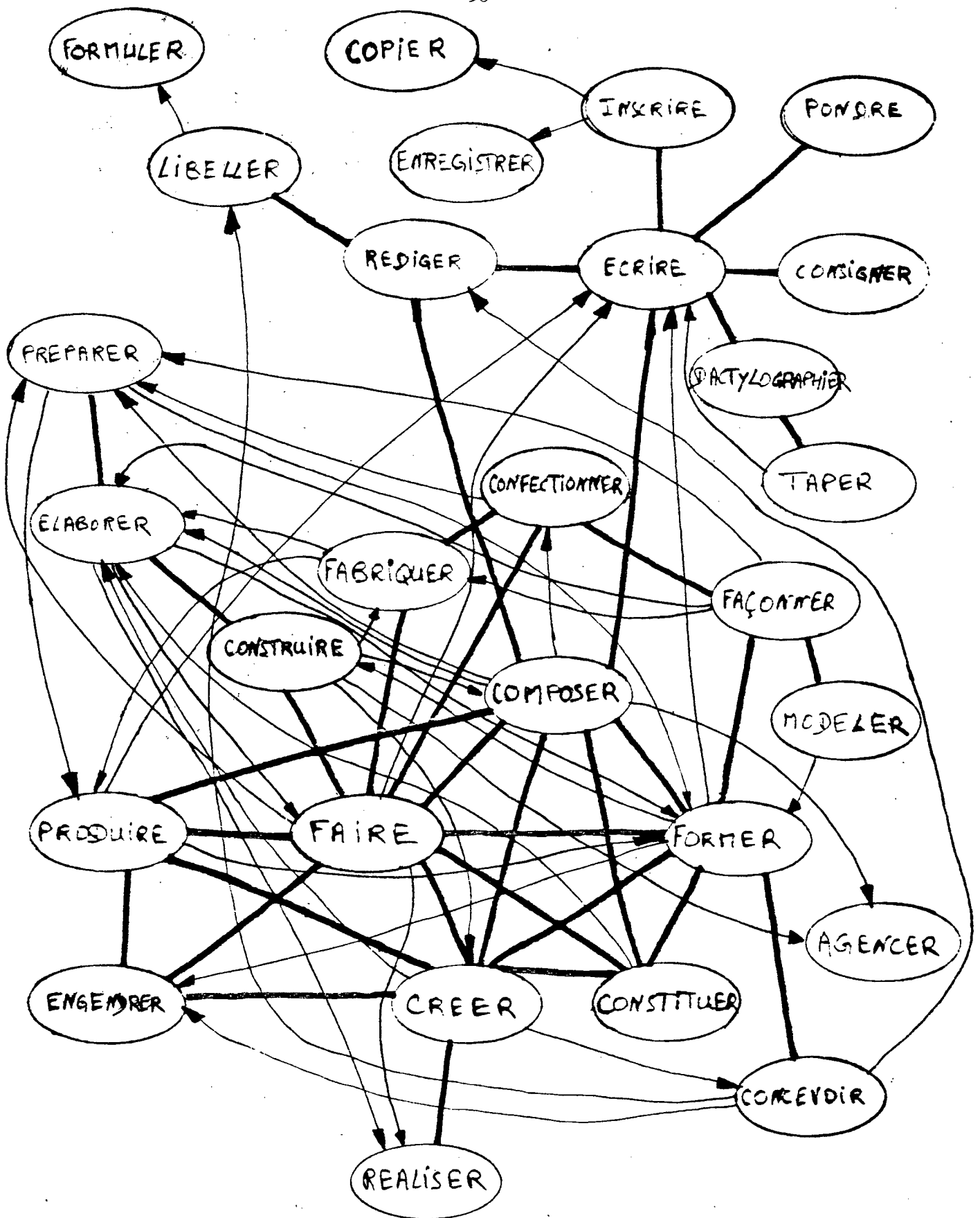


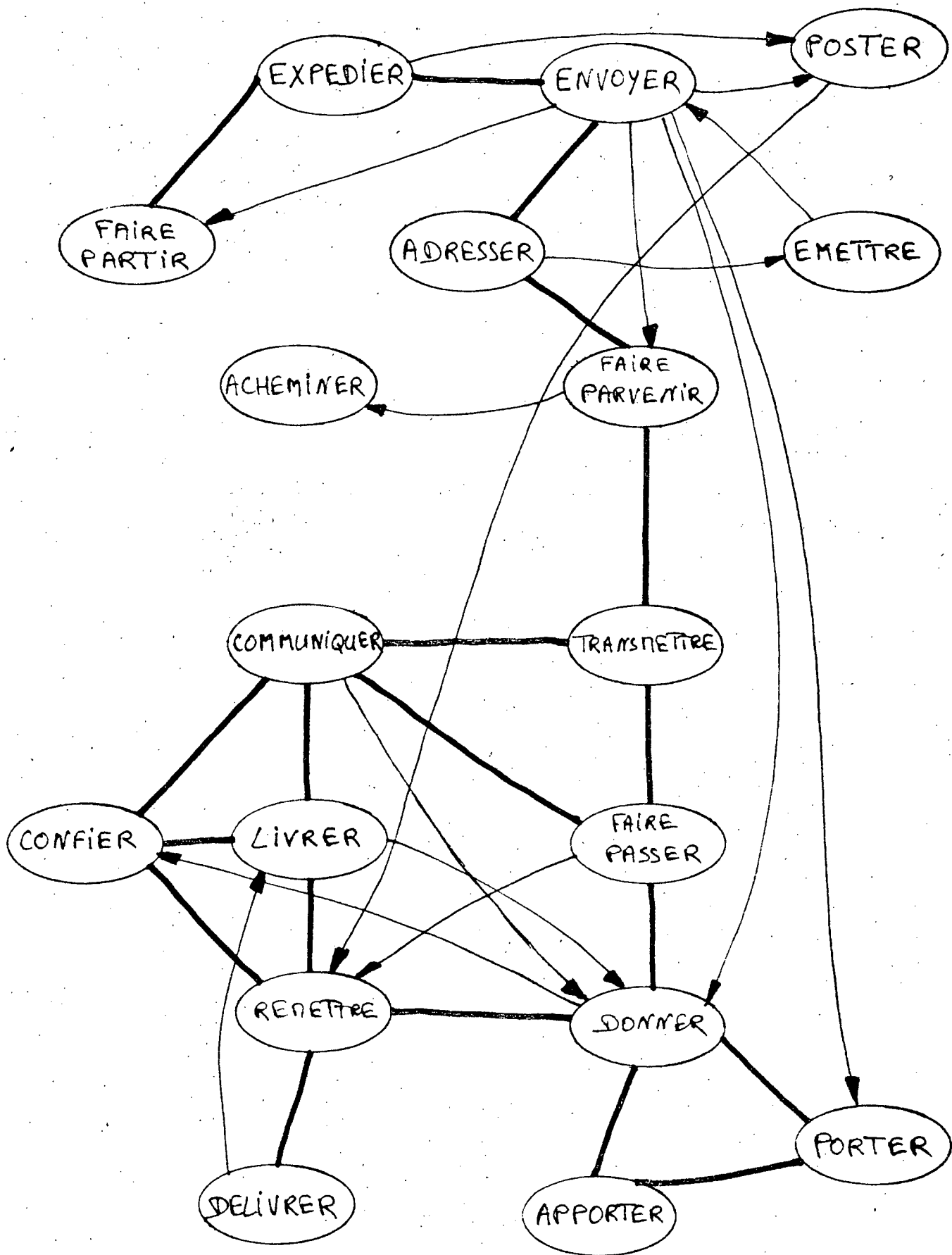


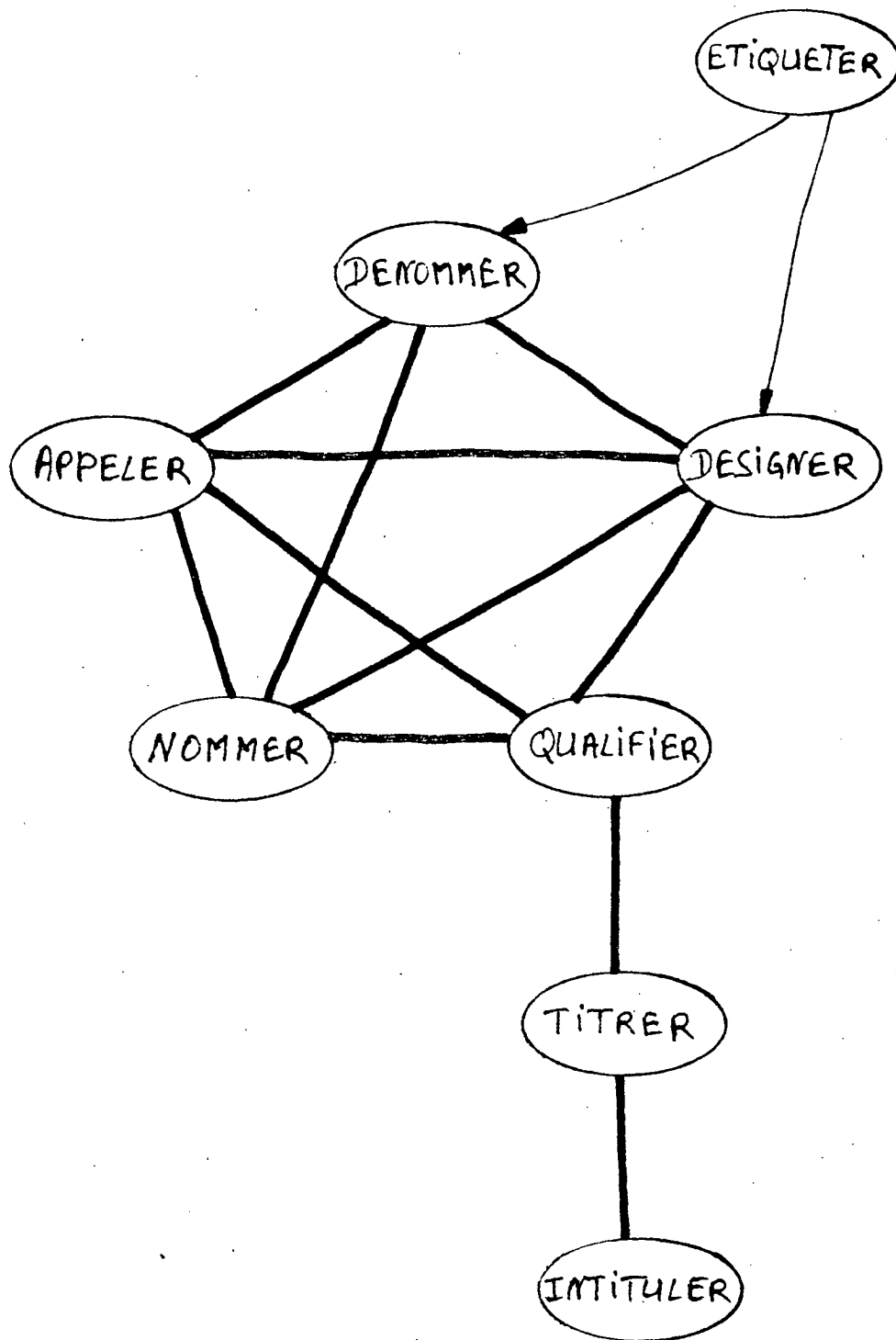


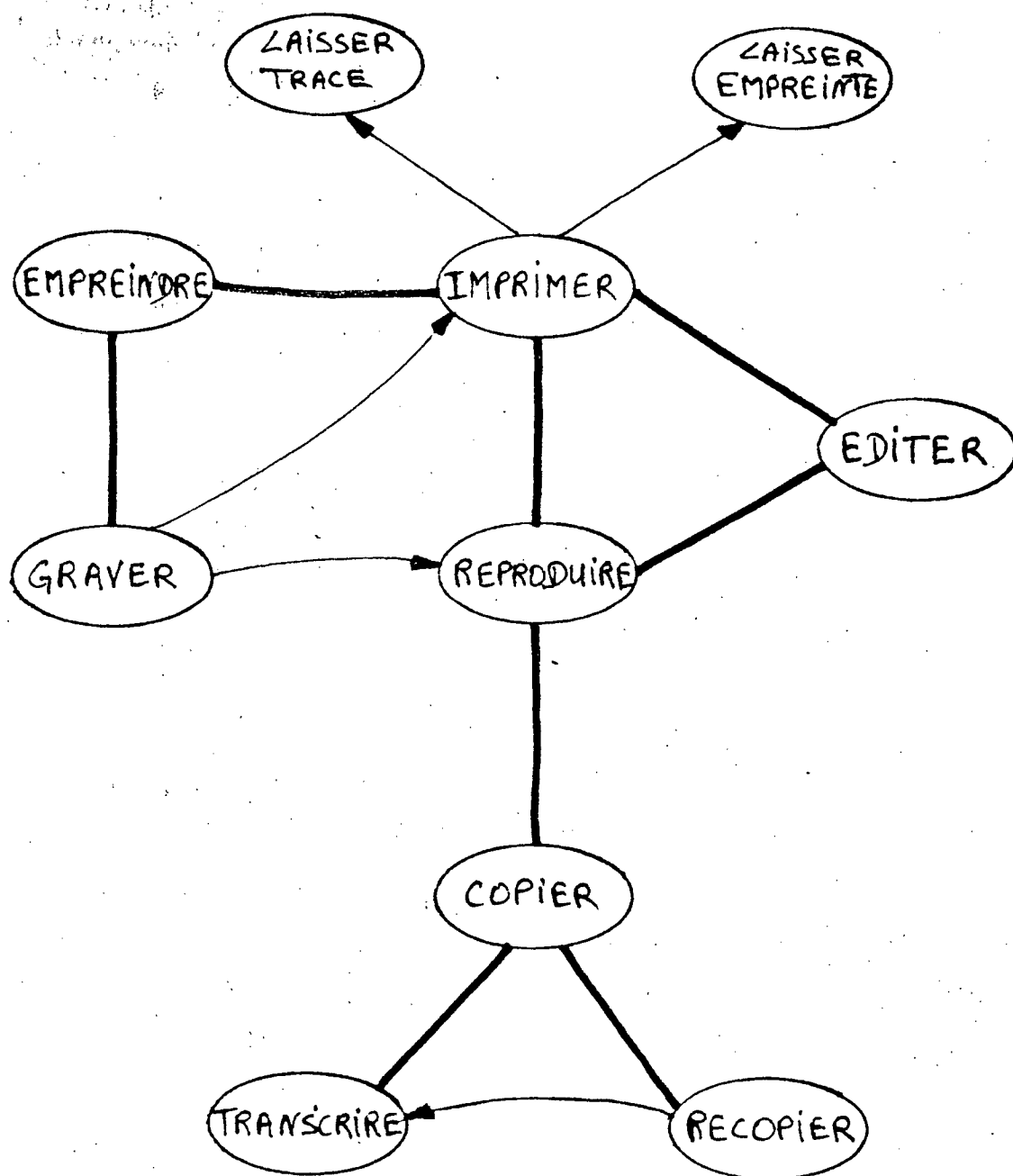


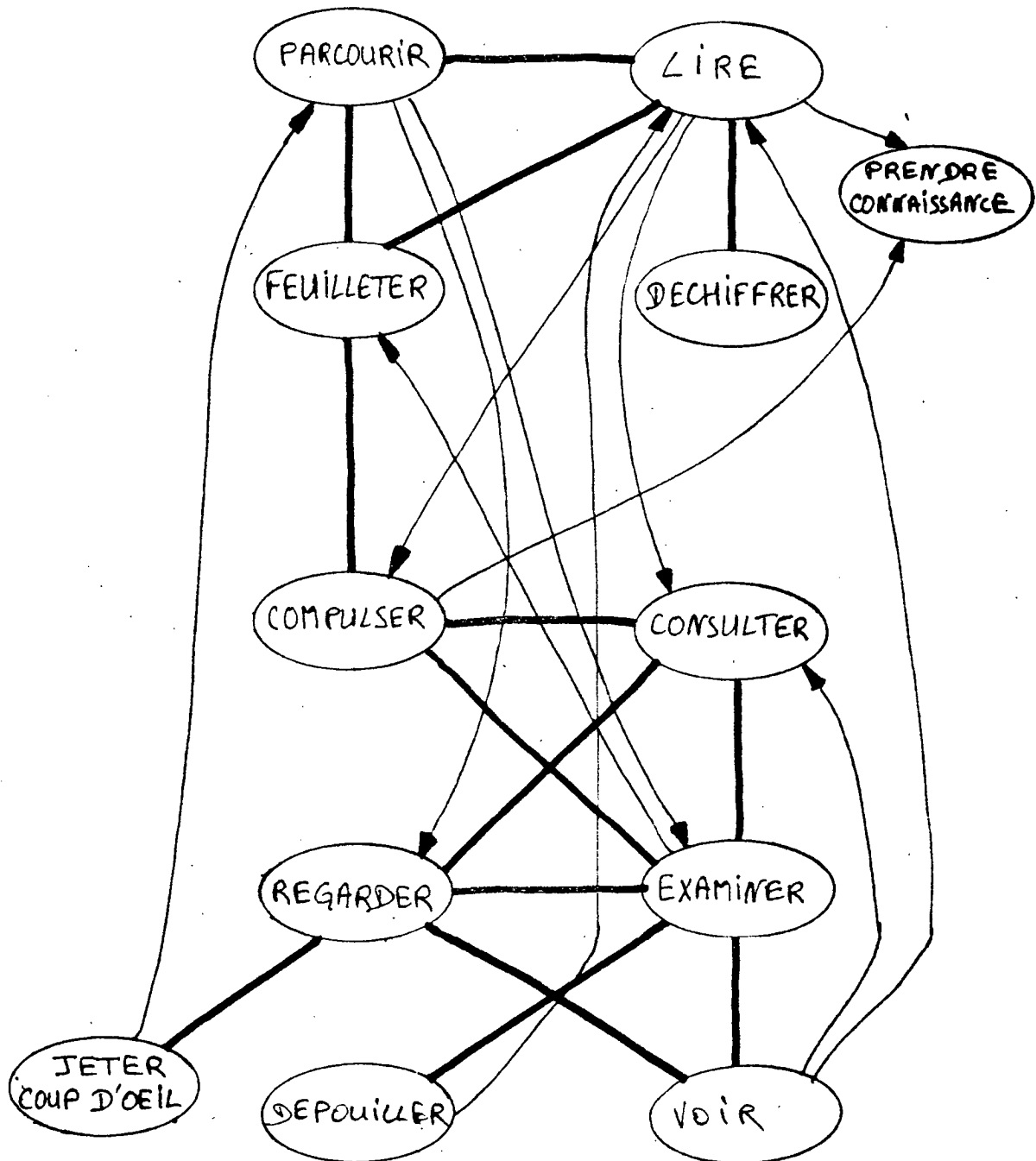


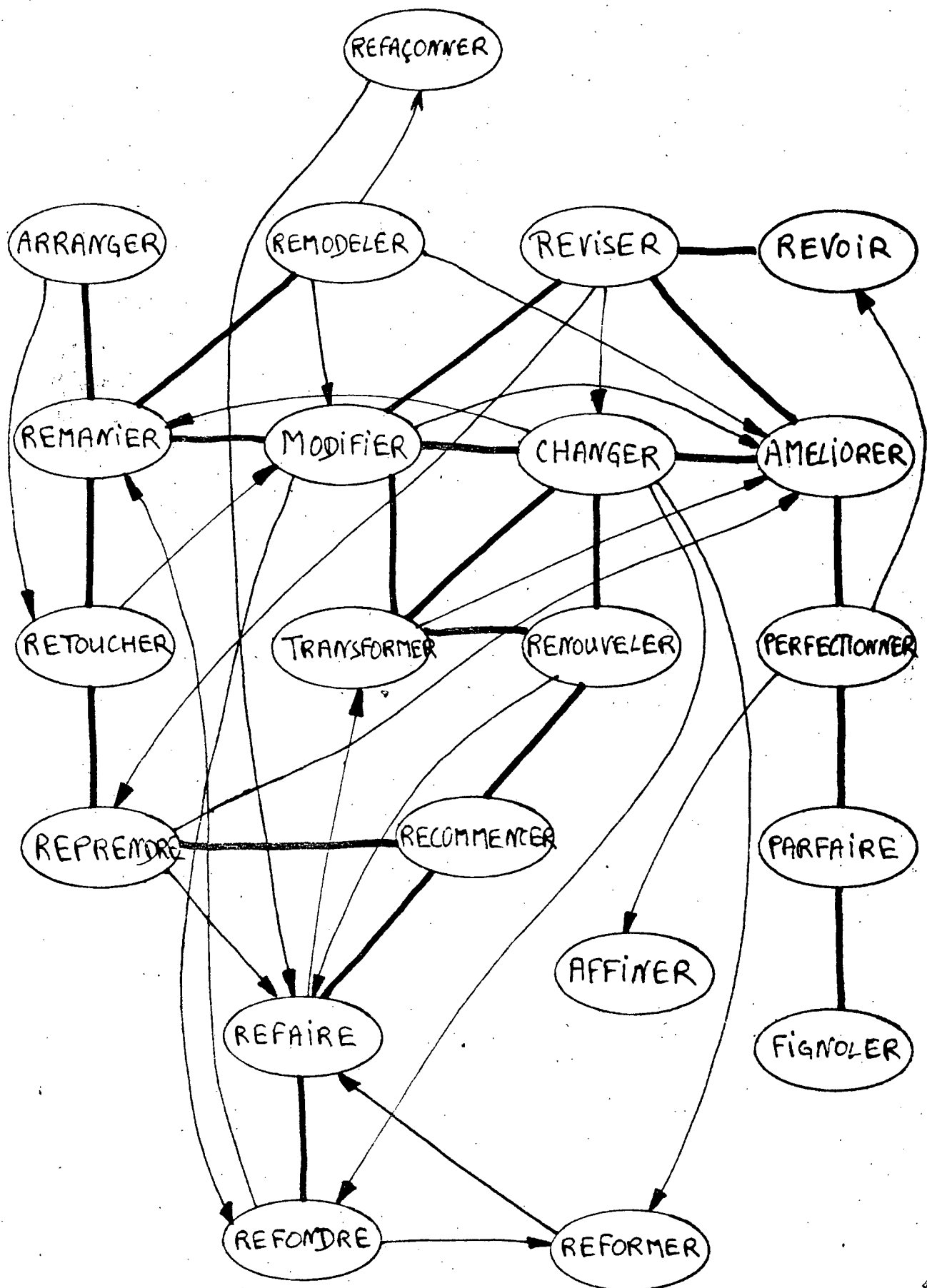




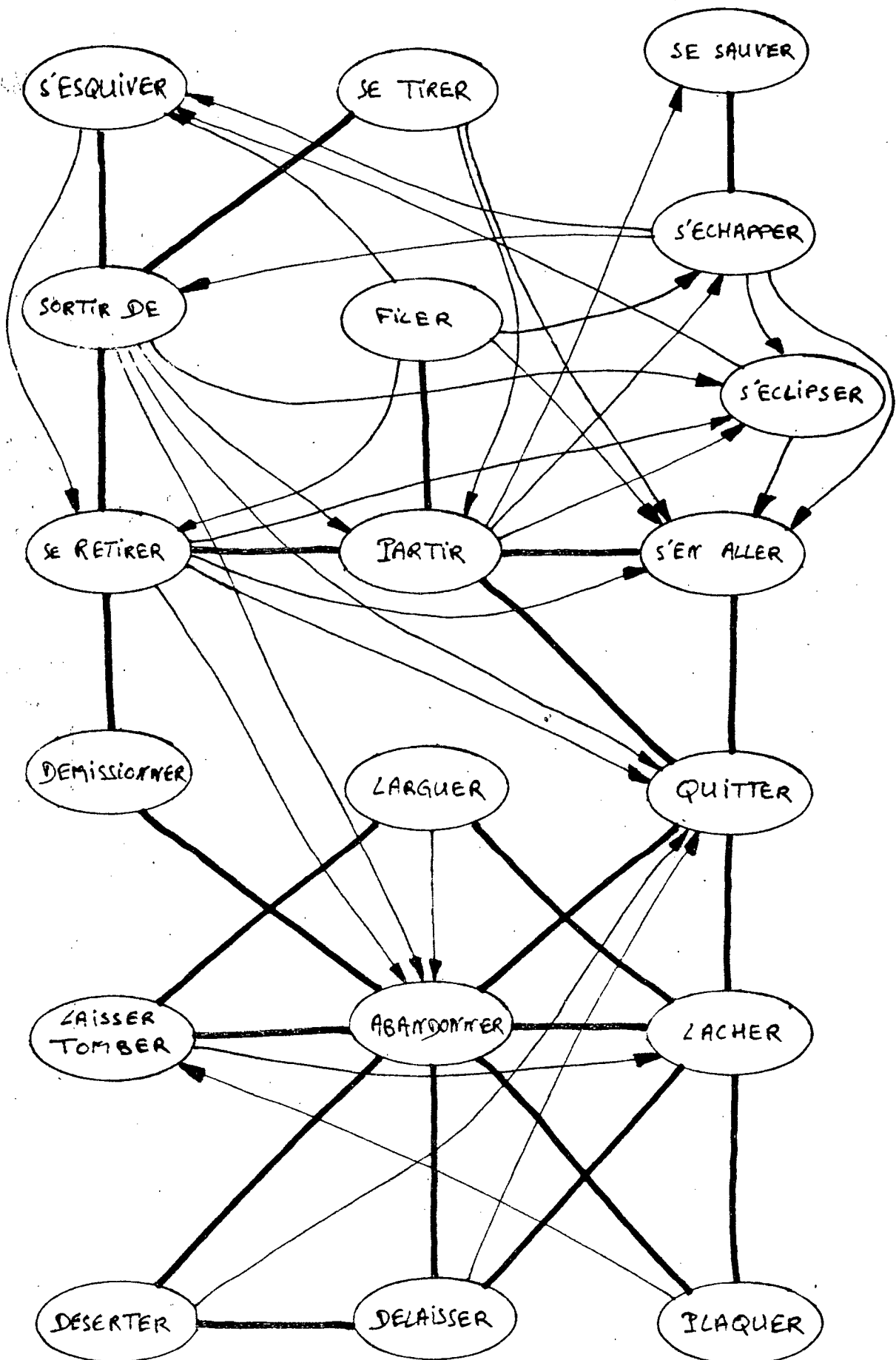


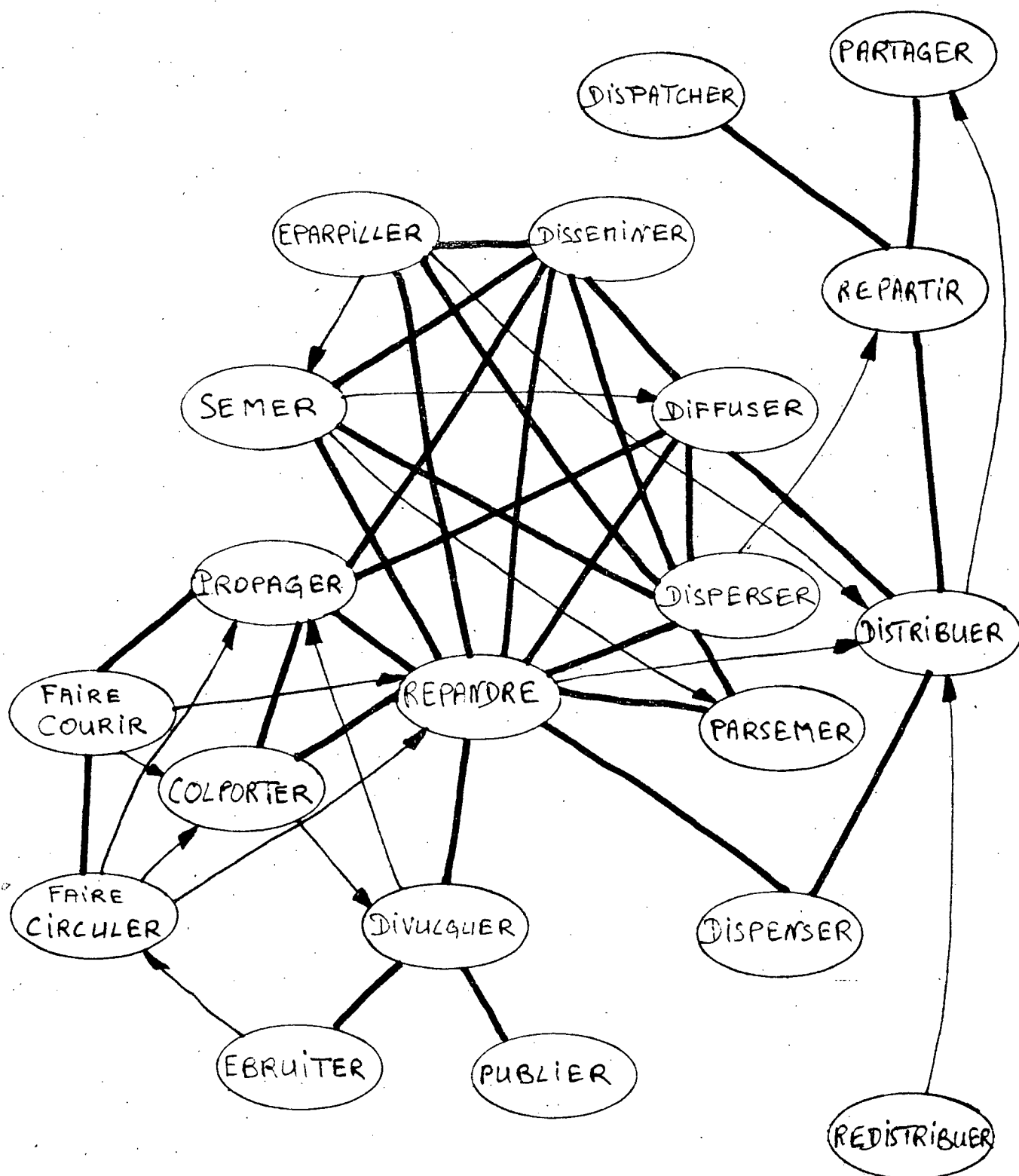


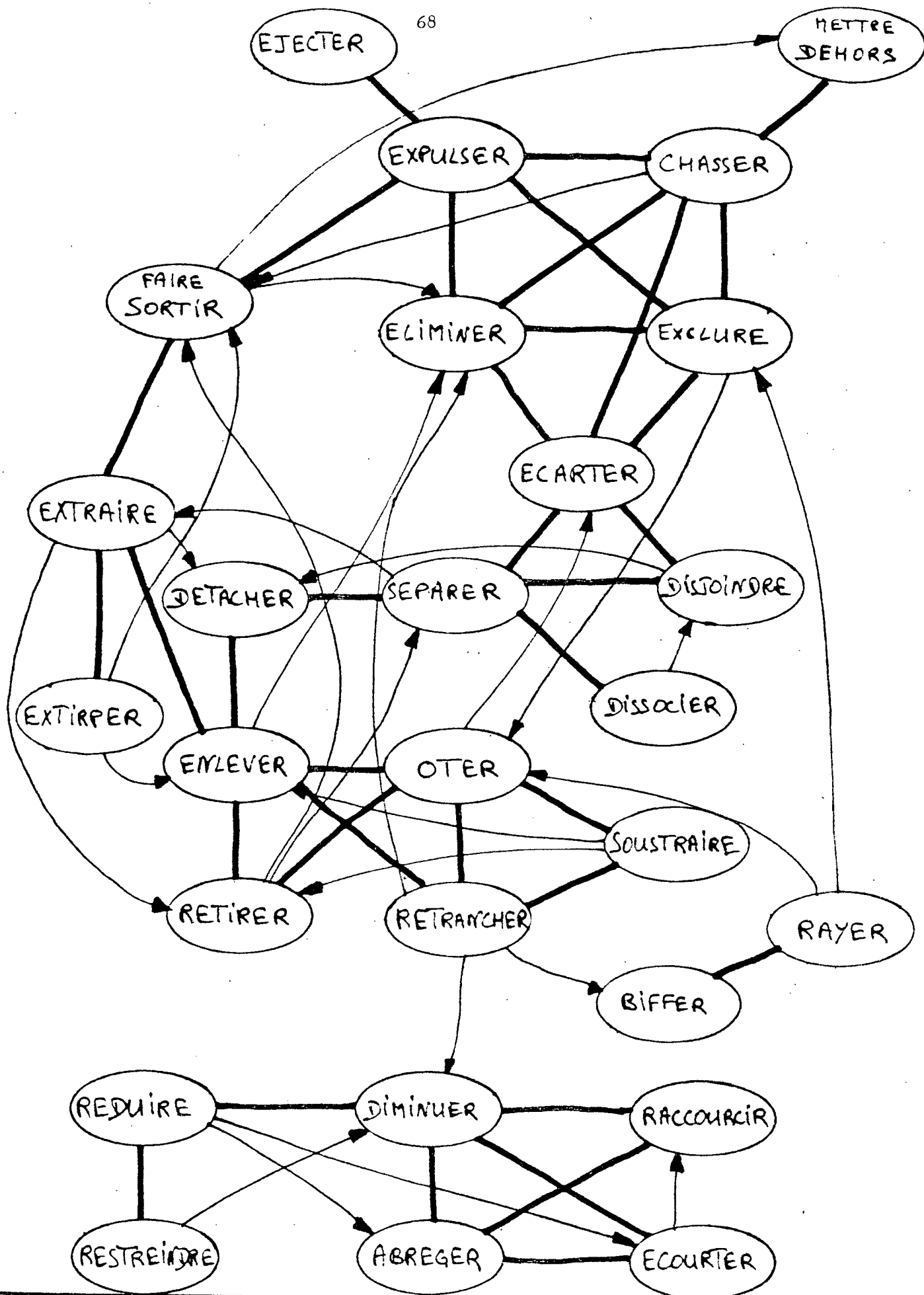


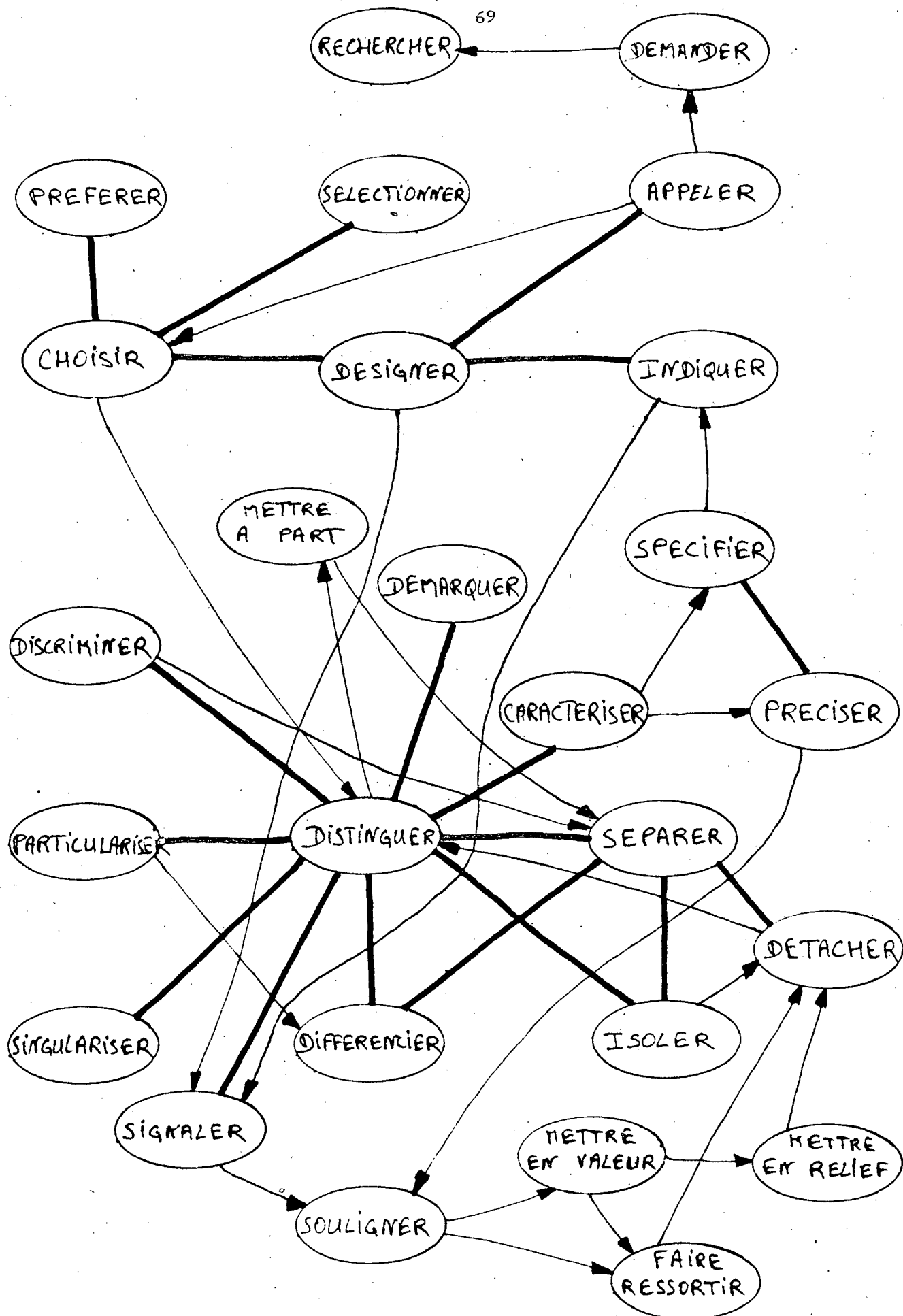


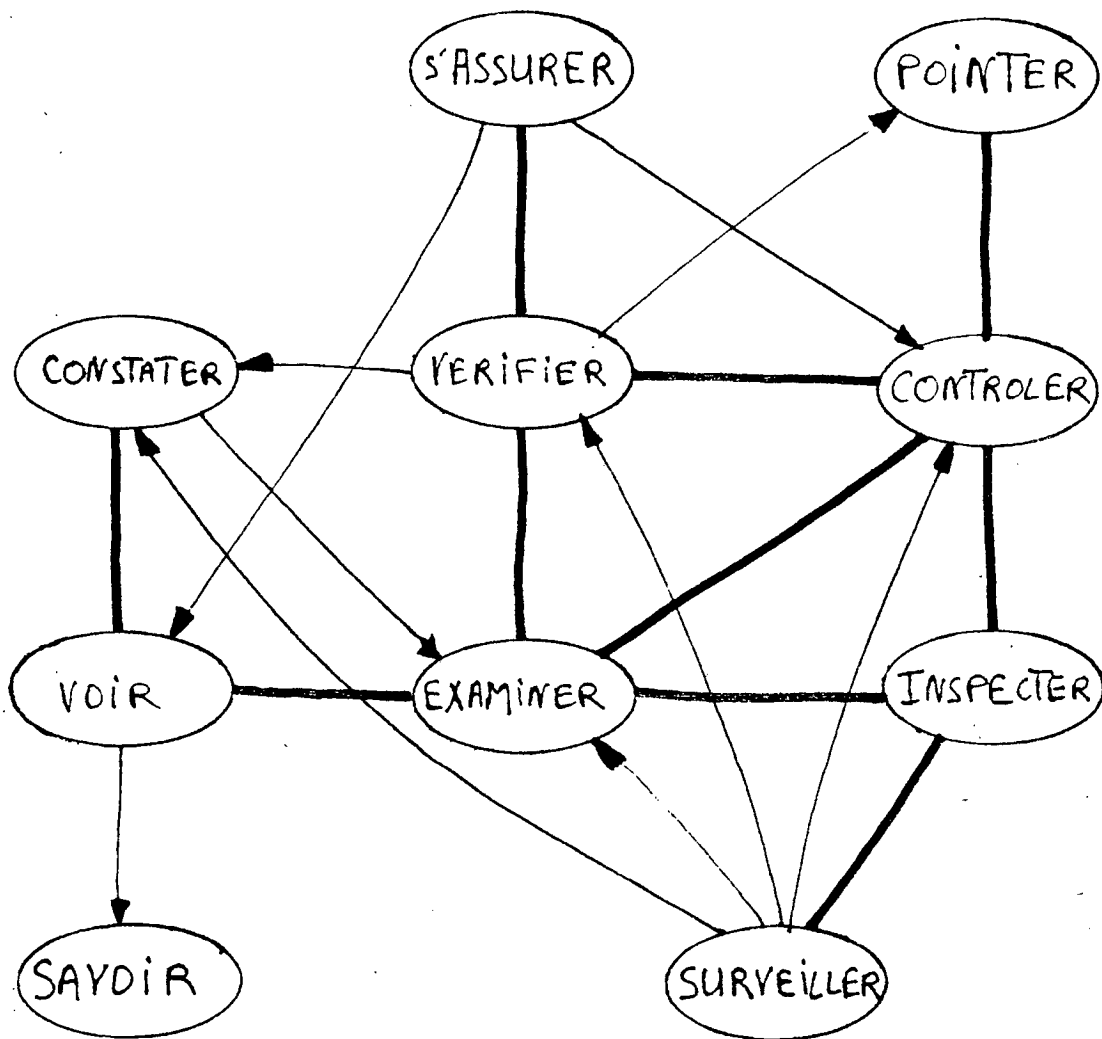
- G R A P H E G 17 -

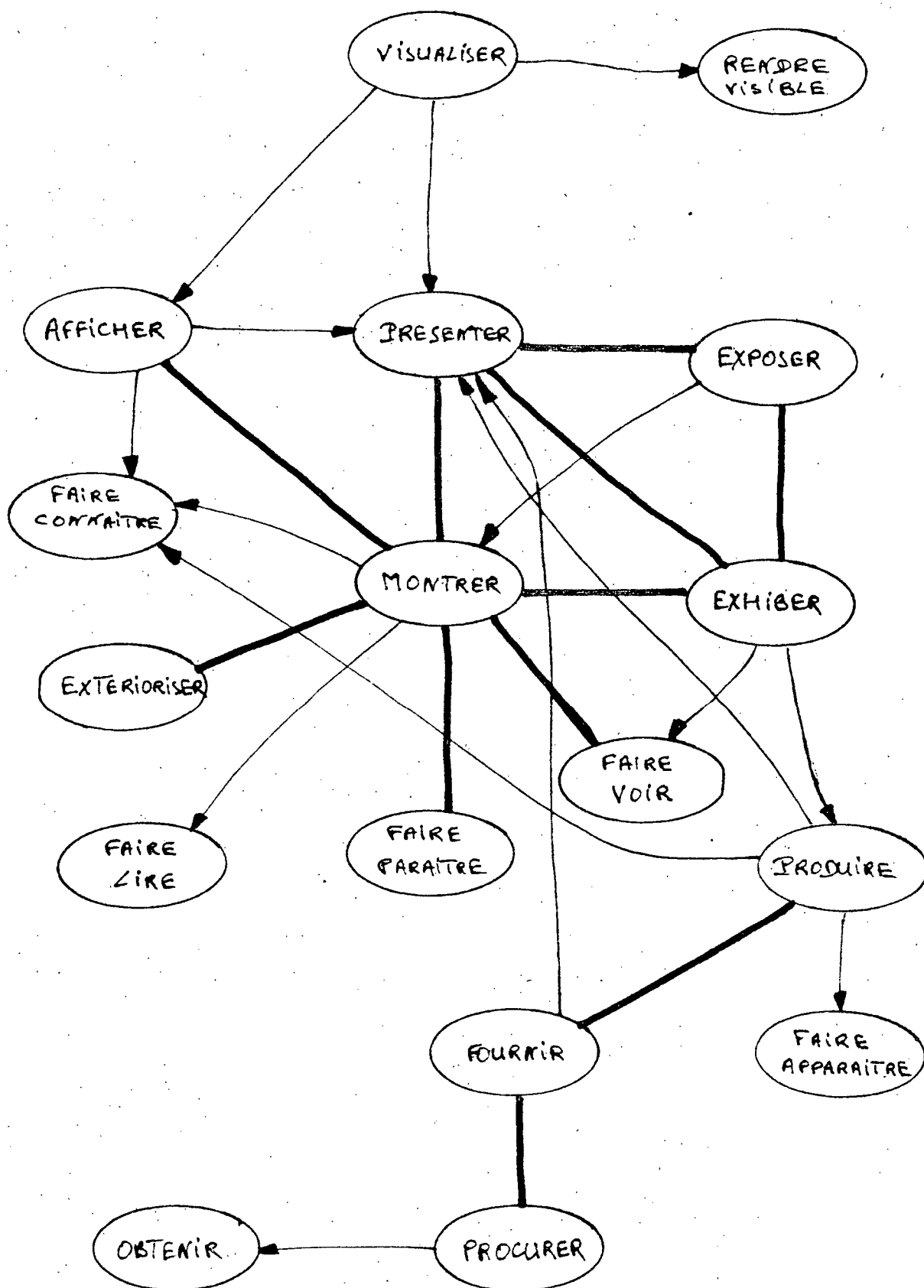












A N N E X E 4

Liste pour chaque graphe des éléments minimaux et maximaux

Elément minimal	Elément maximal
G1 : Abonner	G1 : Faire partie de
G2 : Archiver	G4 : Entreposer, Faire durer, Mémoriser, Mettre à part,
G3 : Etiqueter	G9 : Organiser, Agencer
G4 : Redistribuer	G10 : Quantifier
G25 : Visualiser	G12 : Copier, Enregistrer, Agencer, Formuler
	G13 : Acheminer
	G15 : Laisser trace, Laisser empreinte
	G16 : Prendre connaissance
	G17 : Affiner
	G20 : Assister, Se faire admettre
	G23 : Rechercher
	G24 : Savoir
	G25 : Faire lire, Faire connaître, Faire apparaître, Obtenir, Rendre visible

A N N E X E 5

Intersections lexicales entre matrices 2 à 2

Graphes		Items lexicaux en commun
G1	G20	ENTRER, FAIRE PARTIE DE
G2	G5	MODIFIER
G2	G8	MODIFIER
G2	G10	ADDITIONNER
G2	G17	MODIFIER
G2	G20	JOINDRE
G3	G8	SUPPRIMER
G3	G11	ANNULER, DEFAIRE, DETRUIRE, SUPPRIMER
G3	G22	DISSOCIER
G4	G23	METTRE A PART
G5	G8	CHANGER, MODIFIER
G5	G17	CHANGER, MODIFIER
G7	G16	COMPULSER, CONSULTER, DEPOUILLER, EXAMINER, FEUILLETER, JETER COUP D'OEIL, PARCOURIR, REGARDER, VOIR
G7	G23	RECHERCHER
G7	G24	EXAMINER, INSPECTER, VOIR
G8	G11	FAIRE DISPARAITRE, EFFACER, ELIMINER, RETIRER, SUPPRIMER

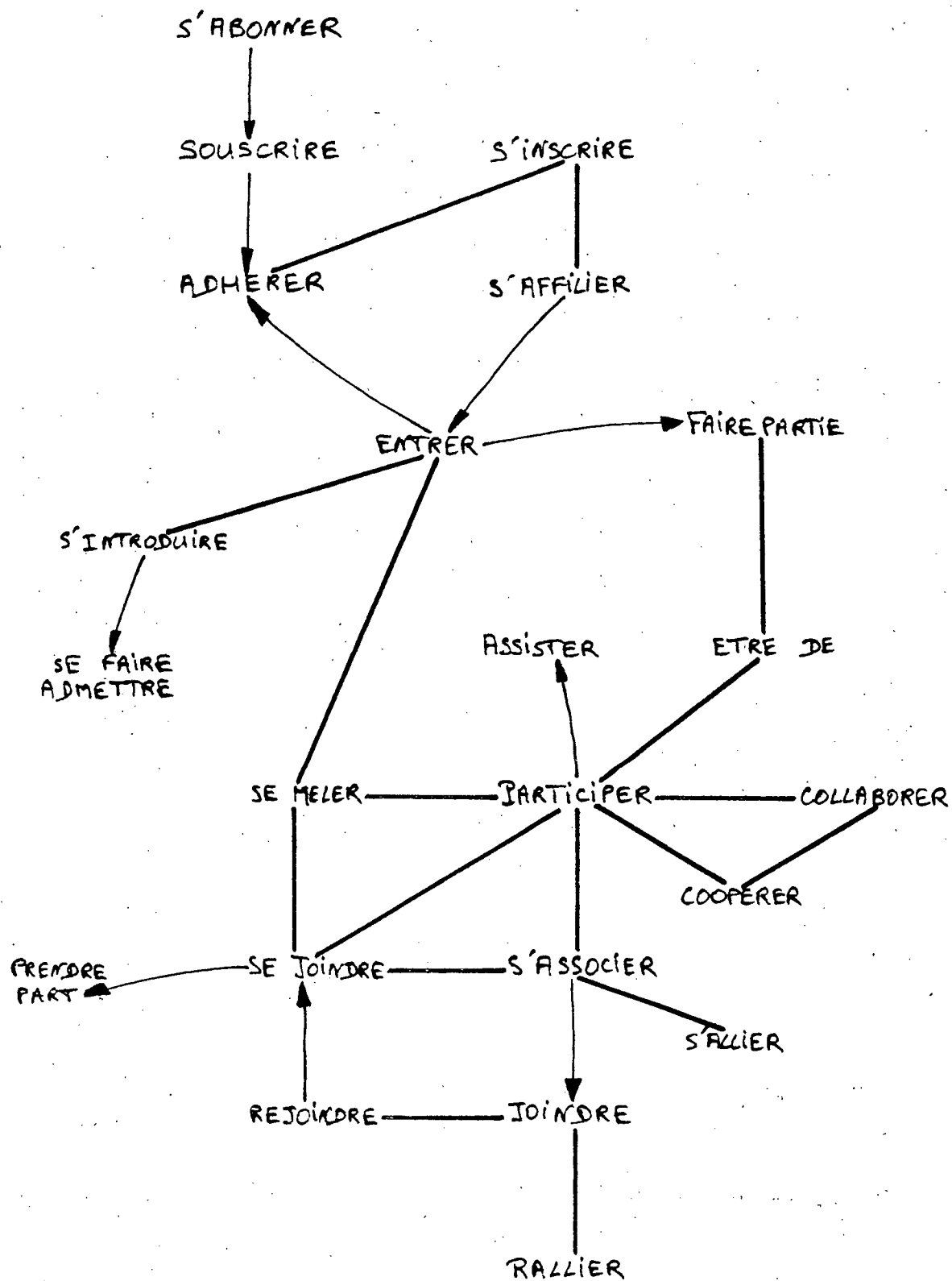
Graphes		Items lexicaux en commun
G8	G17	AMELIORER, CHANGER, MODIFIER, PERFECTIONNER, REPENDRE, RETOUCHER, REVISER, REVOIR
G8	G22	ELIMINER, ENLEVER, OTER, RETIRER
G9	G12	AGENCER, COMPOSER, CONCEVOIR, CONFECTIONNER, CONSTITUER, CONSTRUIRE, CREER, ELABORER, ENGENDRER, FABRIQUER, FACONNER, FAIRE, FORMER, MODELER, PREPARER, PRODUIRE, REALISER
G9	G25	PRODUIRE
G11	G22	ELIMINER, RETIRER
G12	G15	COPIER
G12	G25	PRODUIRE
G14	G23	APPELER, DESIGNER
G16	G24	EXAMINER, VOIR
G22	G23	DETACHER, SEPARER

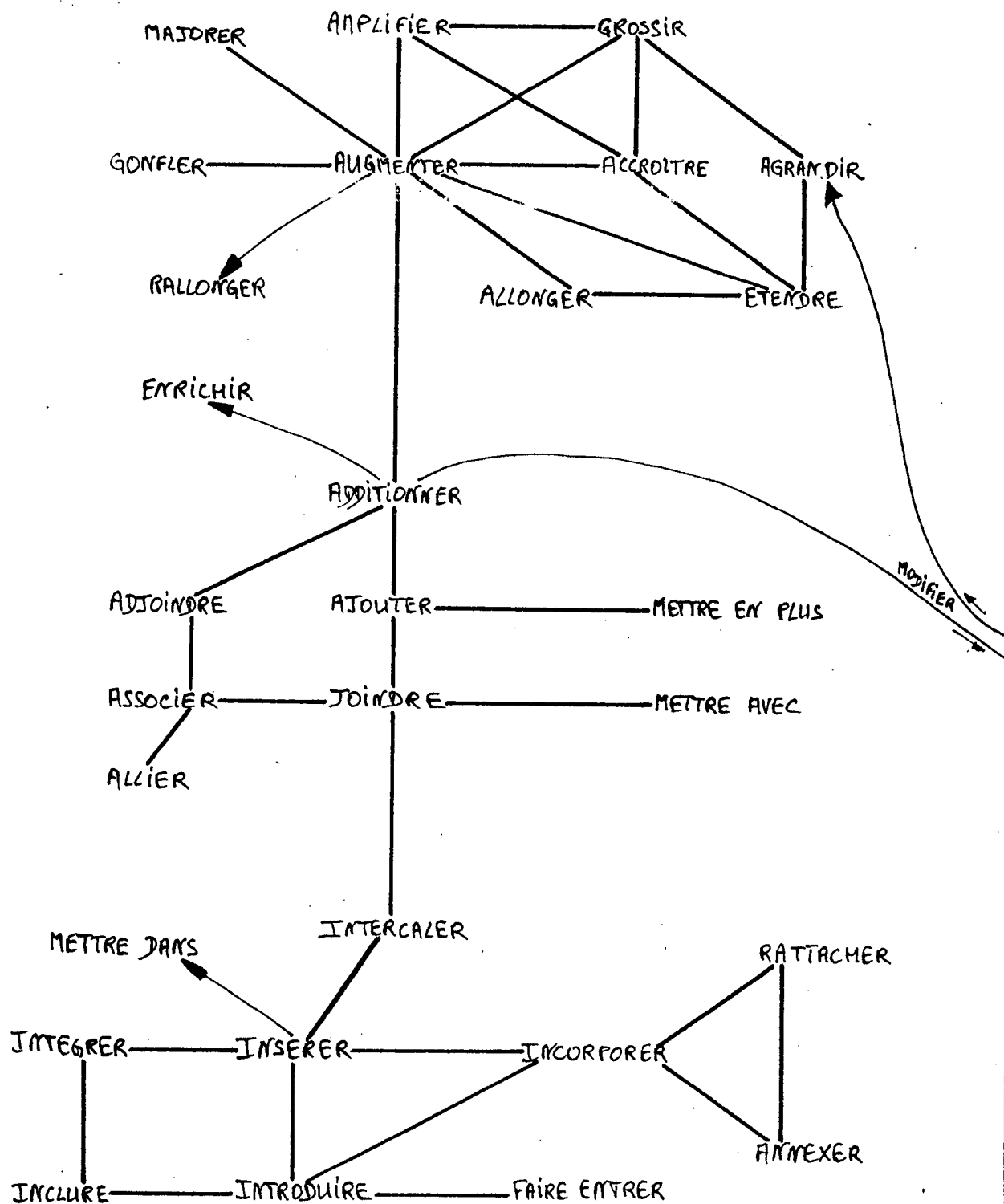
A N N E X E 6

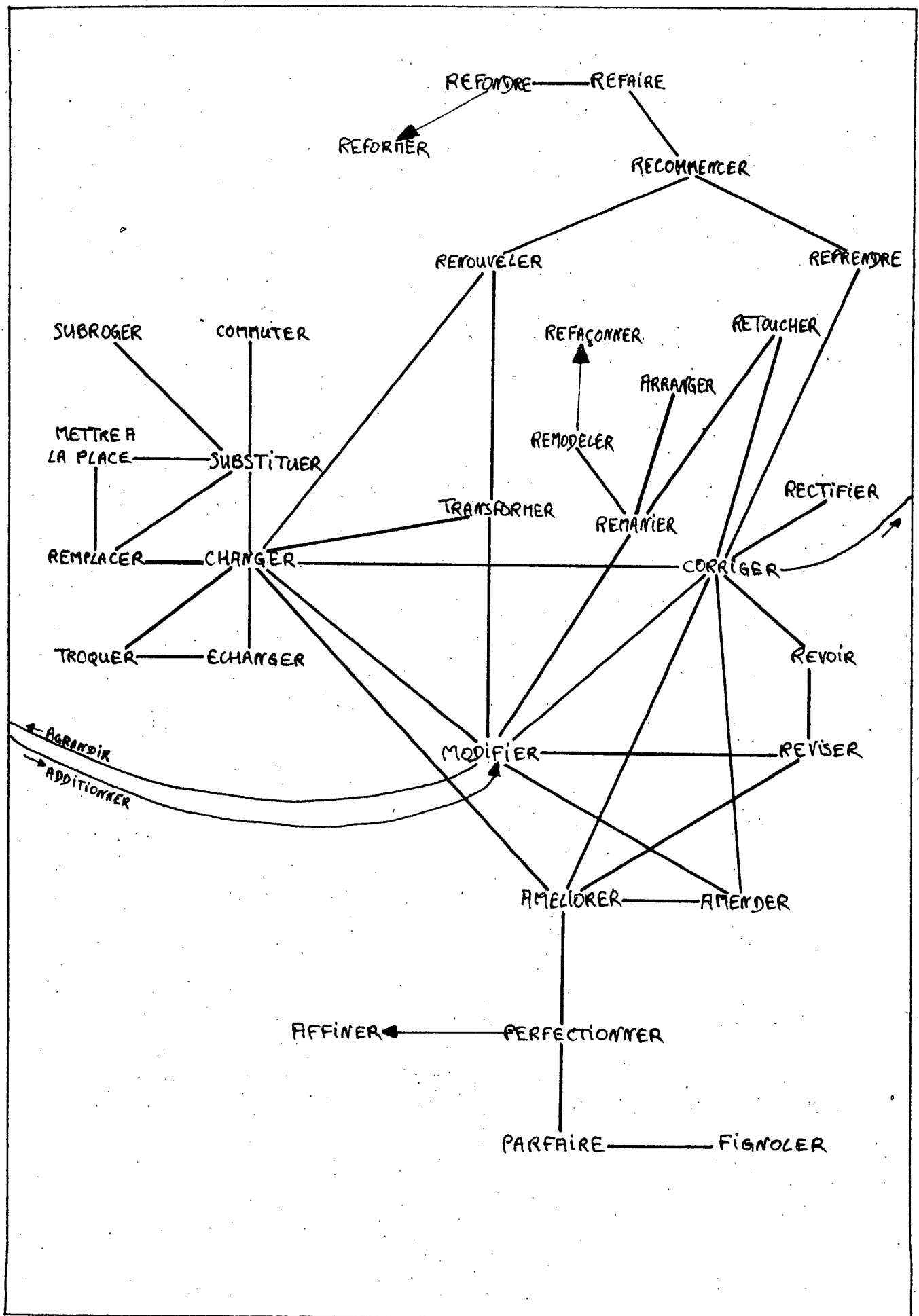
Représentation graphique du T-Graphe

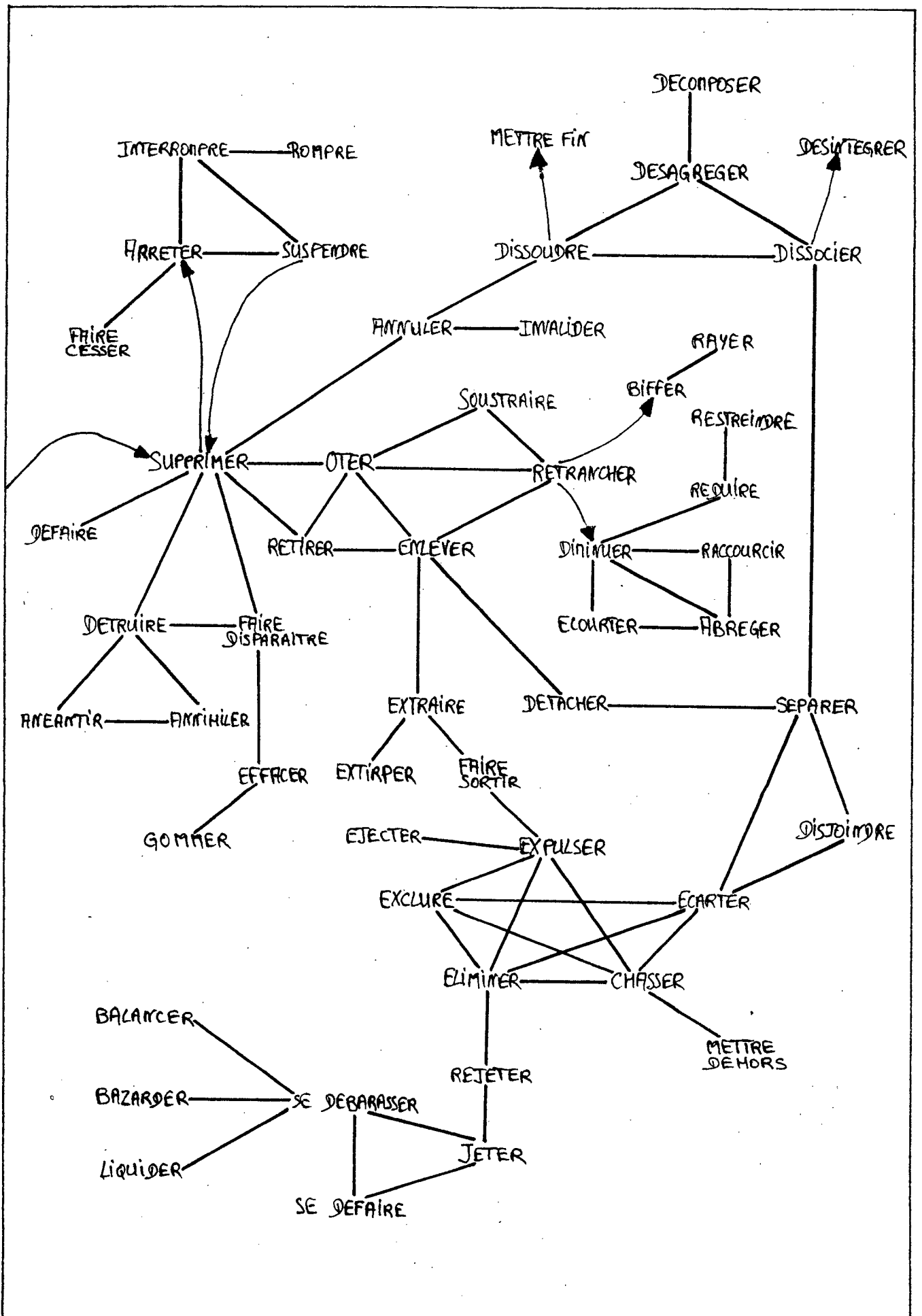
(On peut reconstituer les groupements de plusieurs graphes en mettant les feuilles bout à bout).

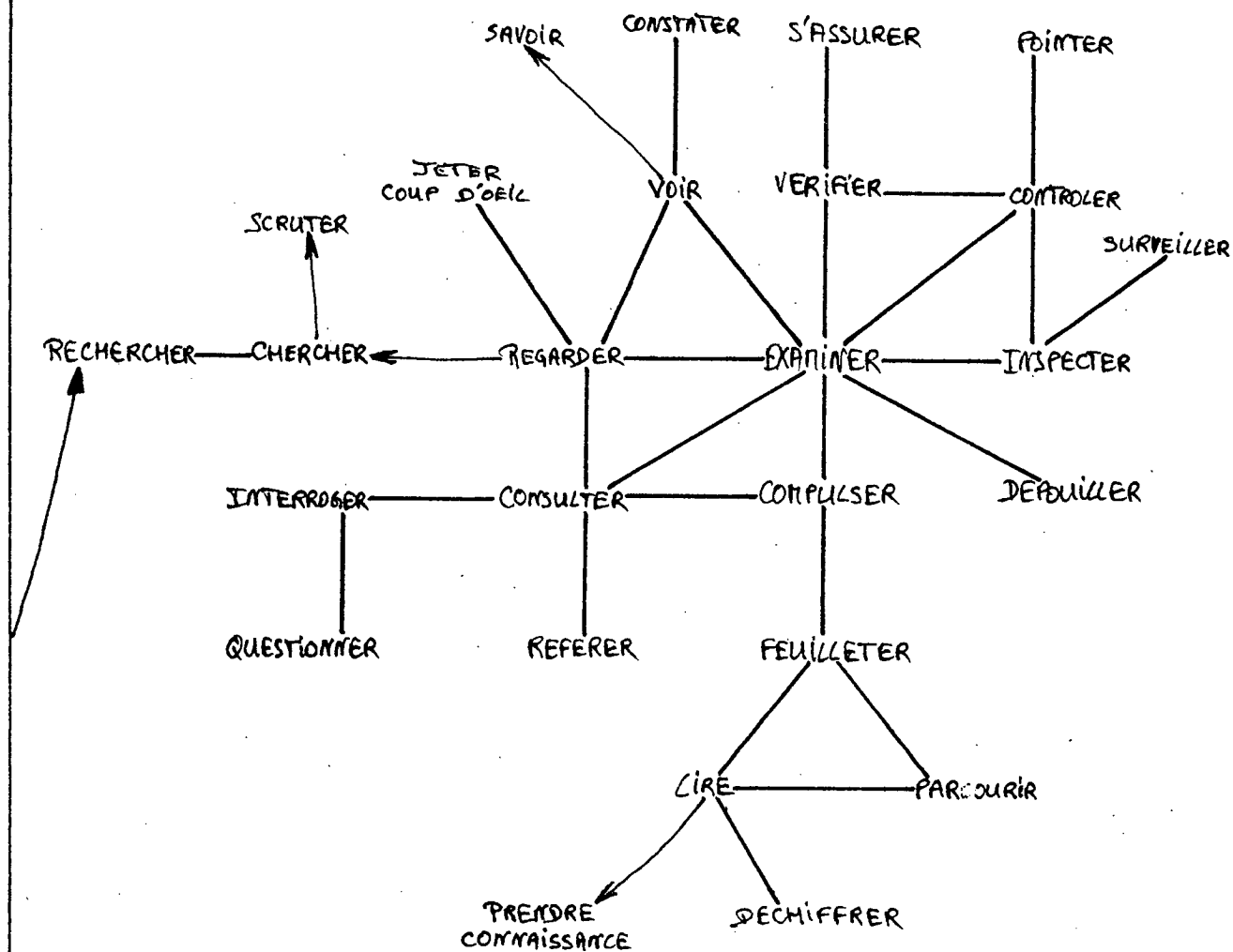
Remarque : si nous observons le "graphe des interactions entre les matrices", que nous figurons page 24, nous observons qu'il existe 4 groupements de plusieurs graphes et 8 graphes isolés. Nous n'avons pas inclu dans la représentation du T-graphe les 8 graphes isolés, ceux-ci étant déjà présentés en annexe 3 (avec pour seule différence qu'en annexe 3 tous les arcs sont représentés, qu'ils soient symétriques ou non symétriques)..

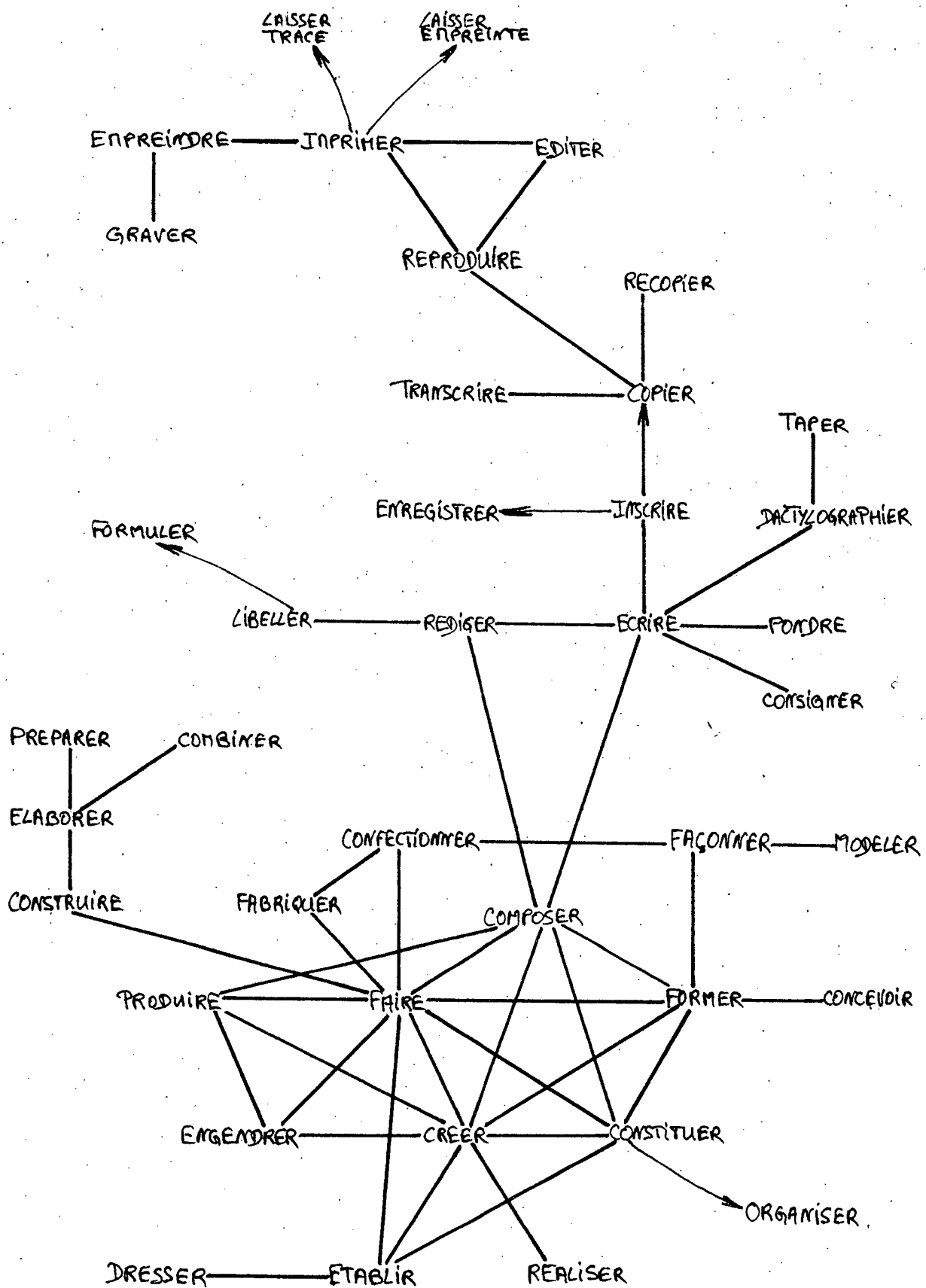












Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique