CHAPTER 7

CONCLUSIONS AND FURTHER RESEARCH

What we have described here are some results of a design study for what we envisage as a much larger-scale project, to design, construct, implement and test an IR system based specifically and explicitly upon the recognition that the representation of information need is the central problem of IR. We view information needs as Anomalous States of Knowledge and therefore aim to design an IR system in which ASKs have explicit structural representations. Our intention, then, is that the system should consist of a structural text-analysis program, a collection of abstracts, a suite of retrieval algorithms and a mechanism whereby user and system interact. The goals of the system can be described as follows:

(a) To build a representation of the enquirer's state of knowledge in the form of a network of associations between words.

(b) To examine this structure, and to interpret certain characteristics of the network as potential anomalies in the state of knowledge.

(c) To search structured document descriptions with a view to resolving the anomalies; i.e. to modify the searcher's network, using components of document structures, so that the identified characteristics disappear.

These system objectives will be achieved through dialogue with the user, and the system would operate as shown in Fig. 1.

We consider that the design study has at least partly justified our initial premises and assumptions in the following ways:

1. It is possible to obtain problem statements from IR system users which can be used to derive adequate representations of ASKs;
2. Abstracts appear to be sufficient document surrogates for information representation in such a system;
3. Even though there are problems in our representations of need and information, they appear to have identifiable characteristics which are potentially useful for IR purposes.

Some aspects of our system have been rather less well studied than others, and some significant problems remain to be resolved. These include:

1. retrieval mechanisms
2. refinement of the analytic procedures
3. classification of ASKs
4. the interactive environment

In the next stages of our long-term project, we intend to investigate these problems in detail. The first point we hope to resolve by collec-

ting and analysing a number of problem statements and documents judged relevant to them. In this way we hope to find regular relationships between ASK representations and representations of information which have been judged suitable for resolving those ASKs. These data will also be suitable for further refinement of our analysis, which we also intend to augment by the addition of some simple linguistic analysis. Classification of ASKs will be investigated both through these data, and through a separate project which we hope will give us access to something like 3000 problem statements, not necessarily with related documents. This corpus will allow a much more reliable classification and investigation of the characteristics of ASKs than has heretofore been possible. And we intend to investigate the interactive environment, and indeed the proposed system as a system, by constructing a prototype and testing it with both simulated and a small number of natural users. We hope to begin work on stage two of the project in Autumn, 1980.

ACKNOWLEDGEMENTS

We wish to thank Mrs. Alina Vickery for her generous help in this investigation, the British Library Research and Development Department for supporting this project, both financially and morally, and especially Ms. Helen Brooks, our collaborator, whose energy and insight were invaluable in this investigation.

REFERENCES

BELKIN, N.J. 1977 a. A concept of information for information science. PhD Thesis, University of London, 1977.

BELKIN, N.J. 1977 b. Internal knowledge and external information. In: The Cognitive Viewpoint, proceedings of an international workshop, Ghent, 1977. Ghent, The University of Ghent, 1977: 187-194.

BELKIN, N.J. in press. The problem of 'matching' in information retrieval. In: Harbo, O. and Kajberg, L., eds. Theory and applications of information research. Proceedings of the Second International Research Forum on Information Science. London, Mansell, 1980.

BELKIN, N.J., BROOKS, H.M. & ODDY, R.N. 1979. Representation and classification of anomalous states of knowledge and information for use in interactive information retrieval. In: Human aspects of information science. Proceedings of the Third International Research Forum in Information Science. Oslo, Norwegian Library School, 1979 (in press).

BELKIN, N.J. & ROBERTSON, S.E. 1976. Information science and the phenomenon of information. Journal of the ASIS, 27, 4(1976): 197-204.

BROOKS, H. 1978 The knowledge structures underlying information needs. M.Sc. Thesis, Centre for Information Science, The City University, London, 1978.

BROOKS, H. ODDY, R.N. & BELKIN, N.J. 1979 Representing and classifying anomalous states of knowledge. Paper presented at Informatics 5 Oxford, 1979. To be published by Aslib.

CHRISTOFIDES, N. 1975 Graph theory: an algorithmic approach. New York and London, Academic Press, 1975.

CROFT, W.B. & VAN RIJSBERGEN, C.J. (unpublished) A single-link clustering algorithm for use in experimental document retrieval systems. Computer Laboratory, University of Cambridge.

HARBO, O. & KAJBERG, L. in press. Theory and applications of information research. Proceedings of the Second International Research Forum in Information Science. London, Mansell, 1980.

KAHN, R.L. & CANNELL, C.F. 1957 The dynamics of interviewing. New York, John Wiley, 1957.

KISS, G. 1975. An associative thesaurus of english: structural analysis of a large relevance network. In: Kennedy, A. & Wilkes, A., eds. Long-term memory. New York and London, Academic Press, 1975. 103-121.

MANN, M.B. 1944 The quantitative differentiation of samples of written language. Psychological Monographs, 56 (1944): 41-74.

MOSES, E.R. 1959 A study of word diversification. Speech Monographs, 26 (1959): 308-312.

ODDY, R.N. 1975 Reference retrieval based on user-induced dynamic clustering. PhD Thesis, University of Newcastle-upon-Tyne, 1975.

ODDY, R.N. 1977 a. Information retrieval through man-machine dialogue. J. Docum., 33 (1977): 1-14.

ODDY, R.N. 1977 b. Retrieving references by dialogue rather than by query formulation, J. Informatics, 1 (1977): 37-53.

PAYNE, S.L. 1951. The art of asking questions. Princeton, N.J., Princeton University Press, 1951.

PORTNOY, S. 1973. A comparison of oral and written verbal behaviour. In: Studies in verbal behaviour, K. Salzinger and R.S. Feldman, eds., New York, Pergamon, 1973: 95-151.

ROBERTSON, S.E. 1979 Indexing theory and retrieval effectiveness. Drexel Library Quarterly, 14, 2 (1979): 40-56.

ROBERTSON, S.E. and BELKIN, N.J. 1978. Ranking in principle. J. Docum., 34, 2 (1978): 93-100.

ROBERTSON, S.E. and SPARCK JONES, K. 1976. Relevance weighting of search terms, JASIS 27 (1976): 129-146.

SPARCK JONES, K. and VAN RIJSBERGEN, C.J. 1975 Report on the need for and provision of an 'ideal' information retrieval test collection. Cambridge, University Computer Laboratory, December 1975.

TAYLOR, R.S. 1968. Question-negotiation and information seeking in libraries. College and Research Libraries, 29 (1968): 178-194.

VAN RIJSBERGEN, C.J. 1971. An algorithm for information storage and retrieval. Computer Journal 14, 4 (1971): 402-412.

WERSIG, G. 1971. Information - Kommunikation - Dokumentation. Pullach bei Munchen, Verlag Dokumentation, 1971.

APPENDIX A

Evaluation package for problem statements

Dear

You may remember being interviewed by me about the research you were doing and what sort of information you needed in connection with it, while having an on-line search carried out at Central Information Services, Senate House, The University of London. Thank you for your cooperation on this occasion. May I ask you now to help with the second part of this project? This involves you evaluating the results of our analysis of the interview. I have enclosed a transcript of the interview and the analysis - which is presented in two different formats.

About the analysis

This analysis is based on the idea of words as concepts and examines statistically the strength and pattern of relations between concepts (words) present in a piece of text. The analysis will, eventually, be used to construct an internal machine representation based on the problem which brought the person to have an on-line search carried out. Here the analysis has been used to construct representations which are presented in two formats. The two formats are not equivalent representations because although they are formulated from the same initial data they represent manipulations of different parts of the data.

What to do

What I would like you to do is to study the representations and complete the attached questionnaires. Please do not spend more than thirty minutes doing this - it is not necessary. Bear in mind that we would like you to consider the representations, as far as possible, in terms of your problem at the time of the interview, rather than at present.

Could you please return the completed questionnaire to me, as soon as possible, in the stamped addressed envelope provided. (You may retain everything else). Your responses will enable me to get some idea of how useful the method of analysis and formats are as representations of your information problem and ultimately may help improve information retrieval services. I enclose a small fee of £2 to cover your expenses. If you have any queries, please contact me or Dr. Belkin on 01-253 4399 ext 231. Thank you for your cooperation.

Yours faithfully,

Helen Brooks

Format A.M. - Association Map

The association strengths for all the possible pairs of concepts present in the interview were calculated. The higher the score, the more closely two concepts are statistically associated and vice versa.

The 'map' is a graphical interpretation of the network produced by the 40 highest associations. On the map you will see that there are three types of connecting line:

a) A very thick line which represents a very high association strength;
b) A full thin line which represents medium association strengths;
c) A dashed line which represents weak associations.

For your convenience on the reverse of each complete 'map' the network has been redrawn for each of these three levels of association, high (1), medium (2) and low (3), displaying only those concepts associated at that particular level.

PLEASE REMEMBER - the analysis is based on the idea of words as concepts. Sometimes the actual words used in the interview were stemmed so as to produce a concept e.g. the words USE, USER, USES, USED, USING, in the interview would be stemmed to produce the concept 'US'.

QUESTIONS

(i) Looking at the general shape of the 'map', do you think that the compactness and degree of inter-relation between the concepts (or lack of it), is an accurate representation of how these concepts were related in your mind at the time of the interview?_____YES/NO

(ii) Looking at the relations between concepts, are there any which you feel are too strongly associated?_____YES/NO
Any which you feel are too weakly associated?_____YES/NO

If too strongly associated please give examples

If too weakly associated please give examples

(iii) Looking at the concepts themselves, are there any concepts missing from the 'map' which you feel are essential to your problem as it was at the time of the interview?_____YES/NO

Could you please give examples of these missed out concepts?

(iv) Any other comments?

Format A.C. - Association Clusters

This format presents single link clustering of concepts occuring in the interview. Concepts occuring in the same 'box' are strongly associated with each other. All the members of a particular 'box' are then strongly associated with the members of the adjoining 'box' at the next splitting level. Thus in the example given below:

A B C D E F H I J

A & B are highly associated, as are D, E & F with each other and H & I. The group D E F is quite strongly associated with C and at a less strong level with H I. C is even less associated with H I than D E F and neither are very highly associated with J. A B occurs at the same association level as D E F and like D E F is fairly strongly associated with H I. It is not however associated at all strongly with C.

The higher up the page the sets of concepts join up or the further they are separated the less strongly they are associated. The splitting level is given by L followed by the association strength, low numbers representing weak associations and vice versa.

In some cases lack of space has prevented all the concepts from being displayed and so only the more strongly associated are shown.

PLEASE REMEMBER - the analysis is based on the idea of words as concepts. Sometimes, the actual words in the interview were stemmed so as to produce a concept e.g. the words USE, USER, USES, USING, USED in the interview would be stemmed to produce the concept 'US'.

QUESTIONS

i) Looking at each 'box', and remembering that all the concepts within it are closely related, are any concepts grouped together which you feel should not be so strongly related?_____YES/NO
Any which are not grouped together in a 'box' but which you feel should be?_____YES/NO
If too strongly related please give examples:

If not related strongly enough please give examples:

ii) Looking at the association level at which the 'box' appears, do you think that any of the concepts appear at too high a level of association?
YES/NO
If too high please give examples:

If too low please give examples:

iii) Looking at the relationships between the 'boxes' are there any which are not closely enough linked?_____ YES/NO
Any which are too closely linked?_____ YES/NO
If too close please give examples:
If too far apart please give examples:
Any other comments?

FINALLY

We would now like you to compare the formats. In making this comparison, please remember:

- The formats do not display the same information since different parts of the original data were manipulated in different ways to produce each.
- Eventually the analysis of such interviews will be used by a computer to produce an internal machine representation of the user's problem as part of an on-line search procedure, and will therefore not actually be seen by the user.

Having studied both formats:

i) Does one of them seem to you to be a better, more successful representation of your problem, as it was at the time of the interview, than the other?_____YES/NO

If you answered YES -
a) Which did you consider better?_____ Format A.C./Format A.M.

b) In what ways do you consider it better?

If you answered NO -
c) Was this because you think that on the whole both are successful representations or that both are not successful representations of your problem, at the time of the interview?_____SUCCESSFUL
NOT SUCCESSFUL

Any additional comments?

APPENDIX B.

Evaluation package for abstracts

Dear

We are writing to ask your help in a research project which we are conducting with support from the British Library Research and Development Department. This project is a Design Study for an interactive information retrieval system which would be based on structural representations of users' information needs and of the information underlying texts. An important aspect of the Design Study is evaluation of the appropriateness and accuracy of these structural representations.

As the texts for our study we have chosen a number of abstracts from the library and information science literature, including one of yours. Enclosed you will find a copy of that abstract, a representation of it based on our analytic methods, and a brief questionnaire about the representation. We would very much appreciate your considering the representation and completing the questionnaire, which we think should take no longer than one hour. As partial recompense for your trouble, we also enclose a cheque for £5.00. Please try to return the questionnaire in the enclosed stamped envelope as soon as possible, and in any event before June 1st. All replies will be kept fully confidential. If you wish to know the results of this work, please check the box at the bottom of the questionnaire.

Thanking you in advance for your cooperation, and looking forward to your reply, we are,

Sincerely yours

Dr. N.J. Belkin
Lecturer in Information Science
The City University

Dr. R.N. Oddy
Lecturer in Computing
University of Aston

STRUCTURAL REPRESENTATION QUESTIONNAIRE

About the analysis

This analysis is based on the idea of words as concepts, and examines statistically the strength and pattern of relations between concepts in a piece of text. Although in the eventual system this analysis will be used to construct an internal machine representation of the text, here we have used it to construct a simplified network or 'association map' representation which we hope is easier for humans to interpret.

About the representation

Association strengths for all the possible pairs of concepts in the abstract were calculated. The higher the score, the more closely the two concepts are statistically associated, and vice versa.

The 'association map' is a graphical interpretation of the data, produced by using only a limited number of associations and by reducing the variety of association strengths. On the 'map' you will see that there are two or three types of connecting line:
(a) a very thick line which represents a high association strength;
(b) a full thin line which represents an intermediate association strength;
(c) a dashed line which represents a weak association.

PLEASE REMEMBER the analysis is based on the idea of words as concepts. To accomplish this some words have not been included in the analysis (e.g. function words such as A, OF, etc.) while others have been stemmed so as to produce a single concept from several words. For instance, the words USE, USER, USES, USED, USING in the text would be stemmed to produce the concept 'US'. This stemming will result in some ambiguous nodes in the 'association map'.

QUESTIONS (Please circle YES or NO for each)

(i) Looking at the general shape of the 'map' do you think that the compactness and degree of inter-relation between the concepts (or lack of it) is an accurate reflection of how these concepts were related in your mind at the time you wrote the article?_____YES/NO

(ii) Looking at the relations between the concepts, are there any which you feel are:
too strongly associated?_____YES/NO
too weakly associated?_____YES/NO
If too strongly associated, please give examples:

If too weakly associated, please give examples:

(iii) Looking at the concepts themselves, are there any concepts missing from the 'map' which you feel are essential to the article?___YES/NO
Could you please give examples of these missed-out concepts?

(iv) If you answered NO to question (i), do you think that the abstract itself is an accurate representation of your article?_____YES/NO

(v) Any other comments? (Please continue overleaf if necessary)

Please tick here if you would like details of results to be sent to you ☐
Thank you for your cooperation.

APPENDIX C.

Summaries of problem statements.

(From Brooks, 1978)

1-27 Oral statements

30-37 Written statements

No. 1

-Looking at discrimination and reversal learning in severely subnormal children and also the development of verbal mediation in ESN children.

-Wants papers that are to do with verbal regulation behaviour in severely subnormals, but not necessarily behaviour modification, e.g. papers on the use of conceptual terms and the latest work on reversal learning.

-Previous studies on verbalisation have tended to treat subjects as an homogenous group with respect to their language ability. Is looking at the ways in which tests of mediation can be developed and use these rather than treating the subjects as an homogenous group.

No. 2

-Interested in the sociological aspects of the effects of art education.

-Looking at the way teachers operate in 4 schools in Hounslow - 2 in the multiracial centre of Hounslow and 2 in the more English West Hounslow. Taken a junior and a secondary school in each area and looked at their art. Talked to kids; interviewed parents; used a questionnaire.

- Wants to find out if any similar work done in Britain or U.S.A. (or related research in marginal areas).

- Wants references which could support (his theories / results he has obtained) or which give something to react to - for or against.

No.3

-Problem is to do with extra corporal membrane oxygenation.

-Want to find out if there is some answer to the clotting problems involved.

-Want to find what effects the position of the cannulars has on cardiac output.

- So far unable to find any references to experiments done cannulating various blood vessels in dogs and looking at cardiac output but feels sure that such experiments must have been carried out.

No. 4

-Looking at the corrosion aspects of dental implants.

- Wants bibliographies on alloys, especially chromo-cobalt dental markings on stainless steel and the corrosion of these in biological tissues.

No. 5

- Had a case where a total hip replacement was carried out two years ago but since then the patient has complained of painful hips and so a revision of the orthoplasty was done. This revealed lesion in the bone and the presence of a carcinoma

- Heard of similar incidents. Wants to know if anything written about possible connections / correlation between the cement used in hip replacements and the secondary change into malignant lesions.

No. 6

-Looking at types of classroom environments and their relationship with student outcomes; particularly on student self-concepts, anxiety and locus of control.

- Wants to find out what research done in this area.

-Wants empirical rather than philosophical or polemical papers.

No. 7

-About to start a project analysing membrane proteins during and after fusion in rat muscle skeletal cells.

- Wants literature searched to see if there has been any work done on various aspects of changes in membrane proteins during fusion.

No. 8

- Trying to relate verbal comprehension to a number of variable e.g. intelligence, ego involvement, extroversion , anxiety etc.

-Wanted to check on what work done in this field over the last twelve months.

No. 9

-Looking at factors affecting germination on the plant of wheat grains.

-Wanted to find out what other people had done and whether their results agreed with what had been found.

- Bibliography wanted partly because writing a thesis on the subject and partly because going to give a talk at an agricultural college. Interested in agricultural aspects therefore. (which had not really considered up to then.)

No. 10

- Looking at the economics of fish farming , e.g. inland fisheries as a means of increasing GNP in third world countries.

- Wants to adopt a systems approach and look at a community as a whole,e.g. waste from the scheme could be fed to fish in the fish farms which in turn could be fed to chickens or people, which produce waste which could be recycled.

-Information wanted on the economics of fisheries and inland fresh water fisheries in the UK or the third world.

No. 11

-Writing an introductory text on growth in farm animals.

-Wants to check up on some facts and on some ideas other people have suggested e.g. the influence of light on growth; of gonadal hormones; of vegetarian diet etc.

No. 12

-Wants to investigate patterns of nonverbal communication.

-Project arose from the observation that Swahili speaking students used gestures when speaking Swahili which they omitted when speaking English.

-Wants to establish that there are language-related gestures and also investigate what the relationship is between them and the language.

No. 13

-Looking at the education voucher scheme just introduced in Ashford in the UK and comparing it with a scheme run at Allan Rock in California.

-Wants an up to date assessment of how the scheme at Allan Rock is going.

No. 14

- Looking at computerised hospital drug information systems.

- Wants two basic sorts of information. What systems are in operation and what the doctors require from these systems, and secondly what sort of information on the drugs themselves one would want to put onto a drug information system.

No. 15

- Evaluating Synopsis Journals. Looking at personal opinions of authors, publishers and readers. Comparing factors like delay times between synopsis and conventional journals.

- Knows that other evaluations have been carried out and would like details.

No. 16

-Looking at the information services of professional institutes in business management - libraries, publications, seminars , conferences etc.

- Would like to know if similar evaluations done of libraries or information services in this field.

No. 17

-Wanted to do a comprehensive review of all the literature concerned with lactose. Interested in its chemistry and the synthesis of new derivatives which might have biochemical applications.

- Not interested in the biochemistry of lactose.

No. 18

-Looking at management information systems - which are designed to give easy access to a wide variety of information to people not experienced in information handling.

-Not sure what information on such systems available. Systems tend to be set up which are specific to a particular firm and are therefore not well documented.

No. 19

-Wants to keep up to date on parasitic diseases because has to write regular chapters for text books on them.

No. 20

-Looking at cancer of the prostate gland with respect to the mechanism of action of androgens.

-The project is a joint clinical and scientific one. The basis for the clinical side is the case history

(No. 20 cont.)

of a patient with a terminal case in whom they have found that giving a combination of androgens and anabolic steroids seemed to produce some relief .

-Wanted to check whether any drug treatments of this particular disease had been tried before, in animal or human systems.

No. 21

-Interested in the concept of plasticity in recovery from brain damage when the brain is young. Is writing a review article on this subject.

- Wants data on hemispherectomy or hemidecortication operations which hopes will support his theory of plasticity.

No.22

- -Looking at motivation for change in patients and how this affects their therapeutic outcome.

-Wants to know if similar work done , correlating motivation of patients with therapeutic results.

No. 23

-Starting research on Echinococcosis / Hydatid disease.

-Particularly interested in reports of the disease in libya. Thinks not much work done there.

-Wants information on the epidemiology and methods of diagnosis.

NO. 24

-Doing research on the Bender-gestalt motor test.

-Problem because different testers often get different results

-Wants to find reliability studies of interscorer marking.

-Main interest is in the Koppits scoring system

No. 25

-Looking at changes in children's use of strategies with development. Trying to link this with the development of the frontal lobes.

-Has heard of some work done by the Moscow Brain Institute on the development of frontal lobes

-Wants to find where the Institute got its data from and how the data might relate to what doing.

No. 26

-Looking at the susceptibility of anopheline mosquitoes to insecticide; particularly with respect to the genetics of insecticide resistance. Project involves trying to establish an acceptable level of low insecticide resistance in the mosquito population.

-Also wants to do a biochemical study of anopheline mosquitoes so can try to develope alternative methods of control.

No. 27

- Doing research on detection of viruses inside cells using gas liquid chromatography.

-Wants to find out if any similar work done.

No. 30

I want a comprehensive search of recent literature concerning the noise (i.e. ground vibration) which is present at any particular time within the earth. The source of this seismic noise will be either human activities or geological events depending on the locality considered. I am interested in surveys of both types of noise.

Given that a certain source produces a certain frequency spectrum of noise, the spectrum of the noise as perceived at some distance from the source depends on the transfer function of the intervening materials i.e. the extent to which the different frequencies are attenuated in travelling from source to receiver.

Hence papers concerning both the spectra of emitted noise and the attenuation of such noise in the ground are of equal interest.

No. 31

I am working in the field of general relativity, in particular on junction conditions for boundaries between coordinate patches on a manifold and how these give rise to so called "admissible" coordinate systems. One problem, for which I hope to find a solution in terms of these coordinates , is that of a collapsing sphere of ideal fluid. An early formulation of junction conditions (conditions de raccordment) was by A. Lichner-owicy in 'Theories relativistes de la gravitation et de l'electromagnetisme.' but since then many authors have avoided them as being too difficult to implement.

No. 32

Hearing tests are currently performed by a trained operator using specialist equipment. I am interested in the application of computers and micro-processors to automate this mainly routine testing to enable larger scale screening of hearing defects to be economically performed.

Any literature relating to computer assisted or automated hearing tests would be useful and also the application of integrated circuits and circuit techniques in this field would be appropriate.

Papers in the field of general computer assisted diagnosis are not required, only those relating to hearing or audiometry would be useful.

No. 33

Chromobacterium Violaceum synthesises and excretes a purple pigment (violacein) during growth as well as cyanide. Violacein has antibiotic properties.It is a derivative of the amino-acid Tryptophan.

Our research is to study the chemical and antibiotic properties of violacein, the growth conditions for its synthesis , the relationship of its excretion to that of cyanide, and the metabolic pathway of formation.

No. 34

Interested in how both microbial and higher eukaryotes regulate nuclear RNA synthesis. I am not at this stage concerned with mitochondrial or chloroplast RNA synthesis nor in viral RNA synthesis. My principle interest is in the coordination of transcription with translation- that is to say, I am interested in the regulation of RNA synthesis as an aspect of growth control. This means that studies involving nutritional transitions, starvations , shifts-up etc. are of particular importance. More attention should be paid to the synthesis of stable RNA species -rRNA and tRNA, than to the synthesis of mRNA. Changes in the synthesis of RNA in response to hormonal stimuli are not of interest.

No. 35

The object of this study is the biochemistry and physiology of vascular wilt pathogens. The areas I will need to cover are for the pathogens Verticillium albo-atrum , Ceratocystis ulmi, Fusarium (cubense, vasinfectum, lycopersici, conglutinans, pisi). I think this will cover the main Fusarium wilt pathogens and titles should include the terms wilt, wilt pathogens, wilt diseases, wilt disease physiology or any combination of these with the specific pathogen in question, or banana wilt, tomato wilt, cotton wilt, Dutch elm disease or wilt involvement with other non-named hosts.

No. 36

The major aspect of my research work is concerned with the functions of a hind-limb preparation., and I shall be using this preparation for the next few (at least 3) years. I am interested in obtaining references on this topic which covers the last six years or as far back as your records will go. I'm not sure how this will be formulated but basically the hind-limb is a muscle preparation and is perfused. A wide variety of animals are used to study the functioning of this hind -limb or isolated limb preparation.

No. 37

I would like you to send for me any references and any articles which deal with bleeding in early pregnancy and the effect on pregnancy, labour and the baby. Also if there is any relation of placental site to the bleeding in early pregnancy.

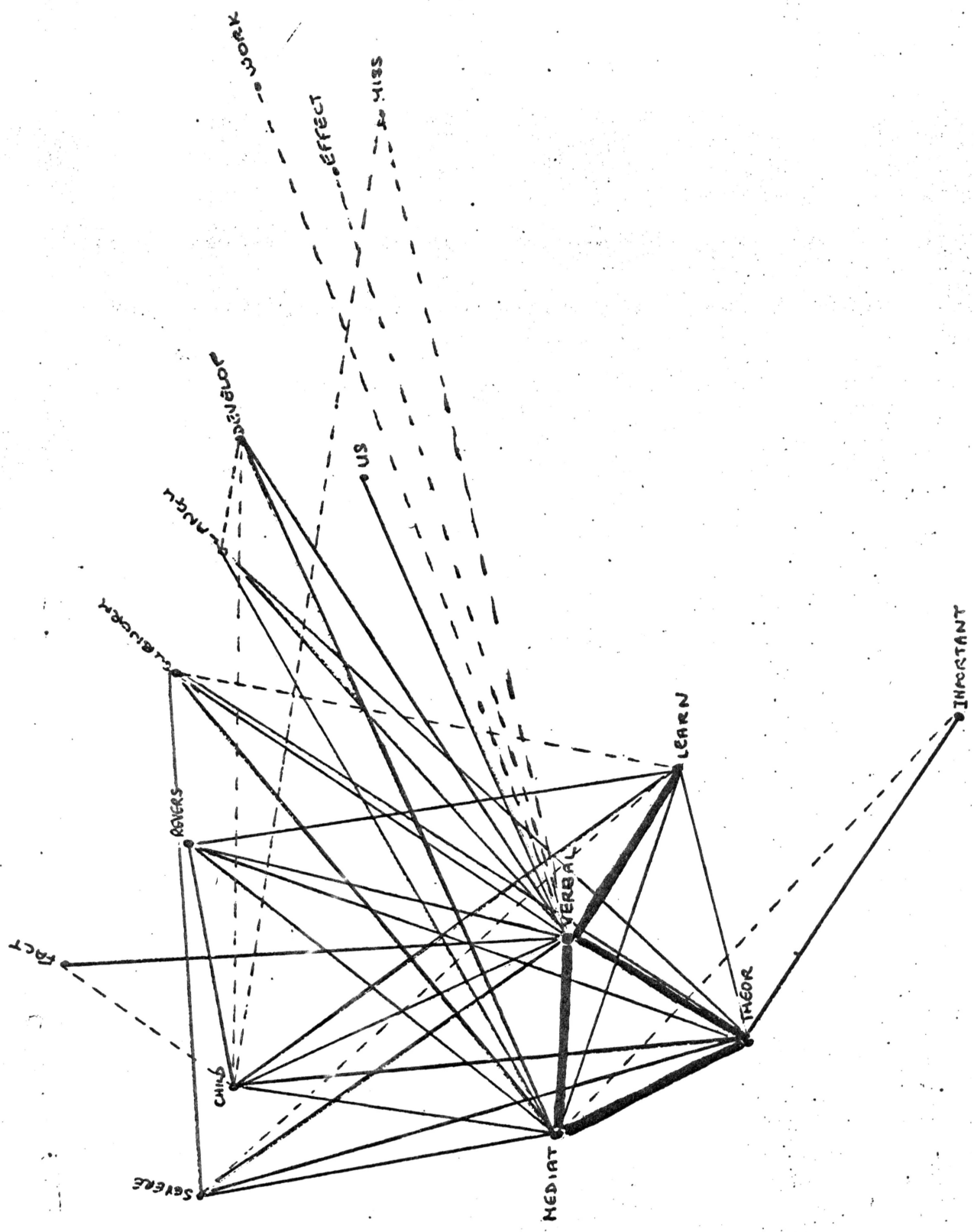
I am doing a survey for the outcome of a patient who had bleeding in early pregnancy.

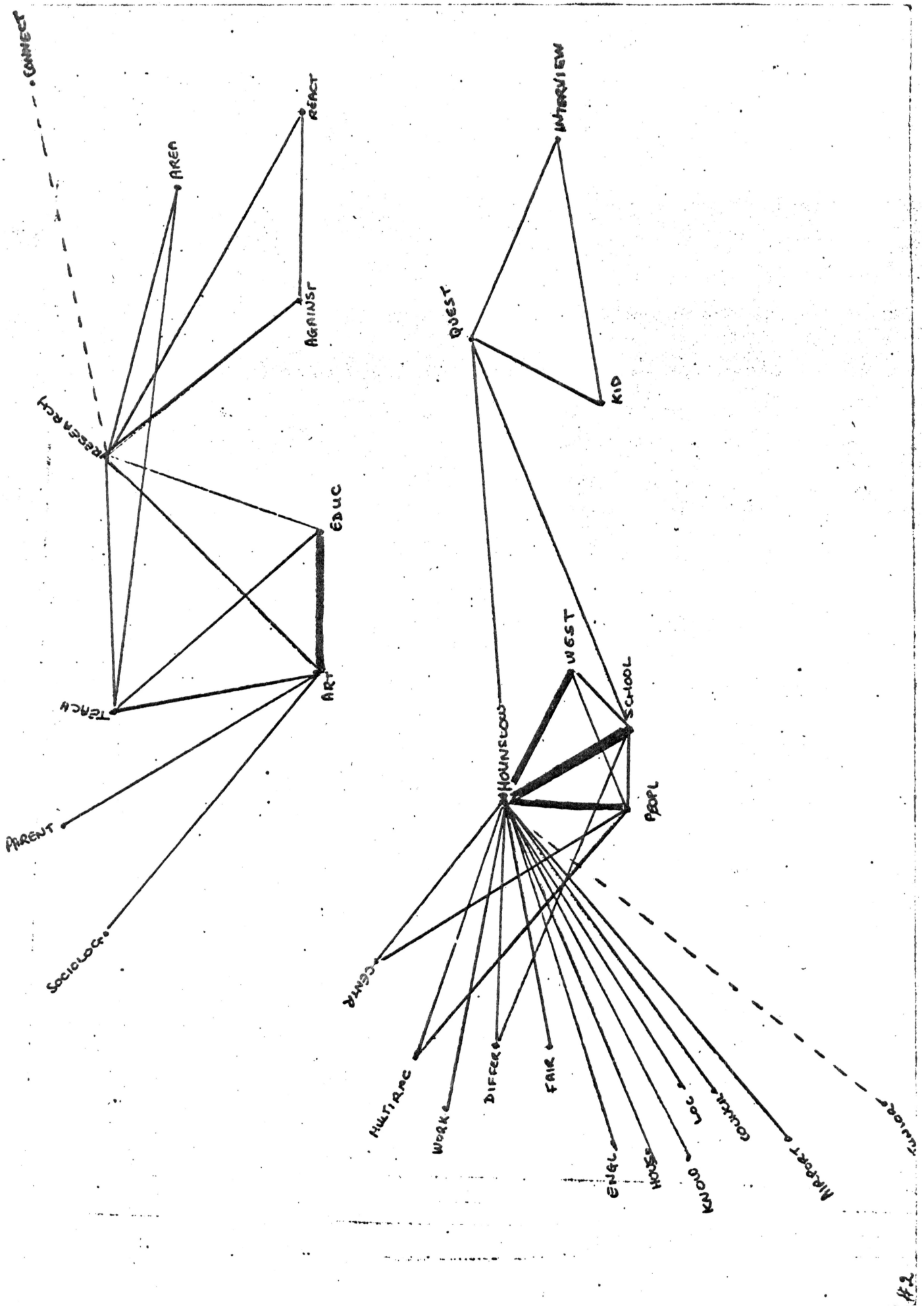
APPENDIX D.

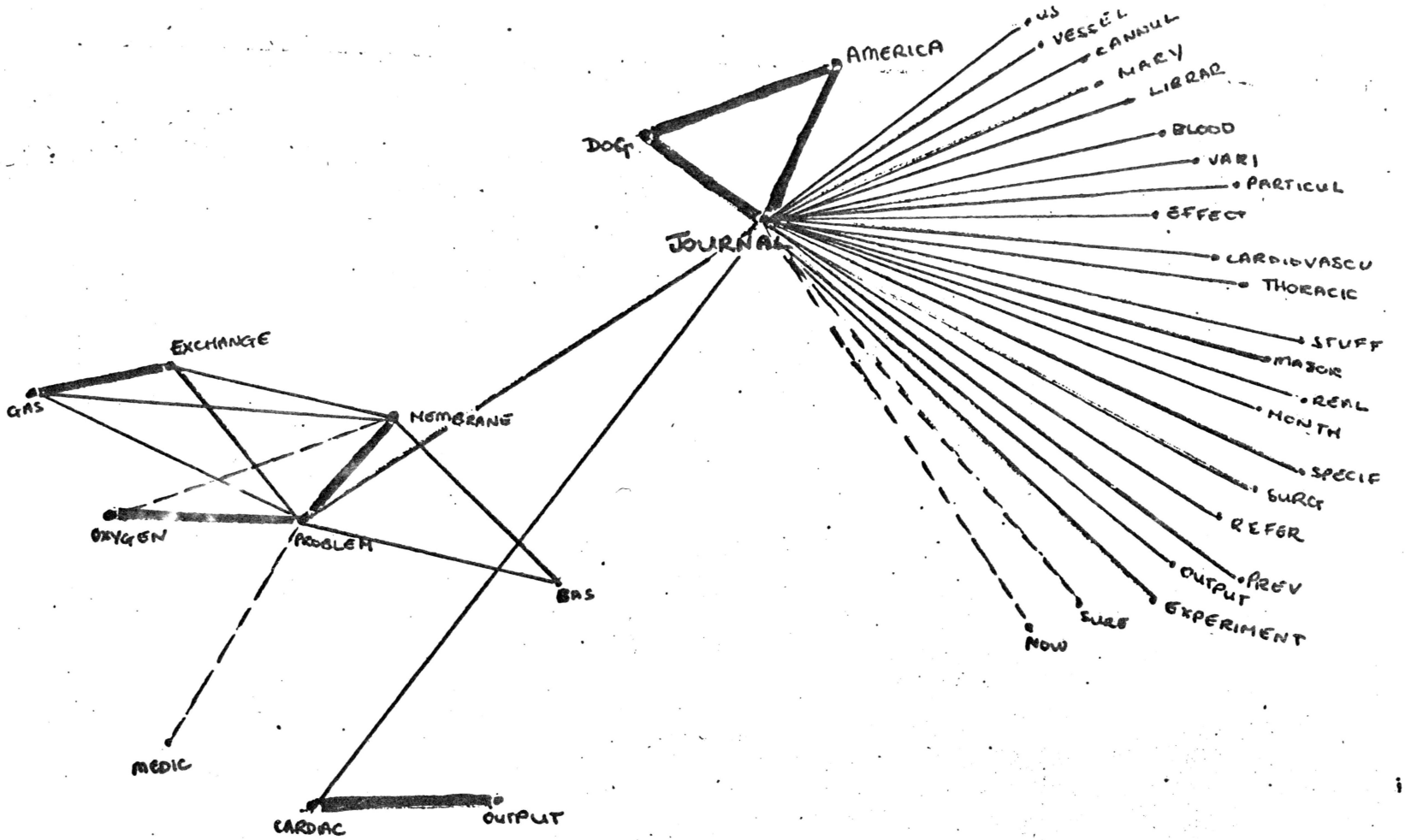
Association-map representations of problem statements.

1-27 Oral statements

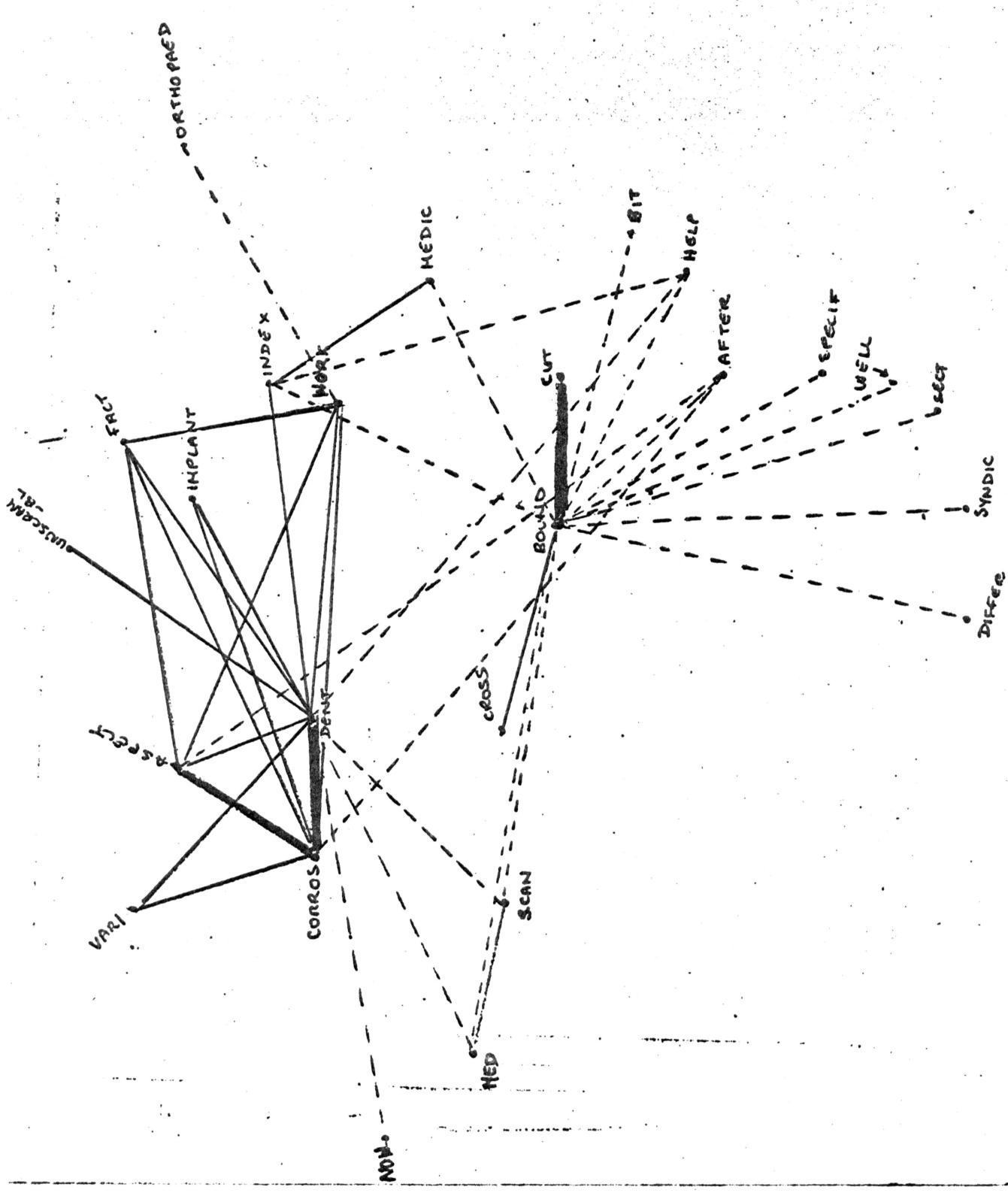
30-37 Written statements

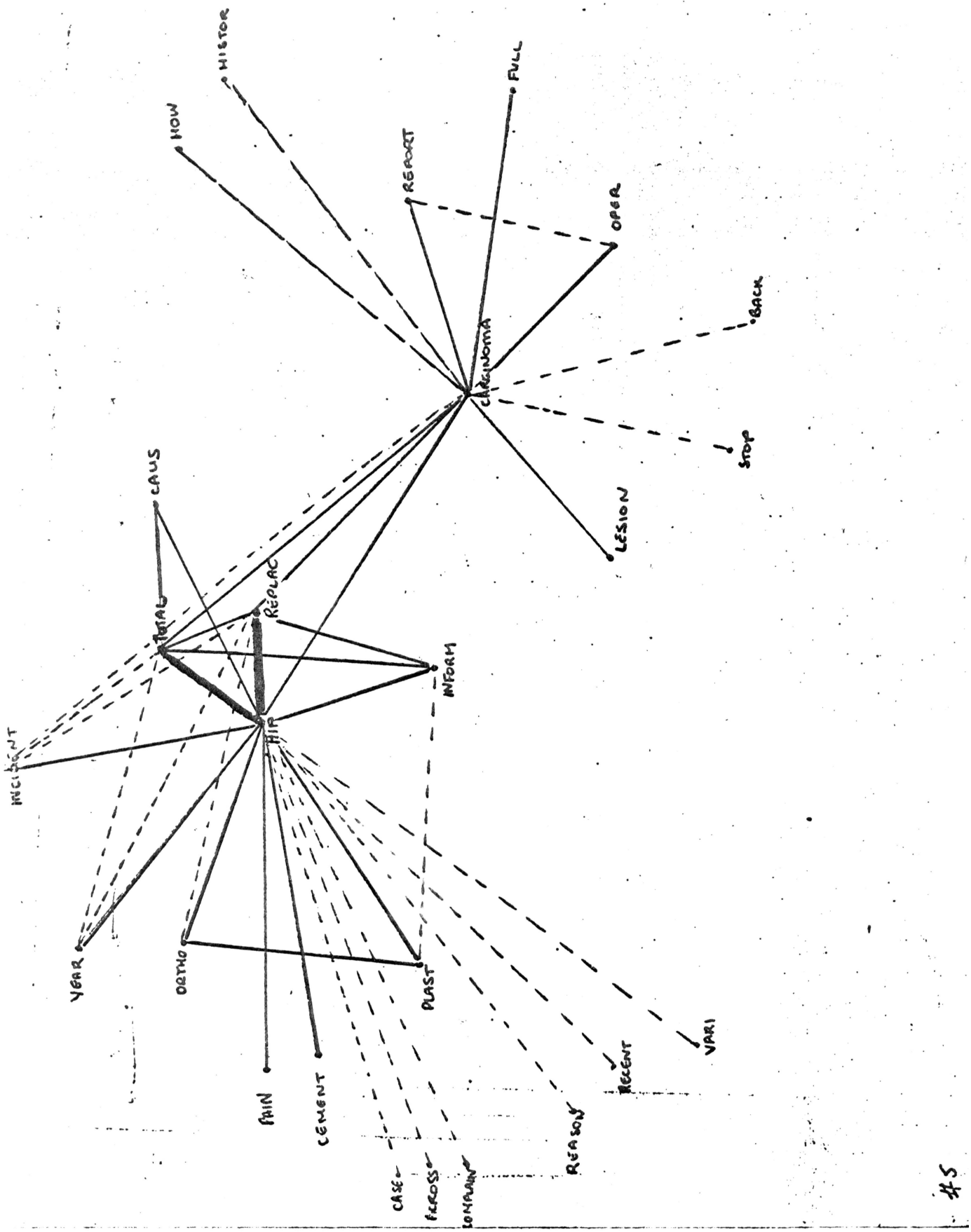

WORK
EFFECT
HIGH
US
ADVERSE
RIVERS
FACT
CHILD
SEVERE
MEDIAT
VERBAL
LEARN
THEOR
IMPORTANT

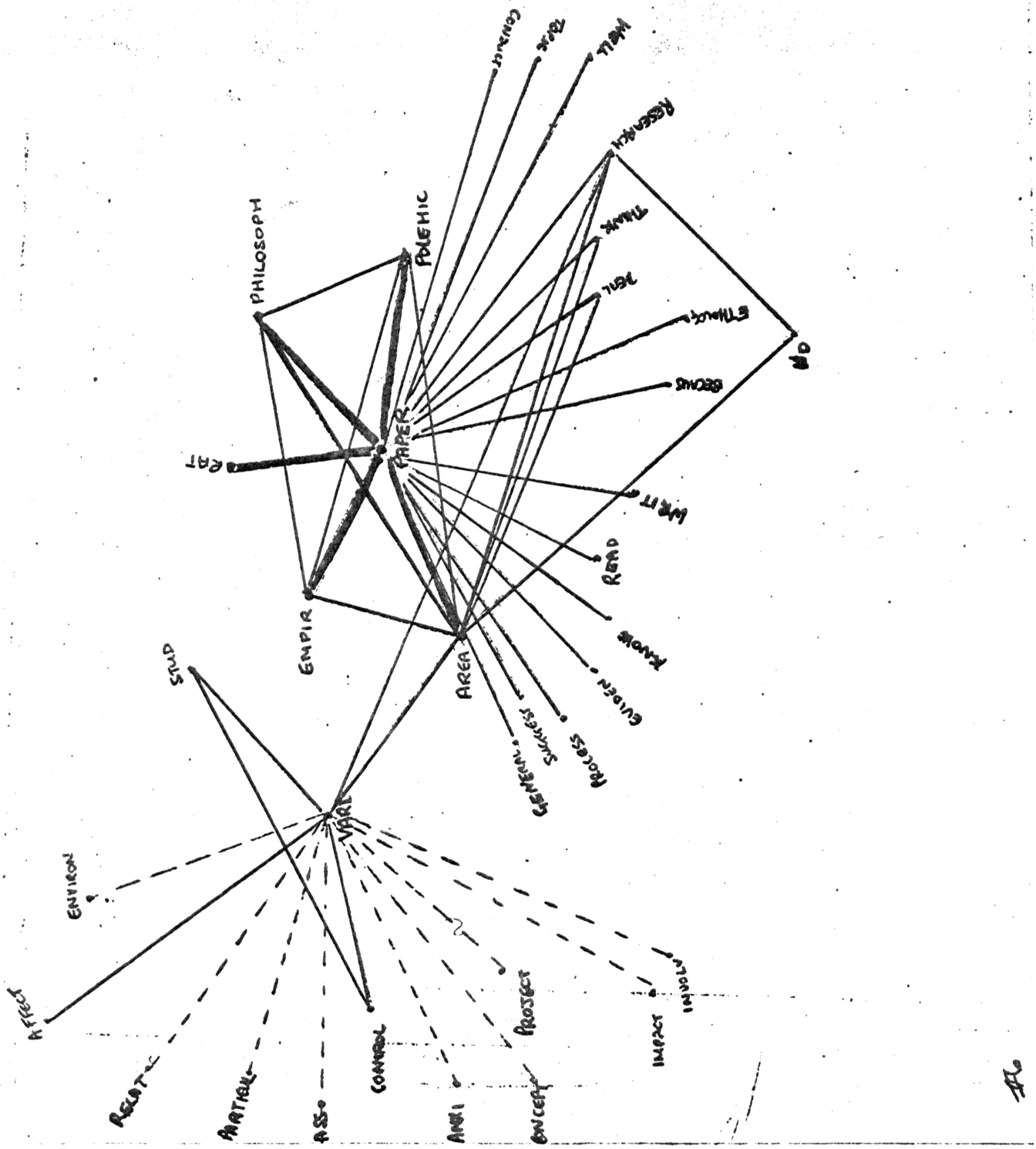

CONNECT
REACT
AREA
AGAINST
RESEARCH
EDUC
ART
TEACH
PARENT
SOCIOLOGY
INTERVIEW
QUEST
KID
WEST
SCHOOL
HOUNSLOW
PEOPL
RANGE
MULTIRAC
WORK
DIFFER
FAIR
ENGL
HOUSE
KNOWLEDG
LOC
COMMUNITY
ALBERT
#2

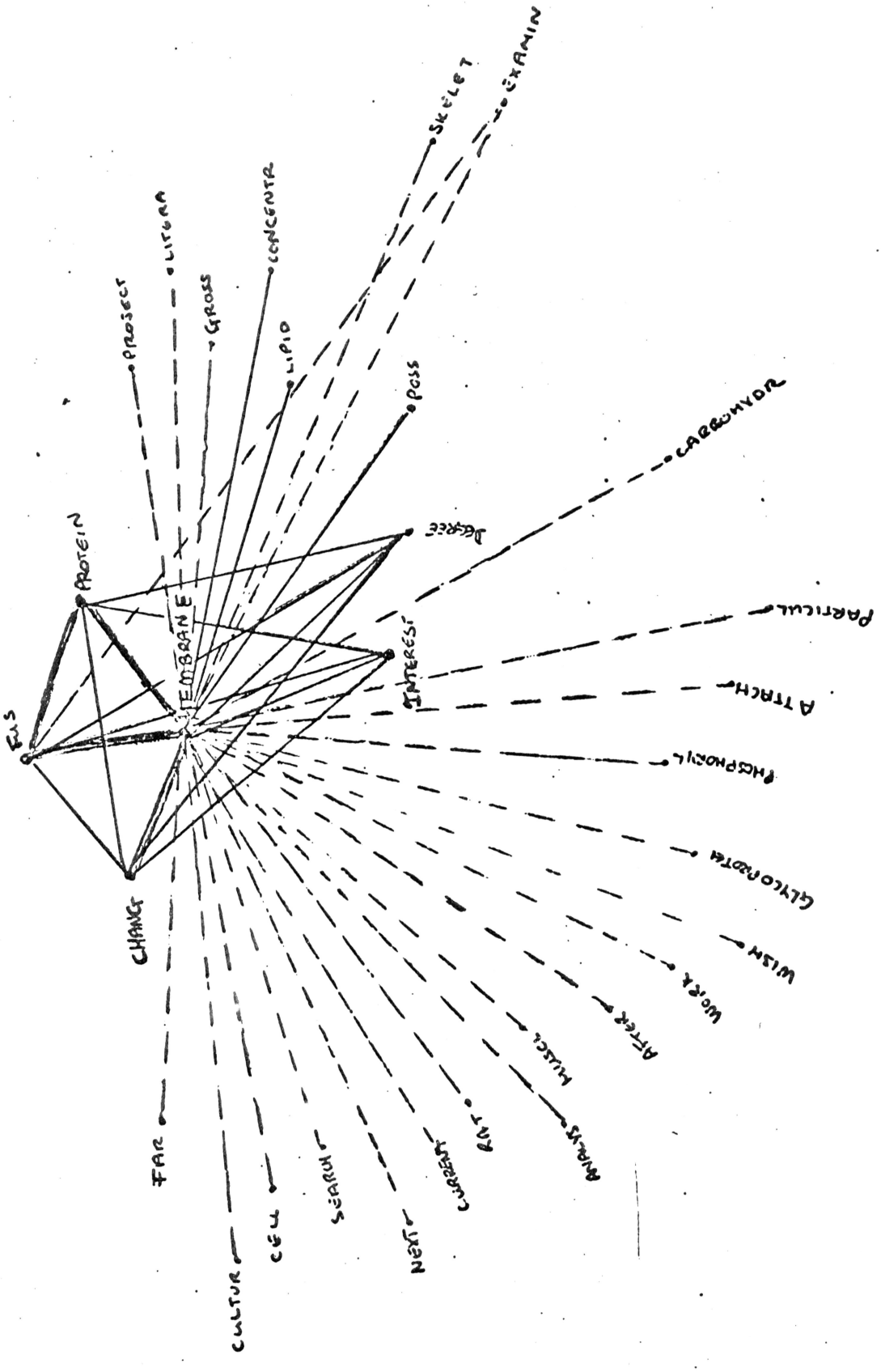

AMERICA
DOG
JOURNAL
EXCHANGE
GAS
MEMBRANE
OXYGEN
PROBLEM
MEDIC
CARDIAC
OUTPUT
BAS
US
VESSEL
CANNUL
MARY
LIBRAR
BLOOD
VARI
PARTICUL
EFFECT
CARDIOVASCU
THORACIC
STUFF
MAJOR
REAL
MONTH
SPECIF
SURG
REFER
PREV
OUTPUT
EXPERIMENT
SURG
NOW

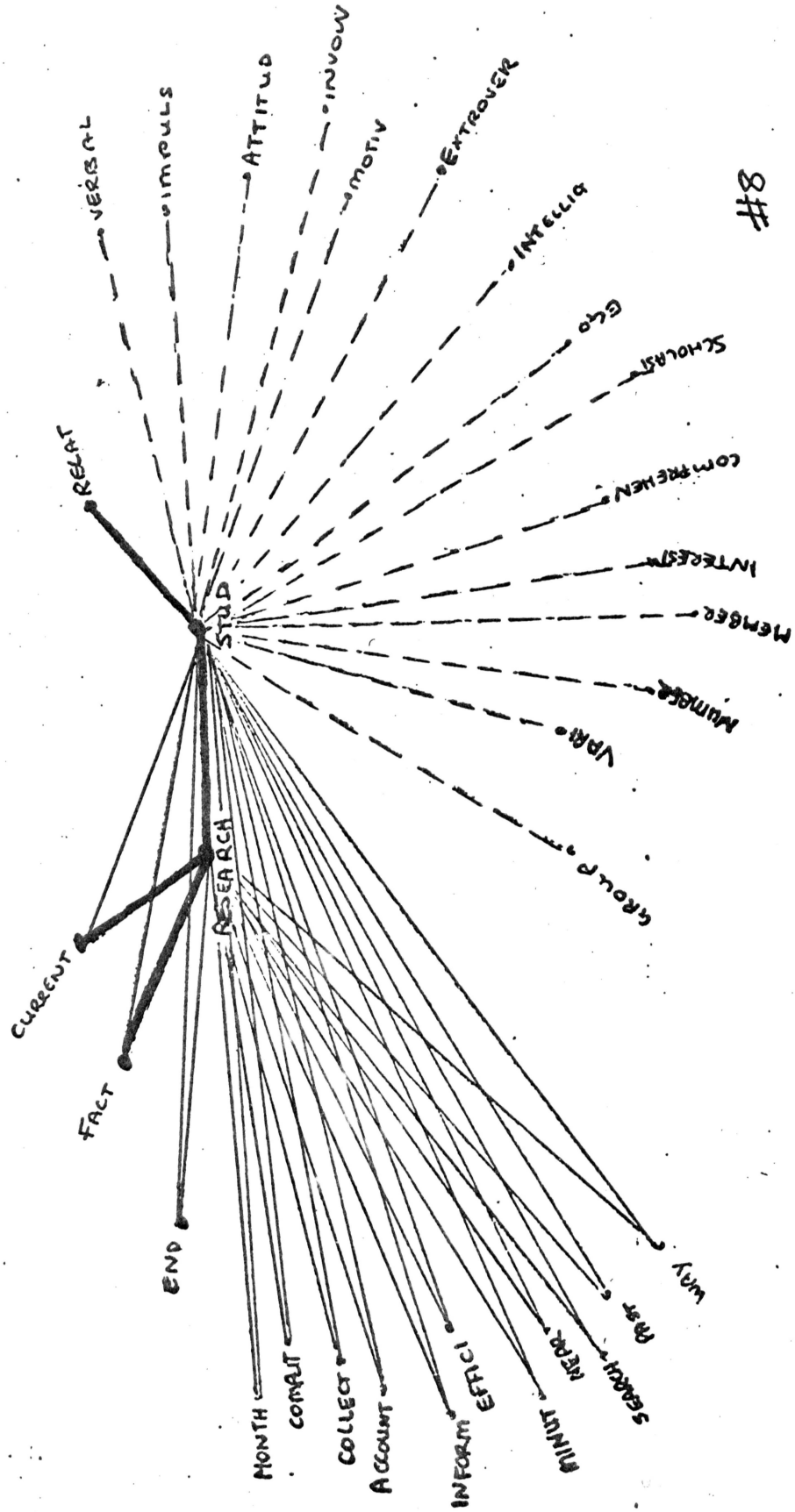
#3


ORTHOPAED
MEDIC
INDEX
WORK
FACT
INPLANT
MUSCARAM
-BL
RESPECT
VARI
CORROS
DENT
SCAN
NED
NDW
CROSS
BOUND
CUT
BIT
HELP
AFTER
EFFECTIF
WELL
SYNDIC
DIFFER

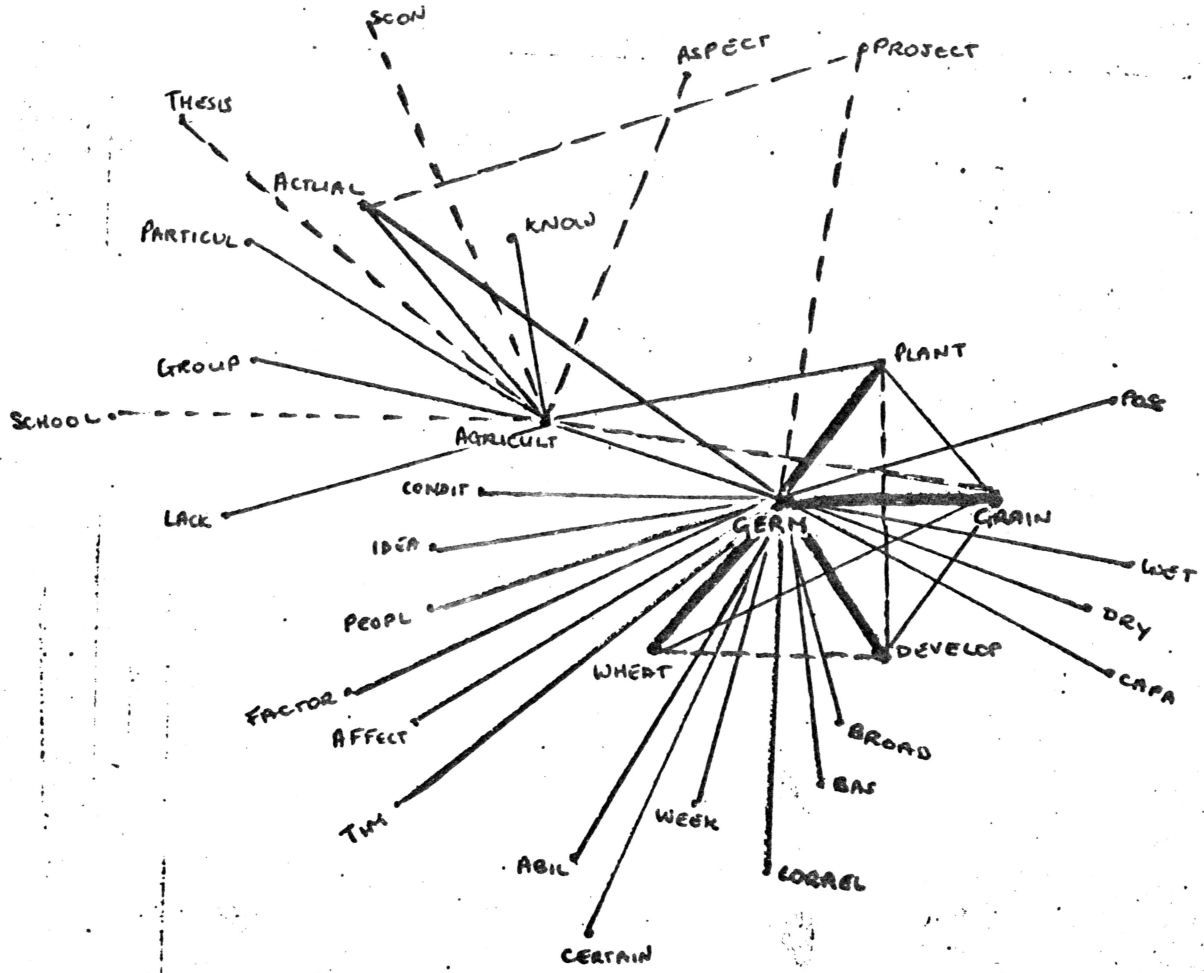

HISTOR
HOW
FULL
REPORT
OPER
BACK
CARCINOMA
STOP
LESION
CAUS
REPLAC
TOTAL
INFORM
HIP
INCIDENT
YEAR
ORTHO
PLAST
THIN
CEMENT
CASE
FEMOSS
REASON
RECENT
VARI


PHILOSOPH
POLEMIC
PAPER
EMPIR
AREA
STUD
VARI
EAT
WRIT
READ
ENVIRON
AFFECT
CONTROL
PROTECT
IMPACT


FAR
CULTUR
CELL
SERUM
NEXT
CURRENT
ANY
MUSCL
AFTER
WORK
WISH
GLYCOPROTE
PHOSPHOLIP
ATTACH
PARTICUL
CARBOHYDR
EXAMIN
SKELET
CONCENTR
LIPID
GROSS
LITERA
PROJECT
PROTEIN
FLS
CHANG
MEMBRANE
INTERESI
DEGREE
POSS


VERBAL
IMPULS
ATTITUD
INVOW
MOTIV
EXTROVER
INTELLIG
EGO
SCHOLAST
COMPREHEN
INTEREST
MEMBER
NUMBER
VPA
GROUP
RELAT
STUD
RESEARCH
CURRENT
FACT
END
MONTH
COMMIT
COLLECT
ACCOUNT
INFORM
EFFICI
MINUT
REPOR
SEARCH
MAST
WAY

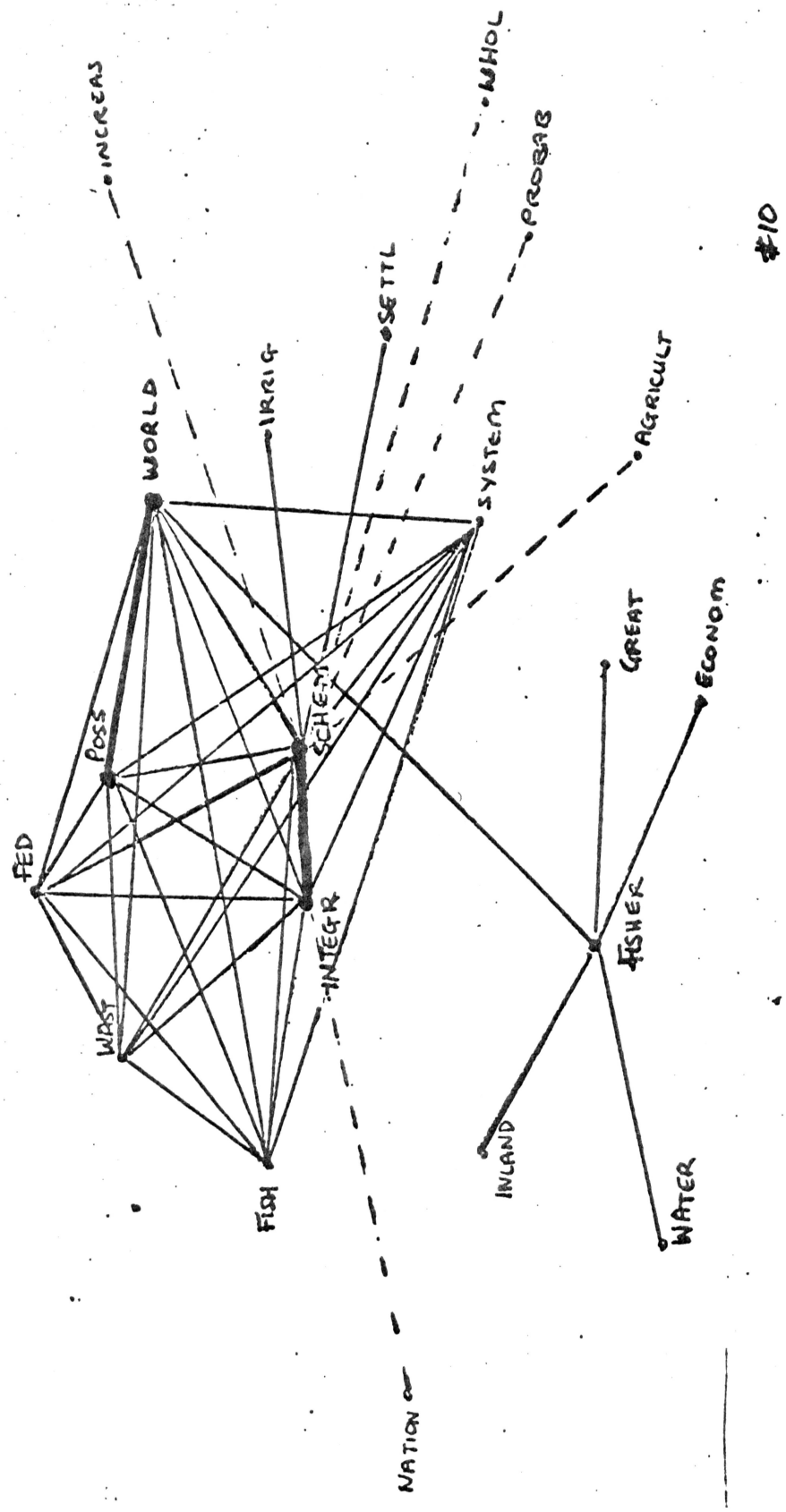
88


SCON
ASPECT
PROJECT
THESIS
ACTUAL
KNOW
PARTICUL
GROUP
SCHOOL
AGRICULT
PLANT
LACK
CONDIT
GERM
GRAIN
IDEA
PEOPL
DEVELOP
WHEAT
FACTOR
AFFECT
BROAD
BAS
WEEK
ABIL
LOGREL
CERTAIN

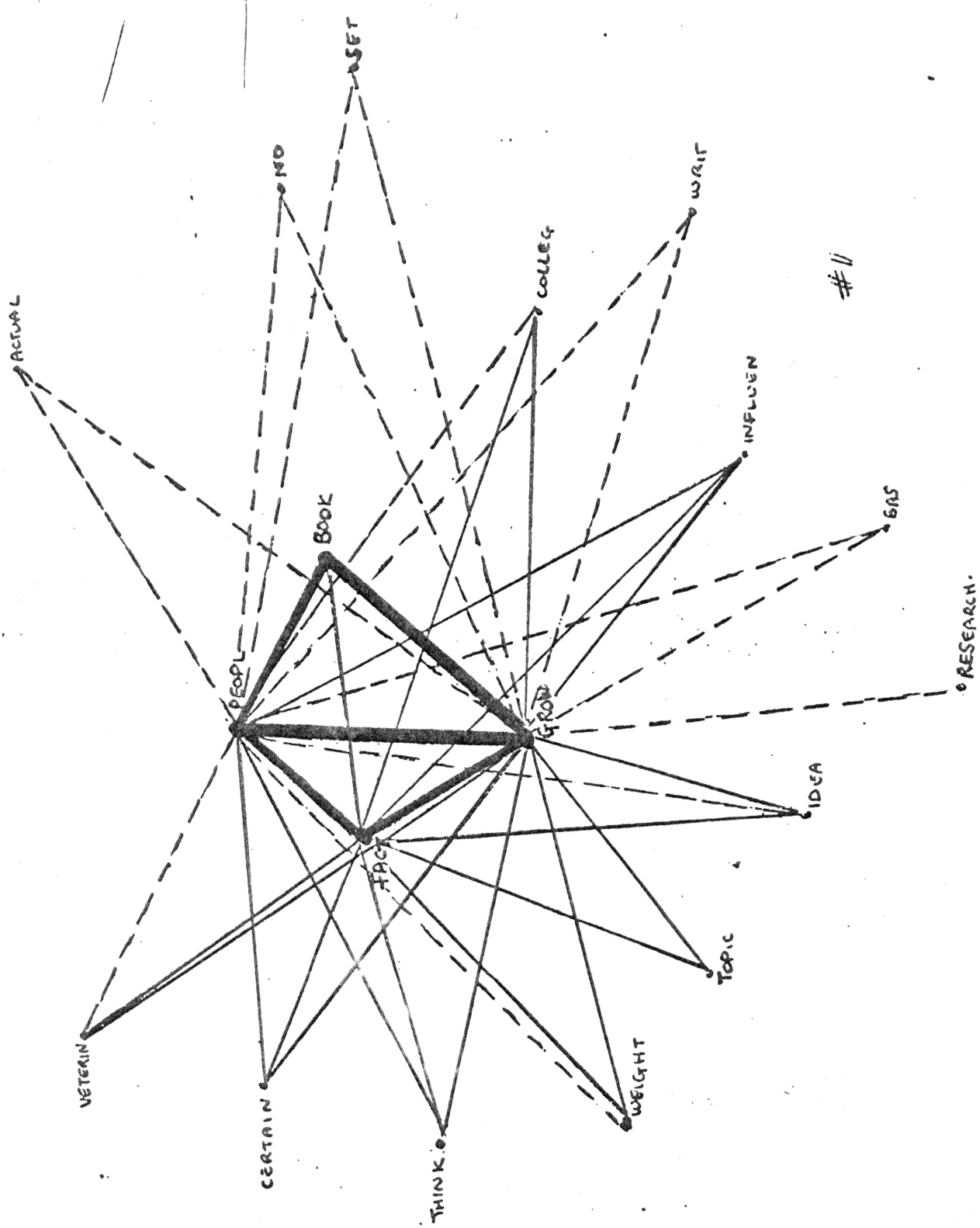
F

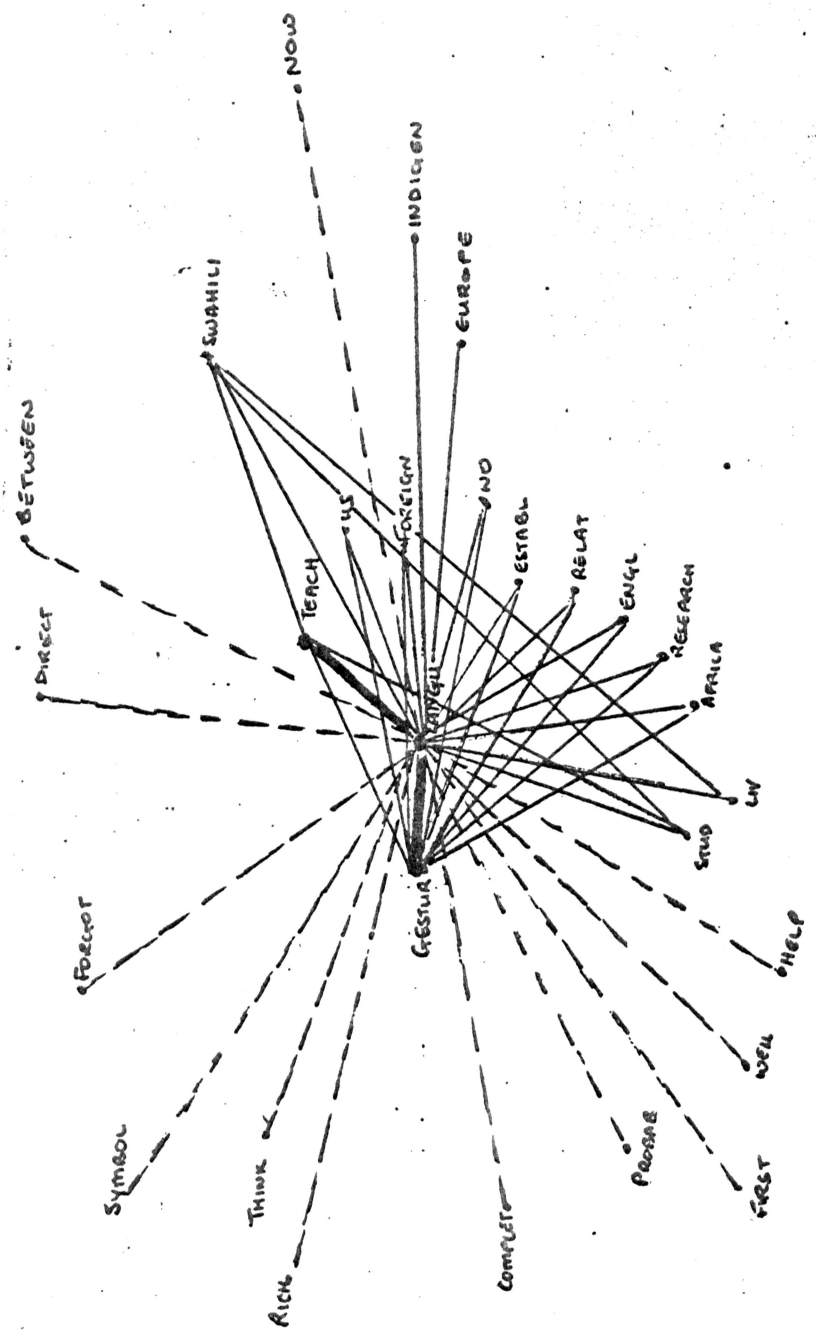
-96-

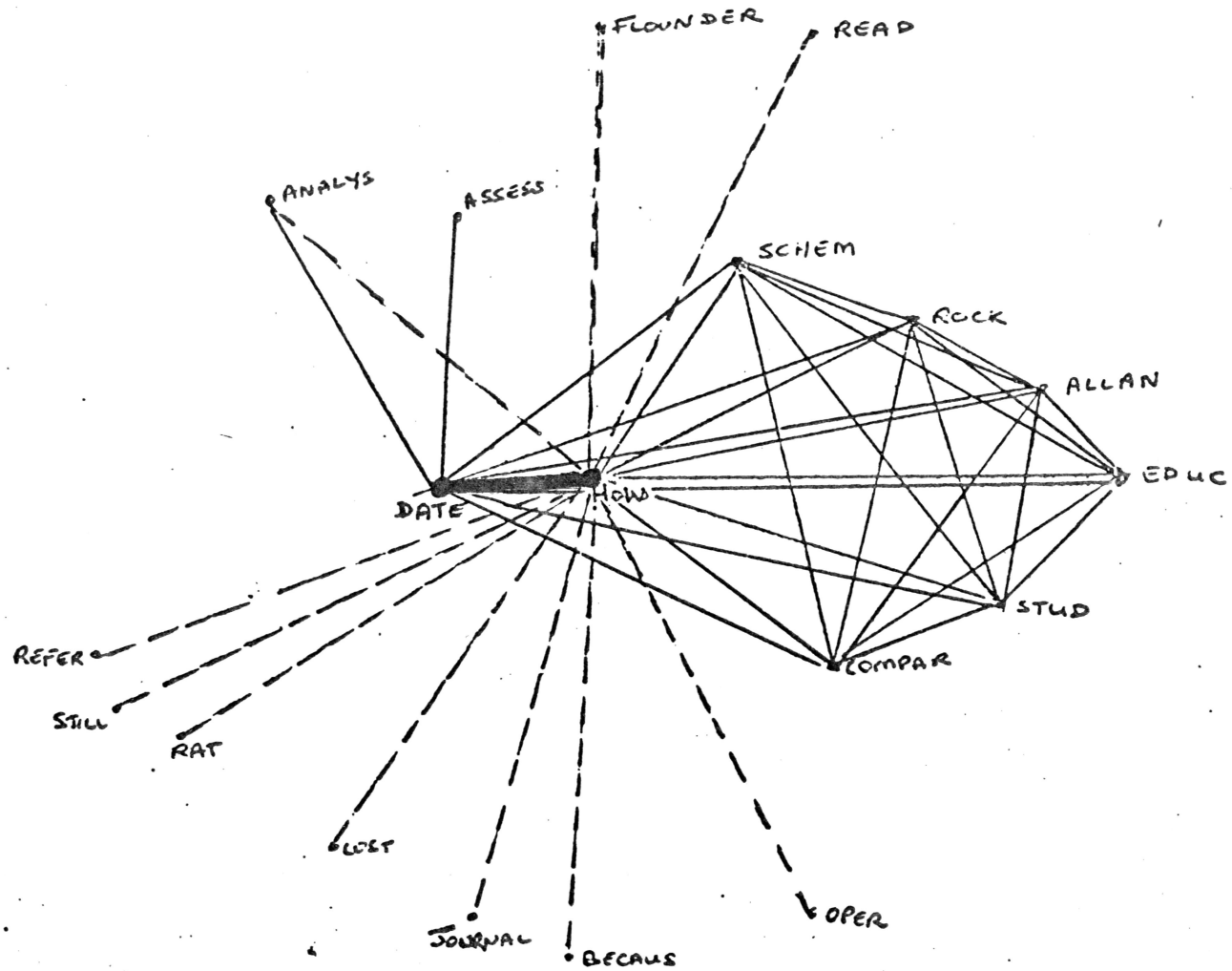
#9.


INCREAS
WHOL
PROGRA
SETTL
IRRIG
WORLD
SYSTEM
AGRICULT
GREAT
ECONOM
POSS
SCHEM
FED
INTEGR
FISHER
WAST
FISH
INLAND
WATER
NATION

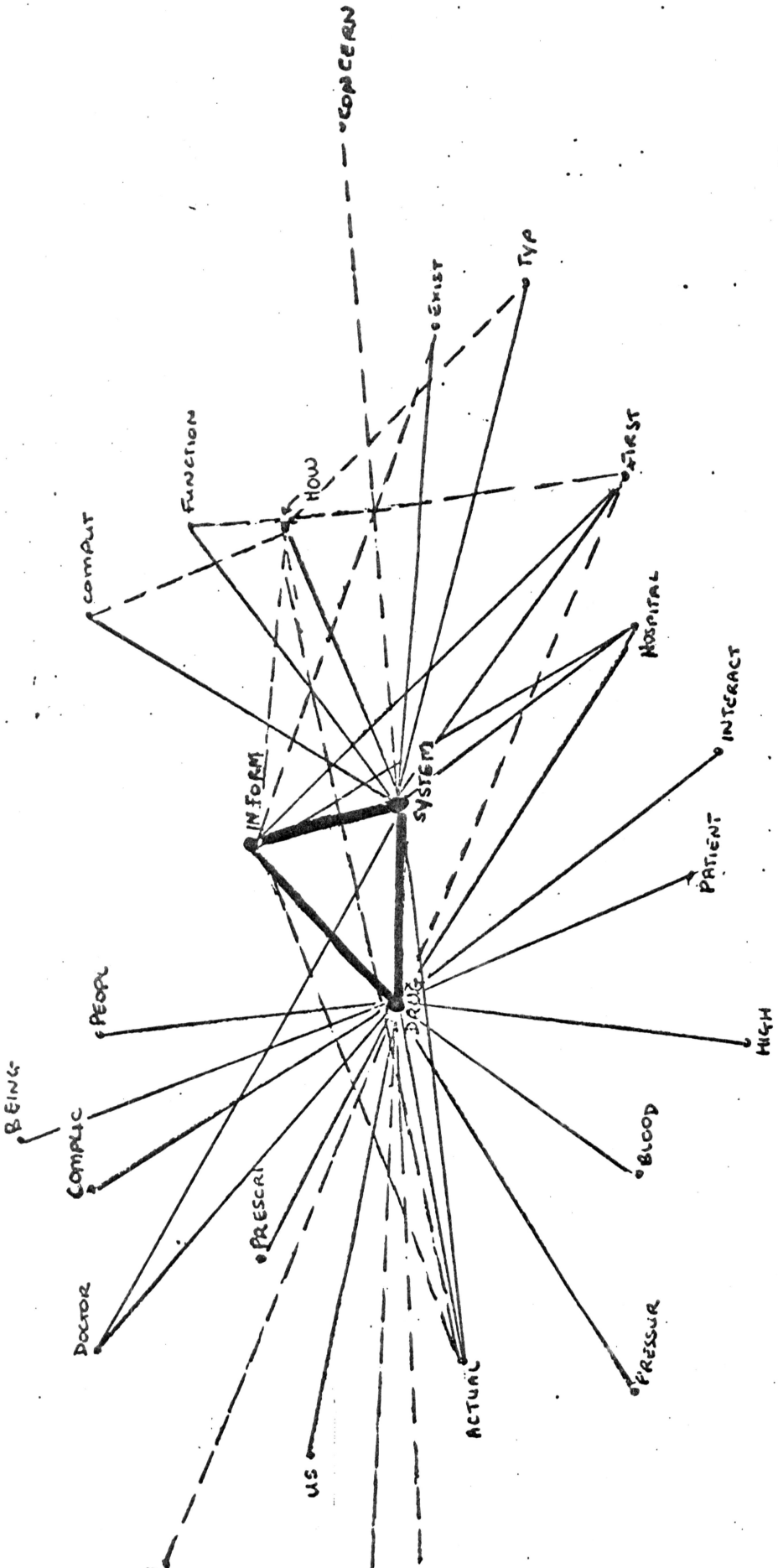
#10

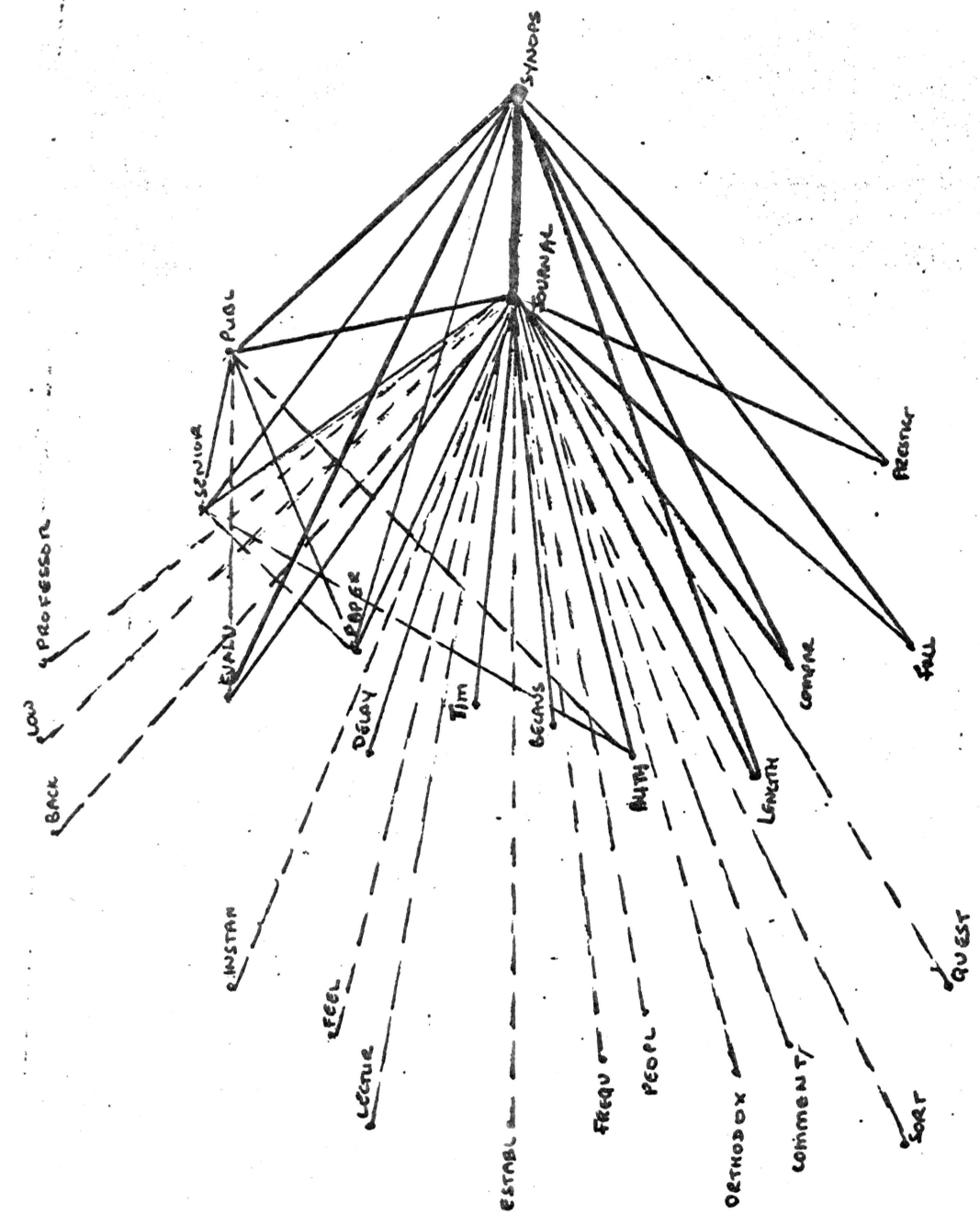

ASET
NO
WRIT
COLLEG
ACTUAL
#11
INFLUEN
BOOK
GAS
RESEARCH.
PEOPL
GROW
IDEA
FACT
TOPIC
VETERIN
CERTAIN
THINK.
WEIGHT

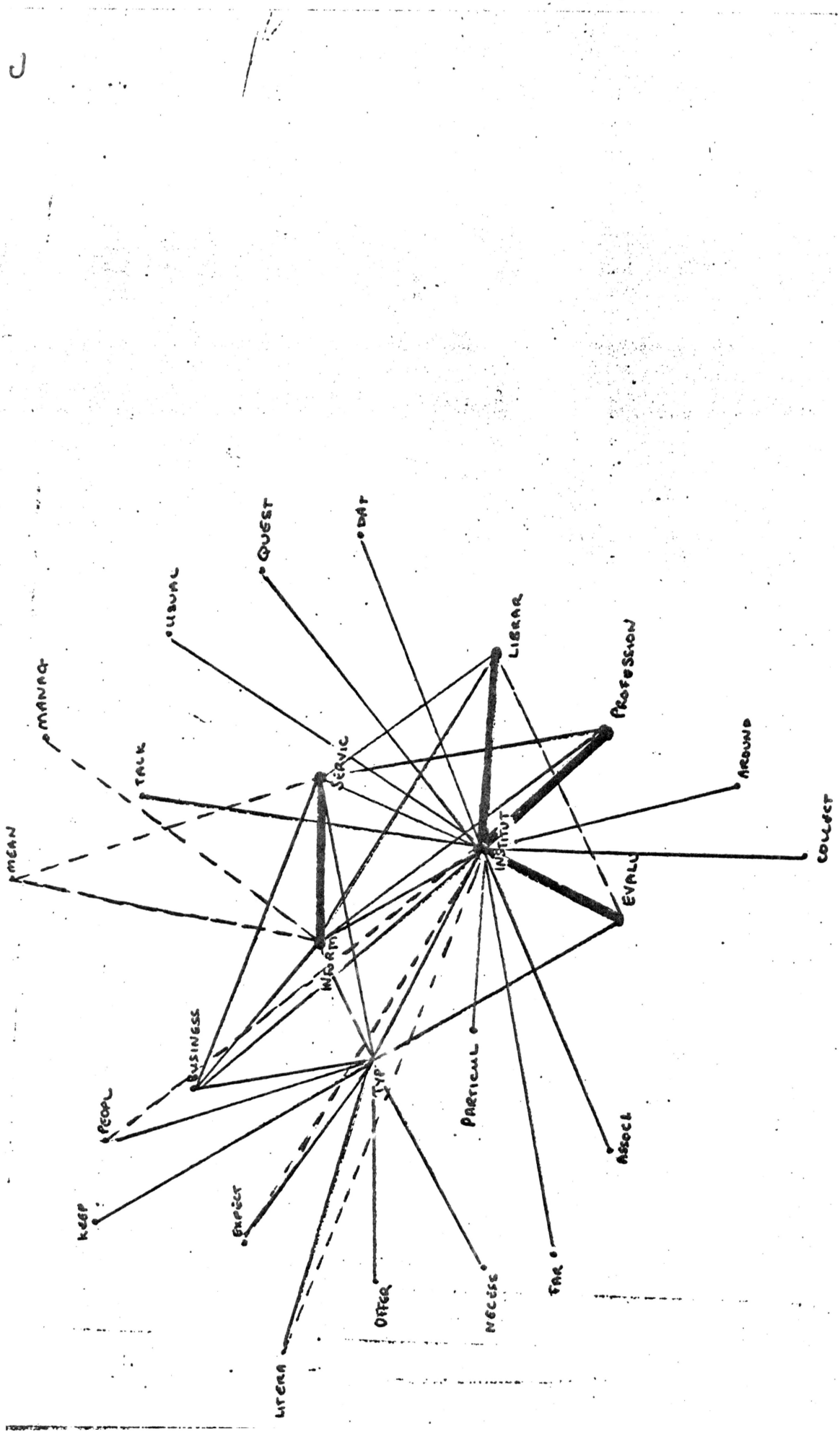

NOUN
INDIGEN
EUROPE
SWAHILI
BETWEEN
DIRECT
TEACH
FOREIGN
RELAT
ESTABL
ENGL
RESEARCH
AFRICA
UN
STUD
HELP
USIL
FIRST
PROGRAM
COMPUTER
RICH
THINK
SYMBOL
FORCAST


FLOUNDER
READ
ANALYS
ASSESS
SCHEM
ROCK
ALLAN
EDUC
HOW
DATE
STUD
COMPAR
REFER
STILL
RAT
LOST
JOURNAL
BECAUS
OPER

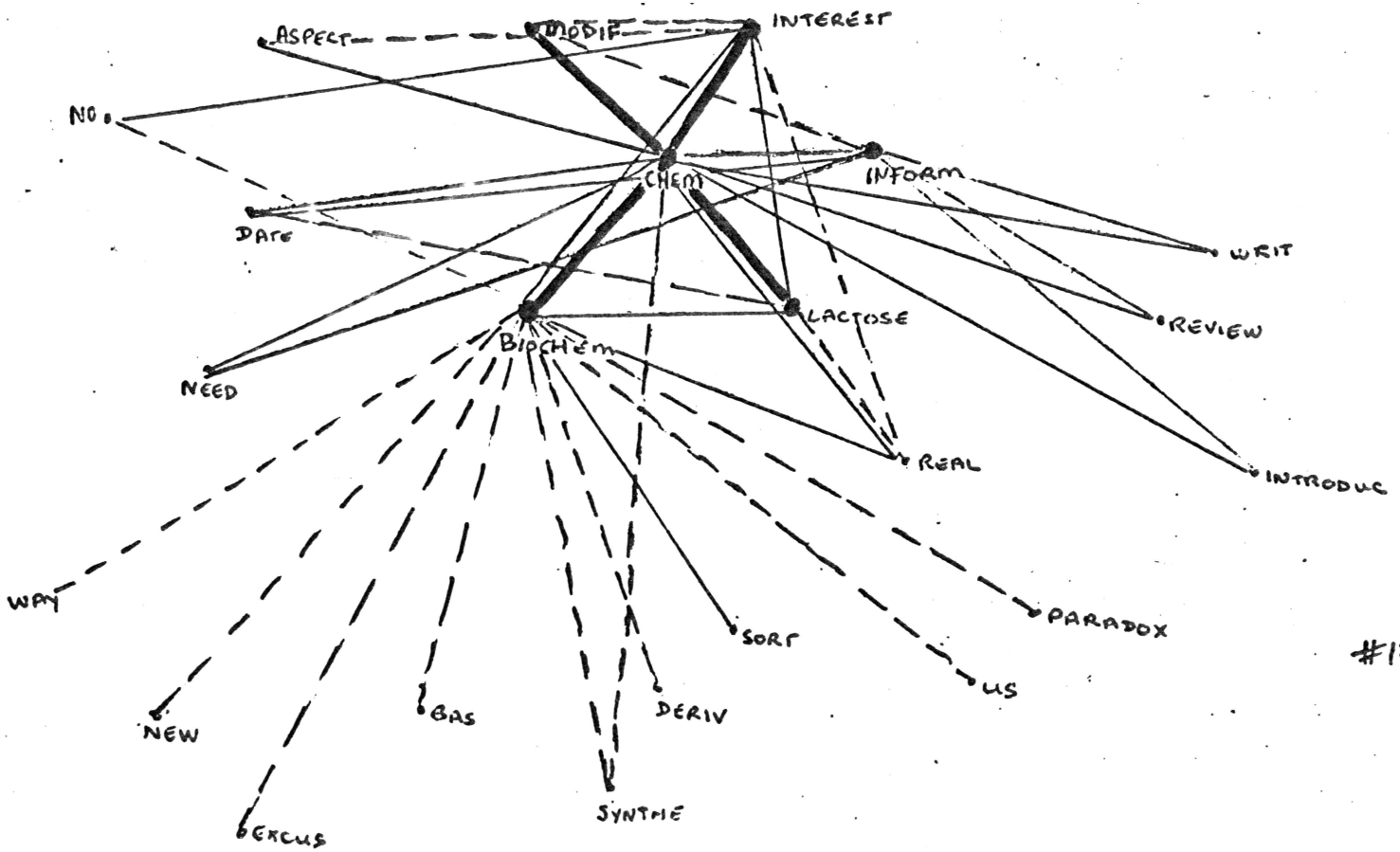
#13


#14
TOPICERN
EXIST
TYP
FUNCTION
HOW
FIRST
COMPLT
HOSPITAL
INTERACT
INFORM
SYSTEM
PATIENT
PEOPL
DRUG
HIGH
BEING
COMPLIC
BLOOD
PRESCRI
DOCTOR
PRESSUR
ACTUAL
US


SYNOPS
JOURNAL
PUBL
SESSION
PROFESSOR
LOU
BACK
EVALU
PAPER
DELAY
TIM
BECAUS
FREQU
BUTH
LENGTH
COMPAR
FALL
FRESHER
QUEST
SORT
COMMENT
ORTHODOX
PEOPL
ESTABL
LEGNE
FEEL
R.INSTAN


QUEST
DAY
USUAL
MANAG
LIBRAR
PROFESSION
TALK
SERVIC
AROUND
MEAN
INSTITUT
COLLECT
EVALU
INFORM
BUSINESS
PEOPL
PARTICUL
TYP
ASSOCI
KEEP
EXPECT
OFFER
NECESS
FAR
LITERA

#16

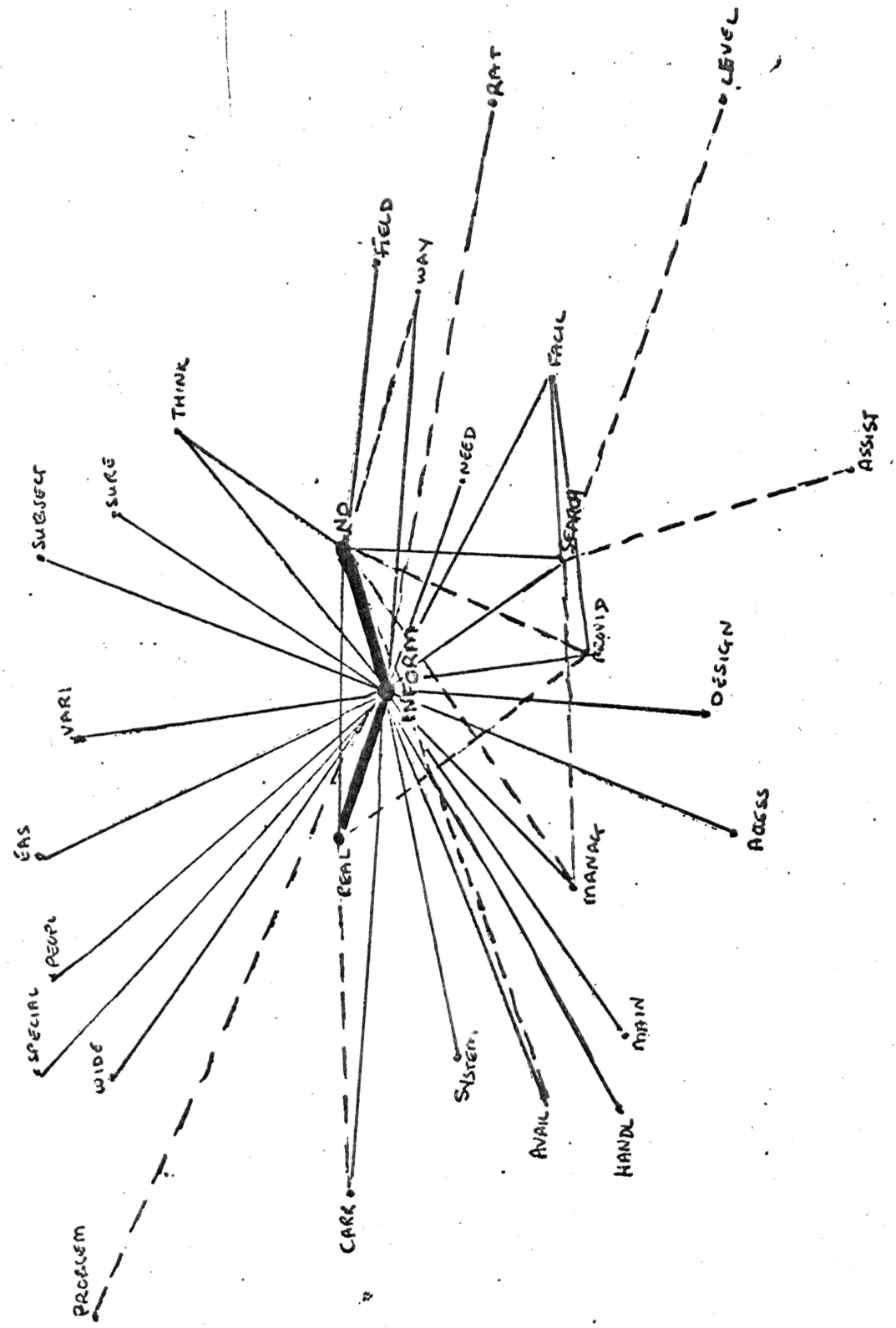

ASPECT
MODIF
INTEREST
NO
CHEM
INFORM
DATE
WRIT
REVIEW
LACTOSE
BIOCHEM
NEED
REAL
INTRODUC
WAY
PARADOX
SORT
US
NEW
BAS
DERIV
SYNTHE
EXCUS

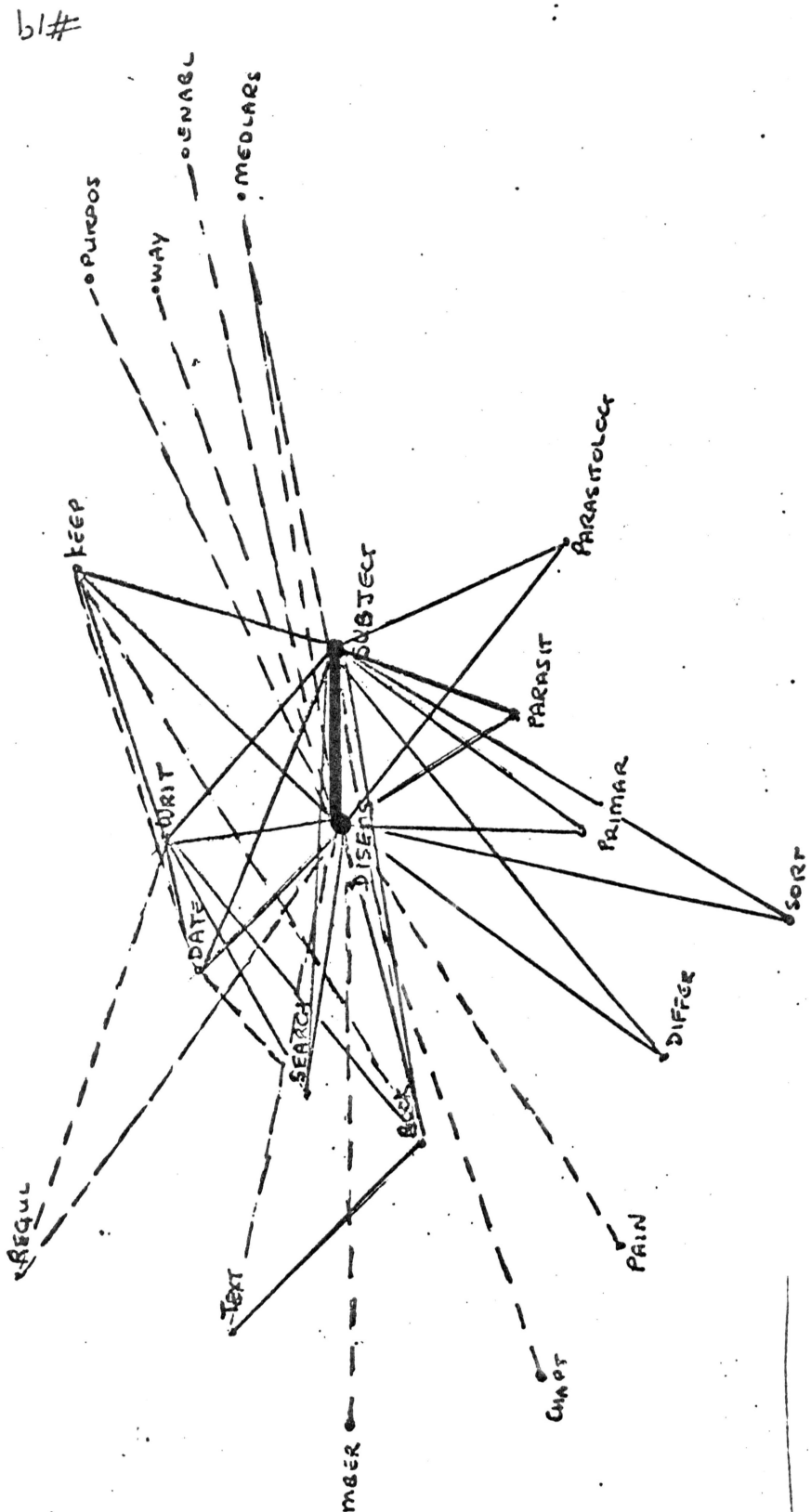
#17

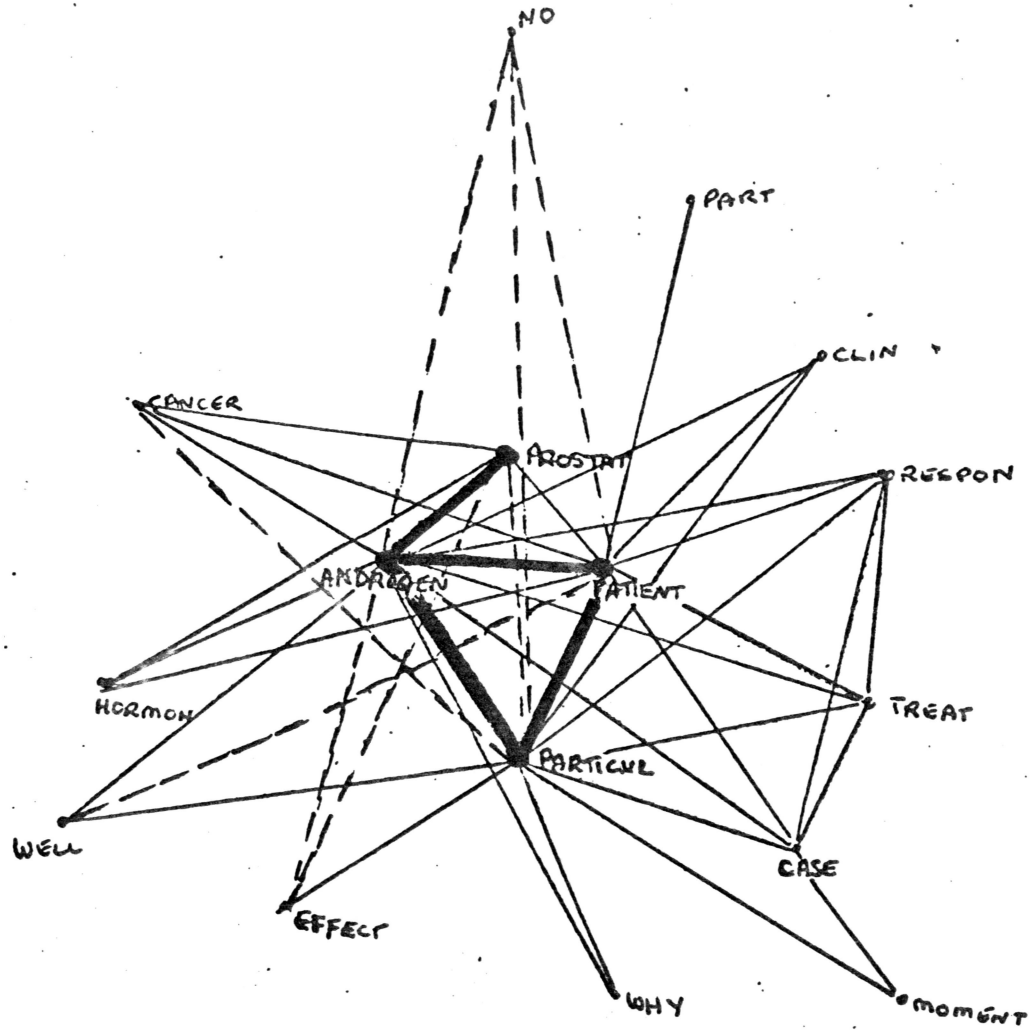
#17

81#

#18


LEVEL
RAT
FIELD
WAY
FACIL
NEED
THINK
ASSIST
SUBJECT
SURE
ND
SEARCH
INFORM
DESIGN
VARI
ACCESS
EAS
MANAG
REAL
PEOPL
SPECIAL
WIDE
SYSTEM
MAIN
AVAIL
HANDL
CARE
PROBLEM


b1#
PURPOS
WAY
ENABL
MEDLARS
KEEP
PARASITOLOGY
SUBJECT
PARASIT
PRIMAR
SORT
DIFFER
PAIN
CHAPT
NUMBER
TEXT
REGUL
BACK
SEARCH
DATE
DISEASE
WRIT

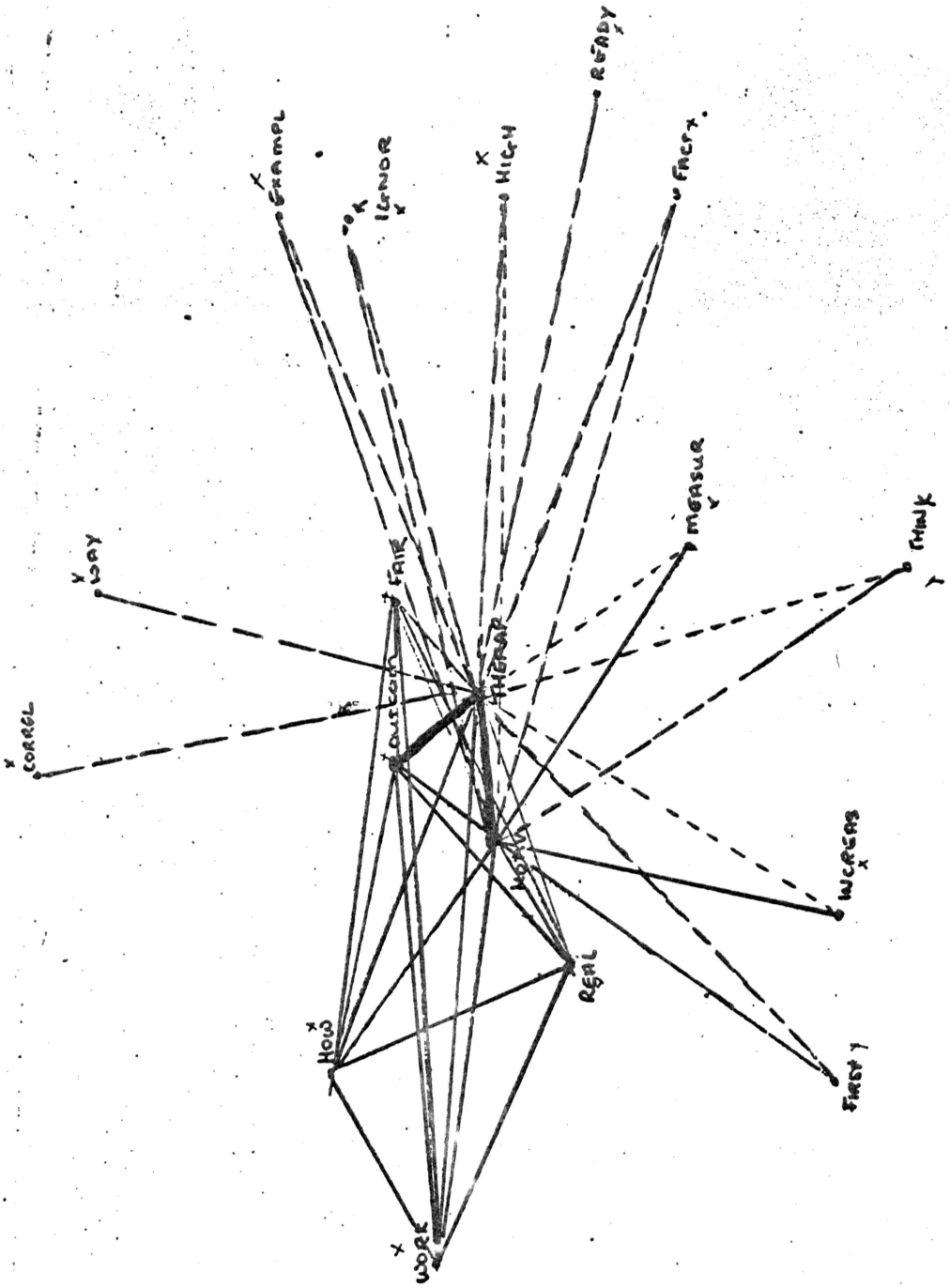

NO
PART
CLIN
CANCER
PROSTAT
RESPON
ANDROGEN
PATIENT
HORMON
TREAT
PARTICUL
WELL
CASE
EFFECT
WHY
MOMENT

#20

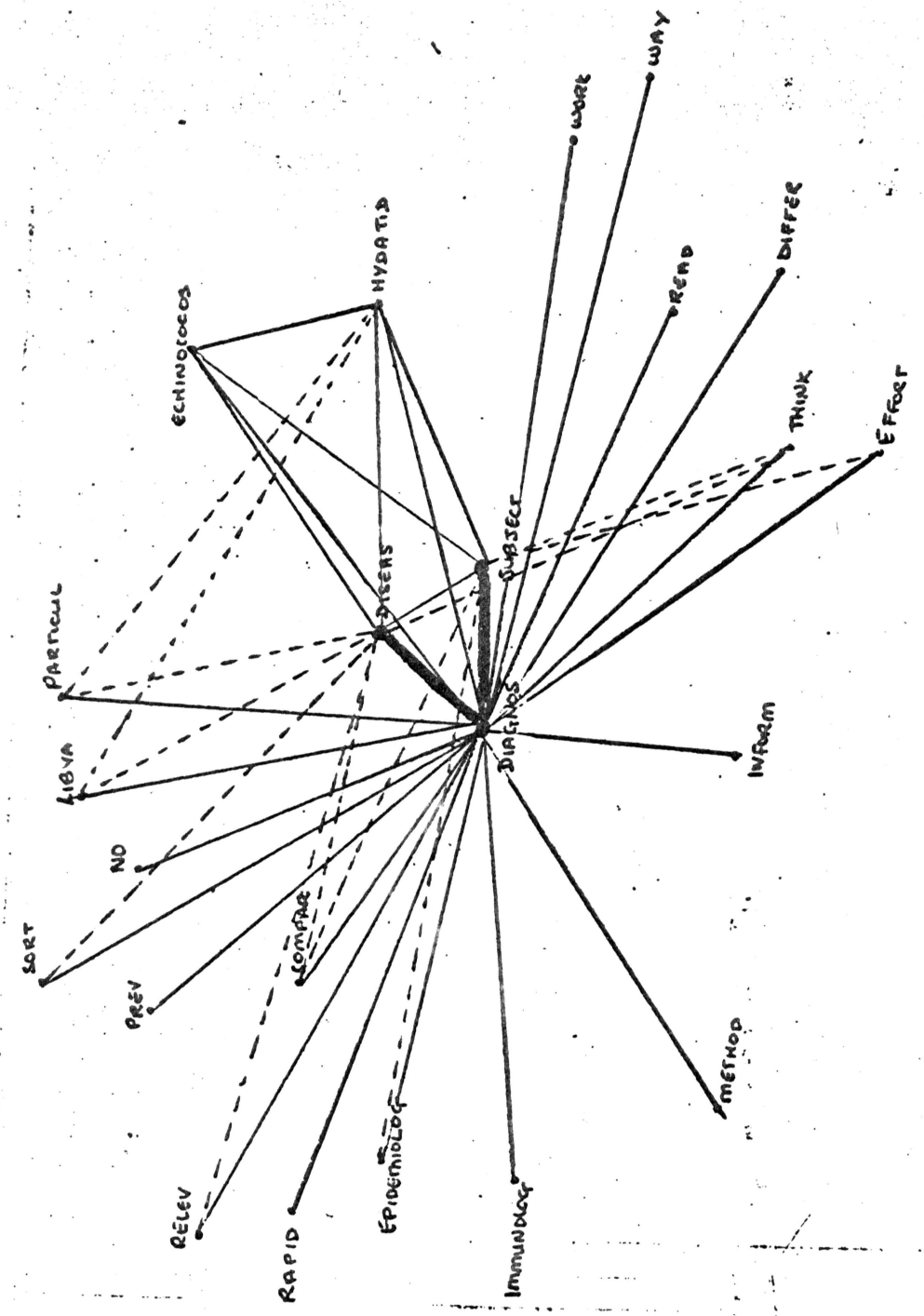
#20

#31

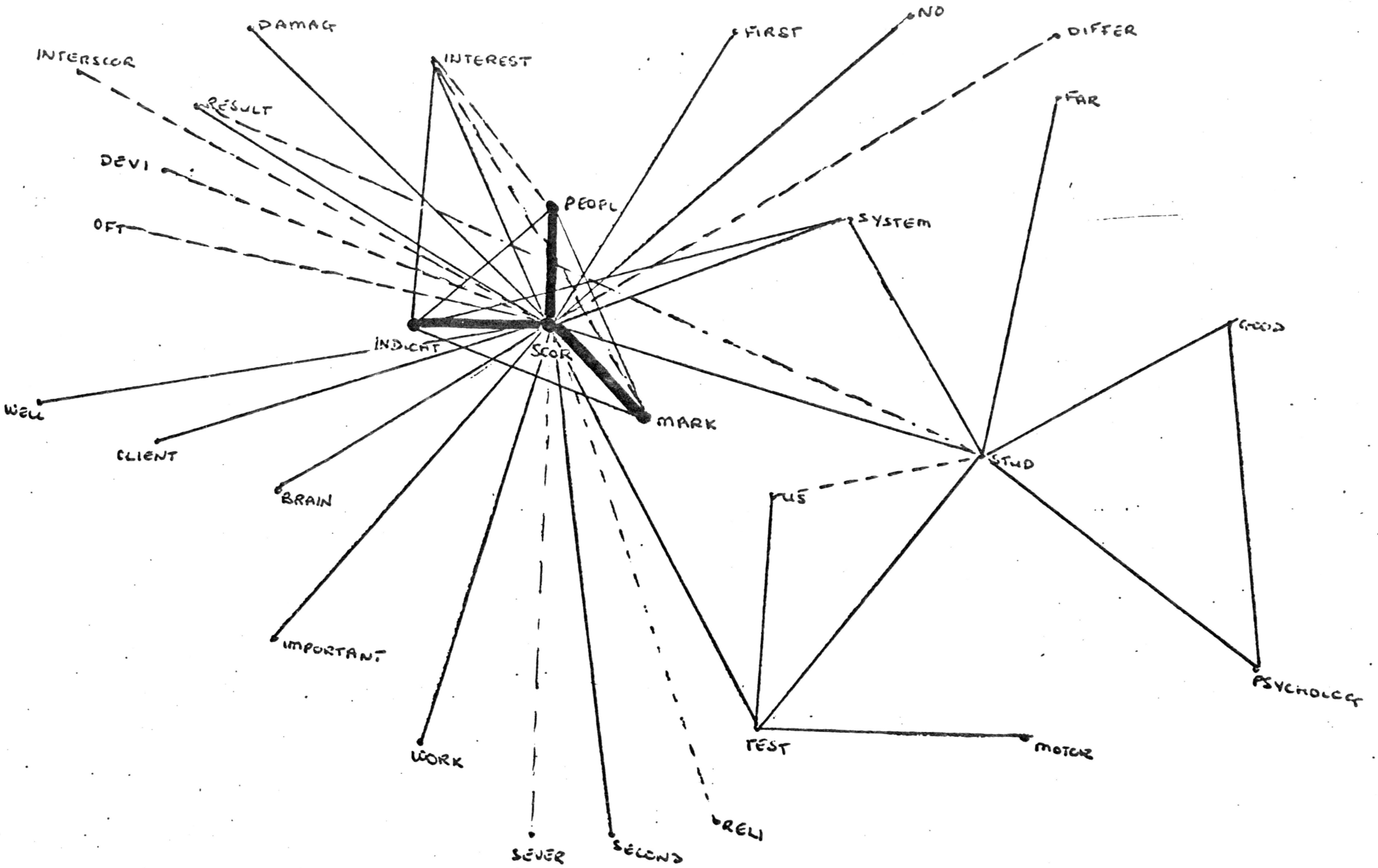
CONCEPT
CONCERN
HEMISPHEREC
PLAST
DAT
RECOV
YOUNG
BRAIN
HUMAN
REQUIR

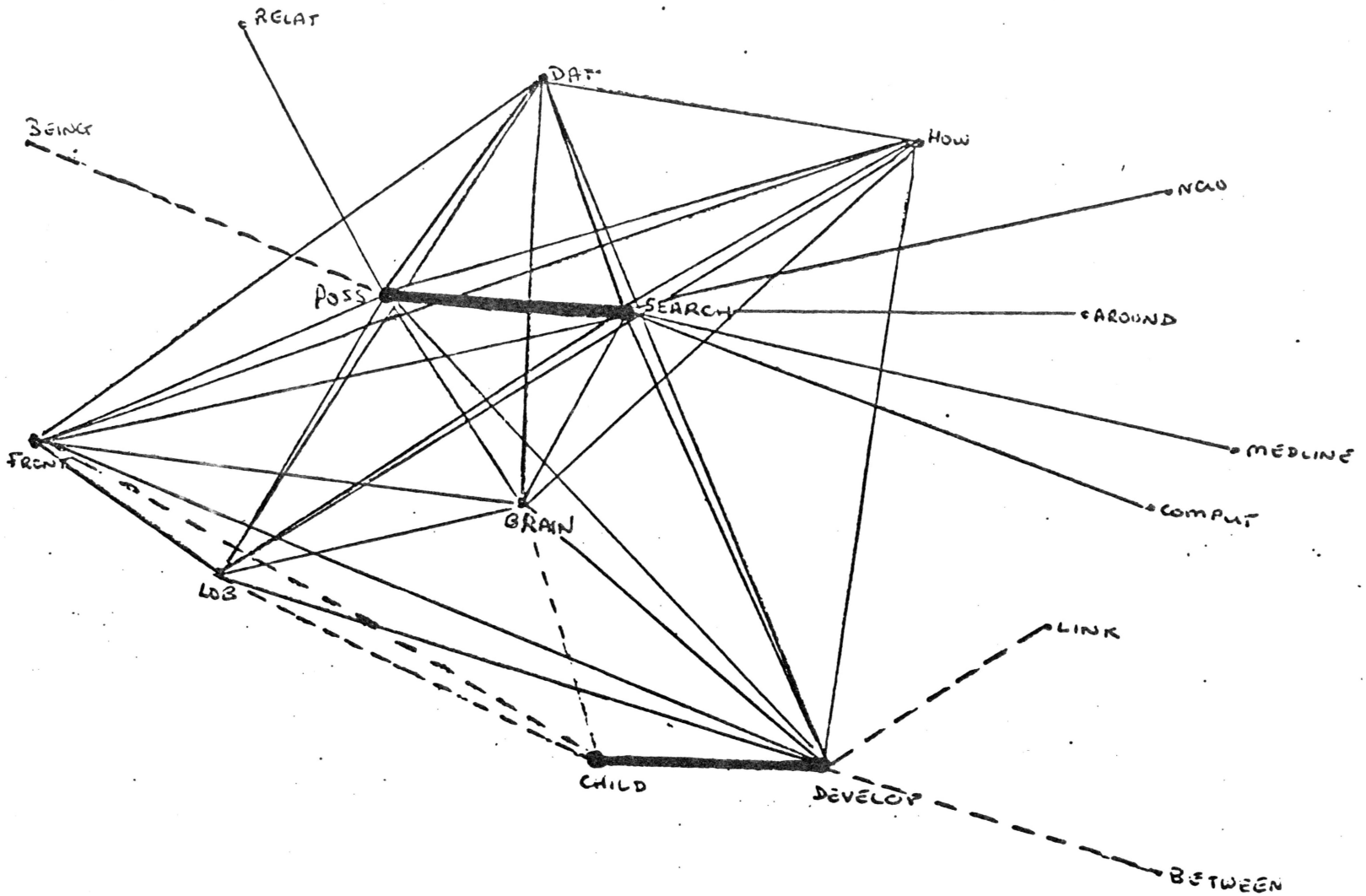

EXAMPL
ILENOR
HIGH
READY
FACTX
MEASUR
THINK
WAY
FAIR
CORREL
CUSTOM
HOW
REAL
WORK
WCASES
FIRST

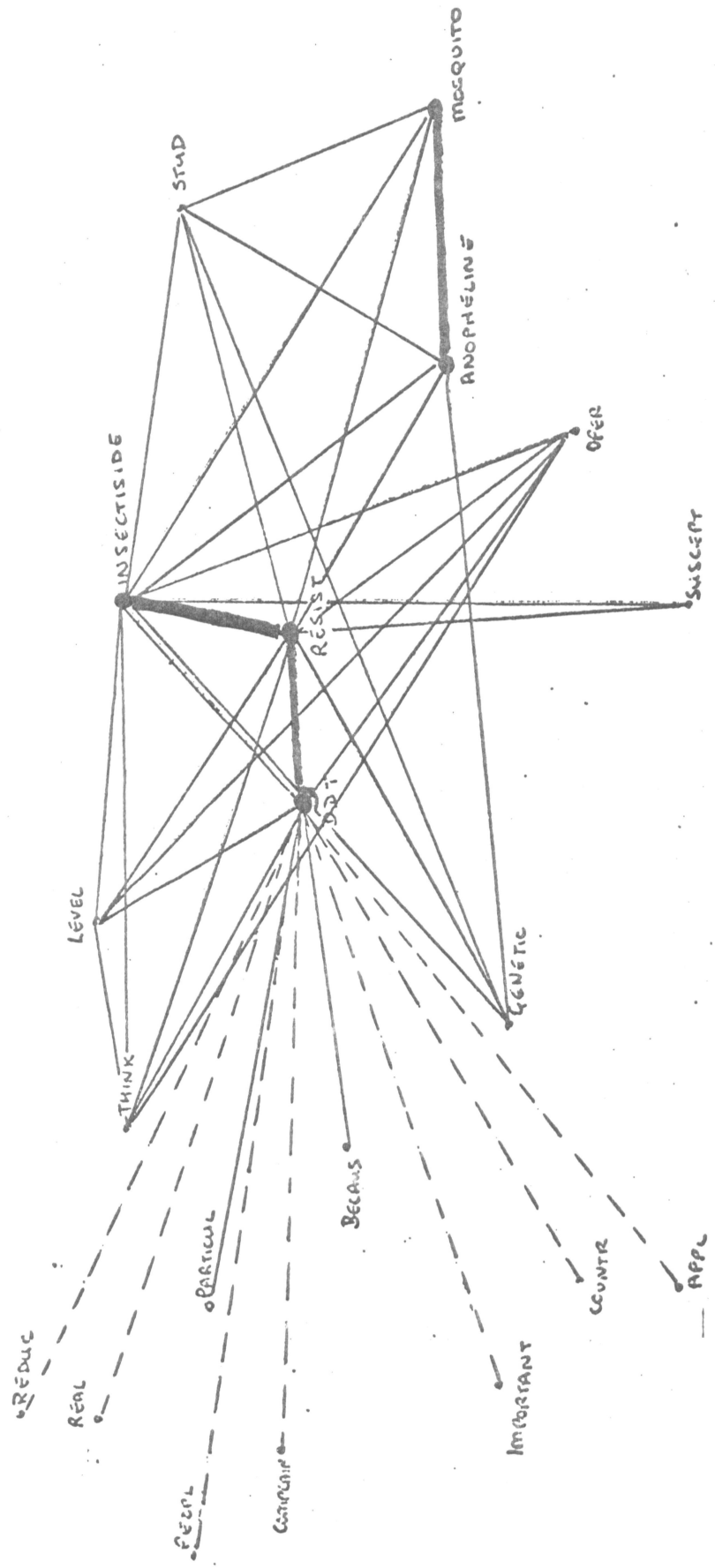
#29

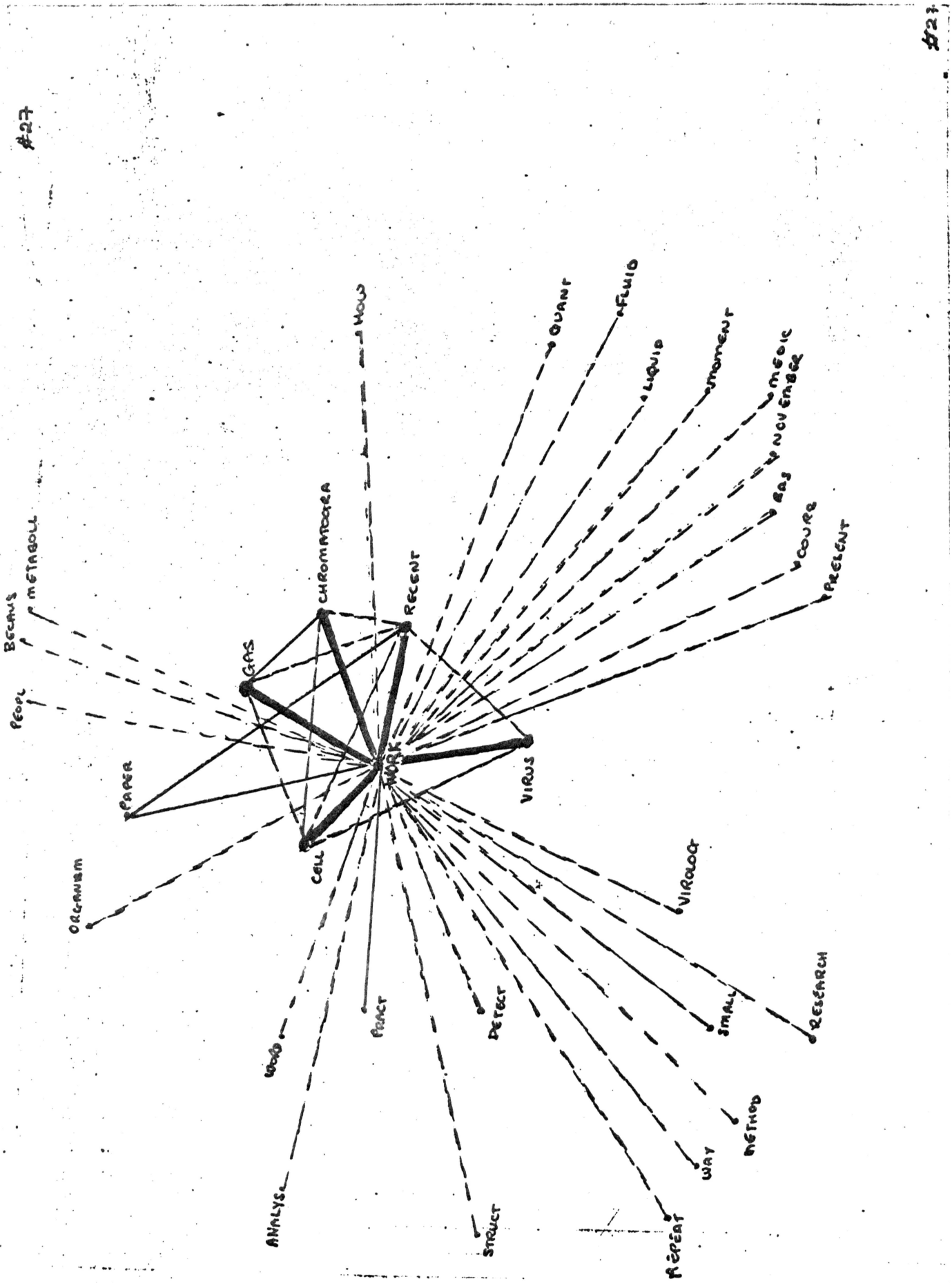

HYDATID
ECHINOCOCCUS
PARTICUL
LIBYA
ND
SORT
PREV
RELEV
RAPID
EPIDEMIOLOG
IMMUNOLOGY
METHOD
IVFORM
DIAGNOS
DICERS
COMPAR
SUBSEC
WORK
WAY
READ
DIFFER
THINK
EFFECT

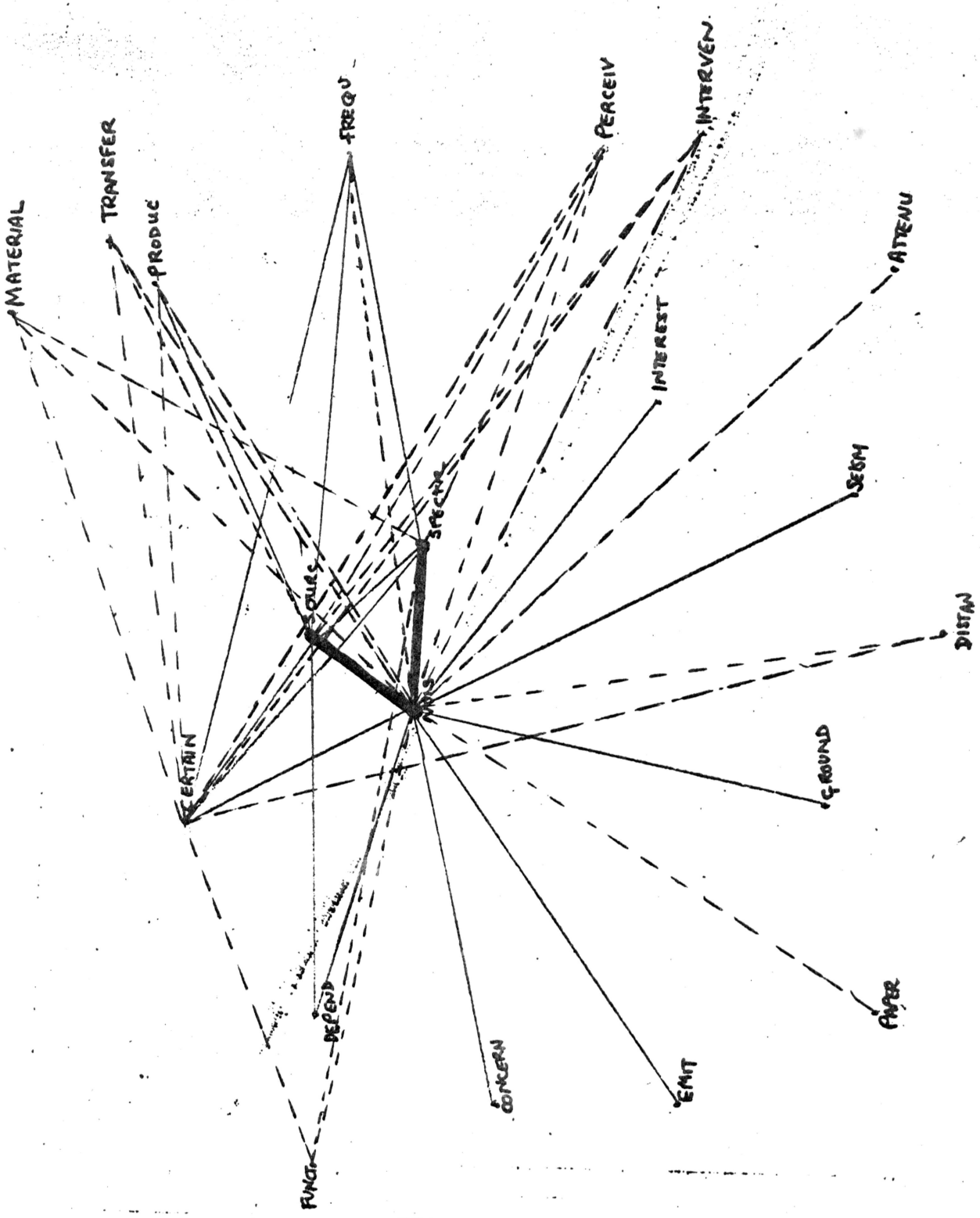
23

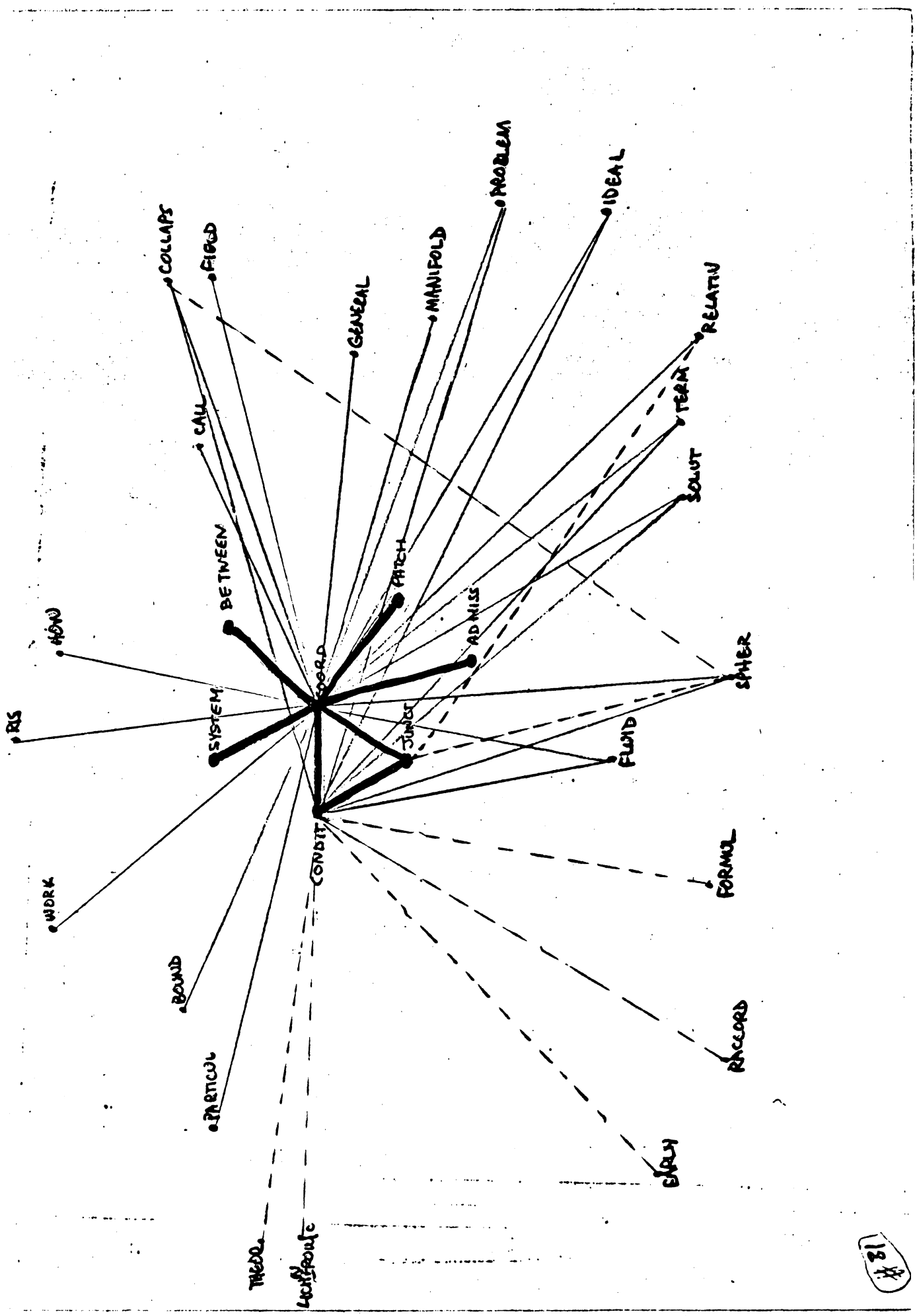

DAMAG
INTERSCOR
INTEREST
FIRST
ND
DIFFER
RESULT
FAR
DEVI
PEOPL
OFT
SYSTEM
INDICAT
SCOR
GOOD
WELL
MARK
CLIENT
STUD
BRAIN
US
IMPORTANT
PSYCHOLOGY
WORK
TEST
MOTOR
SEVER
SECOND
RELI

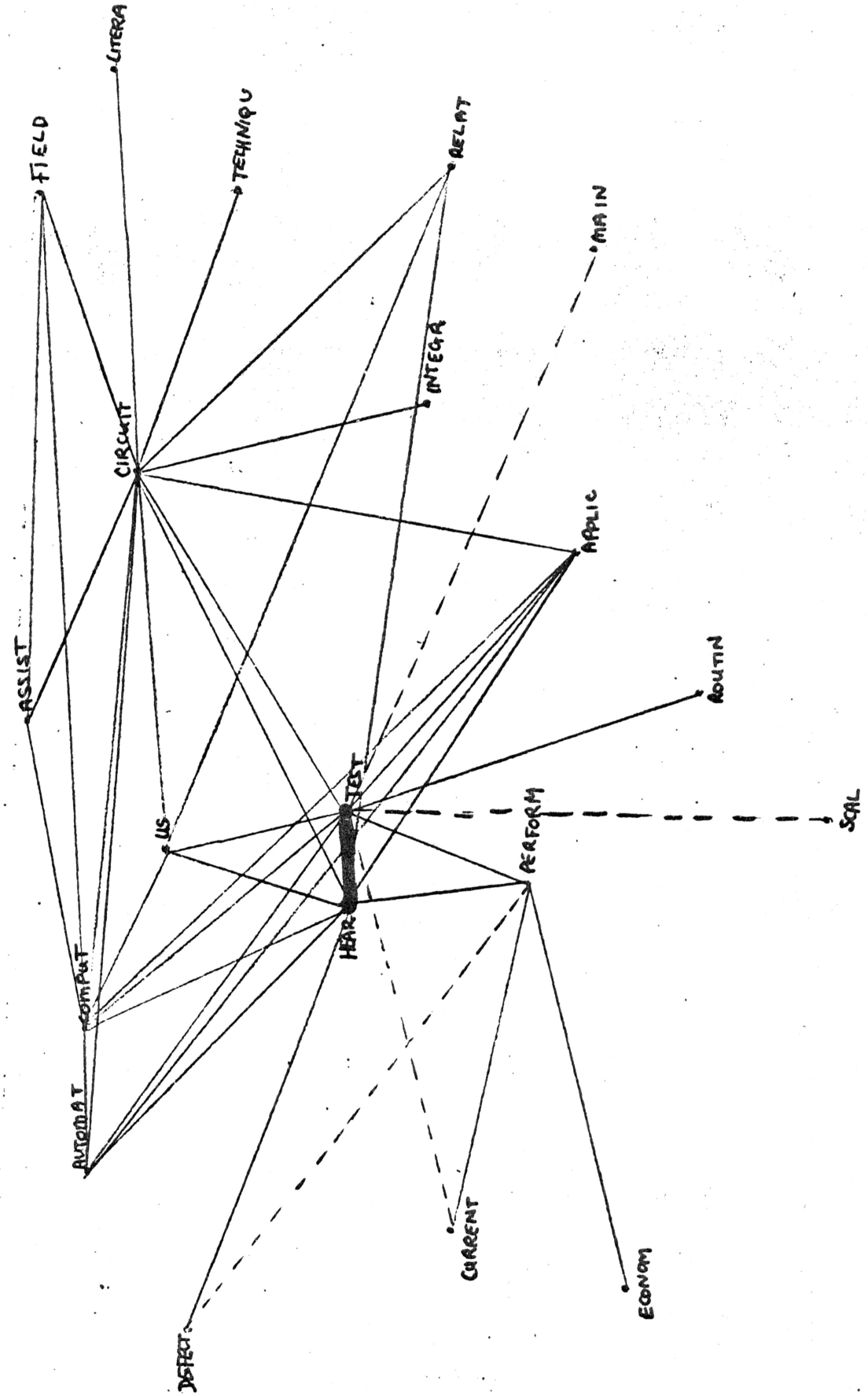

RELAT
DAT
BEING
HOW
NOW
POSS
SEARCH
AROUND
MEDLINE
COMPUT
FRONT
BRAIN
LOB
LINK
CHILD
DEVELOP
BETWEEN


MOSQUITO
STUD
ANOPHELINE
OFER
INSECTICIDE
SUSCEPT
RESIST
LEVEL
GENETIC
THINK
BECAUS
REDUC
REAL
PARTICUL
IMPORTANT
COUNTR
APPL


#27
#27
BEGINS
METABOLL
PEOPL
CHROMATOGRA
RECENT
GAS
QUANT
FLUID
LIQUID
MOMENT
MEDIC
NOVEMBER
EAS
COURS
PRELENT
PAPER
WORK
VIRUS
CELL
ORGANISM
VIROLOG
RESEARCH
SMALL
METHOD
WAY
REPEAT
STRUCT
ANALYS
PERCT
PETECT
WORD


MATERIAL
TRANSFER
PRODUC
FREQU
PERCEIV
INTERVEN
ATTENU
INTEREST
SEISM
SPECTR
COURS
DISTAN
CERTAIN
GROUND
PAPER
EMIT
CONCERN
DEPEND
FUNCT


COLLAPS
FIBRD
CALL
GENERAL
MANIFOLD
PROBLEM
IDEAL
RELATIV
PERM
SOLUT
BETWEEN
PATCH
ADMISS
HOW
RIS
SYSTEM
COORD
JUNCT
SPHER
FLUID
CONDIT
FORMUL
WORK
BOUND
RECORD
PARTIAL
EARLY
THEORE
LOCAL


LITERA
FIELD
TEHNIQU
RELAT
MAIN
INTEGR
CIRCUIT
APPLIC
ASSIST
ROUTIN
TEST
US
SCAL
REFORM
HEAR
COMPUT
AUTOMAT
CURRENT
ECONOM
DSTRCT

#33

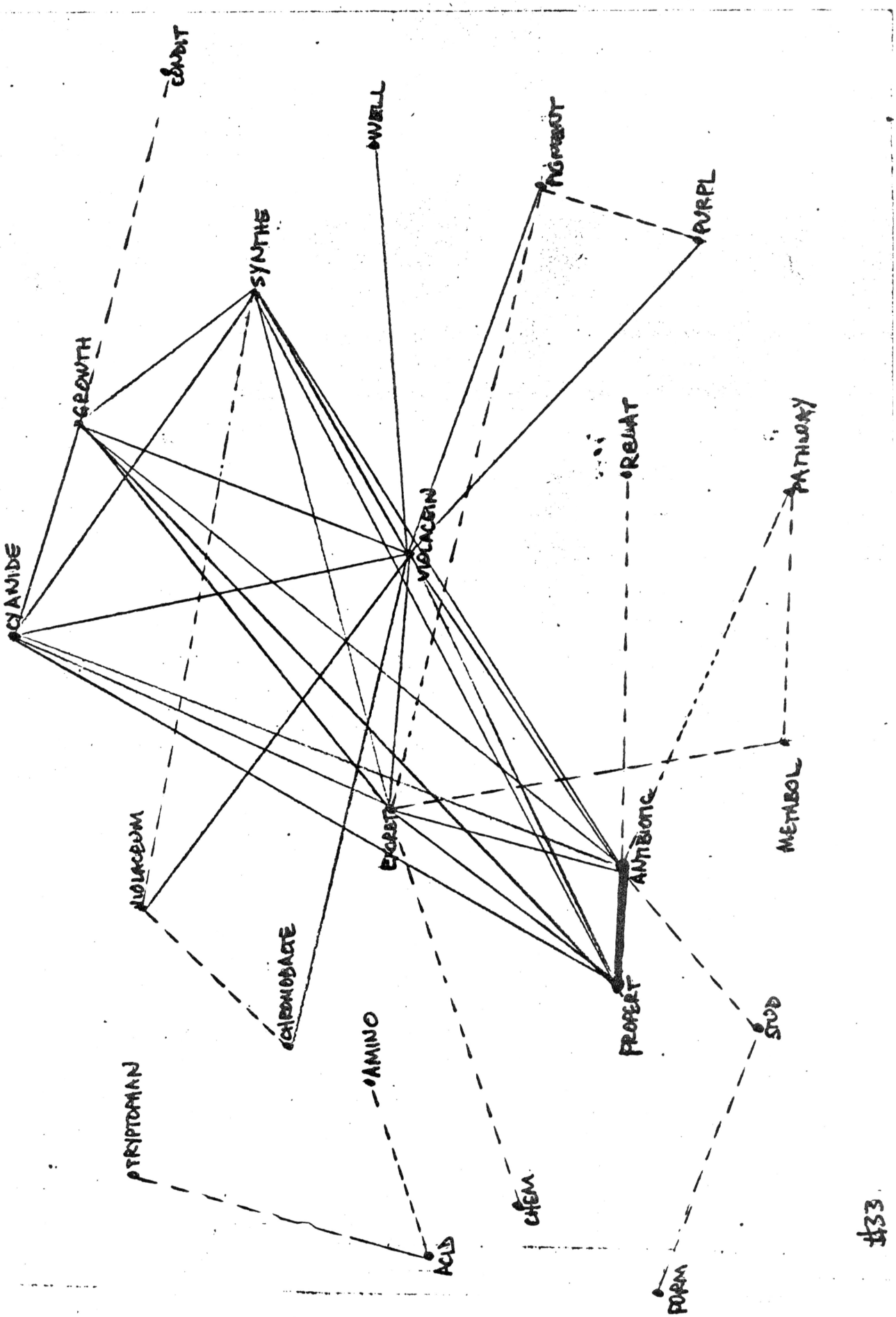
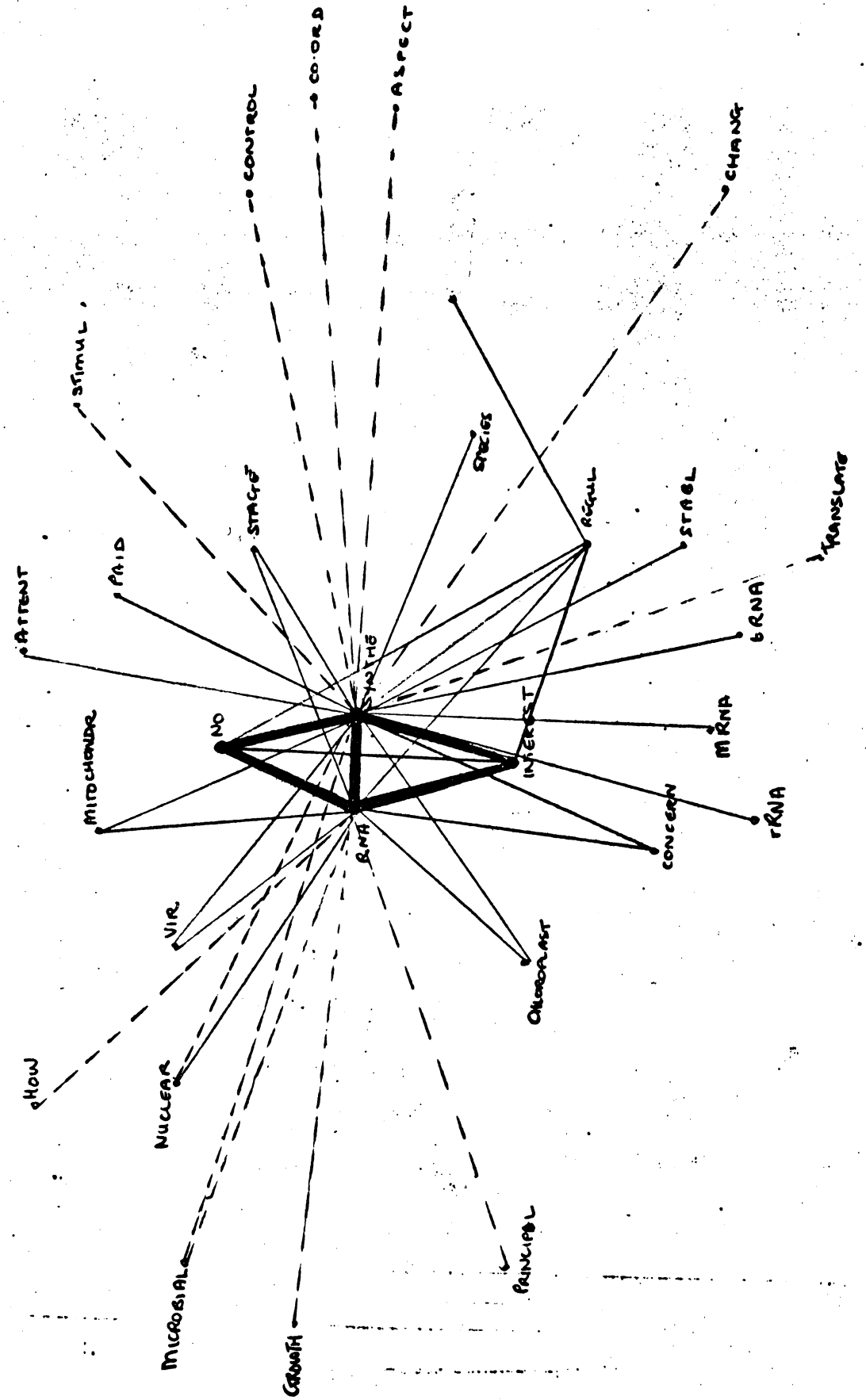
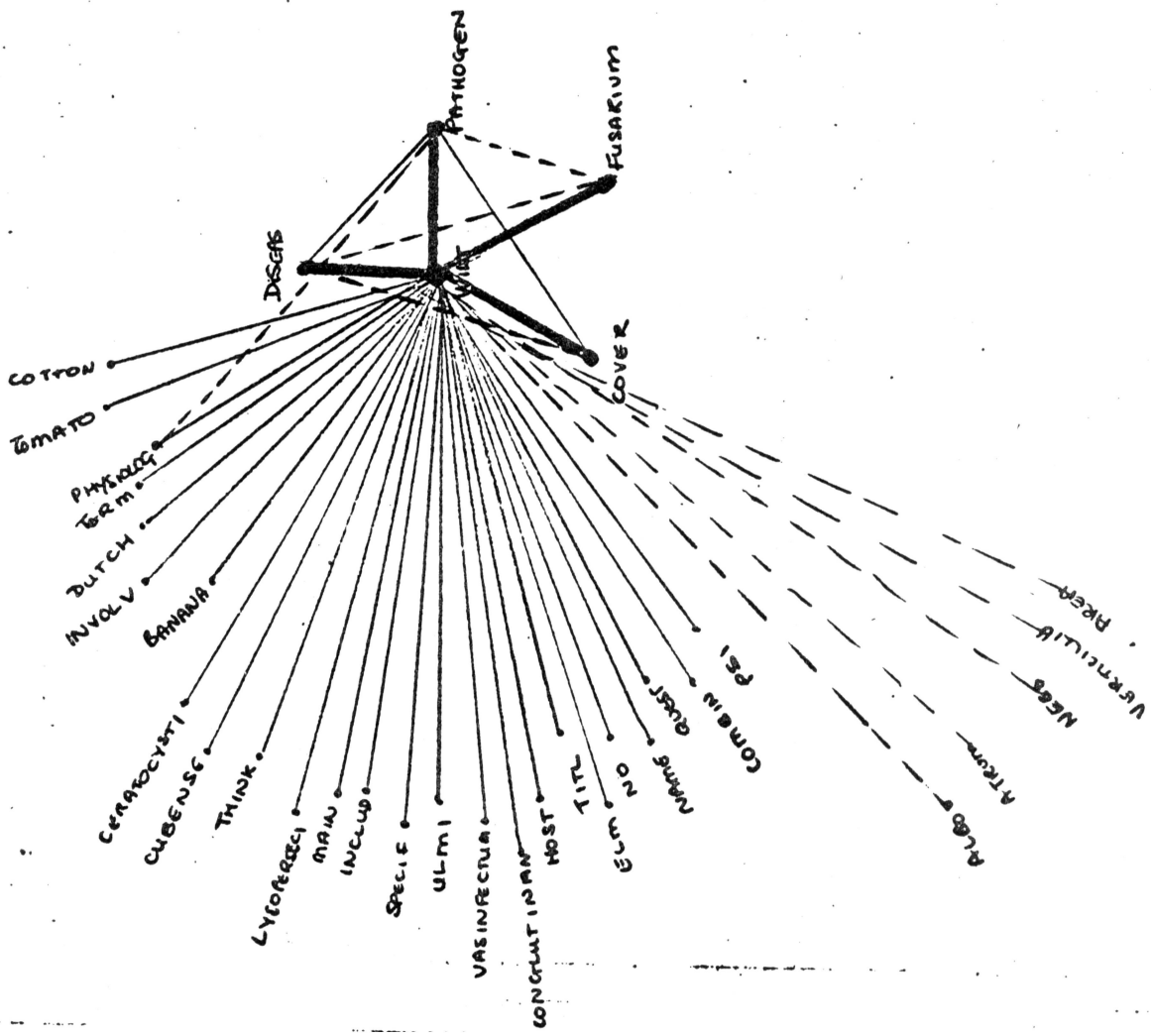
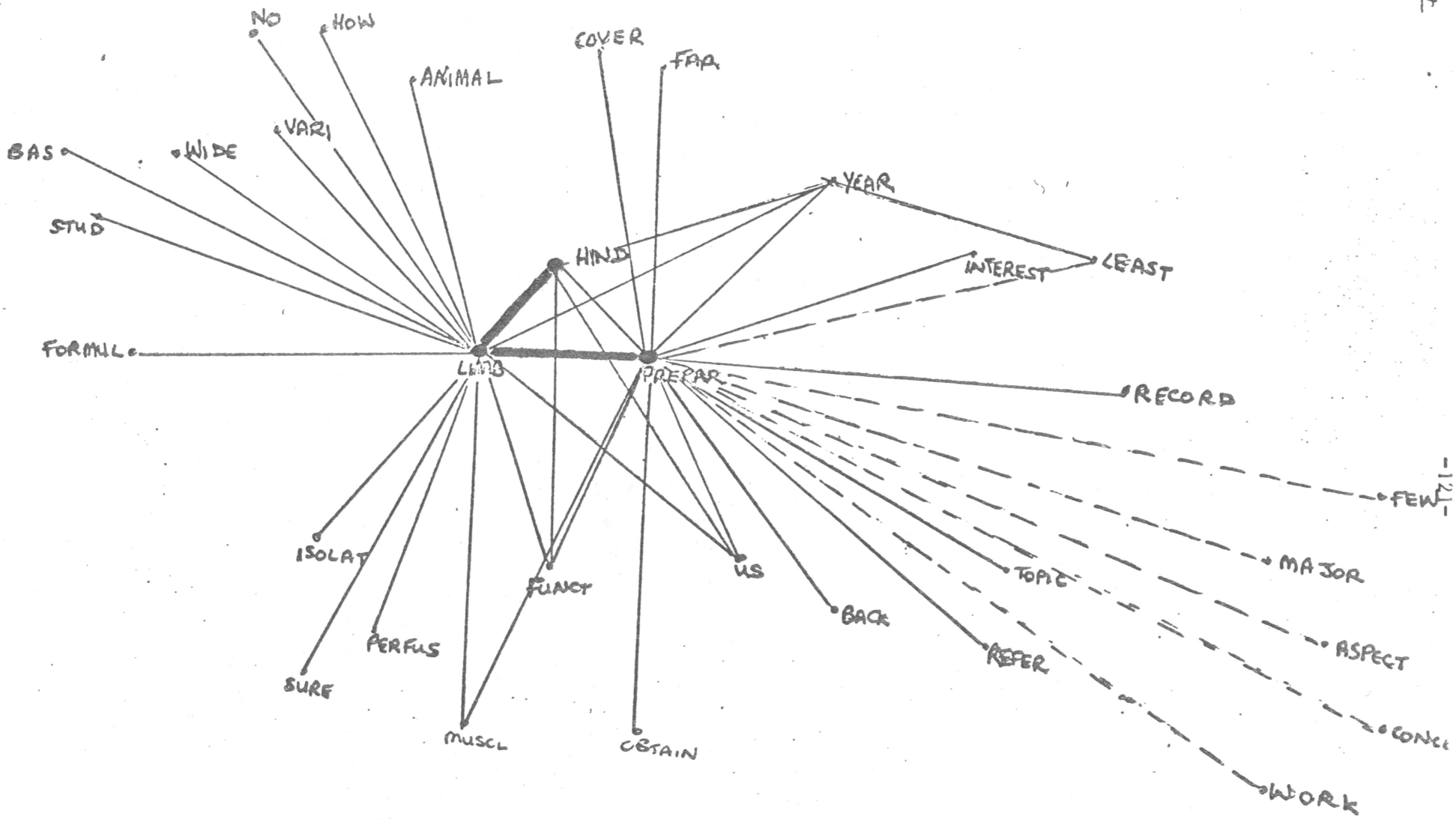


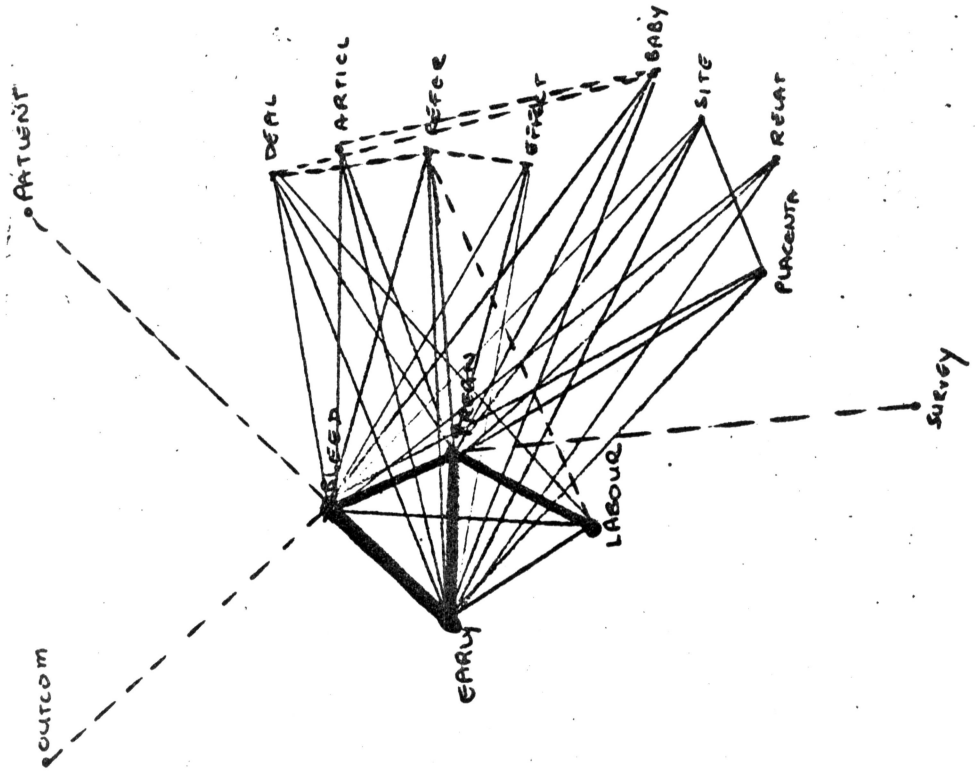
EXHIBIT
SYNTHE
GROWTH
CYANIDE
ALUMINUM
CHROMOBACTE
TRYPTOPHAN
AMINO
ACID
CHEM
POLYM
STUD
METABOL
PATHWAY
ANTIBIOTIC
PROPERT
EXPERT
VIOLACEIN
RELAT
PURPL
#33


CONTROL
CO-ORD
ASPECT
STIMUL
CHANG
SPECIES
STAGES
REGUL
STABIL
TRANSLATE
ATTENT
PAID
SYNTHS
tRNA
mRNA
NO
INTEREST
MITOCHONDR
CONCERN
rRNA
RNA
VIR
CHLOROPLAST
HOW
NUCLEAR
PRINCIPAL
MICROSIAL
GROWTH

#34


PATHOGEN
FUSARIUM
DISEAS
COVER
COTTON
TOMATO
PHYSIOLOG
TEAM
DUTCH
INVOLV
BANANA
CERATOCYSTI
CUBENSIS
THINK
LYCOPERSCI
MAIN
INCLUD
SPECIF
ULMI
VASINFECTUM
CONGLUTINANS
HOST
TITL
ELM
NO
MANAG
QUEST
COMBIN
PEI
ALBO
ATRUM
NEES
VERTICILLIU
PRIEM


NO
HOW
ANIMAL
COVER
FAR
VARI
WIDE
BAS
STUD
YEAR
HIND
INTEREST
LEAST
FORMUL
LEARN
PREPAR
RECORD
FEW
MAJOR
TOPIC
ASPECT
CONCL
WORK
REFER
BACK
US
FUNCT
ISOLAT
PERFUS
SURE
MUSCL
OBTAIN


PATIENT
OUTCOM
DEAL
ARTICL
REFER
BABY
SITE
REHAB
PLACENTA
SURGY
LABOUR
EARLY

#37.

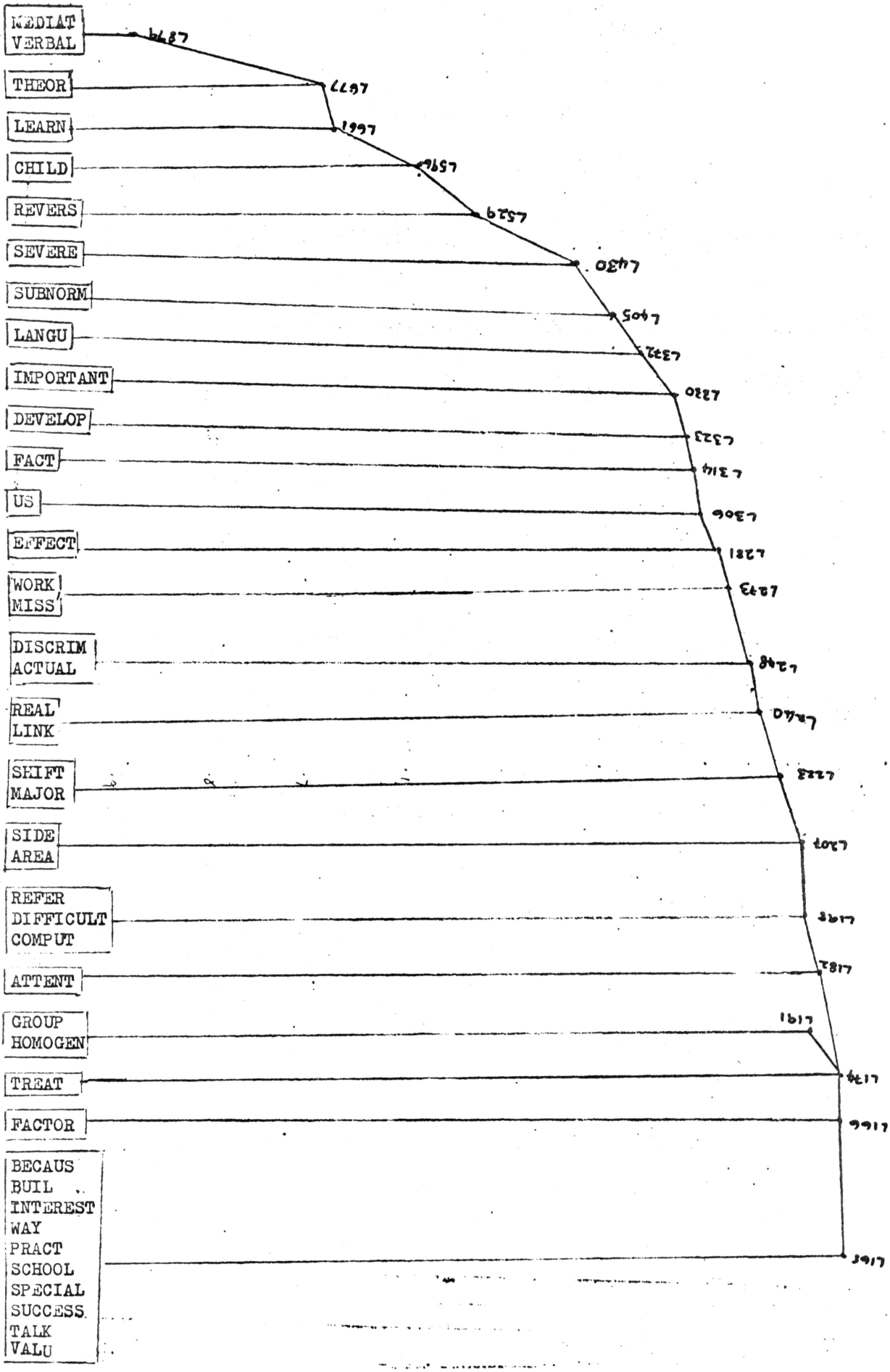
APPENDIX E.

Single-link clustering representation of problem statements

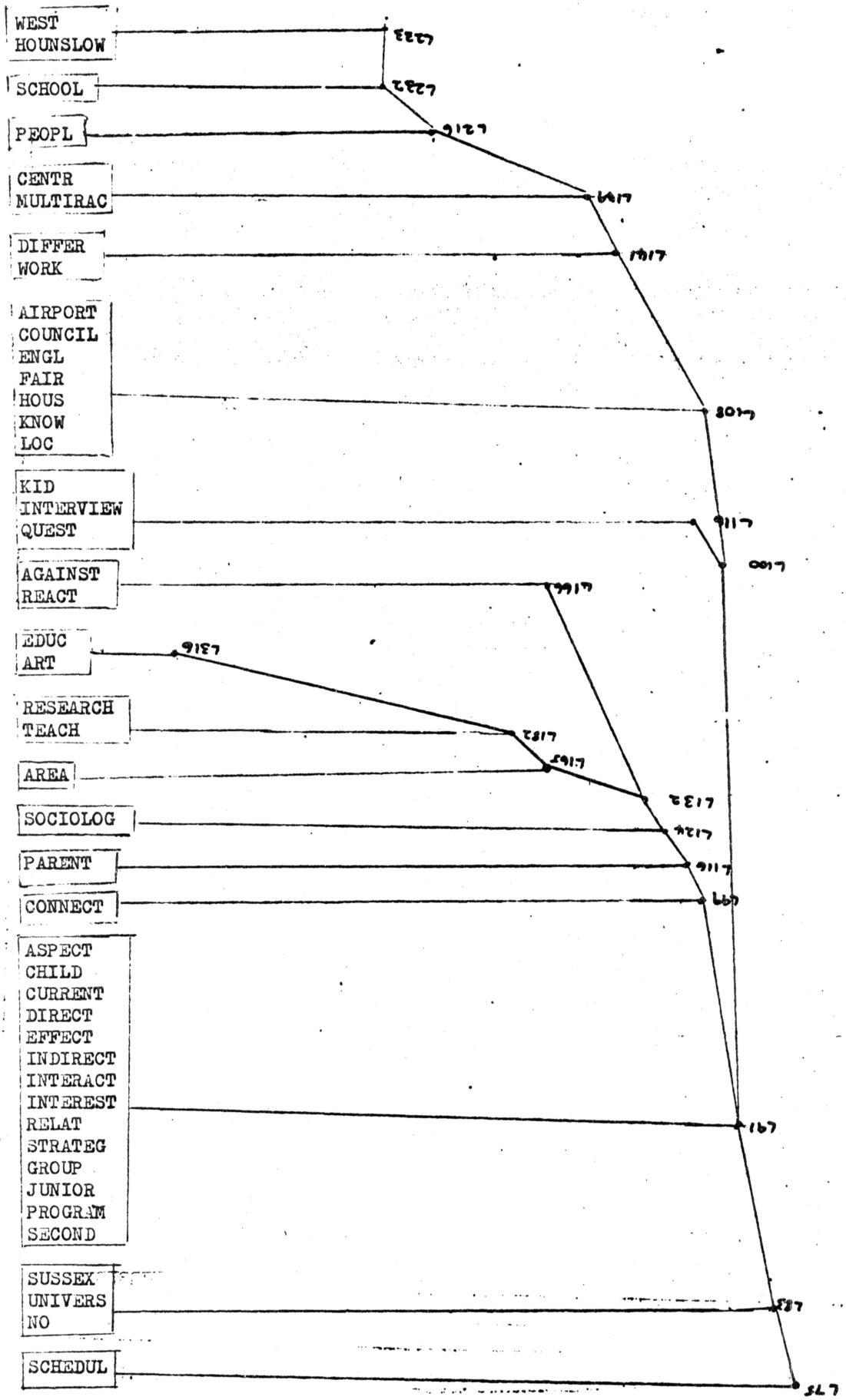
1.27 Oral statements

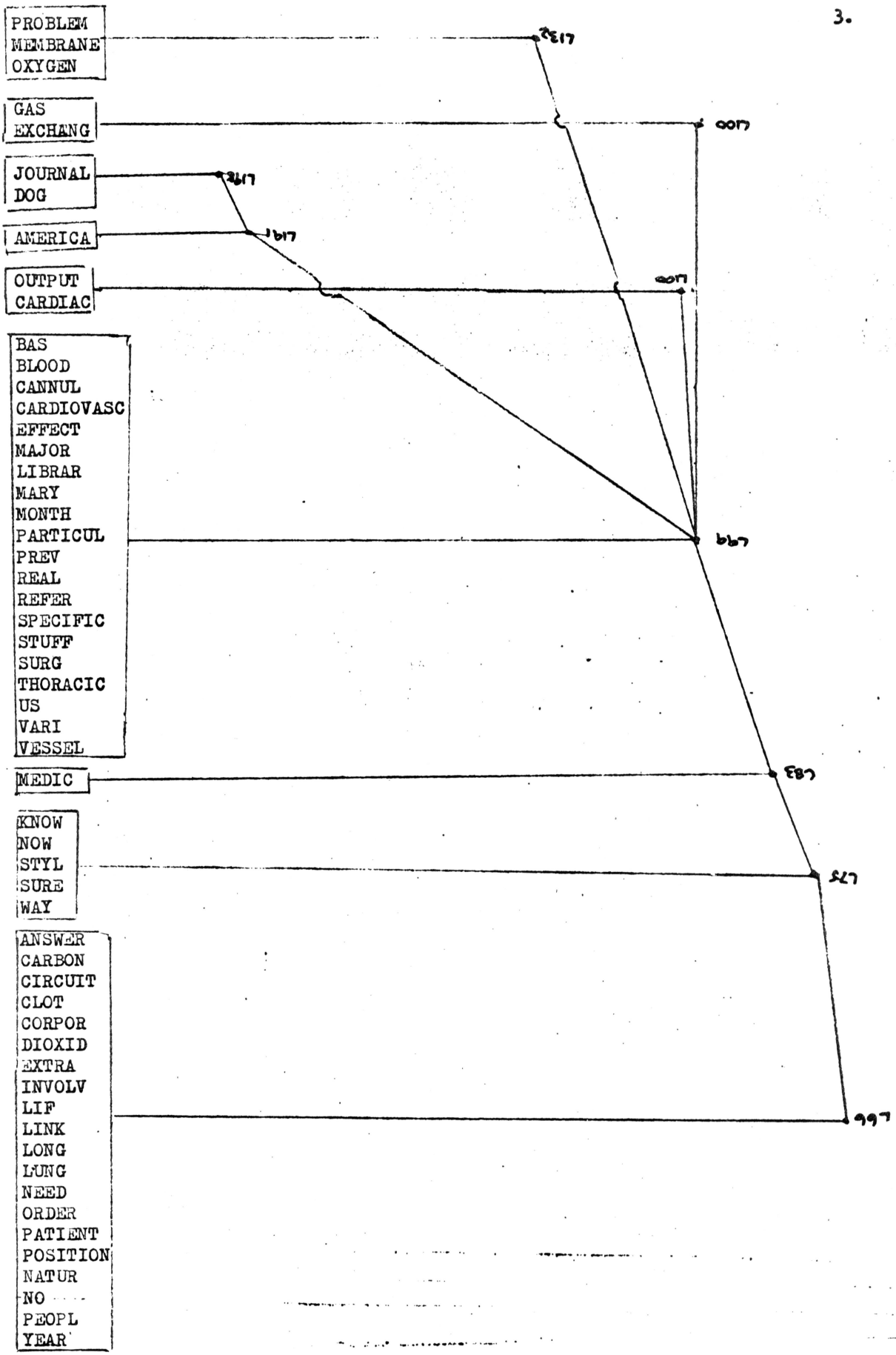
30-37 Written statements

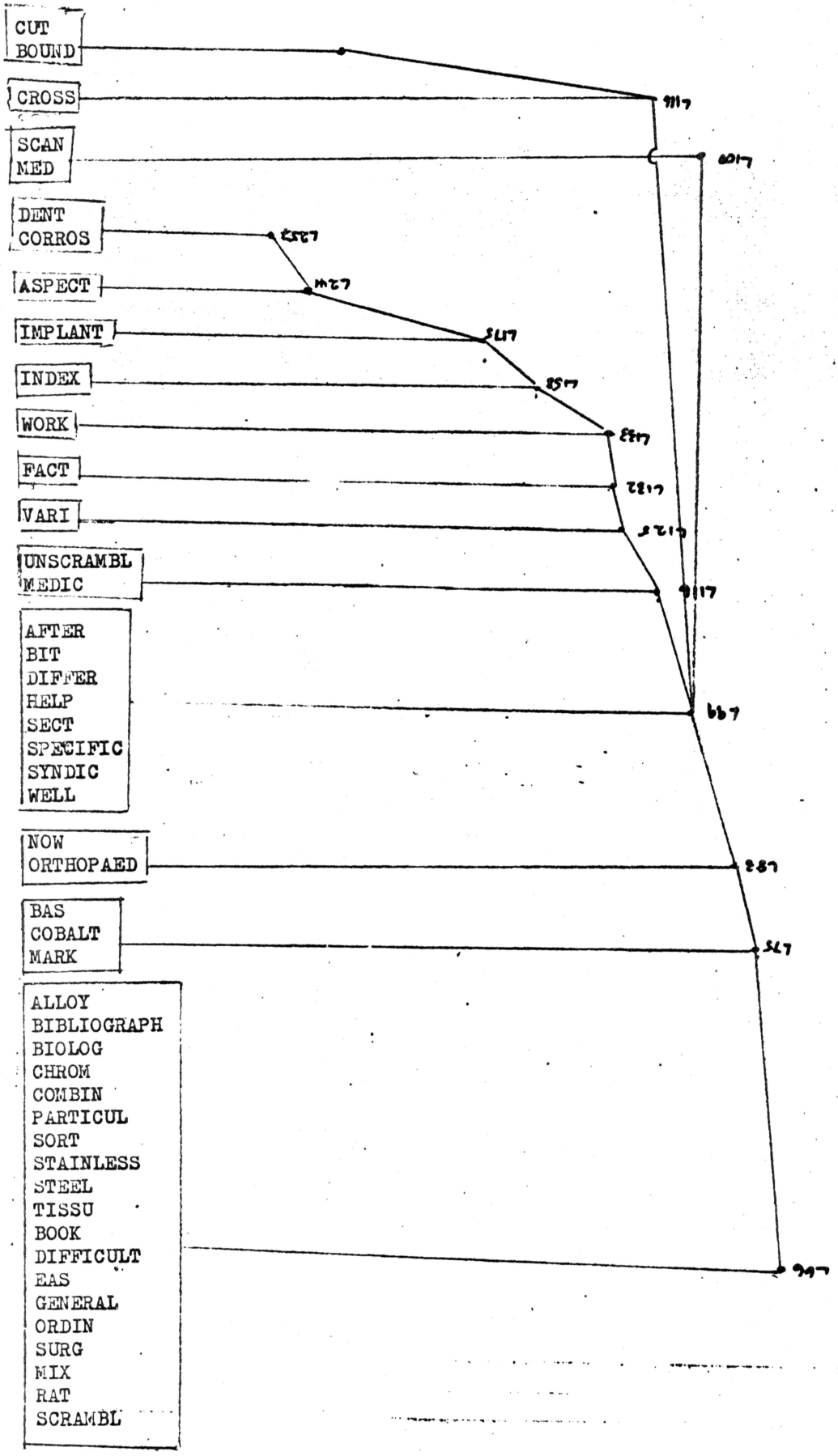
I.


MEDIAT VERBAL
.6879
THEOR
.6677
LEARN
.6661
CHILD
.5965
REVERS
.5529
SEVERE
.4930
SUBNORM
.4905
LANGU
.4332
IMPORTANT
.4330
DEVELOP
.4323
FACT
.4314
US
.4306
EFFECT
.4281
WORK MISS
.4243
DISCRIM ACTUAL
.4248
REAL LINK
.4240
SHIFT MAJOR
.4223
SIDE AREA
.4201
REFER DIFFICULT COMPUT
.4192
ATTENT
.4182
GROUP HOMOGEN
.4191
TREAT
.4179
FACTOR
.4166
BECAUS
BUIL
INTEREST
WAY
PRACT
SCHOOL
SPECIAL
SUCCESS
TALK
VALU
.4167

2.

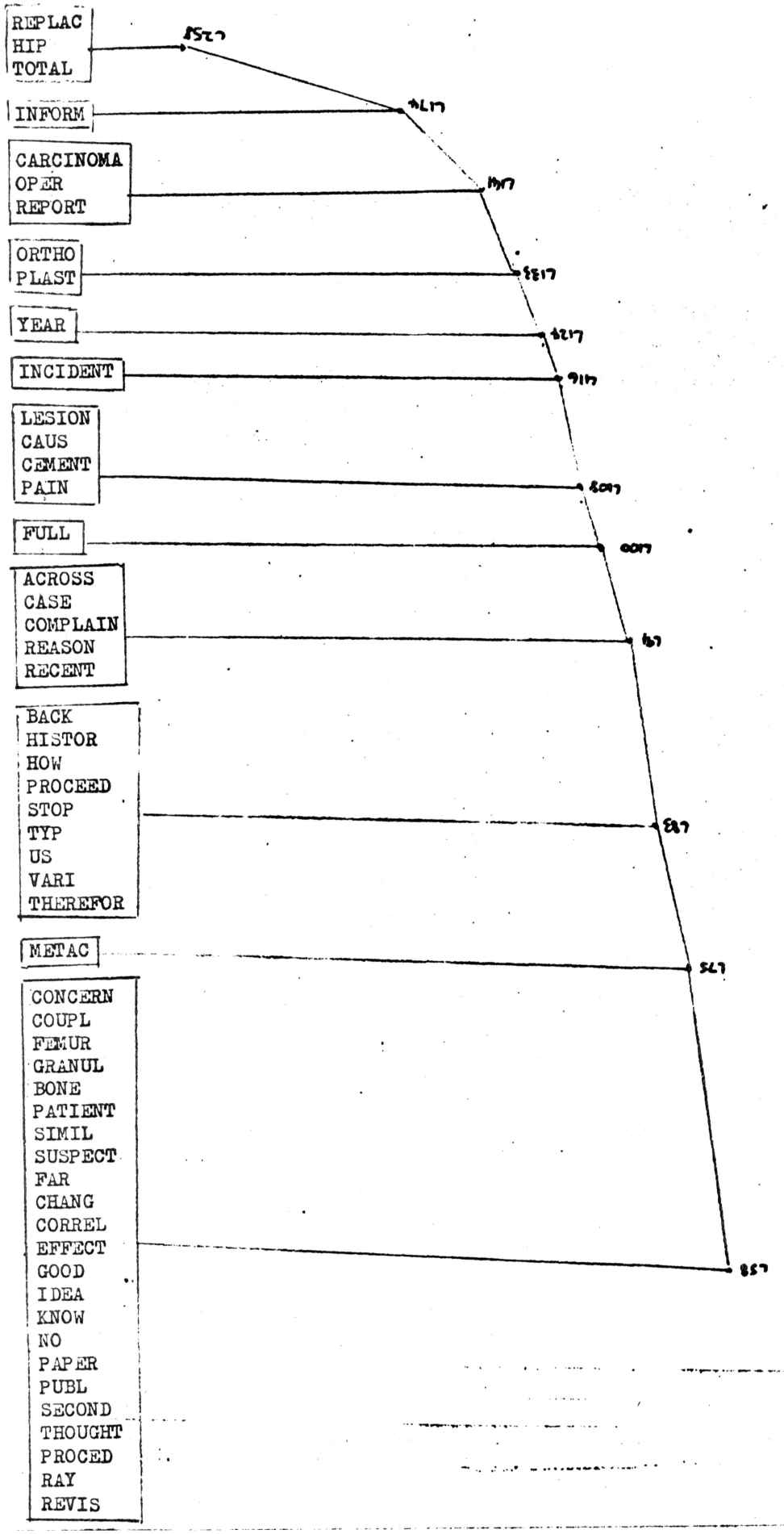

WEST
HOUNSLOW
SCHOOL
PEOPL
CENTR
MULTIRAC
DIFFER
WORK
AIRPORT
COUNCIL
ENGL
FAIR
HOUS
KNOW
LOC
KID
INTERVIEW
QUEST
AGAINST
REACT
EDUC
ART
RESEARCH
TEACH
AREA
SOCIOLOG
PARENT
CONNECT
ASPECT
CHILD
CURRENT
DIRECT
EFFECT
INDIRECT
INTERACT
INTEREST
RELAT
STRATEG
GROUP
JUNIOR
PROGRAM
SECOND
SUSSEX
UNIVERS
NO
SCHEDUL


PROBLEM
MEMBRANE
OXYGEN
GAS
EXCHANG
JOURNAL
DOG
AMERICA
OUTPUT
CARDIAC
BAS
BLOOD
CANNUL
CARDIOVASC
EFFECT
MAJOR
LIBRAR
MARY
MONTH
PARTICUL
PREV
REAL
REFER
SPECIFIC
STUFF
SURG
THORACIC
US
VARI
VESSEL
MEDIC
KNOW
NOW
STYL
SURE
WAY
ANSWER
CARBON
CIRCUIT
CLOT
CORPOR
DIOXID
EXTRA
INVOLV
LIF
LINK
LONG
LUNG
NEED
ORDER
PATIENT
POSITION
NATUR
NO
PEOPL
YEAR

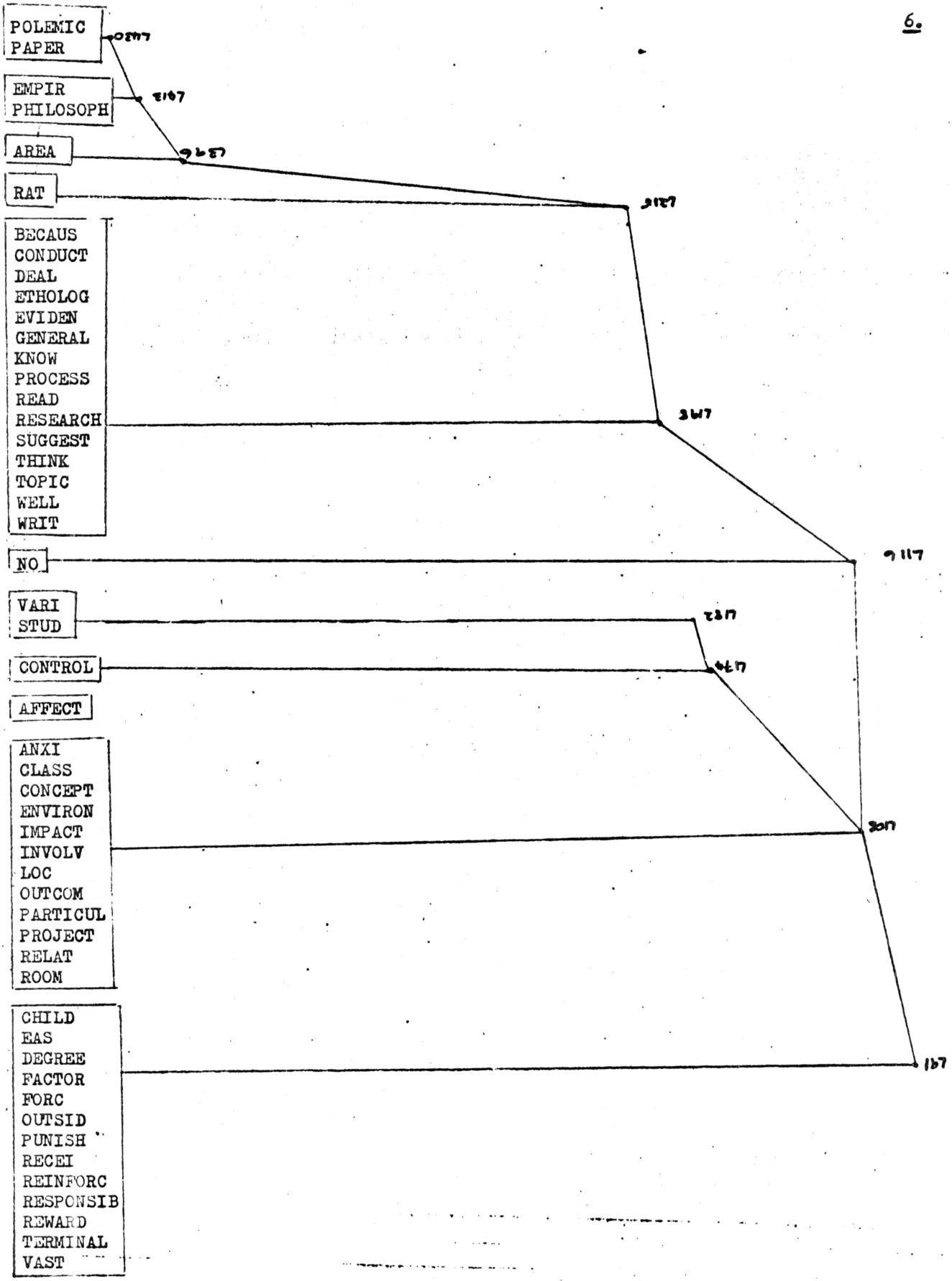

CUT
BOUND
CROSS
SCAN
MED
DENT
CORROS
ASPECT
IMPLANT
INDEX
WORK
FACT
VARI
UNSCRAMBL
MEDIC
AFTER
BIT
DIFFER
HELP
SECT
SPECIFIC
SYNDIC
WELL
NOW
ORTHOPAED
BAS
COBALT
MARK
ALLOY
BIBLIOGRAPH
BIOLOG
CHROM
COMBIN
PARTICUL
SORT
STAINLESS
STEEL
TISSU
BOOK
DIFFICULT
EAS
GENERAL
ORDIN
SURG
MIX
RAT
SCRAMBL

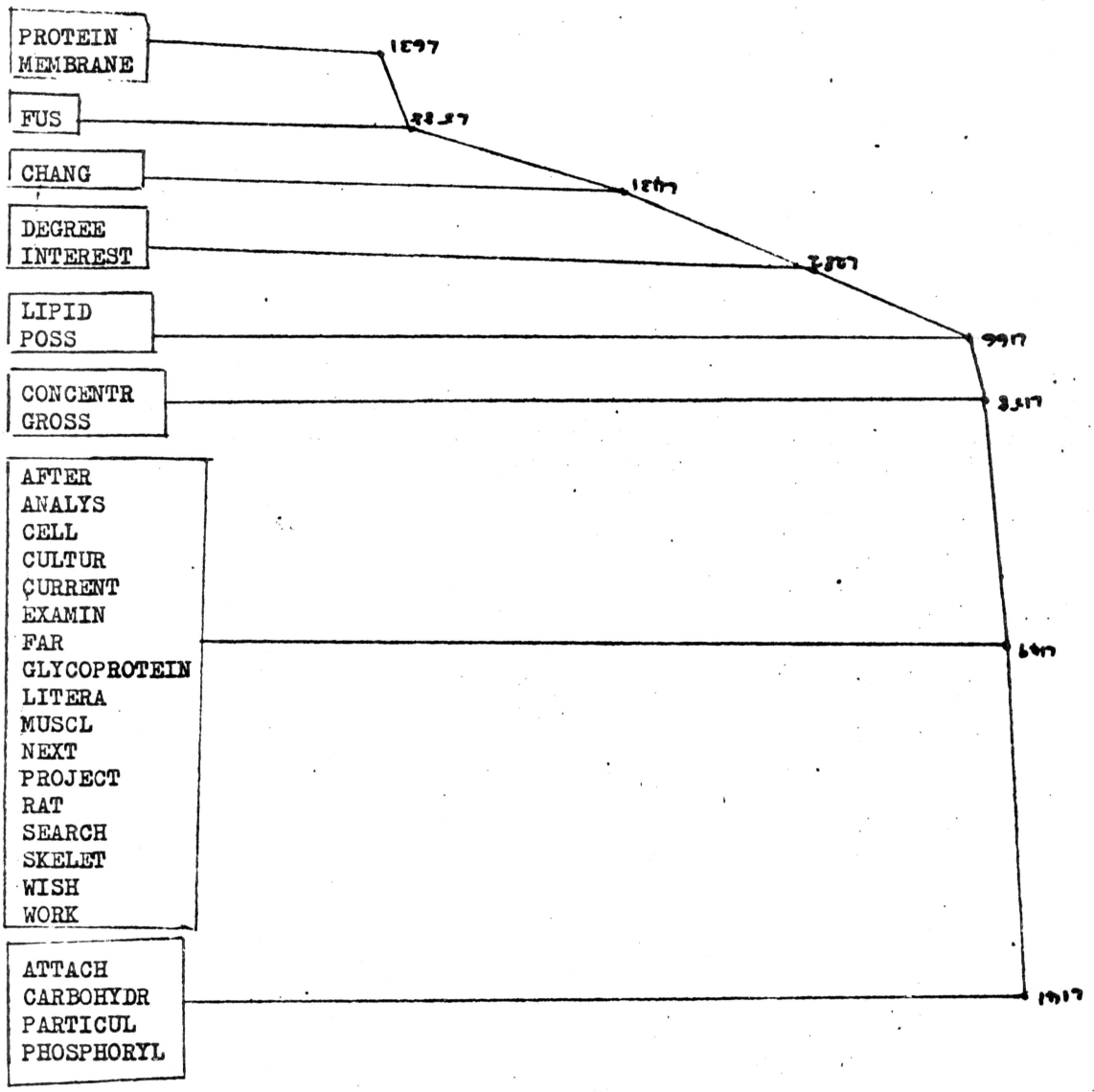
4.

5.


REPLAC
HIP
TOTAL
INFORM
CARCINOMA
OPER
REPORT
ORTHO
PLAST
YEAR
INCIDENT
LESION
CAUS
CEMENT
PAIN
FULL
ACROSS
CASE
COMPLAIN
REASON
RECENT
BACK
HISTOR
HOW
PROCEED
STOP
TYP
US
VARI
THEREFOR
METAC
CONCERN
COUPL
FEMUR
GRANUL
BONE
PATIENT
SIMIL
SUSPECT
FAR
CHANG
CORREL
EFFECT
GOOD
IDEA
KNOW
NO
PAPER
PUBL
SECOND
THOUGHT
PROCED
RAY
REVIS

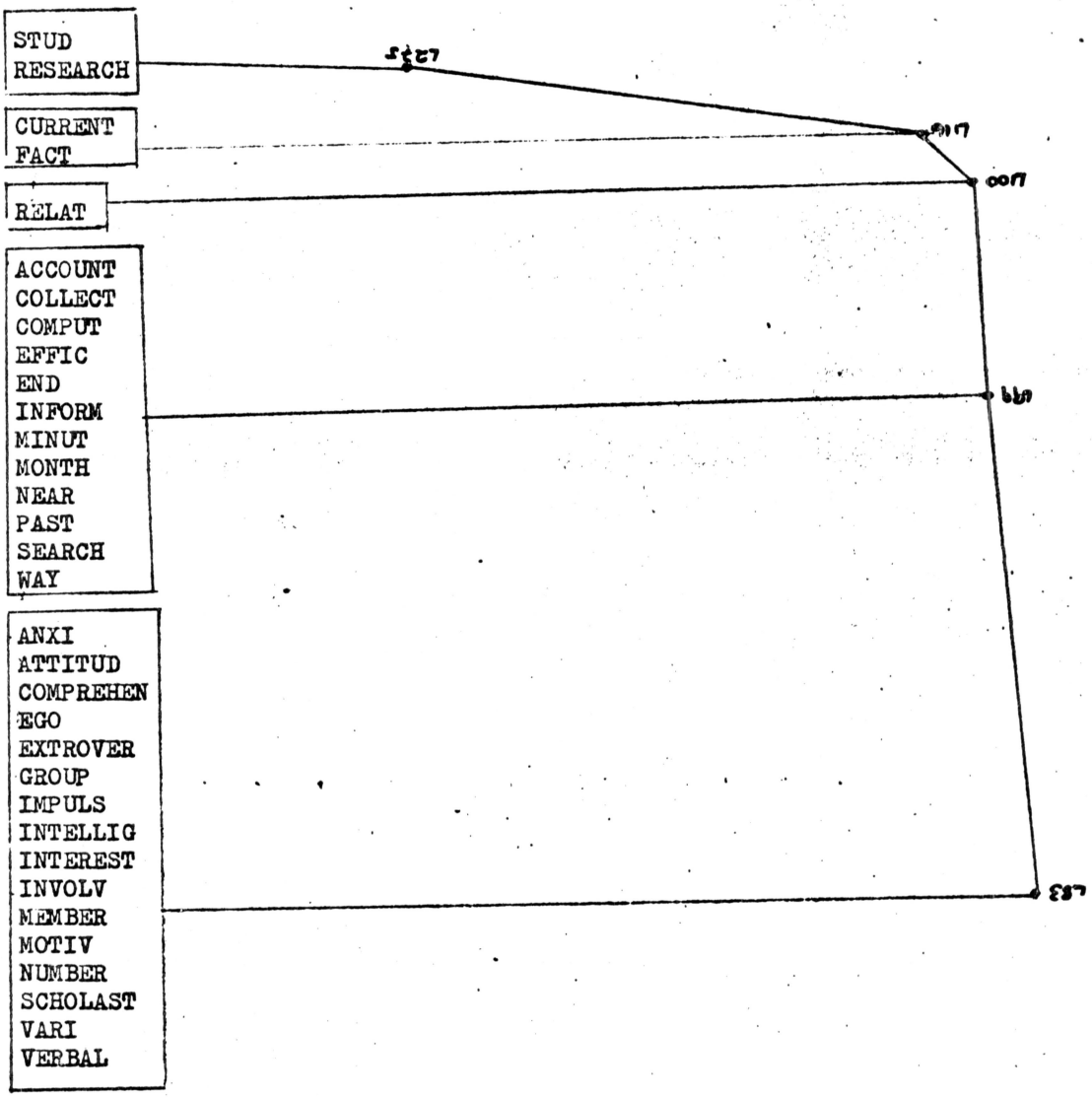
6.


POLEMIC
PAPER
EMPIR
PHILOSOPH
AREA
RAT
BECAUS
CONDUCT
DEAL
ETHOLOG
EVIDEN
GENERAL
KNOW
PROCESS
READ
RESEARCH
SUGGEST
THINK
TOPIC
WELL
WRIT
NO
VARI
STUD
CONTROL
AFFECT
ANXI
CLASS
CONCEPT
ENVIRON
IMPACT
INVOLV
LOC
OUTCOM
PARTICUL
PROJECT
RELAT
ROOM
CHILD
EAS
DEGREE
FACTOR
FORC
OUTSID
PUNISH
RECEI
REINFORC
RESPONSIB
REWARD
TERMINAL
VAST

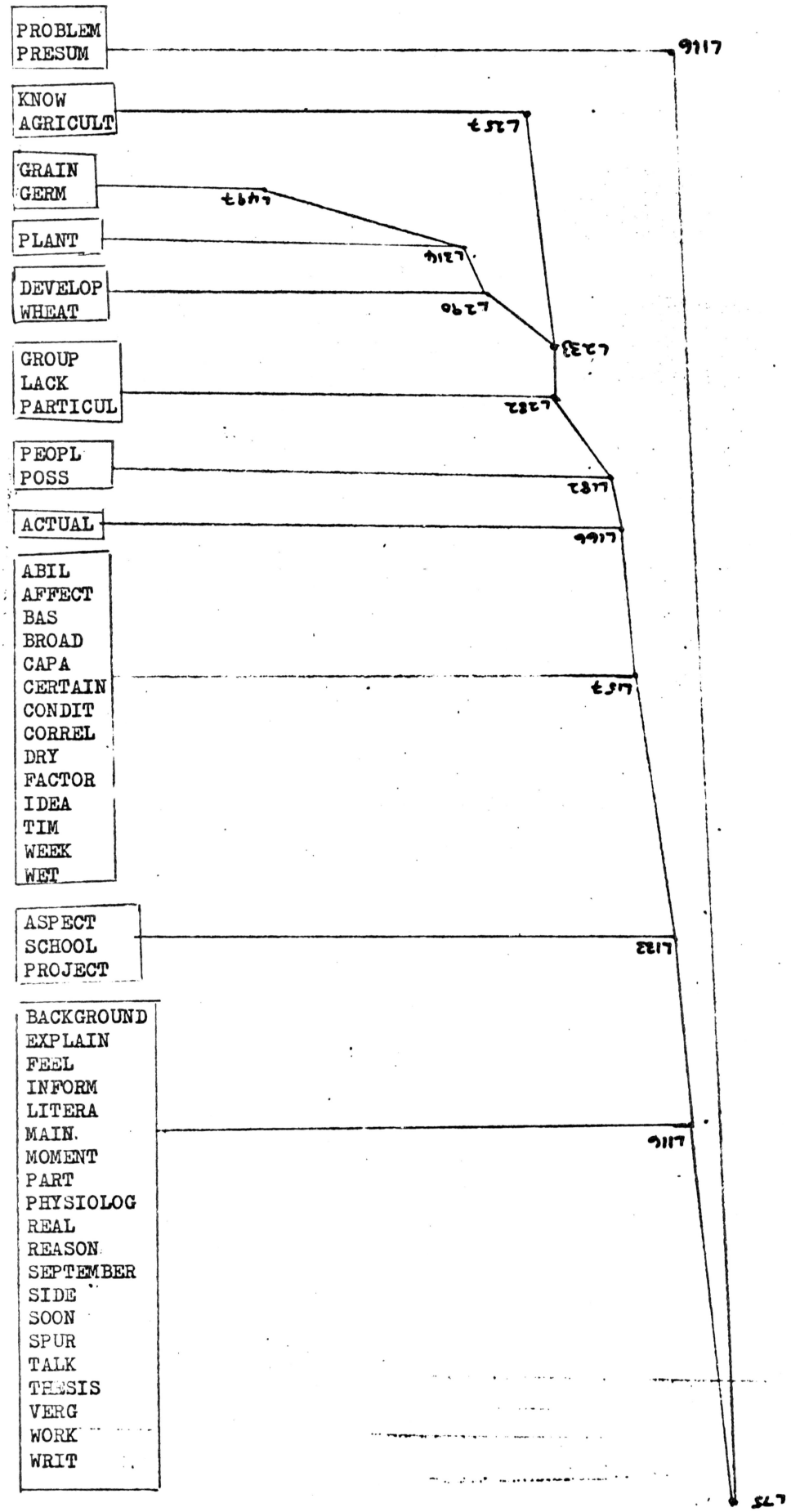

PROTEIN
MEMBRANE
FUS
CHANG
DEGREE
INTEREST
LIPID
POSS
CONCENTR
GROSS
AFTER
ANALYS
CELL
CULTUR
CURRENT
EXAMIN
FAR
GLYCOPROTEIN
LITERA
MUSCL
NEXT
PROJECT
RAT
SEARCH
SKELET
WISH
WORK
ATTACH
CARBOHYDR
PARTICUL
PHOSPHORYL

7.

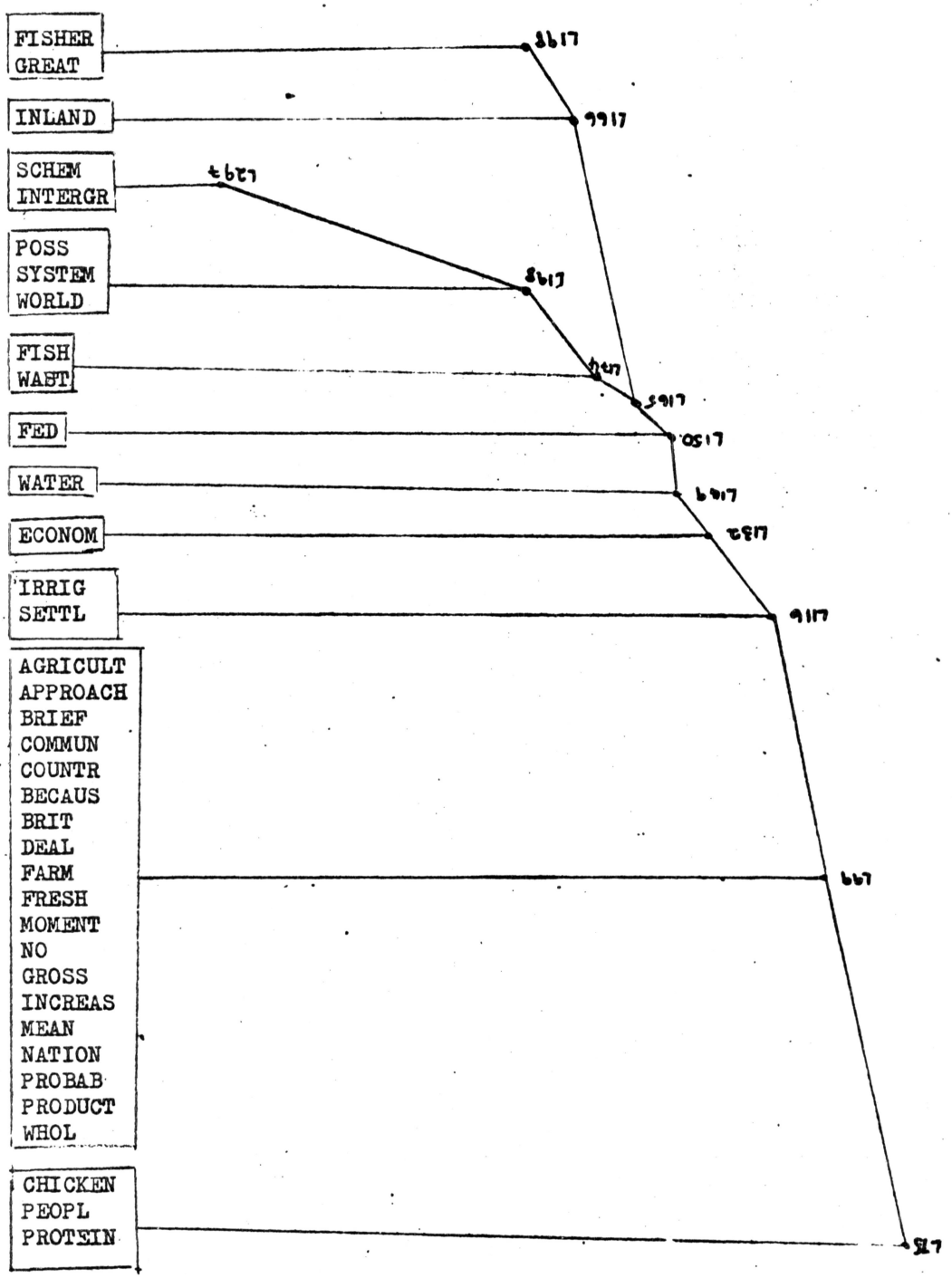
8.


STUD
RESEARCH
CURRENT
FACT
RELAT
ACCOUNT
COLLECT
COMPUT
EFFIC
END
INFORM
MINUT
MONTH
NEAR
PAST
SEARCH
WAY
ANXI
ATTITUD
COMPREHEN
EGO
EXTROVER
GROUP
IMPULS
INTELLIG
INTEREST
INVOLV
MEMBER
MOTIV
NUMBER
SCHOLAST
VARI
VERBAL

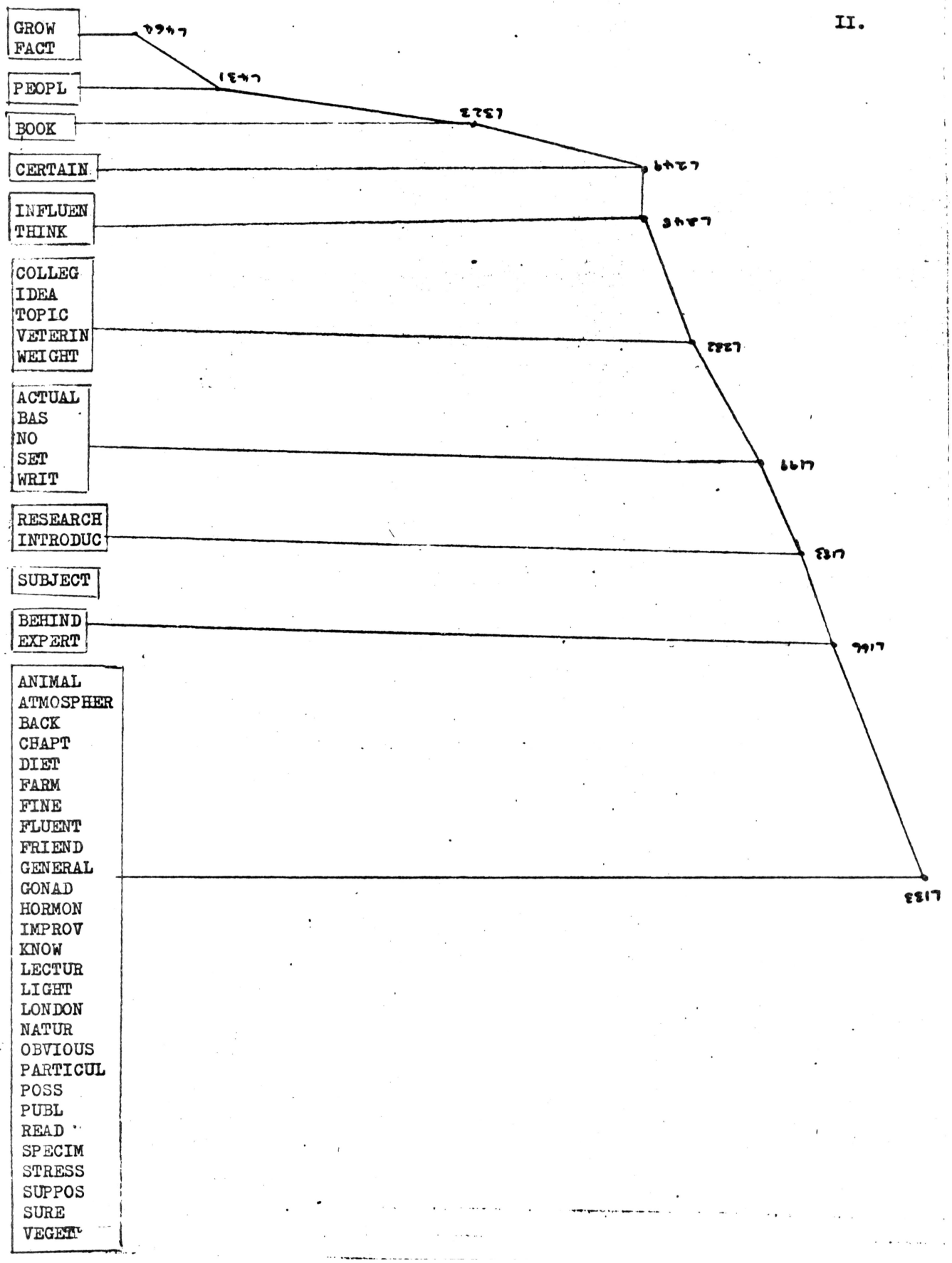
9.


PROBLEM
PRESUM
9117
KNOW
AGRICULT
4527
GRAIN
GERM
4649
PLANT
4314
DEVELOP
WHEAT
4290
4233
GROUP
LACK
PARTICUL
4282
PEOPL
POSS
4182
ACTUAL
9117
ABIL
AFFECT
BAS
BROAD
CAPA
CERTAIN
CONDIT
CORREL
DRY
FACTOR
IDEA
TIM
WEEK
WET
4557
ASPECT
SCHOOL
PROJECT
4132
BACKGROUND
EXPLAIN
FEEL
INFORM
LITERA
MAIN.
MOMENT
PART
PHYSIOLOG
REAL
REASON
SEPTEMBER
SIDE
SOON
SPUR
TALK
THESIS
VERG
WORK
WRIT
9117
527

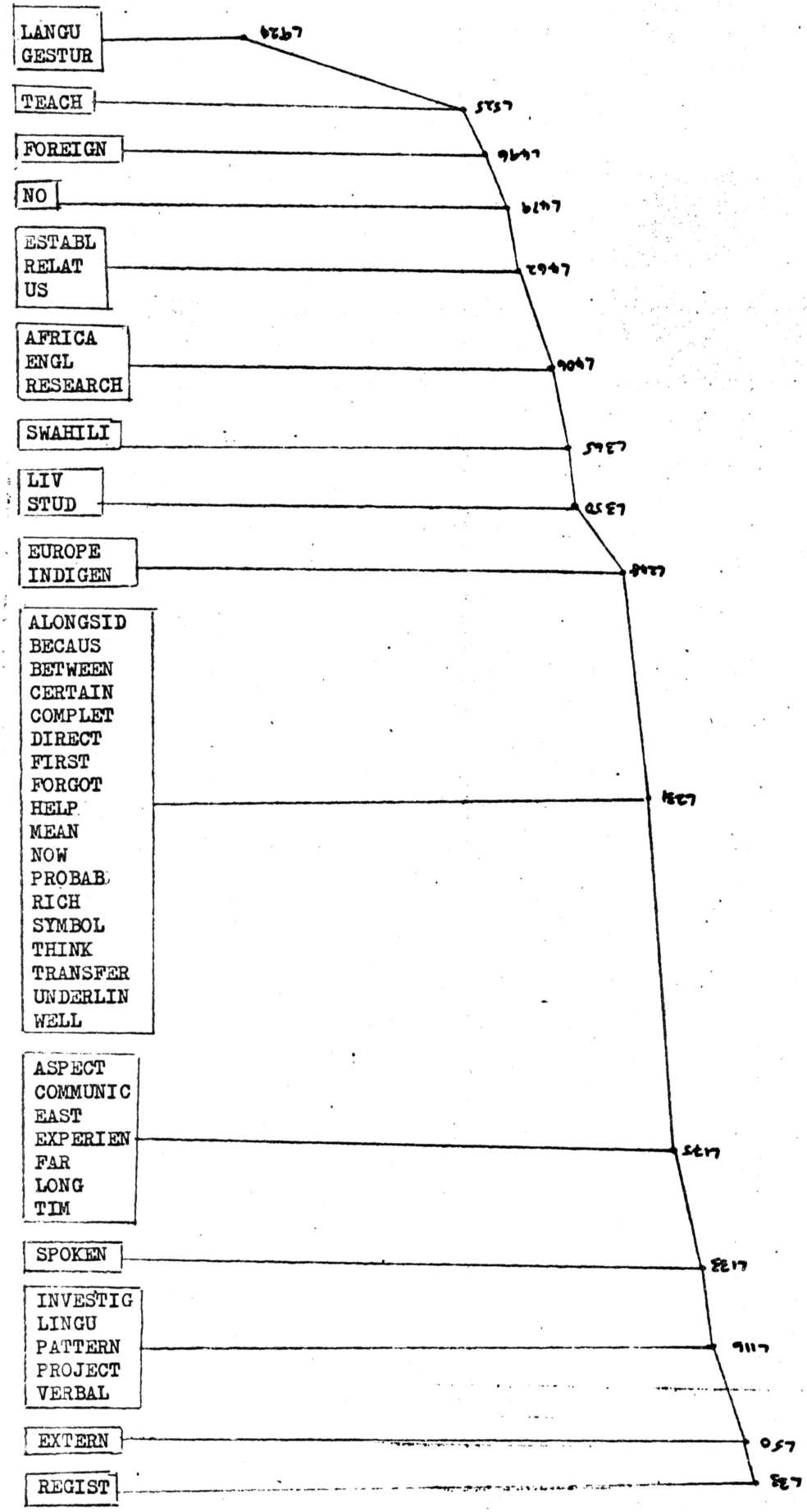
IO.


FISHER
GREAT
INLAND
SCHEM
INTERGR
POSS
SYSTEM
WORLD
FISH
WAFT
FED
WATER
ECONOM
IRRIG
SETTL
AGRICULT
APPROACH
BRIEF
COMMUN
COUNTR
BECAUS
BRIT
DEAL
FARM
FRESH
MOMENT
NO
GROSS
INCREAS
MEAN
NATION
PROBAB
PRODUCT
WHOL
CHICKEN
PEOPL
PROTEIN

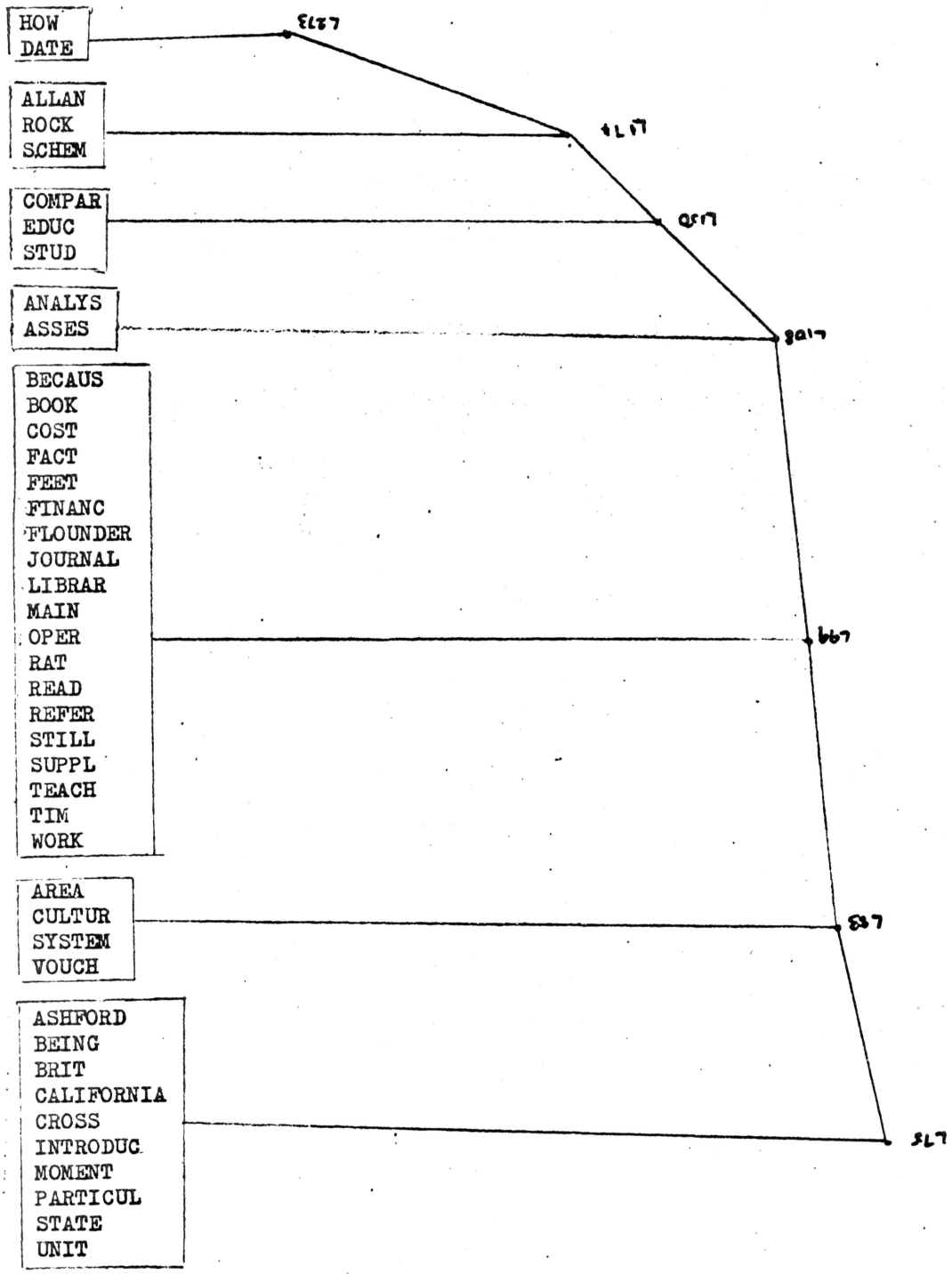
II.


GROW
FACT
PEOPL
BOOK
CERTAIN
INFLUEN
THINK
COLLEG
IDEA
TOPIC
VETERIN
WEIGHT
ACTUAL
BAS
NO
SET
WRIT
RESEARCH
INTRODUC
SUBJECT
BEHIND
EXPERT
ANIMAL
ATMOSPHER
BACK
CHAPT
DIET
FARM
FINE
FLUENT
FRIEND
GENERAL
GONAD
HORMON
IMPROV
KNOW
LECTUR
LIGHT
LONDON
NATUR
OBVIOUS
PARTICUL
POSS
PUBL
READ
SPECIM
STRESS
SUPPOS
SURE
VEGET

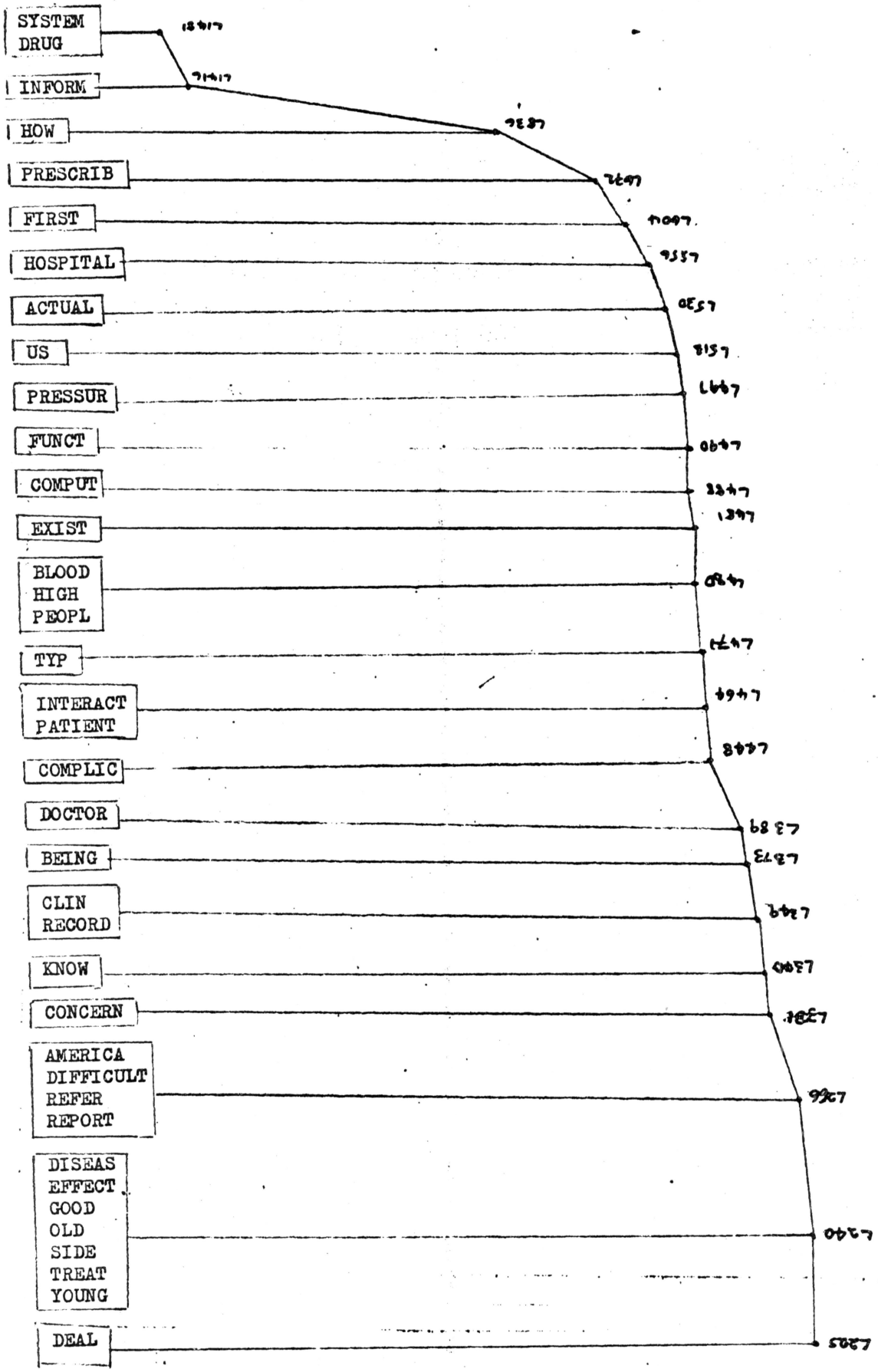
I2.


LANGU
GESTUR
TEACH
FOREIGN
NO
ESTABL
RELAT
US
AFRICA
ENGL
RESEARCH
SWAHILI
LIV
STUD
EUROPE
INDIGEN
ALONGSID
BECAUS
BETWEEN
CERTAIN
COMPLET
DIRECT
FIRST
FORGOT
HELP
MEAN
NOW
PROBAB
RICH
SYMBOL
THINK
TRANSFER
UNDERLIN
WELL
ASPECT
COMMUNIC
EAST
EXPERIEN
FAR
LONG
TIM
SPOKEN
INVESTIG
LINGU
PATTERN
PROJECT
VERBAL
EXTERN
REGIST

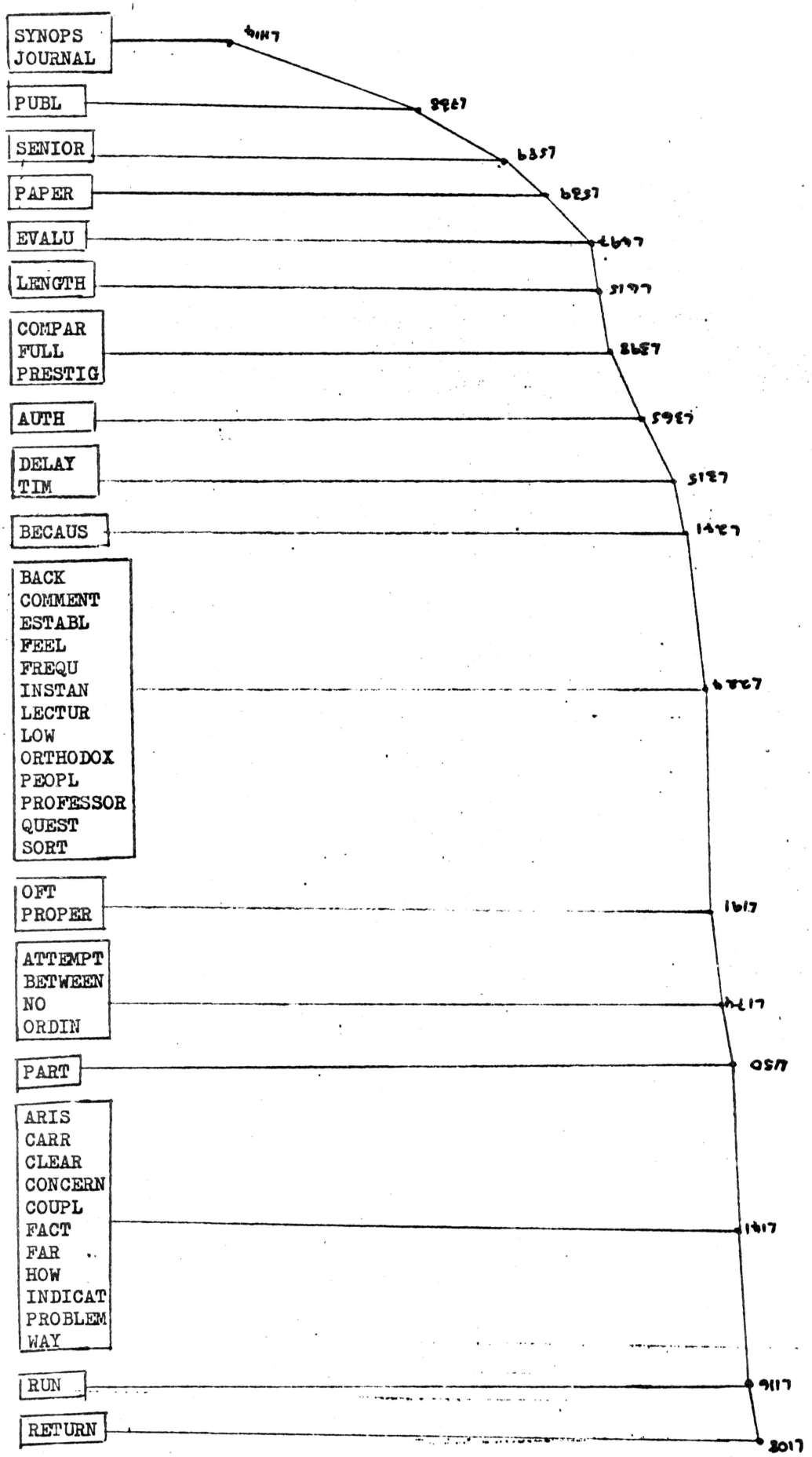
I3.


HOW
DATE
ALLAN
ROCK
SCHEM
COMPAR
EDUC
STUD
ANALYS
ASSES
BECAUS
BOOK
COST
FACT
FEET
FINANC
FLOUNDER
JOURNAL
LIBRAR
MAIN
OPER
RAT
READ
REFER
STILL
SUPPL
TEACH
TIM
WORK
AREA
CULTUR
SYSTEM
VOUCH
ASHFORD
BEING
BRIT
CALIFORNIA
CROSS
INTRODUC
MOMENT
PARTICUL
STATE
UNIT

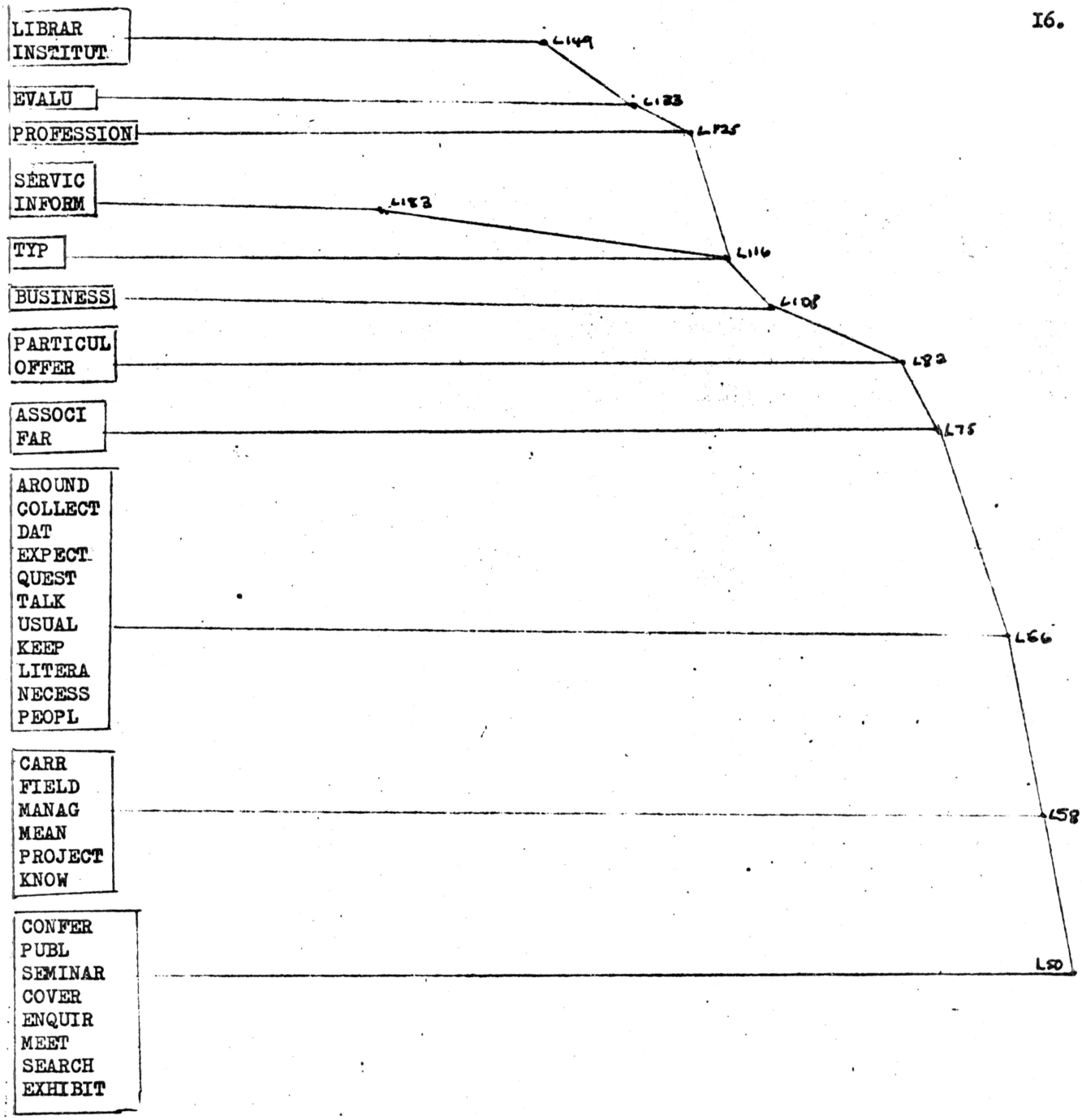
I4.


SYSTEM
DRUG
INFORM
HOW
PRESCRIB
FIRST
HOSPITAL
ACTUAL
US
PRESSUR
FUNCT
COMPUT
EXIST
BLOOD
HIGH
PEOPL
TYP
INTERACT
PATIENT
COMPLIC
DOCTOR
BEING
CLIN
RECORD
KNOW
CONCERN
AMERICA
DIFFICULT
REFER
REPORT
DISEAS
EFFECT
GOOD
OLD
SIDE
TREAT
YOUNG
DEAL

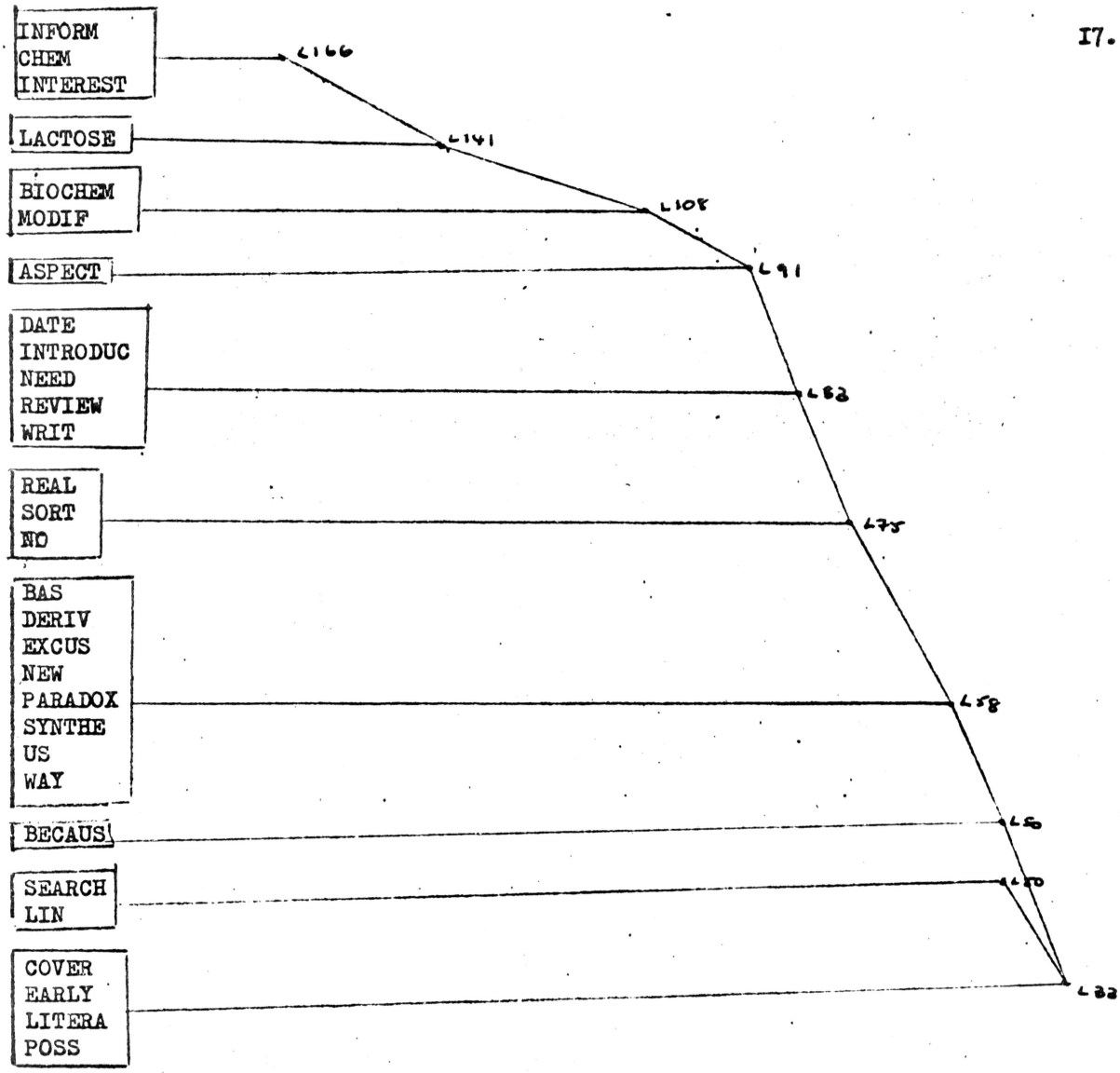
I5.

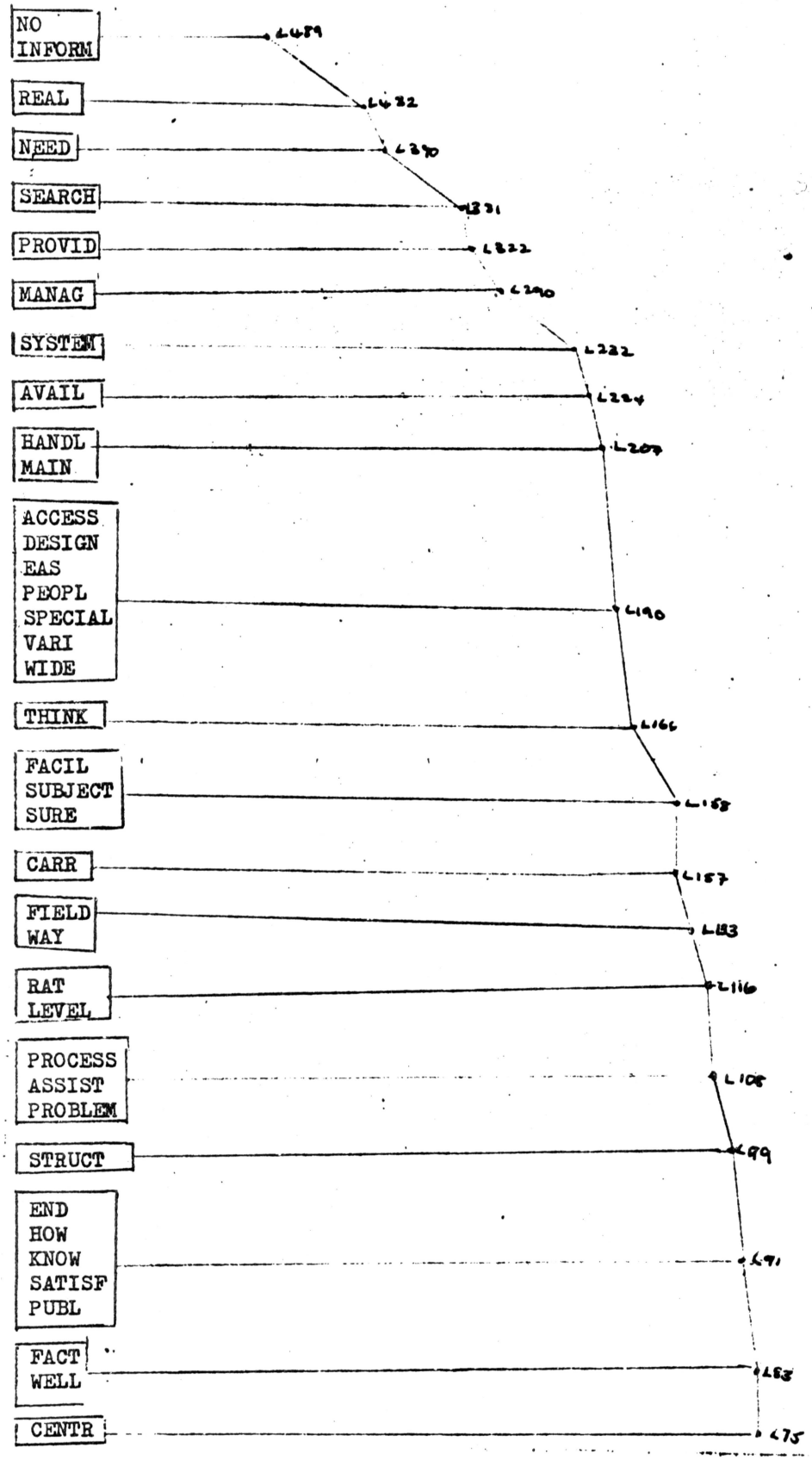

SYNOPS
JOURNAL
PUBL
SENIOR
PAPER
EVALU
LENGTH
COMPAR
FULL
PRESTIG
AUTH
DELAY
TIM
BECAUS
BACK
COMMENT
ESTABL
FEEL
FREQU
INSTAN
LECTUR
LOW
ORTHODOX
PEOPL
PROFESSOR
QUEST
SORT
OFT
PROPER
ATTEMPT
BETWEEN
NO
ORDIN
PART
ARIS
CARR
CLEAR
CONCERN
COUPL
FACT
FAR
HOW
INDICAT
PROBLEM
WAY
RUN
RETURN

I6.

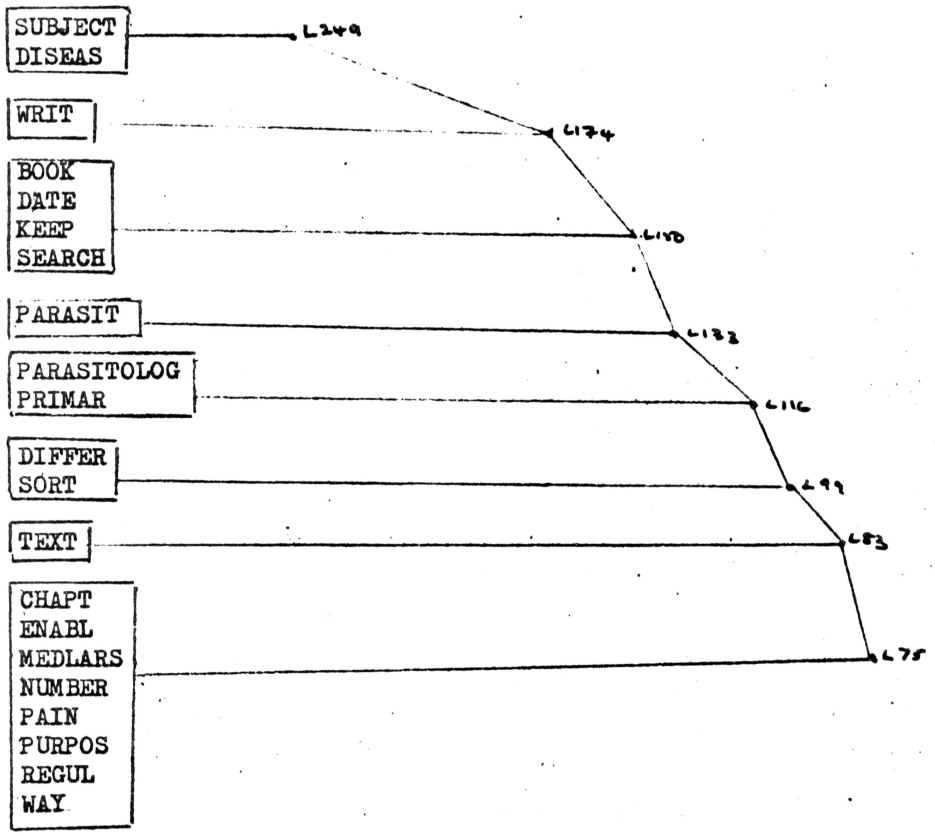

LIBRAR
INSTITUT
EVALU
PROFESSION
SERVIC
INFORM
TYP
BUSINESS
PARTICUL
OFFER
ASSOCI
FAR
AROUND
COLLECT
DAT
EXPECT
QUEST
TALK
USUAL
KEEP
LITERA
NECESS
PEOPL
CARR
FIELD
MANAG
MEAN
PROJECT
KNOW
CONFER
PUBL
SEMINAR
COVER
ENQUIR
MEET
SEARCH
EXHIBIT
L149
L133
L125
L83
L116
L108
L82
L75
L66
L58
L50

I7.

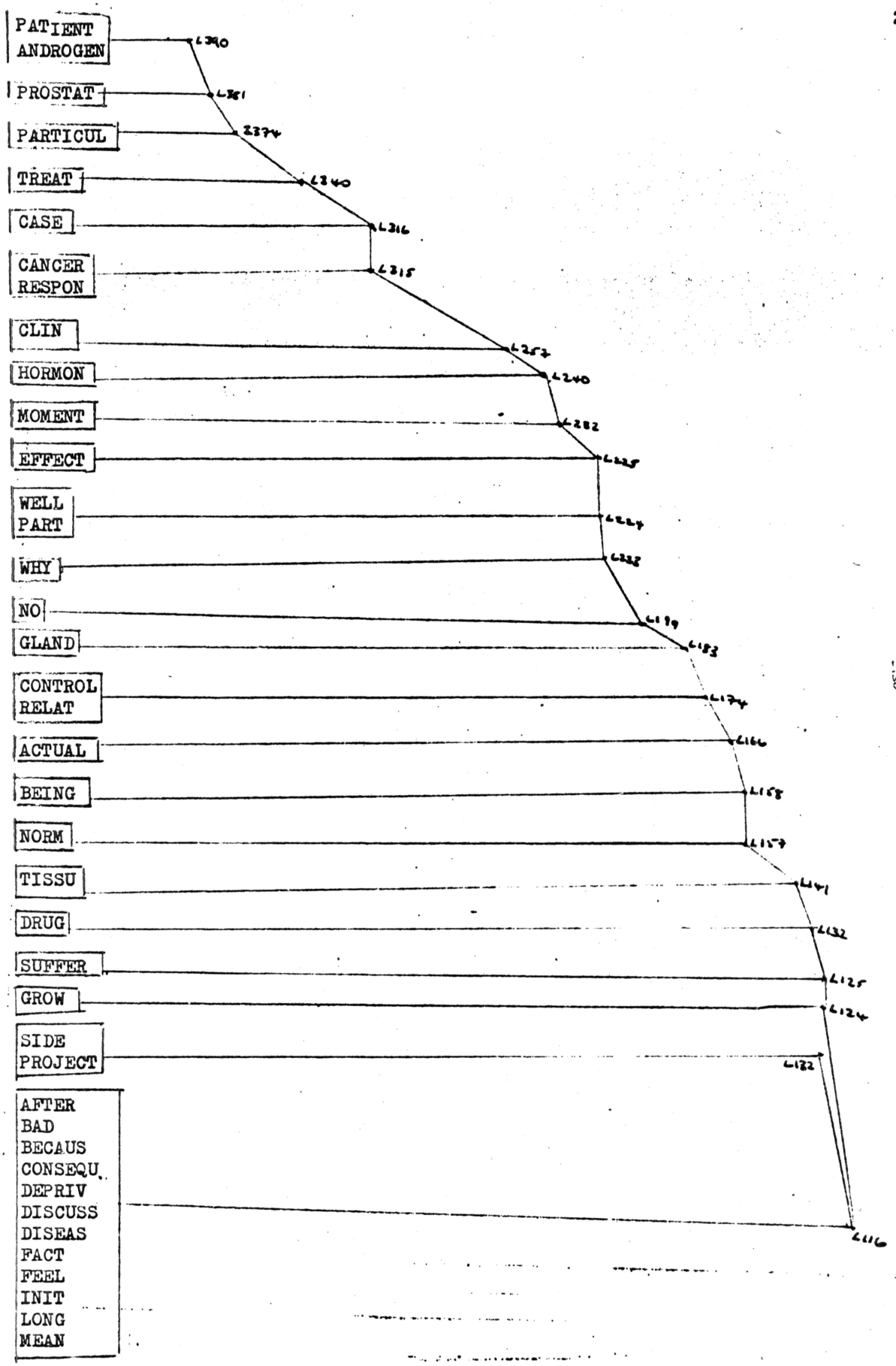

INFORM
CHEM
INTEREST
LACTOSE
BIOCHEM
MODIF
ASPECT
DATE
INTRODUC
NEED
REVIEW
WRIT
REAL
SORT
NO
BAS
DERIV
EXCUS
NEW
PARADOX
SYNTHE
US
WAY
BECAUS
SEARCH
LIN
COVER
EARLY
LITERA
POSS
166
141
108
91
82
75
58
50
40
32

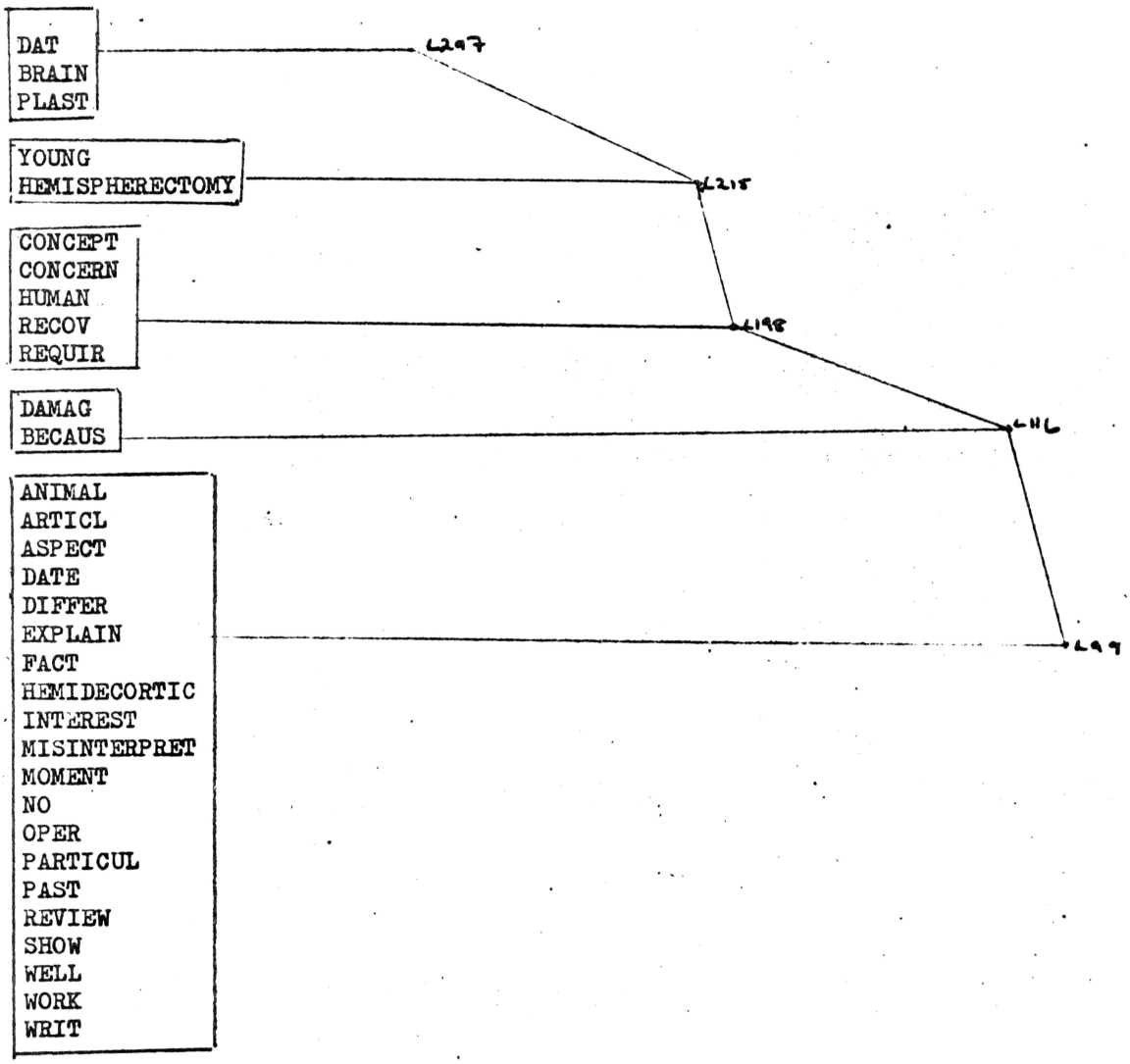

NO INFORM
REAL
NEED
SEARCH
PROVID
MANAG
SYSTEM
AVAIL
HANDL MAIN
ACCESS DESIGN EAS PEOPL SPECIAL VARI WIDE
THINK
FACIL SUBJECT SURE
CARR
FIELD WAY
RAT LEVEL
PROCESS ASSIST PROBLEM
STRUCT
END HOW KNOW SATISF PUBL
FACT WELL
CENTR

I9.


SUBJECT
DISEAS
L249
WRIT
L174
BOOK
DATE
KEEP
SEARCH
L150
PARASIT
L132
PARASITOLOG
PRIMAR
L116
DIFFER
SORT
L99
TEXT
L83
CHAPT
ENABL
MEDLARS
NUMBER
PAIN
PURPOS
REGUL
WAY
L75

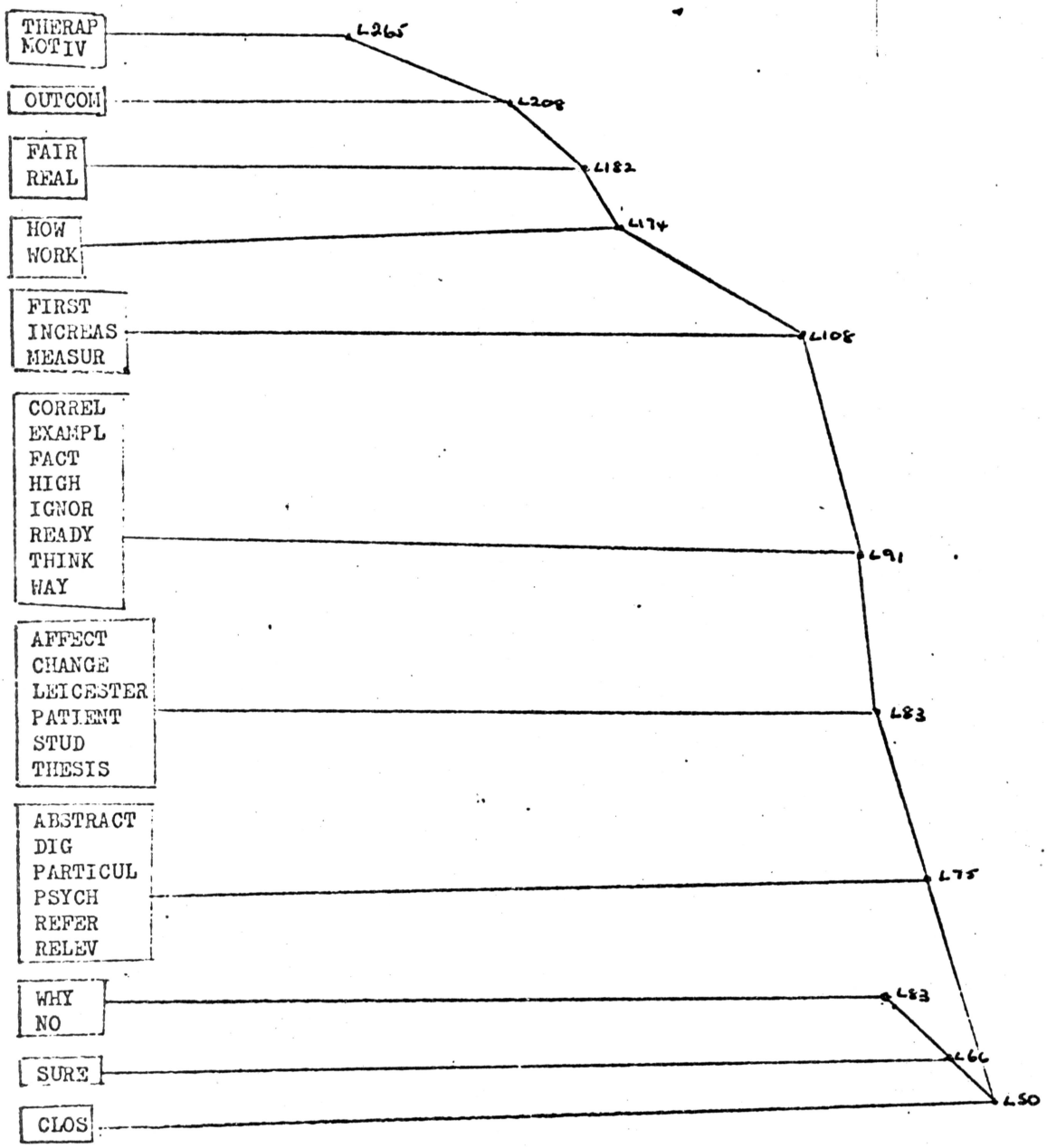
20.


PATIENT ANDROGEN
PROSTAT
PARTICUL
TREAT
CASE
CANCER RESPON
CLIN
HORMON
MOMENT
EFFECT
WELL PART
WHY
NO
GLAND
CONTROL RELAT
ACTUAL
BEING
NORM
TISSU
DRUG
SUFFER
GROW
SIDE PROJECT
AFTER
BAD
BECAUS
CONSEQU
DEPRIV
DISCUSS
DISEAS
FACT
FEEL
INIT
LONG
MEAN

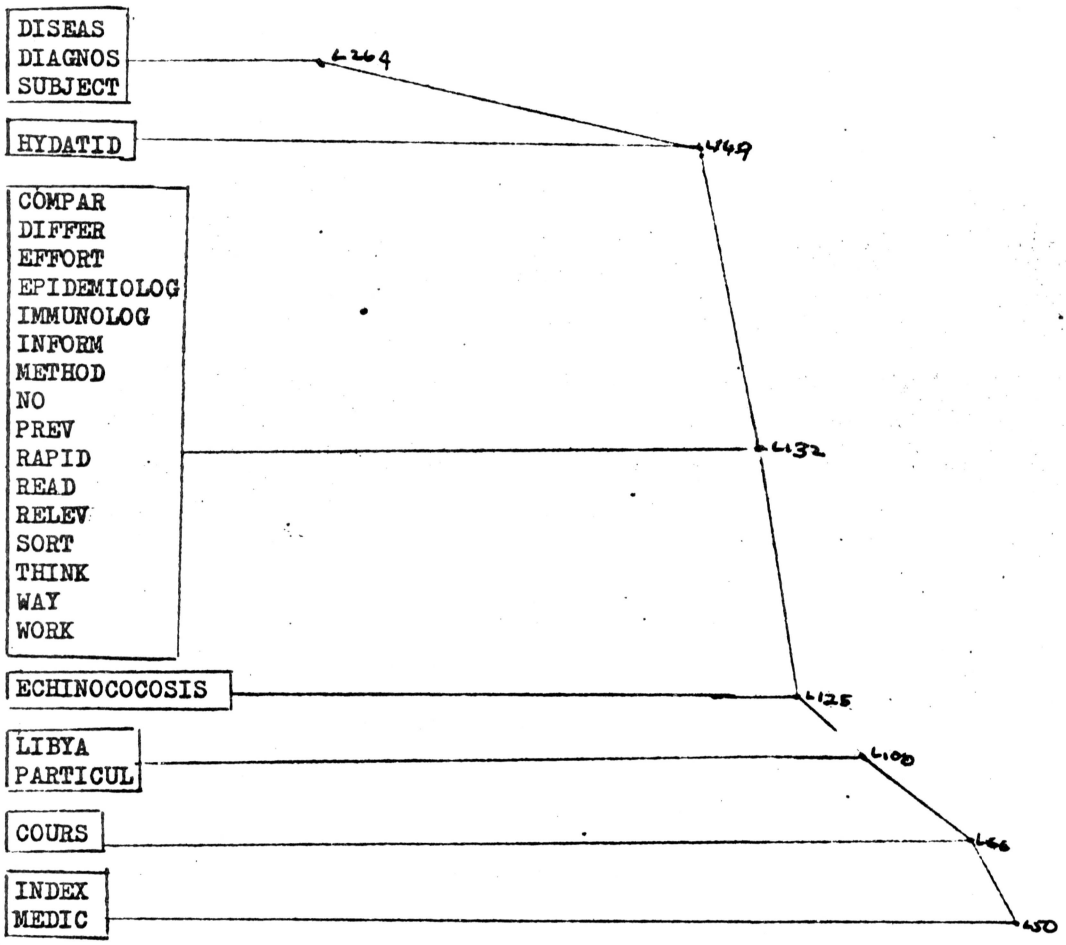

DAT
BRAIN
PLAST
L297
YOUNG
HEMISPHERECTOMY
L215
CONCEPT
CONCERN
HUMAN
RECOV
REQUIR
L198
DAMAG
BECAUS
L116
ANIMAL
ARTICL
ASPECT
DATE
DIFFER
EXPLAIN
FACT
HEMIDECORTIC
INTEREST
MISINTERPRET
MOMENT
NO
OPER
PARTICUL
PAST
REVIEW
SHOW
WELL
WORK
WRIT
L99

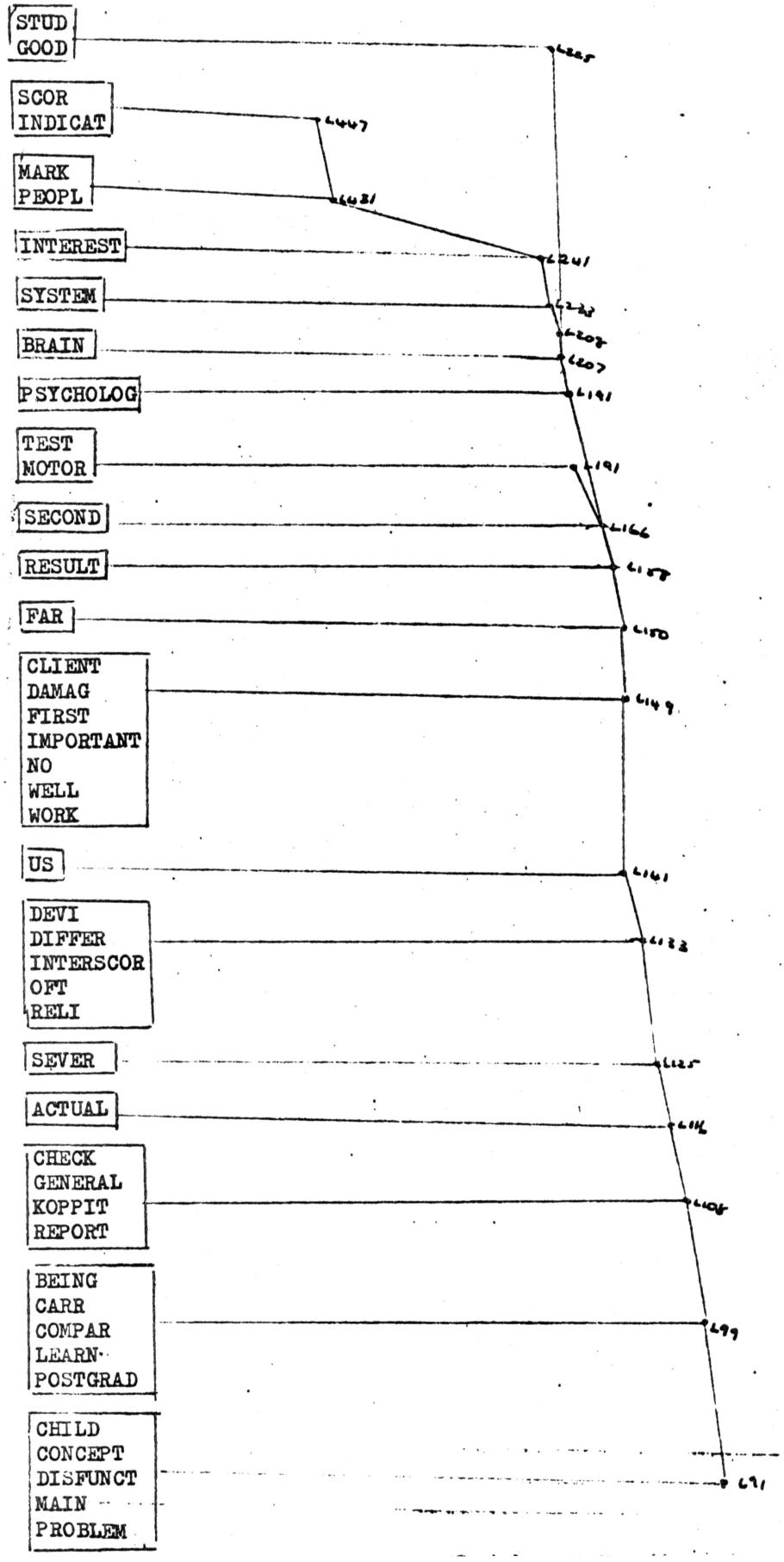
21.

22.

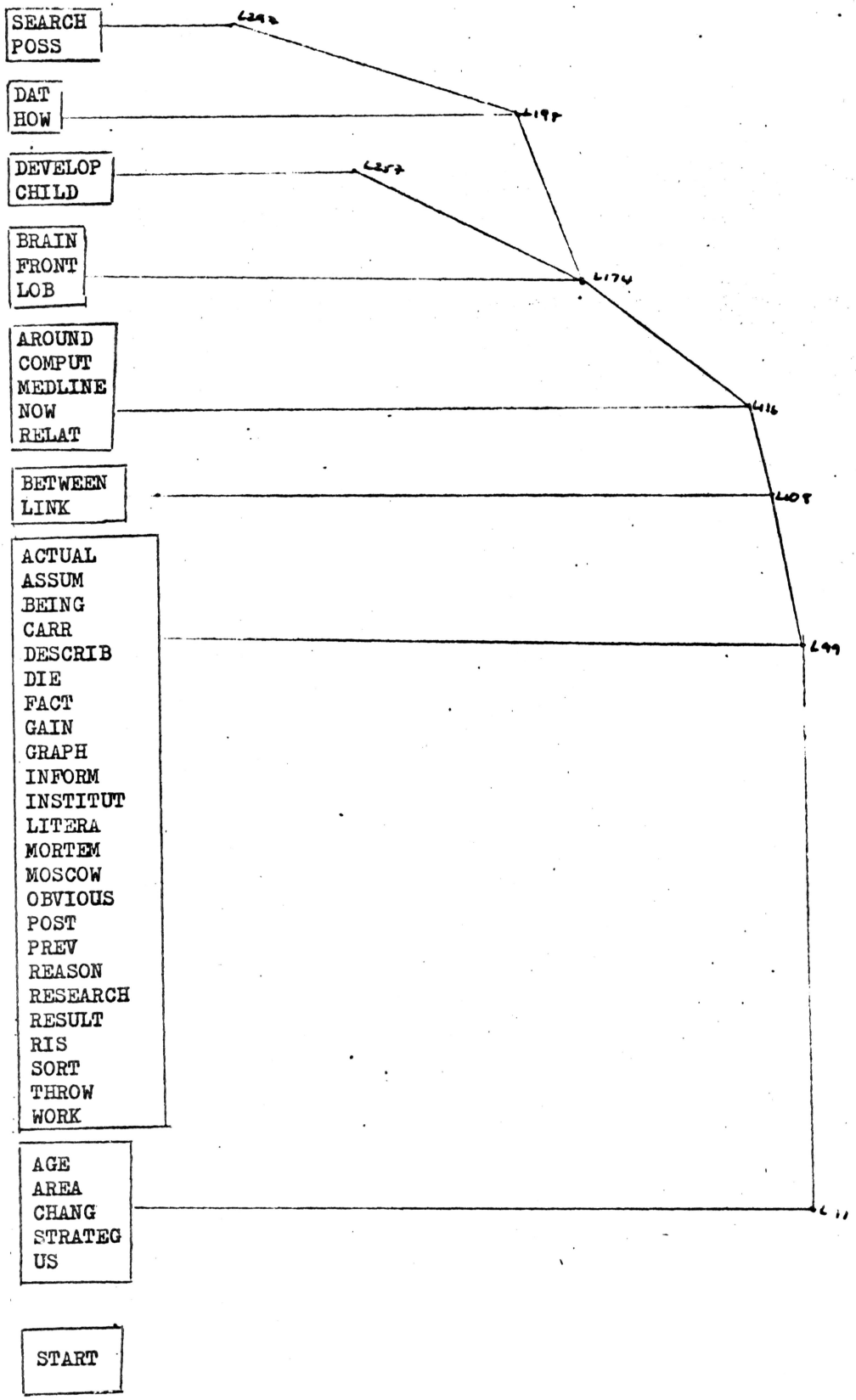

THERAP
MOTIV
OUTCOM
FAIR
REAL
HOW
WORK
FIRST
INCREAS
MEASUR
CORREL
EXAMPL
FACT
HIGH
IGNOR
READY
THINK
WAY
AFFECT
CHANGE
LEICESTER
PATIENT
STUD
THESIS
ABSTRACT
DIG
PARTICUL
PSYCH
REFER
RELEV
WHY
NO
SURE
CLOS
L265
L208
L182
L174
L108
L91
L83
L75
L83
L66
L50

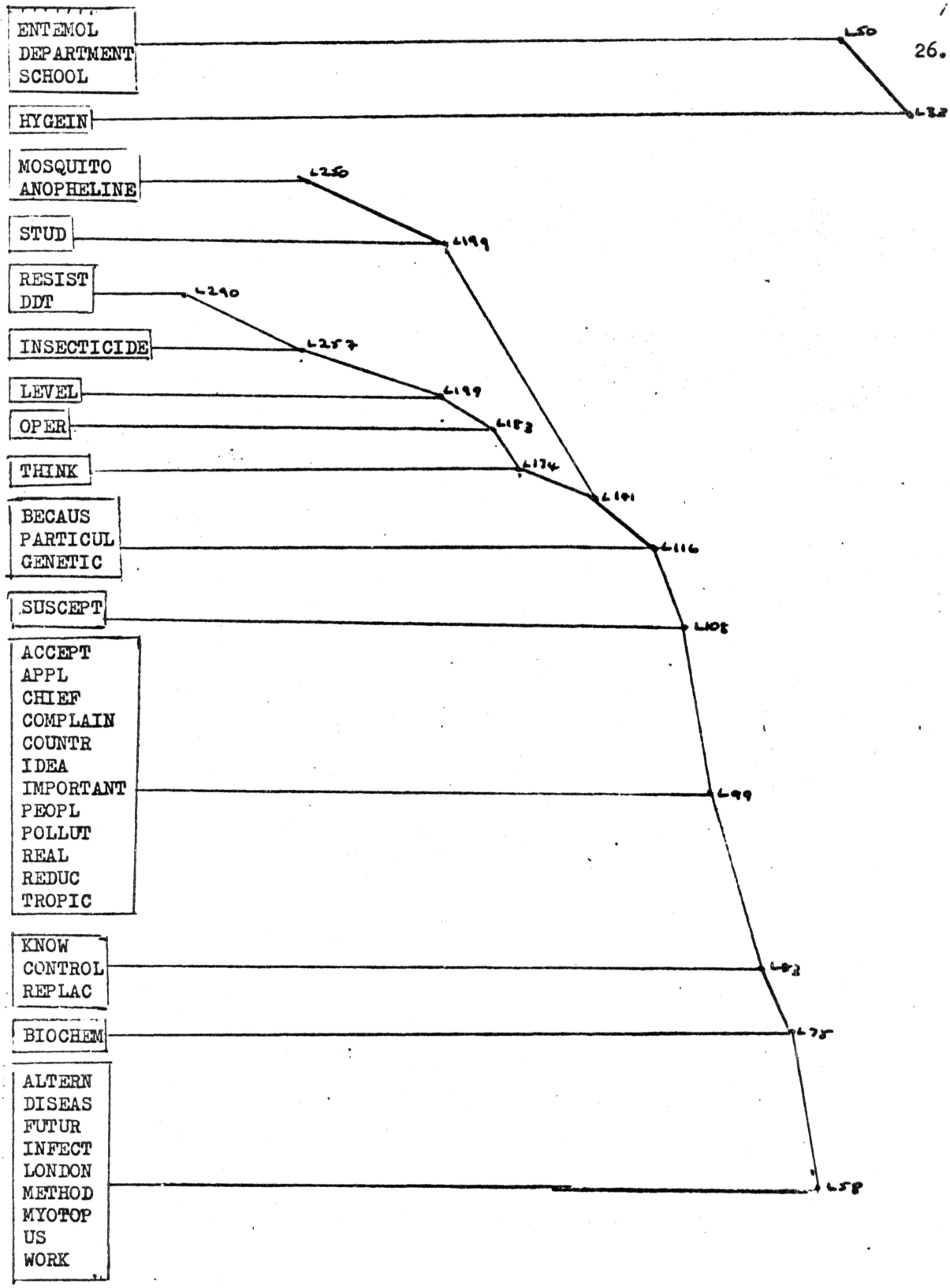
23.

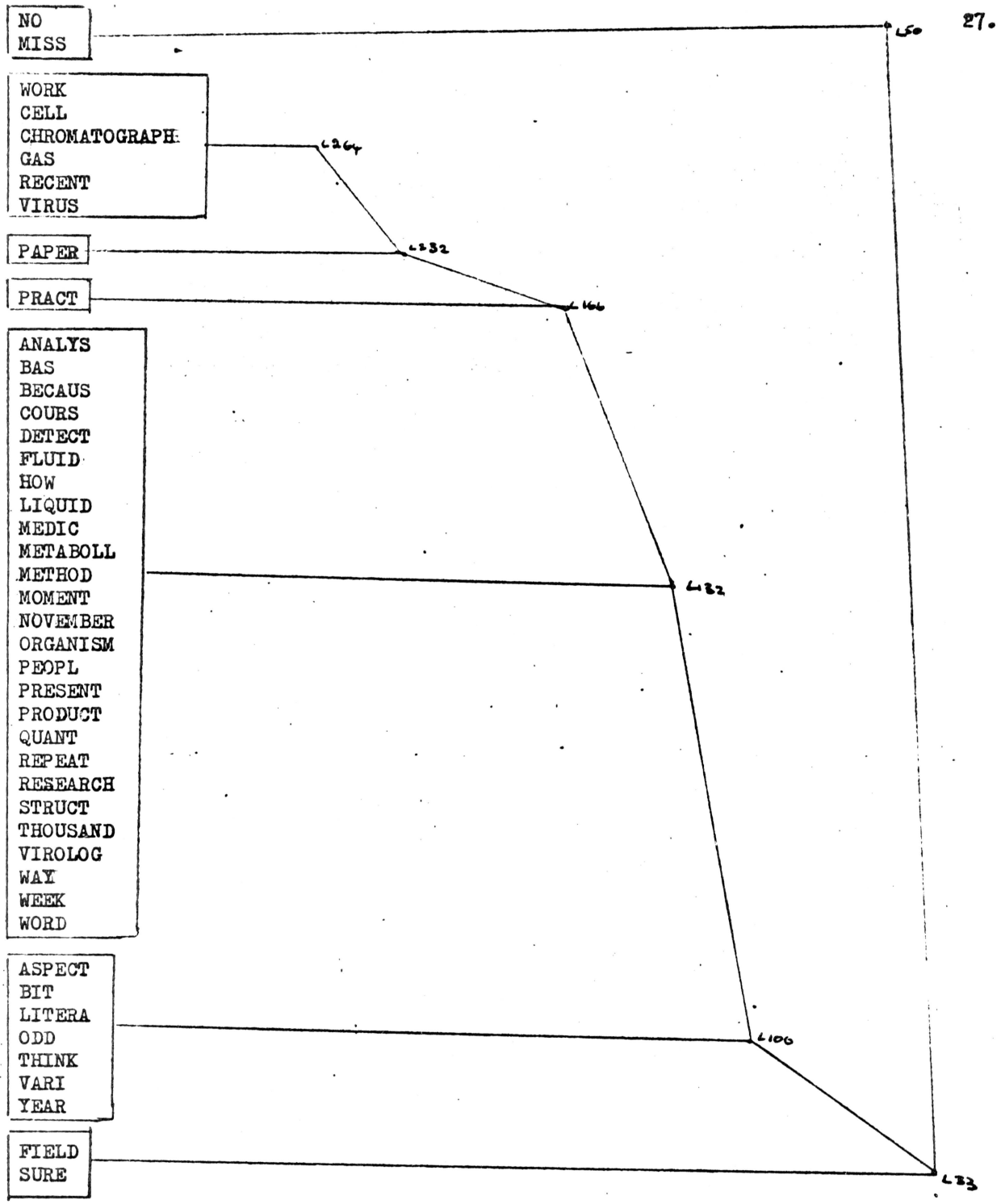

DISEAS
DIAGNOS
SUBJECT
HYDATID
COMPAR
DIFFER
EFFORT
EPIDEMIOLOG
IMMUNOLOG
INFORM
METHOD
NO
PREV
RAPID
READ
RELEV
SORT
THINK
WAY
WORK
ECHINOCOCOSIS
LIBYA
PARTICUL
COURS
INDEX
MEDIC
L264
L469
L132
L125
L100
L66
450


STUD
GOOD
SCOR
INDICAT
MARK
PEOPL
INTEREST
SYSTEM
BRAIN
PSYCHOLOG
TEST
MOTOR
SECOND
RESULT
FAR
CLIENT
DAMAG
FIRST
IMPORTANT
NO
WELL
WORK
US
DEVI
DIFFER
INTERSCOR
OFT
RELI
SEVER
ACTUAL
CHECK
GENERAL
KOPPIT
REPORT
BEING
CARR
COMPAR
LEARN
POSTGRAD
CHILD
CONCEPT
DISFUNCT
MAIN
PROBLEM

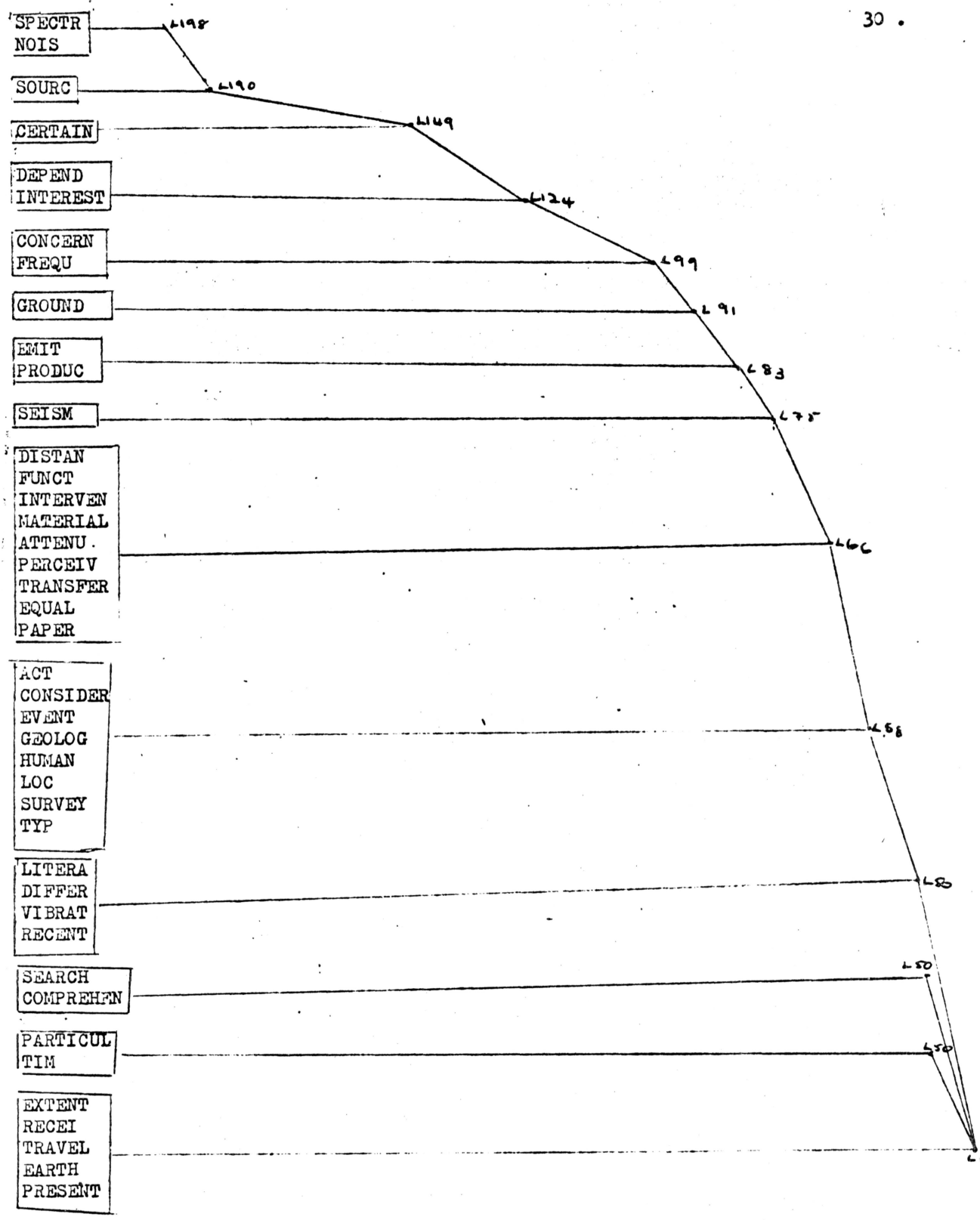
25.

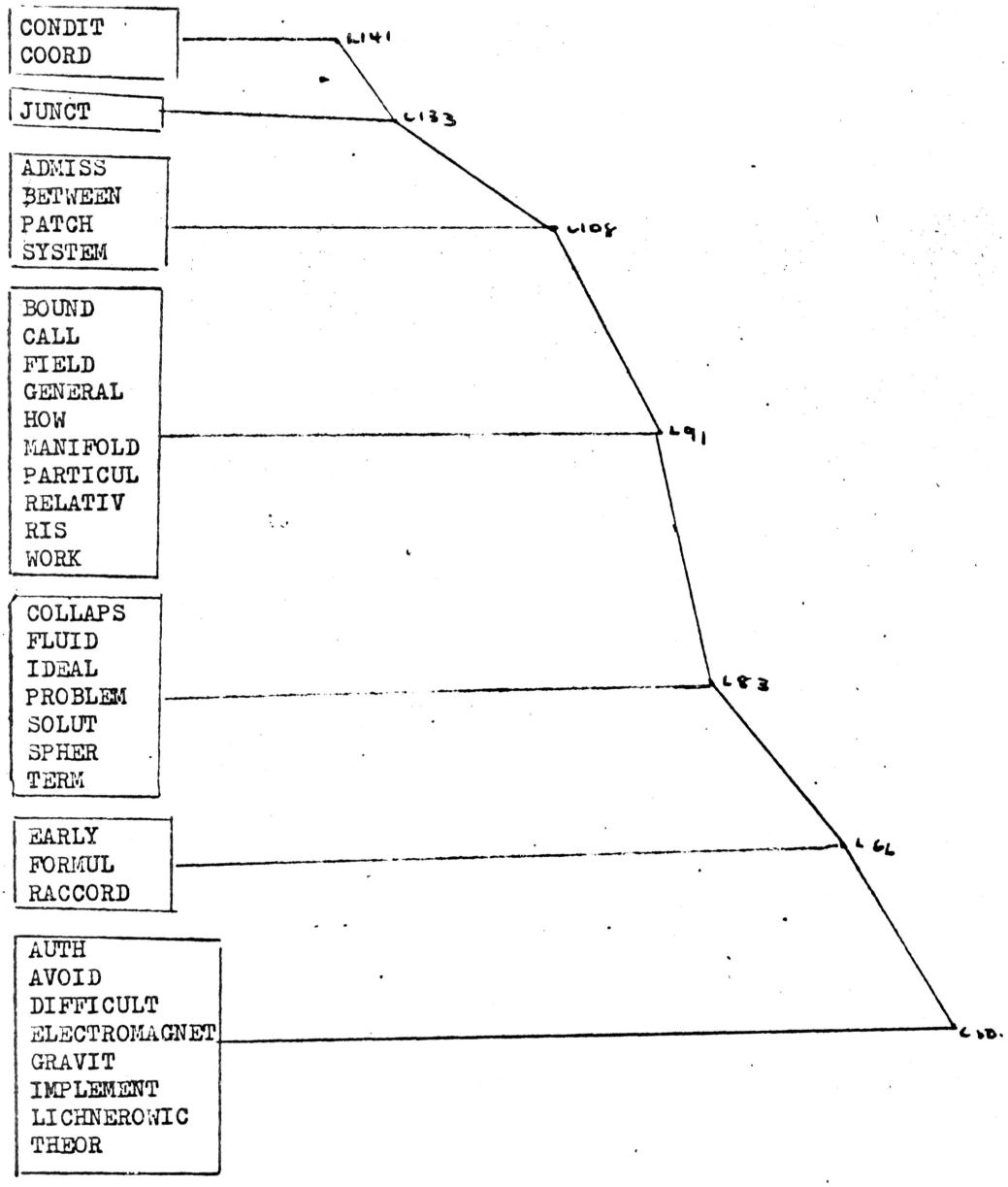

SEARCH
POSS
DAT
HOW
DEVELOP
CHILD
BRAIN
FRONT
LOB
AROUND
COMPUT
MEDLINE
NOW
RELAT
BETWEEN
LINK
ACTUAL
ASSUM
BEING
CARR
DESCRIB
DIE
FACT
GAIN
GRAPH
INFORM
INSTITUT
LITERA
MORTEM
MOSCOW
OBVIOUS
POST
PREV
REASON
RESEARCH
RESULT
RIS
SORT
THROW
WORK
AGE
AREA
CHANG
STRATEG
US
START


ENTEMOL
DEPARTMENT
SCHOOL
L50
26.
HYGEIN
L32
MOSQUITO
ANOPHELINE
L250
STUD
L199
RESIST
DDT
L290
INSECTICIDE
L257
LEVEL
L199
OPER
L183
THINK
L174
L191
BECAUS
PARTICUL
GENETIC
L116
SUSCEPT
L108
ACCEPT
APPL
CHIEF
COMPLAIN
COUNTR
IDEA
IMPORTANT
PEOPL
POLLUT
REAL
REDUC
TROPIC
L99
KNOW
CONTROL
REPLAC
L82
BIOCHEM
L75
ALTERN
DISEAS
FUTUR
INFECT
LONDON
METHOD
MYOTOP
US
WORK
L58

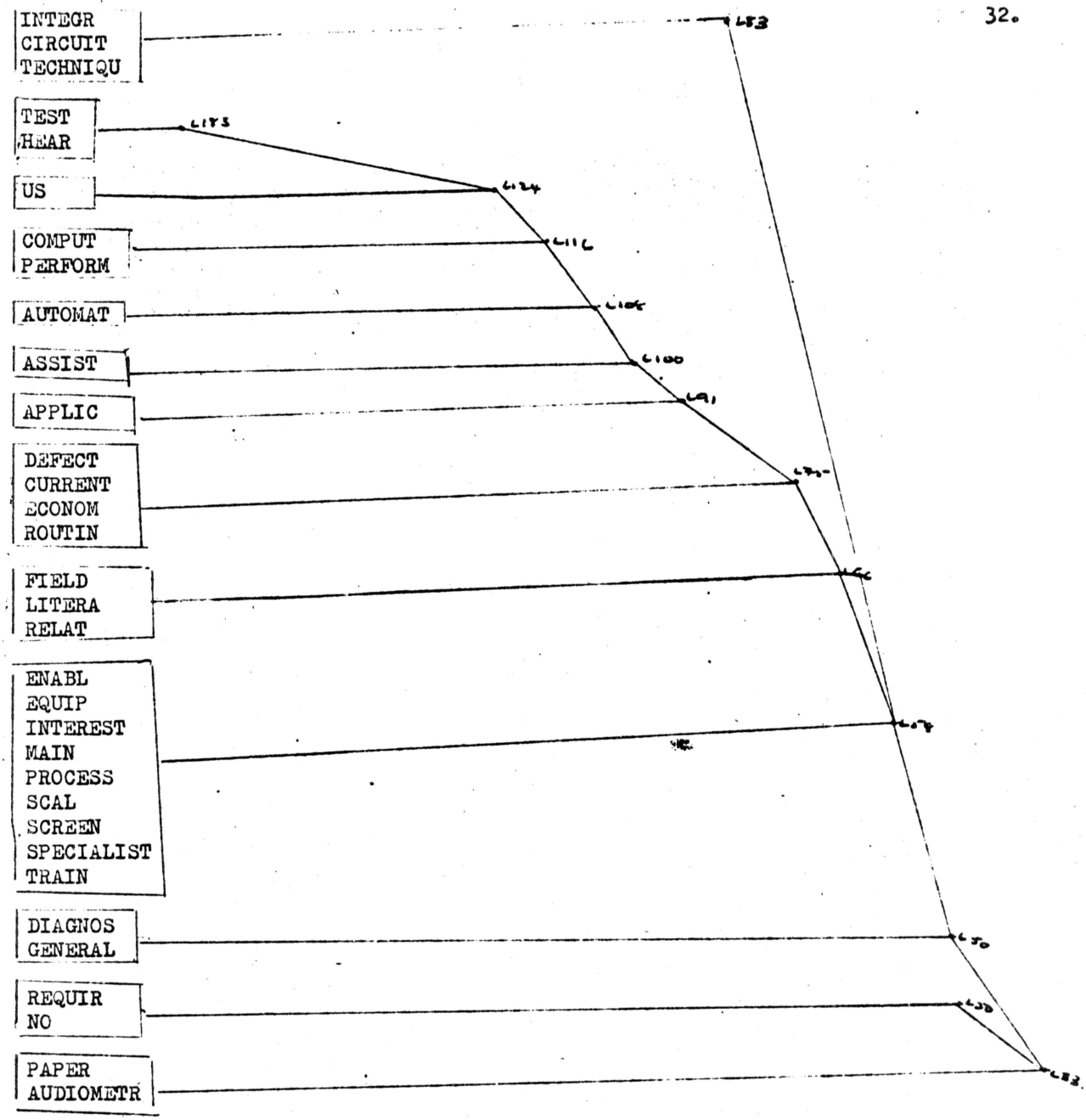

NO
MISS
WORK
CELL
CHROMATOGRAPH
GAS
RECENT
VIRUS
PAPER
PRACT
ANALYS
BAS
BECAUS
COURS
DETECT
FLUID
HOW
LIQUID
MEDIC
METABOLL
METHOD
MOMENT
NOVEMBER
ORGANISM
PEOPL
PRESENT
PRODUCT
QUANT
REPEAT
RESEARCH
STRUCT
THOUSAND
VIROLOG
WAY
WEEK
WORD
ASPECT
BIT
LITERA
ODD
THINK
VARI
YEAR
FIELD
SURE
L50
L264
L232
L166
L132
L100
L33

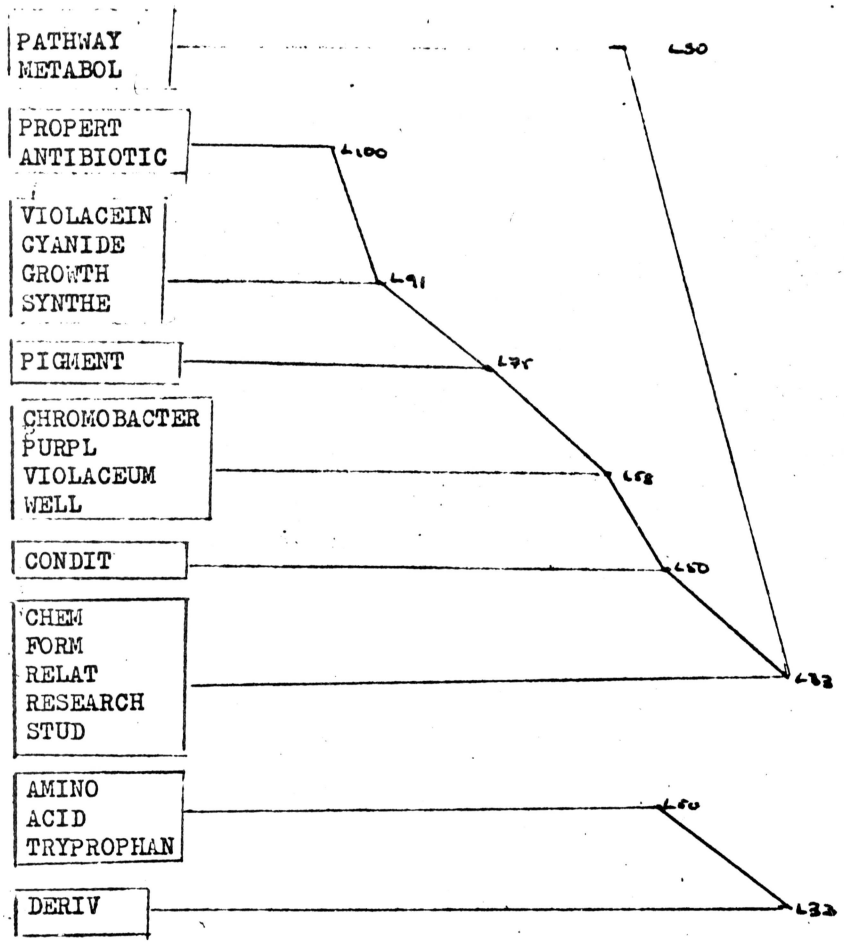
27.


30 .
SPECTR
NOIS
SOURC
CERTAIN
DEPEND
INTEREST
CONCERN
FREQU
GROUND
EMIT
PRODUC
SEISM
DISTAN
FUNCT
INTERVEN
MATERIAL
ATTENU.
PERCEIV
TRANSFER
EQUAL
PAPER
ACT
CONSIDER
EVENT
GEOLOG
HUMAN
LOC
SURVEY
TYP
LITERA
DIFFER
VIBRAT
RECENT
SEARCH
COMPREHEN
PARTICUL
TIM
EXTENT
RECEI
TRAVEL
EARTH
PRESENT


CONDIT
COORD
JUNCT
ADMISS
BETWEEN
PATCH
SYSTEM
BOUND
CALL
FIELD
GENERAL
HOW
MANIFOLD
PARTICUL
RELATIV
RIS
WORK
COLLAPS
FLUID
IDEAL
PROBLEM
SOLUT
SPHER
TERM
EARLY
FORMUL
RACCORD
AUTH
AVOID
DIFFICULT
ELECTROMAGNET
GRAVIT
IMPLEMENT
LICHNEROWIC
THEOR

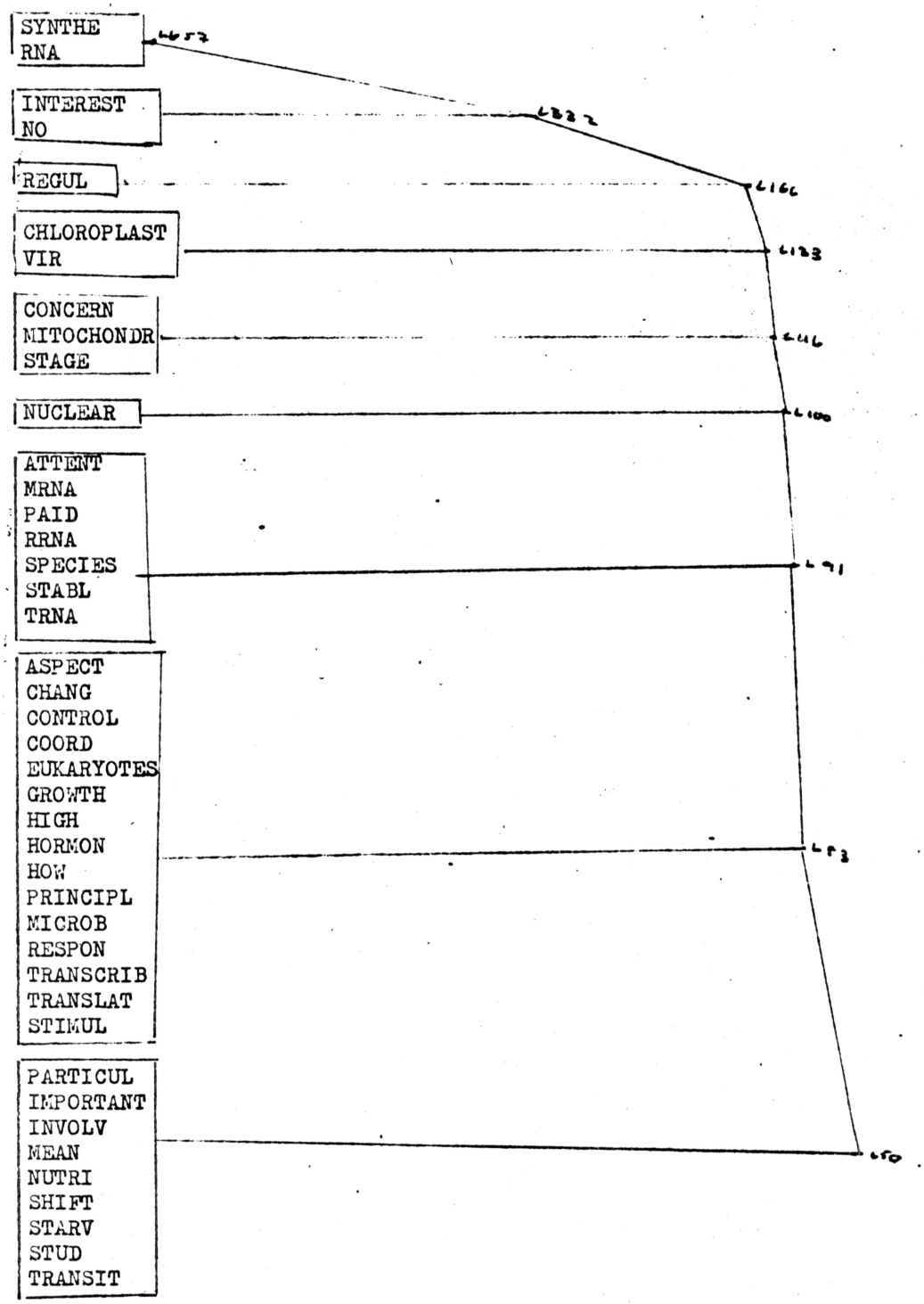
32.

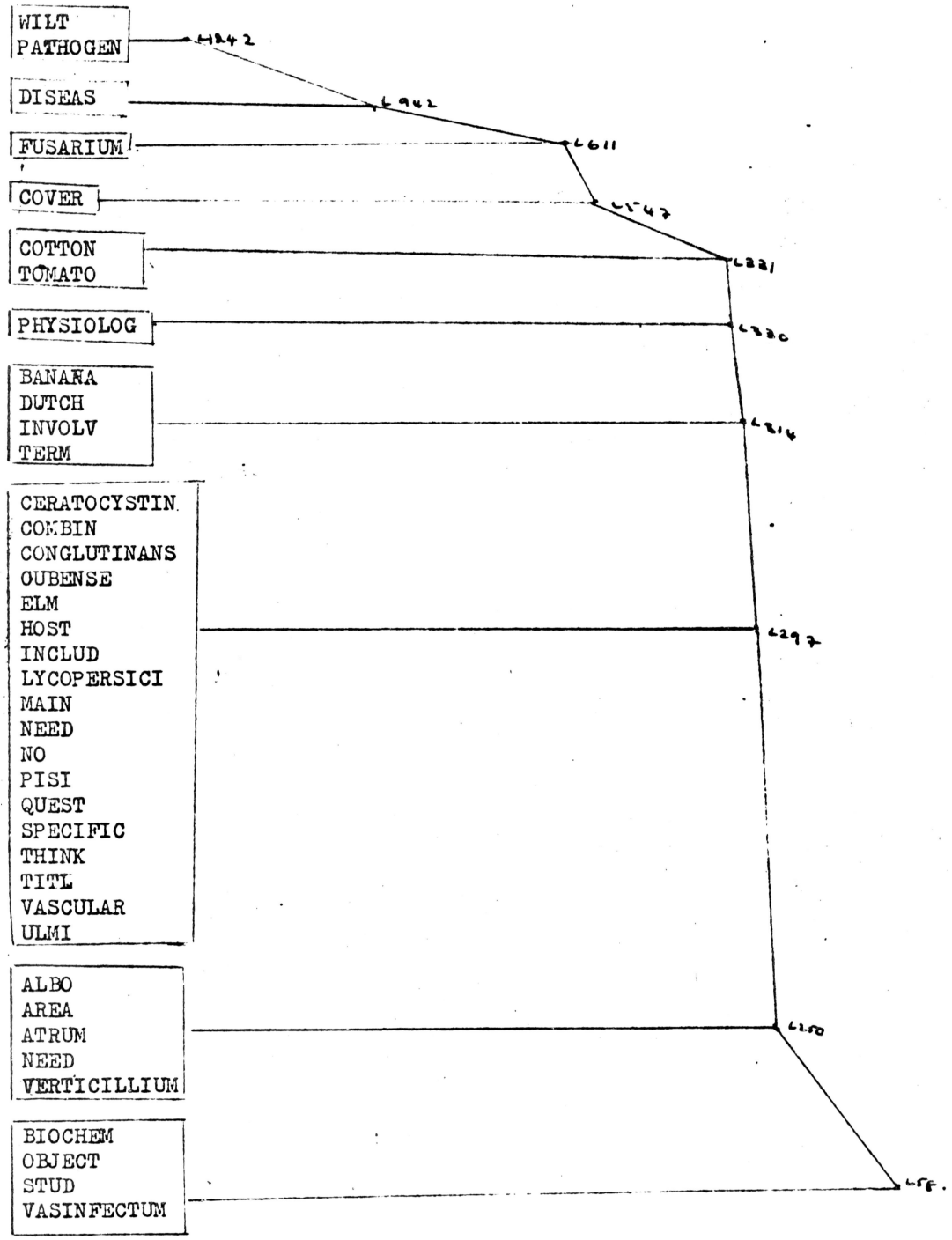

INTEGR
CIRCUIT
TECHNIQU
TEST
HEAR
US
COMPUT
PERFORM
AUTOMAT
ASSIST
APPLIC
DEFECT
CURRENT
ECONOM
ROUTIN
FIELD
LITERA
RELAT
ENABL
EQUIP
INTEREST
MAIN
PROCESS
SCAL
SCREEN
SPECIALIST
TRAIN
DIAGNOS
GENERAL
REQUIR
NO
PAPER
AUDIOMETR


PATHWAY
METABOL
PROPERT
ANTIBIOTIC
VIOLACEIN
CYANIDE
GROWTH
SYNTHE
PIGMENT
CHROMOBACTER
PURPL
VIOLACEUM
WELL
CONDIT
CHEM
FORM
RELAT
RESEARCH
STUD
AMINO
ACID
TRYPROPHAN
DERIV

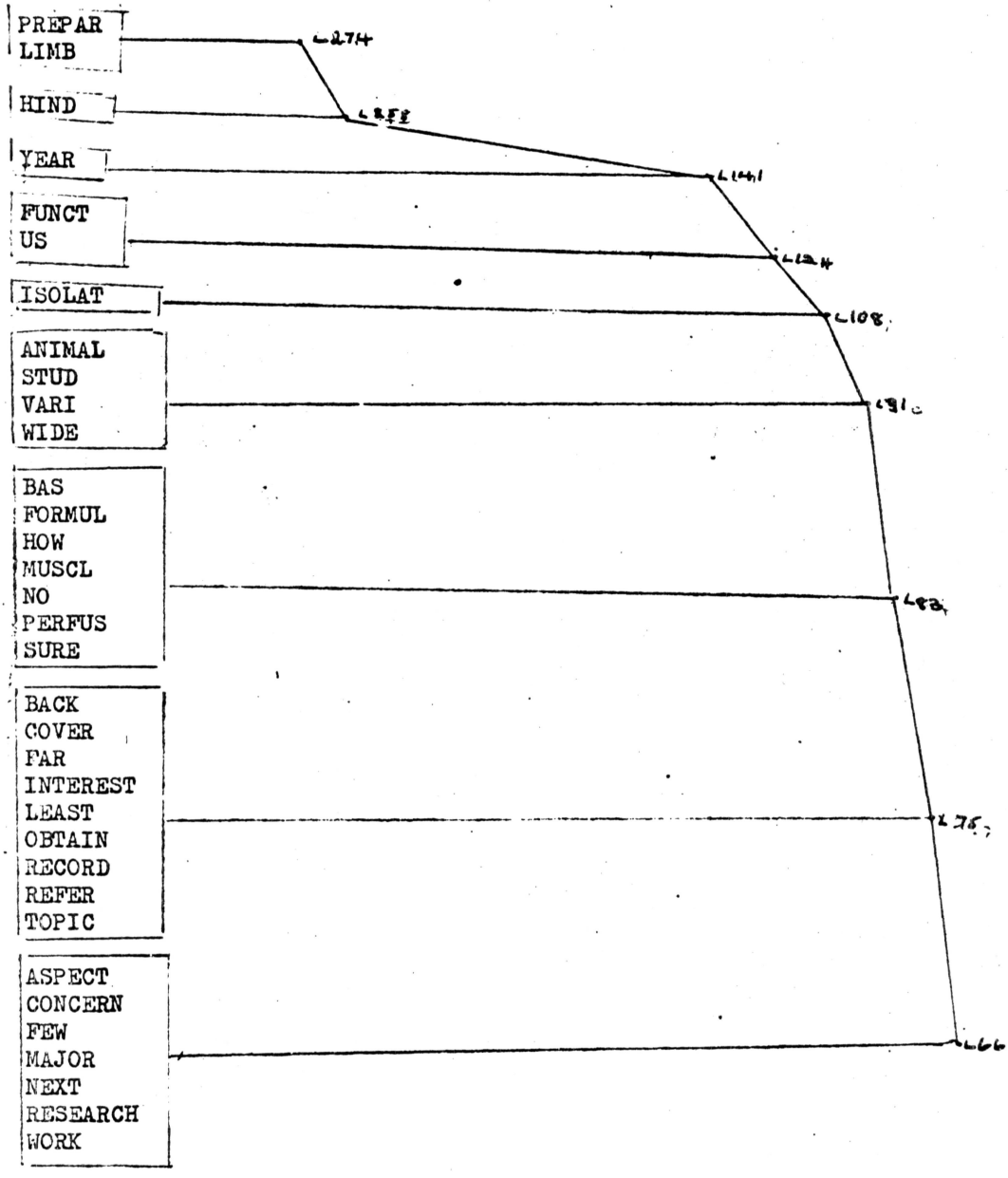
/ 33.

/34.

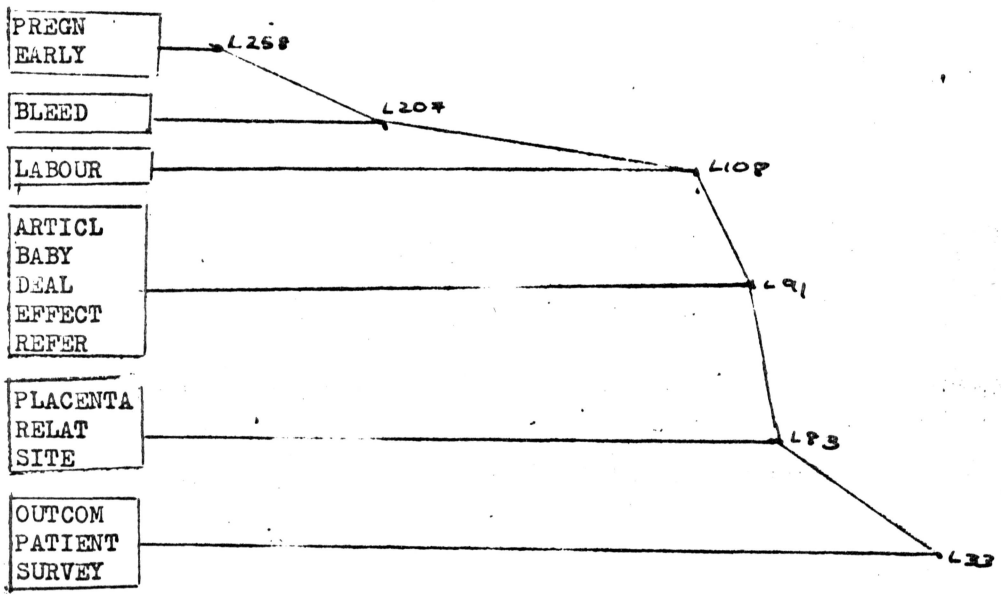

SYNTHE
RNA
INTEREST
NO
REGUL
CHLOROPLAST
VIR
CONCERN
MITOCHONDR
STAGE
NUCLEAR
ATTENT
MRNA
PAID
RRNA
SPECIES
STABL
TRNA
ASPECT
CHANG
CONTROL
COORD
EUKARYOTES
GROWTH
HIGH
HORMON
HOW
PRINCIPL
MICROB
RESPON
TRANSCRIB
TRANSLAT
STIMUL
PARTICUL
IMPORTANT
INVOLV
MEAN
NUTRI
SHIFT
STARV
STUD
TRANSIT


WILT
PATHOGEN
DISEAS
FUSARIUM
COVER
COTTON
TOMATO
PHYSIOLOG
BANANA
DUTCH
INVOLV
TERM
CERATOCYSTIN
COMBIN
CONGLUTINANS
OUBENSE
ELM
HOST
INCLUD
LYCOPERSICI
MAIN
NEED
NO
PISI
QUEST
SPECIFIC
THINK
TITL
VASCULAR
ULMI
ALBO
AREA
ATRUM
NEED
VERTICILLIUM
BIOCHEM
OBJECT
STUD
VASINFECTUM

36.


PREPAR
LIMB
HIND
YEAR
FUNCT
US
ISOLAT
ANIMAL
STUD
VARI
WIDE
BAS
FORMUL
HOW
MUSCL
NO
PERFUS
SURE
BACK
COVER
FAR
INTEREST
LEAST
OBTAIN
RECORD
REFER
TOPIC
ASPECT
CONCERN
FEW
MAJOR
NEXT
RESEARCH
WORK

/37


PREGN
EARLY
BLEED
LABOUR
ARTICL
BABY
DEAL
EFFECT
REFER
PLACENTA
RELAT
SITE
OUTCOM
PATIENT
SURVEY
L268
L204
L108
L91
L83
L33

APPENDIX F.

Abstracts

1

Clustering Large Files of Documents Using the Single-Link Method

A method for clustering large files of documents using a clustering algorithm which takes O(n 2) operations (single-link) is proposed. This method is tested on a file of 11,613 documents derived from an operational system. One property of the generated cluster hierarchy (hierarchy connection percentage) is examined and it indicates that the hierarchy is similar to those from other test collections. A comparison of clustering times with other methods shows that large files can be clustered by single-link in a time at least comparable to various heuristic algorithms which theoretically require fewer operations.

2
Theory of the Bradford Law

The Bradford law is explored theoretically by means of a very mixed Poisson model which, it is claimed, elucidates the uncertainties surrounding the law and its applications. It is argued that Bradford succeeded in formulating an empirical regularity which has pure and hybrid forms but that all the variants can be subsumed under a simple logarithmic law which, for reasons explained, escapes exact expression in conventional frequency terms. The theoretical aspects discussed include the hybridity of form, estimations, sampling problems, the stability of ranks, homogeneity of data, and tests of significance. Some numerical examples, some simulated and some drawn from social contexts outside bibliography, are used both to illustrate theoretical issues and also to indicate the wide generality of the Bradford law. Possible applications and developments of the theory are indicated.

3

Post-coordinate indexing and the prescribed classification and cataloguing curriculum in ALA-accredited library schools:
- national case study together with proposals of general applicability

The introduction of post-coordinate indexing within the mandatory education in subject headings, classification and cataloguing at ALA-accredited library schools was slow to follow the adoption of the technique in the library and information field. Surveys conducted in 1956, 1961 and 1966 produced no positive evidence of the teaching of the method. A recent study has found that rather less than half the schools were teaching coordinate indexing and most of these granted it only slight attention. Those schools that covered coordinate indexing were far more likely than the average school to include UDC in the classification part of the curriculum. Schools which featured modern integrated courses of wide scope nearly all taught coordinate indexing. There are strong vocational and academic arguments to warrant the incorporation of the model in all required instruction in concept and term analysis and special librarians can work towards this end through curricular consultation with the schools.

4

The Probability Ranking Principle in IR

The principle that for optimal retrieval, documents should be ranked in order of the probability of relevance or usefulness has been brought into question by Cooper. It is shown that the principle can be justified under certain assumptions, but that in cases where these assumptions do not hold, the principle is not valid. The major problem appears to lie in the way the principle considers each document independently of the rest. The nature of the information on the basis of which the system decides whether or not to retrieve the documents determines whether the document by document approach is valid.

5

Information Science and the Phenomenon of Information

This paper aims to deduce the fundamental phenomena of information science, starting from two premises: that information science is a problem-oriented discipline concerned with the effective transfer of desired information from human generator to human user, and that the single notion common to all concepts of information now extant is that of change of structure.

From these premises, a spectrum of information concepts is derived, and a partition of that spectrum particular to the purposes of information science is described. From this partition, the terms text and information (both in information science) are defined, and the fundamental phenomena of information science are deduced: the text and its structure, the structure of the recipient and changes in that structure, and the structure of the sender and the structuring of the text.

These phenomena are seen as the basic components of the mechanisms of the channel, which have been the traditional area of interest to information science. Some implications of this approach for research in information science are discussed in this paper.

And, finally, the question of the ethics of theoretical research in information science is raised, and a restrictive condition is proposed.

6

Some Sampling Characteristics of Bibliometric Distributions

Many bibliometric distributions - for example, word occurrences in text, posting density of index terms and classification scheme subject headings, productivity of authors, and Bradford scattering of references - are positively skewed long tailed discrete distributions. A number of theoretical distributions have been hypothesised to fit these distributions (the most popular of which have been the log-normal and the Bradford-Zipf-Pareto type distributions), but none seem completely satisfactory. The actual distributions may however be postulated to represent a random sampling by items from a population. In this distribution-free model, as in actuality, increasing sample size leads to an increasing number of classes represented and an increase in both the mean and variance of the number of items per class, whereas the characteristic constant parameter, known as Yules Characteristic, is a function of these three variables. In this paper data from a number of real bibliometric distributions are presented to show the numerical value of Yules Characteristic and its change with sample size, in order to test the validity of the model. The effect of sample size on the number of classes represented is also discussed in the same light.

7

Relational Indexing Applied to the Selective Dissemination of Information

Farradane's system of relational indexing, which had been previously used in a retrospective search system with good results, was further tested as the indexing language for an experimental S.D.I. system. Sections of Metals Abstracts were used for the data base of 2,820 abstracts, and forty-three volunteer users participated in the experiment which lasted for six months. Performance was assessed by recall, precision and fallout ratios, and the 'coefficient of association' (Q value) and the product (recall by precision) were used as overall measures. The overall average performance was about 75% recall and 75% precision. A failure analysis was also carried out. The browsing strategies incorporated into the system were analysed, as were the profile structure, the distribution of performance measures and possible relationships between recall, precision and generality. Farradane's relational indexing appeared applicable to the different scientific area of the properties of metals and again gave good results with a greater depth of indexing. Some new features of the system were observed.

8

Journal Ranking and Selection: A Review in Physics

Over several decades many ranking techniques have been proposed as aids to journal selection by libraries. We review those closely related to physics and others with novel features. There are three main methods of ranking - citation analysis, use or user judgement, and size or 'productivity'. Citations offer an 'unobtrusive' quantitative measure, but not only is the absolute value of a citation in question, but also there is no consensus on a 'correct' way to choose the citing journals, nor of the ranking parameter. Citations can, however, point out anomalies and show the changing status of journals over the years. Use and user judgement also employ several alternative methods. These are in the main of limited applicability outside the specific user group in question. There is greater 'parochialism' in 'use' ranking then in 'judged value' lists, with citation lists the most international. In some cases the attempted 'quantification' of subjective judgement will be misleading. Size and productivity rankings are normally concerned with one or other formulation of the Bradford distribution. Since the distribution is not universally valid, for library use the librarian must satisfy him/herself that the collection conforms to the distribution, or that his users would be well served by one that did. This may require considerable effort, and statistics gained will then render the Bradford distribution redundant.

Rank correlations are calculated on many of the lists. Correlations between methods are generally low, although there may be a case for a study of various techniques on one journal/article collection and user group.

The lack of an agreed quantitative measure should ensure continued reliance on subjective judgement of librarian and user. The appropriate role for more apparently sophisticated techniques should be a minor, auxiliary one.

9

Communication and Information Needs of Earth Science Engineers

The paper attempts to define the information needs of earth science engineers: mining, engineering geology, soil and rock mechanics. The approach was to investigate these needs in relation to the environment in which the engineer lives and works, his organisation, team and leadership, and partly to his own traits. The roles of engineers as researchers, linkers and practitioners are defined. The information needs and sources are related to the career stages of engineers, their duties and responsibilities. In conclusions, recommendations for a viable information retrieval system are briefly discussed.

The methodology of the work was based on questionnaires, interviews, visits to organisations, study of correspondence and diary, analysis of citations and evaluation of two abstracts journals which exist in the field. The SPSS computer package was used for analysis of the results of questionnaires.

10

The Paradoxical Role of Unexamined Documents in the Evaluation of Retrieval Effectiveness

Traditional measures of retrieval effectiveness, of which the recall ratio is an outstanding example, are strongly influenced by the relevance properties of unexamined documents - documents with which the system user has no direct contact. Such an influence is awkward to explain in traditional terms, but is readily justified within the broader framework of a utility-theoretic approach. The utility-theoretic analysis shows that unexamined documents can be important in theory, but usually are not when it is the statistics of large samples that are of interest. It is concluded that the traditional concern with the relevance or nonrelevance of unexamined documents is misplaced, and that traditional measures of effectiveness should be replaced by estimates of the direct utility of the examined documents.

11

Incorporation of the Age of a Document into the Retrieval Process

A full treatment of the significance of a document for an enquirer should include a joint description of the similarity between the document and the enquiry in a linguistic sense, and the age of the document at the time of the enquiry. The basic variables are identified in terms of a signal detection model. The age variable is related to the phenomenon of obsolescence, which is treated as a perceived, signed attribute of relevant documents. Two retrieval methods that use both index terms and document age are described: one in which a set of documents, first selected by a term-intersection process, is reduced by applying a date of publication criterion (the "subset method"); and one in which a bivariate function attaches a single number to each document, and a retrieved set is defined by a single threshold value (the "bivariate weight method"). In the latter method, discriminant analysis is a useful aid. A model of the retrieval process, based on continuous variables, is described, and the effectiveness of each method is predicted, both in terms of the Precision-Recall graph and language measures. The model suggests that either method can improve retrieval performance but incorrect usage will depress it. The better choice of method will depend on the Recall/Precision mix required by the user, as well as the actual parameters of the distributions. A relationship is hypothesised between the growth rate of a data base and the underlying distributions defined by relevance judgements.

12

On the Problem of 'Aboutness' in Document Analysis

One of the most crucial problem areas of information science concerns the identification of what documents are 'about'. This paper seeks to define the notion of 'aboutness' within the context of recent work in text linguistics. It describes, first, the essential communicational structures of sentences, paragraphs and texts in terms of theme, rheme and thematic progression, connectors of clauses and sentences, and semantic progression. It then identifies the basic features of the global structures of narrative and expository texts, describes the interaction of macro and micro structure in the interpretation of texts and the role of presupposed 'states of knowledge' in both text production and text comprehension. Finally, it is argued that for the purposes of information systems the 'aboutness' of documents is to be found among the presuppositions of authors concerning the knowledge of their potential readers.

13

Information and the Explicitly Performative Verb

The view is taken that unless and until a viable Theory of Language is produced, language-understanding computer programs will remain an unattainable goal. Having rejected Transformational Generative Theory and having noted the absence of a specific theory in the traditional Computational Linguistics/Artificial Intelligence field, the author proposes the Functional Theory of Language (FTL). FTL proceeds from the realization that language is used to convey information from one person to another (not always accepted by philosophers!) A class of verbs, explicitly performative verbs, is distinguished. Use of such verbs in sentences displays unambiguously the intentions of the speaker (the information he wishes to convey). This information is carried by the presuppositions inherent in the verbs, which limit the choice of verbs to that which accurately reveals the attitudes of the speaker (his cognitive structure). Two computer programs have been written as a preliminary test for FTL, both of which programs accept as valid input only those sentences containing explicitly performative verbs. The first program detects inconsistencies in the speech of various "people"; that is to say, it tests how well understanding of English can be achieved according to FTL. For the second program, the computer is given a cognitive structure of its own - its own wishes, its own beliefs, its own emotions. It then "converses" with the user, thus testing how well English can be 'generated' according to FTL. Conclusions are drawn with relevance to both Artificial Intelligence and Informatics.

14

Retrieving References by Dialogue Rather than by Query Formulation

A substantial browsing element is very common in the interaction between research workers and the literature, because they can rarely pose perfectly defined questions. We should take account of this, not only in the way that volumes are arranged on the library shelves, but also in the tools we provide for reference retrieval.

This paper describes a computer program design which aims to satisfy incompletely defined needs through a dialogue between man and machine which does not require the man to formulate a query. The machine builds a model of the user's interest, and chooses references for display according to its state, which varies as a consequence of his reactions to the displays.

15

The Use of On-line Information Retrieval Services

As part of a British Library Research and Development Department research programme for studying the use of on-line information systems, UMIST examined the efficiency and reactions of users. Comparisons were made between search results of users, users assisted by an intermediary experienced in using the system, and the intermediary on his own after discussion with the user. Using the data bases offered through the LOCKHEED DIALOG service, it was found that users were very pleased with on-line literature searching. Over 90% of users were partially or completely satisfied with the results. While 95% of users were satisfied with their own search results, 90% preferred those of the intermediary whose search was shorter but retrieved more references on everage.

16

A Minicomputer Retrieval System with Automatic Root Finding and Roling Facilities

Since 1965, a feature card index has formed an essential component in answering technical enquiries. It now contains over 20,000 items. By 1973 the Mathatron DeskTop computer used to process the paper tape output from an automatic testing machine had become obsolete. It was decided that the replacement should handle the retrieval system input as well as meeting the general needs for scientific calculating. The machine selected was a Varian 620L: a £12,000 machine with 12-16K core and disc storage.

Although the basic notion of an inverted file has been retained, a number of novel automatic features have been incorporated. These include the reduction of index entries (which may be compounds) to their singular root forms, the elimination of redundant words and the auto-roling of words through their morphology. Some measure of heuristic performance is sought in this process. The system can easily contain the entire index to date on a single interchangeable disc and it is expected that subsequent discs will contain at least 3-4 years information.

17

A Theoretical Basis for the Use of Co-occurrence Data in Information Retrieval

This paper provides a foundation for a practical way of improving the effectiveness of an automatic retrieval system. Its main concern is with the weighting of index terms as a device for increasing retrieval effictiveness. Previously index terms have been assumed to be independent for the good reason that then a very simple weighting scheme can be used. In reality index terms are most unlikely to be independent. This paper explores one way of removing the independence assumption. Instead the extent of the dependence between index terms is measured and used to construct a nonlinear weighting function. In a practical situation the values of some of the parameters of such a function must be estimated from small samples of documents. So a number of estimation rules are discussed and one in particular is recommended. Finally the feasibility of the computations required for a non-linear weighting scheme is examined.

19

On the processing of Printed Subject Index Entries During Searching

Reports a laboratory experiment in which verbalized tape-recorded searches on five printed subject indexes reveal something of the linguistic processing that took place. Some 20% of the entries examined were changed grammatically, by word order change or function word supply, according to the linguistic form of the index concerned. Extracts from the search transcriptions are given, and the search processing modes of seeking, scanning, and screening are discussed.

20

Relevance Weighting of Search Terms

This paper examines statistical techniques for exploiting relevance information to weight search terms. These techniques are presented as a natural extension of weighting methods using information about the distribution of index terms in documents in general. A series of retevance weighting functions is derived and is justified by theoretical considerations. In particular, it is shown that specific weighted search methods are implied by a general probabilistic theory of retrieval. Different applications of relevance weighting are illus trated by experimental results for test collections.

21

PRECIS: the Preserved Context Index System

PRECIS was originally developed by the British National Bibliography to provide subject index data for UK/MARC records, and to produce an alphabetical subject index for the national bibliography. The present potential of the system extends beyond these initial objectives. The processes involved in index production are divided between the indexer and the computer: the former undertakes those tasks which require human judgments; the latter carries out clerical operations which are entirely mechanical. Index entries in PRECIS have been planned with certain user criteria in mind: in particular, comprehensibility and co-extensiveness. Basic research suggested that these could be achieved if entries were constructed according to the principle of context-dependency'; that is, if each term in the entry sets the next term into its obvious context. Entries constructed in this way are generated by a computer from input strings of terms and instruction codes written by the indexer. The rules governing the preparation of these strings comprise the syntax of PRECIS, and are embodied in a schema of role operators. Reference Indicator Numbers (RINs) are also appended to strings; these numbers lead to the extraction of appropriate See and See also references from a machine-held thesaurus. The creation and use of this thesaurus constitutes the semantic side of the system. In addition to its semantic and syntactic components, PRECIS has a facility, known as the Subject Indicator Number, which provides an economical means for dealing with 'repeat' subjects.

22

Alternatives to conventional multilingual thesauri

The amount of highly skilled effort required to compile a multilingual thesaurus is enormous. Although the co-operation necessitated by a multinational system may enable the burden to be shared, experience has shown that the multilingual approach complicates and increases the amount of effort required. Any thesaurus or indexing language needs to be maintained and thus requires continuing effort. This is more complicated in a multilingual situation. A controlled language system requires continuing skilled effort to be applied to the input operation. In a multilingual system this is likely to be greater than in a monolingual one. A multilingual thesaurus alone merely enables one to interrogate a data base in several languages. The content of the database must still be translated. Thus it is worthwhile considering the alternatives to conventional multilingual thesauri. Of the alternatives briefly described below, only one is a true alternative to a multilingual thesaurus. However the others involve unconventional thesauri or depend on thesauri together with other devices.

TITUS

This is a system for the automatic translation of text, such as abstracts, into several other languages (currently English, French, German and Spanish) and for retrieving abstracts by searching for Boolean combinations of selected words in any of the languages[4]. A quadrilingual thesaurus is used, together with restricted syntax and grammar. The input is written in a stylized form, using the limited syntax provided, and using only thesaurus vocabulary. In addition, relational words such as prepositions have to be specially indicated, and certain grammatical features have to be tagged. The system will convert this input into a numerical metalanguage and the information is stored in the computer in this form. A search question, using thesaurus terms in any of the four languages,

is converted to the numerical code in which the search is conducted. The full text of the abstracts found is then printed out in the chosen language by machine translation of the metalanguage into the chosen natural language.

The output is quite readable, although somewhat unnatural. Synonyms and near-synonyms of non-technical words are sometimes ill-chosen. The input operation is expensive in skill and time and the input worksheet is a formidable document: Titles of articles are not translated automatically. They are input in their natural form in the original language, and human translations in the other three languages are input at the same time. The effort of creating and maintaining a multilingual thesaurus is not avoided. This system caters for text translation, not merely multilingual access.

Conversion of, or switching between, existing monolingual indexing languages

This may be done in two ways: (a) by using an intermediate language or system through which switching takes place (e.g. one could take two existing thesauri, in different languages, and allocate UDC numbers to all the keywords[11] or (b) by multilateral comparison of existing thesauri to make them compatible - the 'reconciliation' technique.

(a) Intermediate switching language. The use of UDC is possible in theory, but does not appear to have been used as yet. The tendency in information retrieval in the last decade has been away from UDC. If UDC is to be used, why not use it alone and dispense with the thesauri? The 'intermediate lexicon' consists of an organized list of terms in a given subject area, with an equivalence table or concordance linking the terms to each of the thesauri between which switching is to take place[9].

(b) Reconciliation. This is also a switching technique, but it dispenses with a pre-constructed intermediate language[12]. Instead, the keywords of one thesaurus are compared one by one with another thesaurus. Where exact equivalents are found these terms are given identical arbitrary code numbers. Where equivalence is indirect (e.g. one thesaurus has the keyword 'beer' and the other merely the broader term 'alcoholic beverages') the more specific keyword is given a code number, and this number is identified in the other thesaurus by an added 'use' entry. There are a limited number of incompatibility problems, and solutions are possible for all of them.

Both methods enable existing thesauri to continue to be used in a co-operative network, rather than requiring participants to change to a new common thesaurus. The intermediate lexicon has been tested for switching to and from pre-coordinate languages as well as thesauri. Once constructed, the intermediate lexicon can be applied to further thesauri by constructing additional equivalence tables. The reconciliation process is very laborious and becomes more so when more than two thesauri are treated, except on a small experimental scale. It enables compatibility or equivalence problems to be analysed and solved in other multilingual work when reconciliation of complete thesauri is not the aim. Both methods have shown that incompatibility problems are logical rather than linguistic, i.e. switching or reconciliation methods must preserve the most specific practices in the constituent thesauri. Both methods provide multilingual access, but do not translate the contents of a data base.

PRECIS

This method of mechanically producing printed pre-coordinate indexes was developed in its present form for BNB[8], but is applicable to any document

collection. The product is a more or less conventional printed alphabetical index with 'see' and 'see also' references, but the permutations of citation order and insertion of cross-references are logically controlled by a form of syntax together with a controlled, but open-ended, vocabulary or thesaurus. It should be possible in principle to interpret the formalized syntax into other languages and to make the thesaurus multilingual, so that it would be possible to input index data in one language and obtain printed indexes in other languages. Work on this multilingual aspect of PRECIS is beginning.

PRECIS differs from the other systems considered in that it is the only one which produces a printed index. Moreover this printed index is in a conventional form which anyone can use without having to learn a search technique. The construction and maintenance of a multilingual thesaurus is not avoided, but the thesaurus is built up as indexing proceeds. This thesaurus differs from others in its open-endedness, and in not being necessarily confined to a particular subject.

Natural language data bases with free-text searching

These are largely formless systems which move away from thesauri altogether. They also avoid the skilled input effort of indexing, but are, however, inherently monolingual. The only way of constructing a multilingual system on this basis is to translate the entire data base and to operate several parallel data bases in different languages. At first sight this seems to require an unreasonable amount of translating effort. However, in a multilingual thesaurus system the content of the data base has to be translated if a fully multilingual system is to be obtained, so the problem for a natural language system is no different. There is at least one example of a bilingual data base of this kind, using the STATUS[13] system. The content consists of European statutes in English and French.

23

Alternatives to Conventional Multilingual Thesauri

Multilingual thesauri are expensive to construct and maintain. Although they allow a database to be interrogated in several languages, the content of the database must still be translated. Describes 4 approaches which represent alternatives to, or unconventional uses of multilingual thesauri: (1) TITUS (Traitement de Textile Universelle et Selective), a system for the automatic translation of specially constructed stylized abstracts, which allows input, searching and retrieval in 4 languages; (2) Switching between existing monolingual index languages, through the use of an intermediate switching language, or by thesaurus reconciliation; (3) PRECIS (Preserved Context Index System) which, because of its formalised syntax, has the potential to accept input data in one language and produce printed indexes in another; (4) Free text searching, involving the translation of the entire database to produce parallel databases in different languages. This last approach appears unreasonable, but in a fully multilingual system the content of the database has, in any case, to be translated.

24

The Intermediate Lexicon: the possibilities for information exchange networks.

The purpose of an intermediate lexicon is to switch index terms from one index language to another. Such a device allows information centres to exchange indexing decisions without requiring that they standardise their indexing practices. Describes a research project at the Polytechnic of North London (UK) concerned with the design and evaluation of an intermediate lexicon for information science.

25

The Dewey Decimal Classification and automated subject retrieval

A number of workers in the field of library automation have reported on their use of class marks selected from the DDC (including its nearest relation, the UDC) as keys for searching machine-held files of bibliographic data. It is difficult, from a study of these specific instances, to draw general conclusions concerning the relative effectiveness of class numbers used in this way. Some projects appeared to be reasonably successful, whereas others established that class numbers are relatively ineffective, except perhaps as negative discriminators: that is to say, class numbers were used in the first place to screen out parts of the data base which seemed least likely to contain relevant items, and the rest of the file was then searched by a different means (e.g. matching on verbal data). Some selected experiments are considered with a view to isolating those factors which might account for the poor performance of DDC in some of these experiments. It is concluded that, in general, Dewey numbers are not the best tools for searching mechanised files. However, they were not intended for this purpose and the question must be asked whether the editor and publishers should take steps to remedy these faults. It is the author's opinion that they should not.

26

An On-line System for Handling Personal Data Bases on a PDP 11/20 Minicomputer

Paper presented (in part) at the 50th Aslib Annual Conference, University of Exeter, September 76. Computer support for personal data bases can be given through the production of printed indexes or by providing for the searching of machine-readable files. Some systems offering these facilities are briefly reviewed. The userfile system implemented by the Experimental Information Unit on a DEC PDP-11/20 minicomputer, was designed to offer users a 'workspace' for creating, maintaining and storing personal files containing bibliographic and other types of record. The main features of the system are described, including:design principles; hardware; software; file content; record structure; and searching. The staff and research students of the departments of organic and inorganic chemistry at the University of Oxford formed the user population. The experiment revealed a demand for personal file handling systems, indicated the possible role of personal files in investigating information habits, and highlighted the need for flexibility in approach, ease of access and user-oriented design.

27

The Role of A Computer Retrieval System in Information Gathering by Psychologists: A Pilot Survey

A sample of 15 University Lecturers in Psychology from three different Colleges took part in a structured interview based on a 90 item questionnaire to elucidate the ways in which they dealt with psychological literature. They were then invited to make use of a computerised information retrieval system, in which the main file was for Psychological Abstracts, as frequently as they required over a period of three months from the interview.

11 people made at least one visit to the system. Cross-classifications of 84 of the questionnaire items against the criterion of 'None', 'One' and 'More than one' visits indicated 28 attributes which discriminate the 'High' from the 'Non' user. These attributes fell into 5 classes: Personal Characteristics, Information Needs, Information Gathering Habits, Information Output - Publishing and Research, and Prior Experience of Computers.

Those who used the system were given a second interview about two weeks after their first visit. Answers to a 15 item questionnaire were obtained. These showed that the majority of users valued highly the information they had retrieved and that about half saw the system as likely to produce a change in the way they gathered information - better library use, increased use of Inter-Library Loans and a decreased use of bibliographic tools. All users made suggestions for making the system more'client effective'.

Proposals are made for a main study founded upon the indicative factors revealed by this pilot survey.

28

Investigation of Some Aspects of On-line Searching

The fundamental objectives of this project were (1) to determine the extent to which on-line search systems conceived in terms of thesaurus-controlled data bases are applicable to free-text data bases which do not have a thesaurus, (2) to determine the degree to which knowledge of the data base, experience in searching it, and preliminary preparation of a search strategy are necessary for successful on-line searching, (3) to determine the extent to which searches having a high chemical structure content can be successfully performed on an on-line system such as Lockheed Dialog, with no left-hand truncation facility.

Searches of various data bases on the Lockheed Dialog on-line system were performed by several "user centres", and details of the searches of the Chemcon data base were sent to UKCIS. The questionnaires completed by the users were analysed, and the search dialogues were also studied. For several of the searches for which all the appropriate information, including the search question, was available, the searches were repeated by UKCIS staff, and the user centre searches were compared with the UKCIS searches. The results of all these investigations are given in Chapters 2 and 3. It was found that the difference in quality between the searches carried out by UKCIS and those carried out by the user centres was not significant, but the searches which were carried out by the users themselves were very much inferior.

Since few of the user centre searches contained a significant chemical structure element, UKCIS performed some extra searches on chemical structure-based queries. Although the precision of these searches was, in general, very high, the recall is expected to be quite low, owing to the lack of a left-hand truncation facility. Because of this need for full truncation in searching chemical names, the more "chemical" a search is, the less successful it will normally be. Details of the structure-based searches

are given in Chaper 4.

A comparison of searches carried out on three currently-available on-line systems, Lockheed Dialog, SDC Orbit and SDS/Recon, was performed, and details of this appear in Chapter 5. The ability of the systems to perform the searches, the search times and the search performances were investigated, and the good and bad points of each of the three systems in searching the Chemcon data base are listed.

In Chapter 6, the good points and deficiencies of the Lockheed Dialog system, as identified by UKCIS, are first discussed. Then the impressions and comments of the staff of three of the user centres regarding the system are given. The extent to which the Lockheed Dialog system fulfils the three main objectives of the project is then discussed. Finally, the main differences between theLockheed Dialog, SDC Orbit and SDS/Recon on-line systems are discussed, and the facilities required of a good on-line system for searching the Chemcon data base are stated, consisting of particular facilities from each of the three systems investigated, together with additional facilities not present in any of the three systems.

29

Reference Retrieval by User-Negotiated Term Frequency Ordering Within a Dynamically adjusted Notional "Document"

A project is described which sets out to refine the "user-friendly" interactive retrieval system introduced by Oddy. The documents handled by Oddy's system were indexed by binary terms. But a given term may play quite different roles in different documents. Term frequency weights might have suggested an obvious way of quantifying this distinction, were it not for the mediocre performance of term frequency based strategies in tests. Robertson and Sparck Jones have argued that neither indexing policies nor matching functions can be profitably examined in isolation. An iterative matching procedure is explained which presupposes that terms are ranked within a document description as a function of their term frequency: document descriptions are compared with the ordering of terms in a pattern. The model depends on a convenient fiction - the dynamic description of a notional "document", corresponding to the user's information need and constructed in the course of the dialogue. This model is given no interpretation outside the context of interactive retrieval.

The request pattern is adjusted in response to the user's verdict on a subset of the highest ranking documents. In a large collection many documents might attain an equally high rank. A subsidiary refinement endeavors to improve precision by exploiting syntactic features.

30

Machine Architectures Suitable for Information Retrieval Systems

The present status of on-line information retrieval systems is outlined with some particular problems, concerning the implementation of systems, highlighted. It is suggested that the root causes of the difficulties may lie in the inadequacy of present day, general purpose computers for our particular application. A machine architecture is suggested that may enable the problems to be surmounted and an experiment is described which allows testing of the suggested design at a low cost. A full scale, working system based on the ideas presented could also be implemented at an economical cost should the result of the experiments make this desirable.

31

The Application of Minicomputers to Problems of Information Retrieval

Current awareness information retrieval can easily be handled by a mini, but minis are, in fact, rarely used for this type of application. In retrospective searching of large data bases a mini can be used if an overnight turnround is acceptable. Often, a mini is used to connect user terminals and other peripherals to a mainframe as a means of: (1) removing some work from the mainframe; (2) enabling simpler input/output software to be used; (3) allowing alterations to be made to the mainframe configuration without altering the users' image of the system. The most valuable application of minis is in networking, where they can be used to provide access to a range of mainframes providing on-line searches of different data bases. Because of its relatively small main memory a mini is unlikely to be used to support an information retrieval system by itself. It can, however, play a role in providing retrospective search facilities if used in conjunction with other hardware.

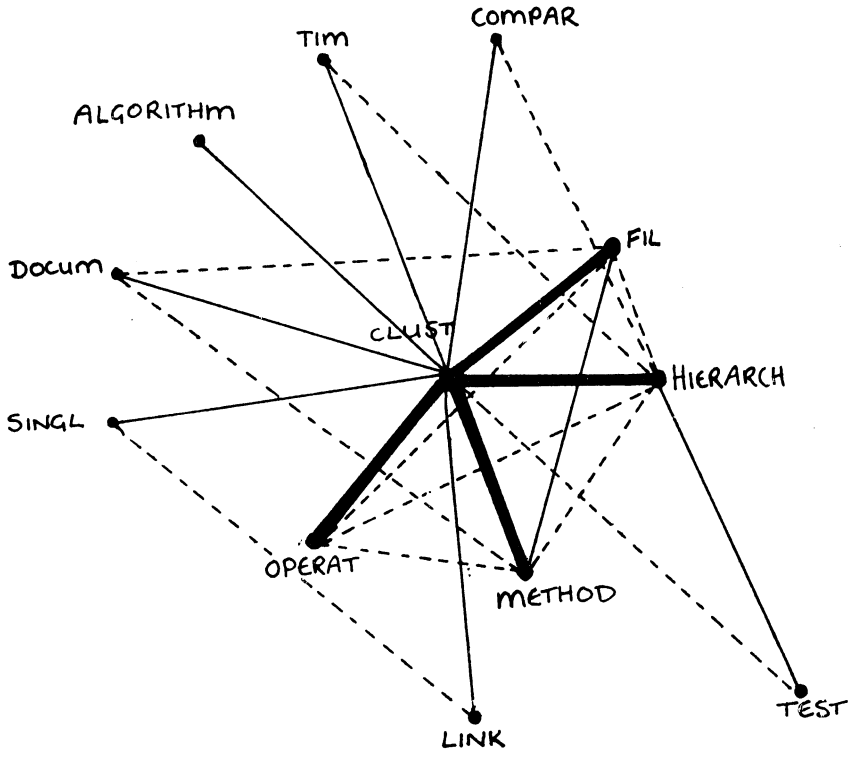
32

Mini-computers and Bibliographic Information Retrieval

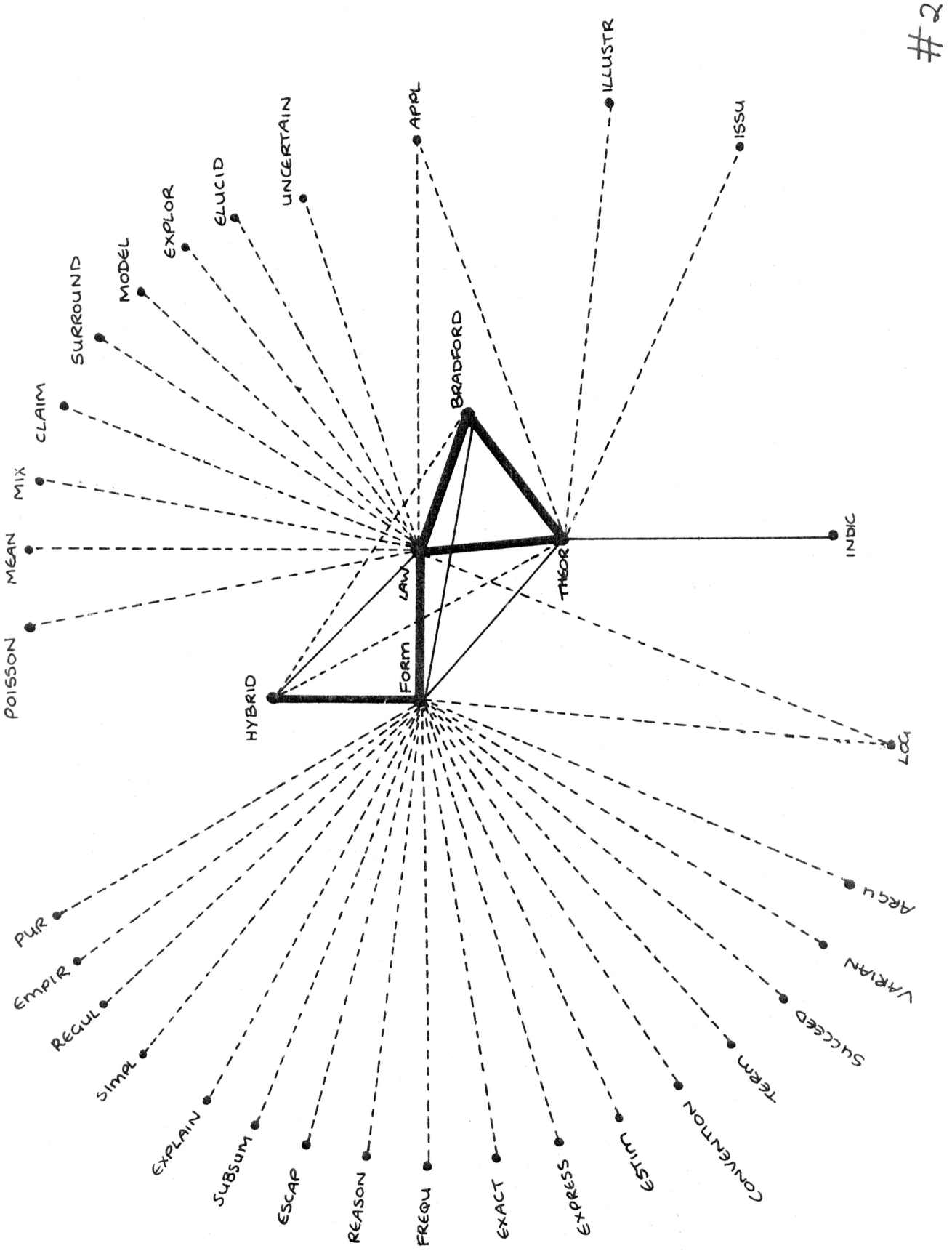
Results of an investigation into several operational and proposed information retrieval systems on minicomputers and a comparison of facilities available on minis with those on mainframe systems. Describes the characteristics of minis, and discusses bibliographic information retrieval, and various questions to be considered before implementing a computerised system, drawing attention to the ways in which the characteristics of minis affect these considerations. Constructs models of configurations suitable for information retrieval on various sizes of data base in order to examine how minis may provide a bibliographic service for which previously a mainframe would have been thought necessary. Concludes that in calculating the value of a computer system it is crucial to establish costs. The advantages of minis in terms of their cheapness, accessibility and communications facilities make them suitable for information retrieval systems.

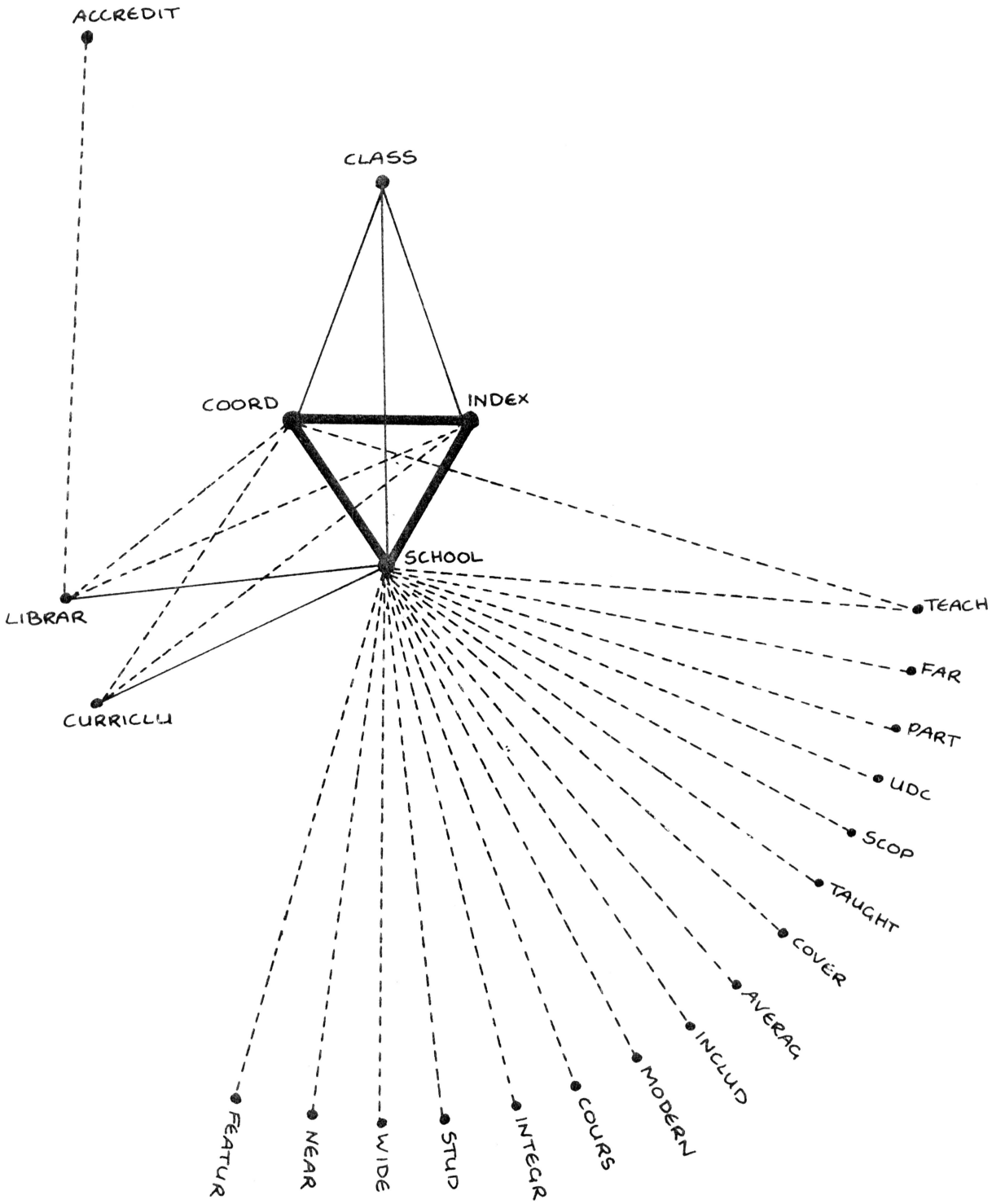
APPENDIX G

Association Map Representations of Abstracts

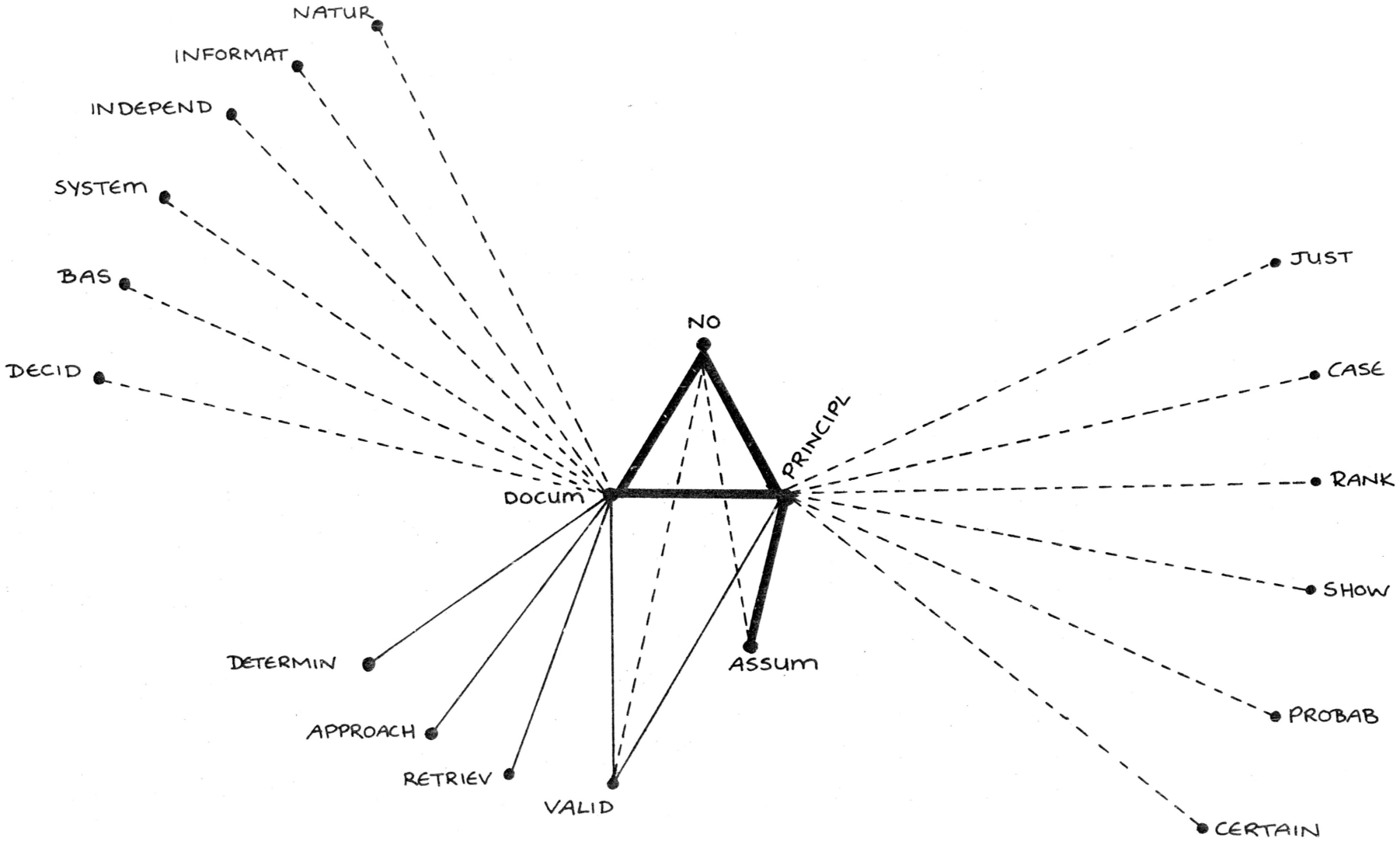

COMPAR
TIM
ALGORITHM
FIL
DOCUM
CLUST
HIERARCH
SINGL
OPERAT
METHOD
TEST
LINK

#1

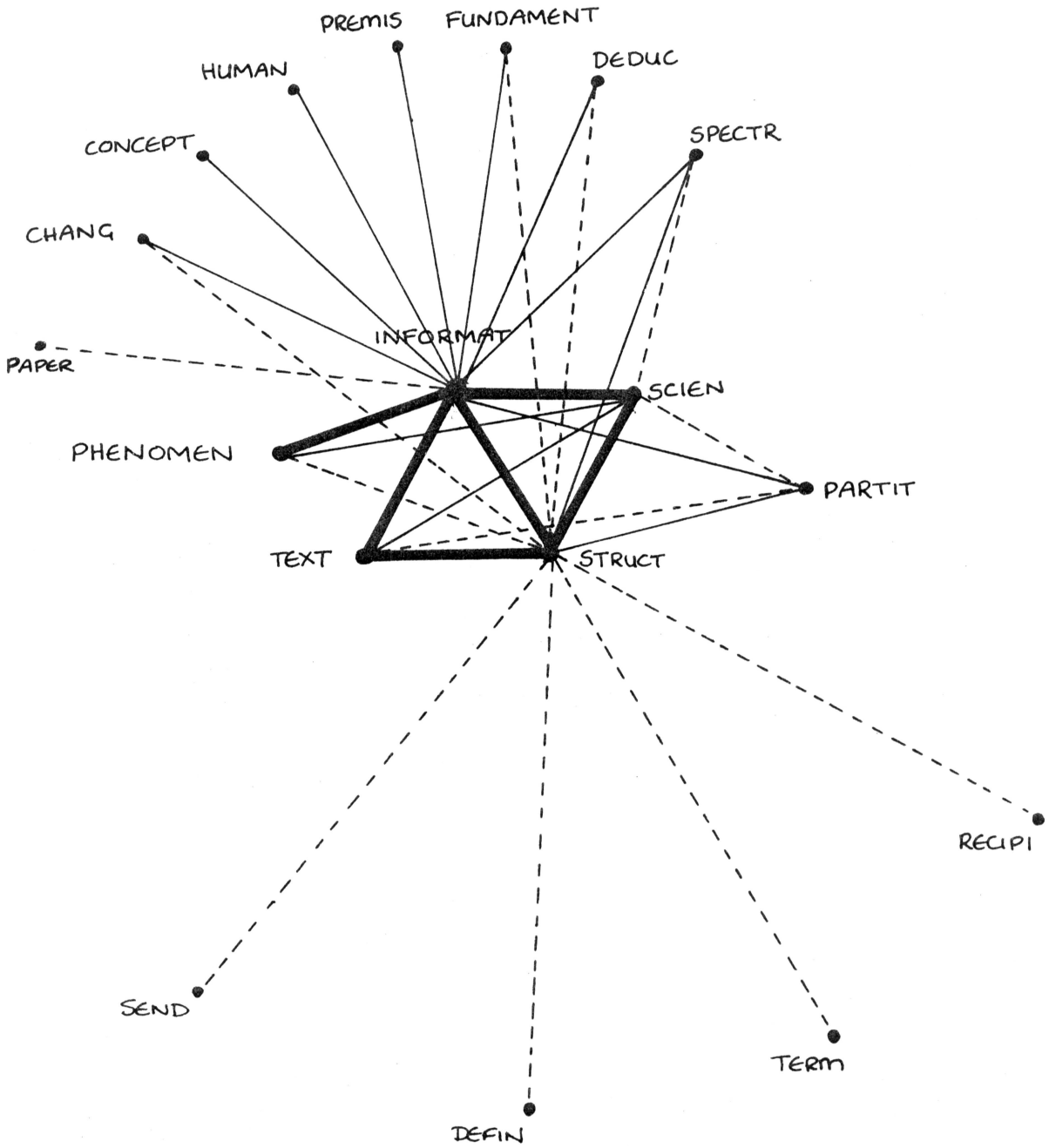

#2
ILLUSTR
ISSU
APPL
UNCERTAIN
ELUCID
EXPLOR
MODEL
SURROUND
CLAIM
MIX
MEAN
POISSON
BRADFORD
LAW
FORM
HYBRID
THEOR
INDIC
LOC
PUR
EMPIR
REGUL
SIMPL
EXPLAIN
SUBSUM
ESCAP
REASON
FREQU
EXACT
EXPRESS
ESTIM
CONVENTION
TERM
SUCCEED
VARIAN
ARGU


ACCREDIT
CLASS
COORD
INDEX
SCHOOL
LIBRAR
CURRICLU
TEACH
FAR
PART
UDC
SCOP
TAUGHT
COVER
AVERAG
INCLUD
MODERN
COURS
INTEGR
STUD
WIDE
NEAR
FEATUR

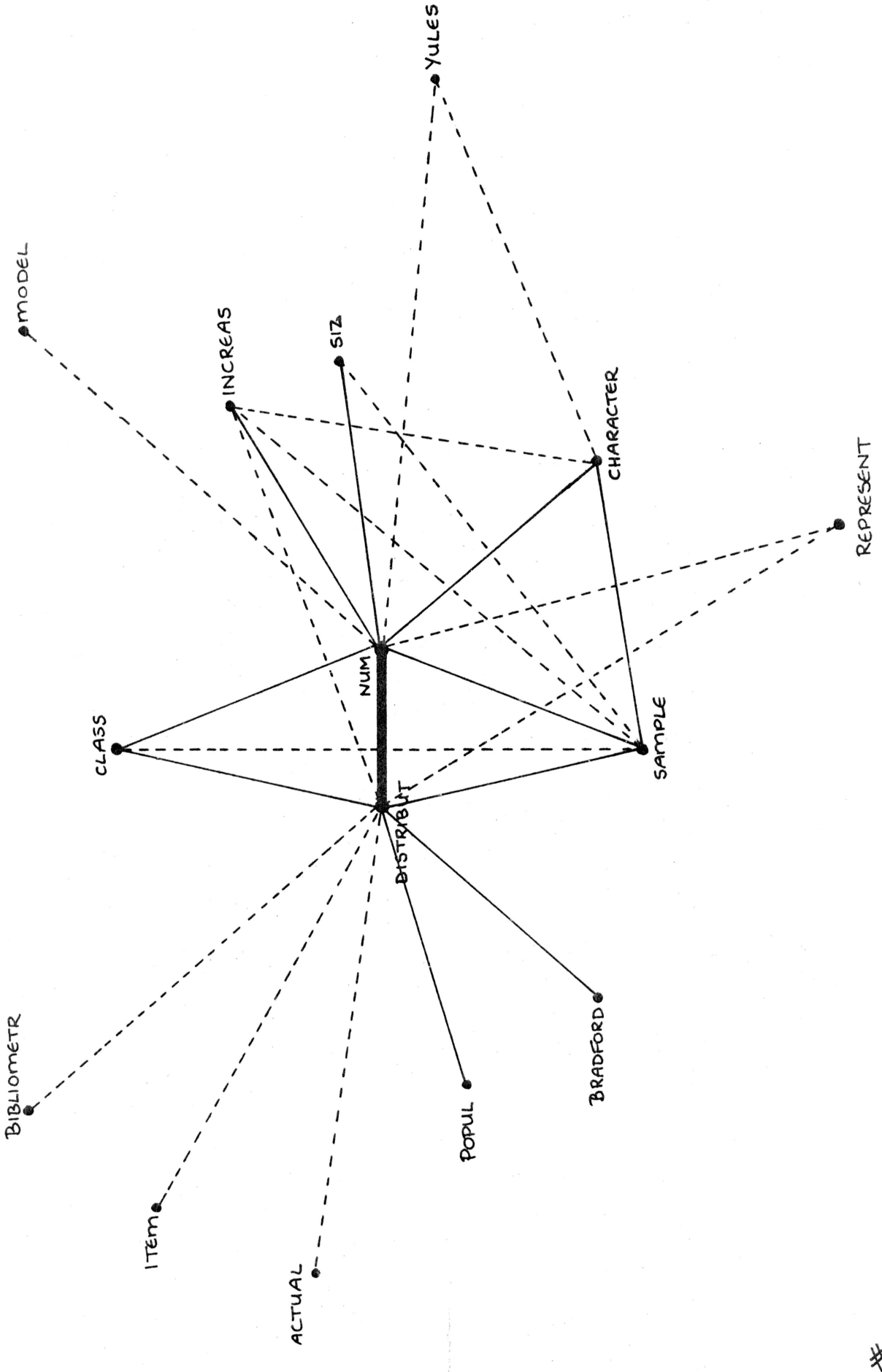
#3


NATUR
INFORMAT
INDEPEND
SYSTEM
BAS
DECID
NO
PRINCIPL
DOCUM
JUST
CASE
RANK
SHOW
PROBAB
CERTAIN
ASSUM
DETERMIN
APPROACH
RETRIEV
VALID

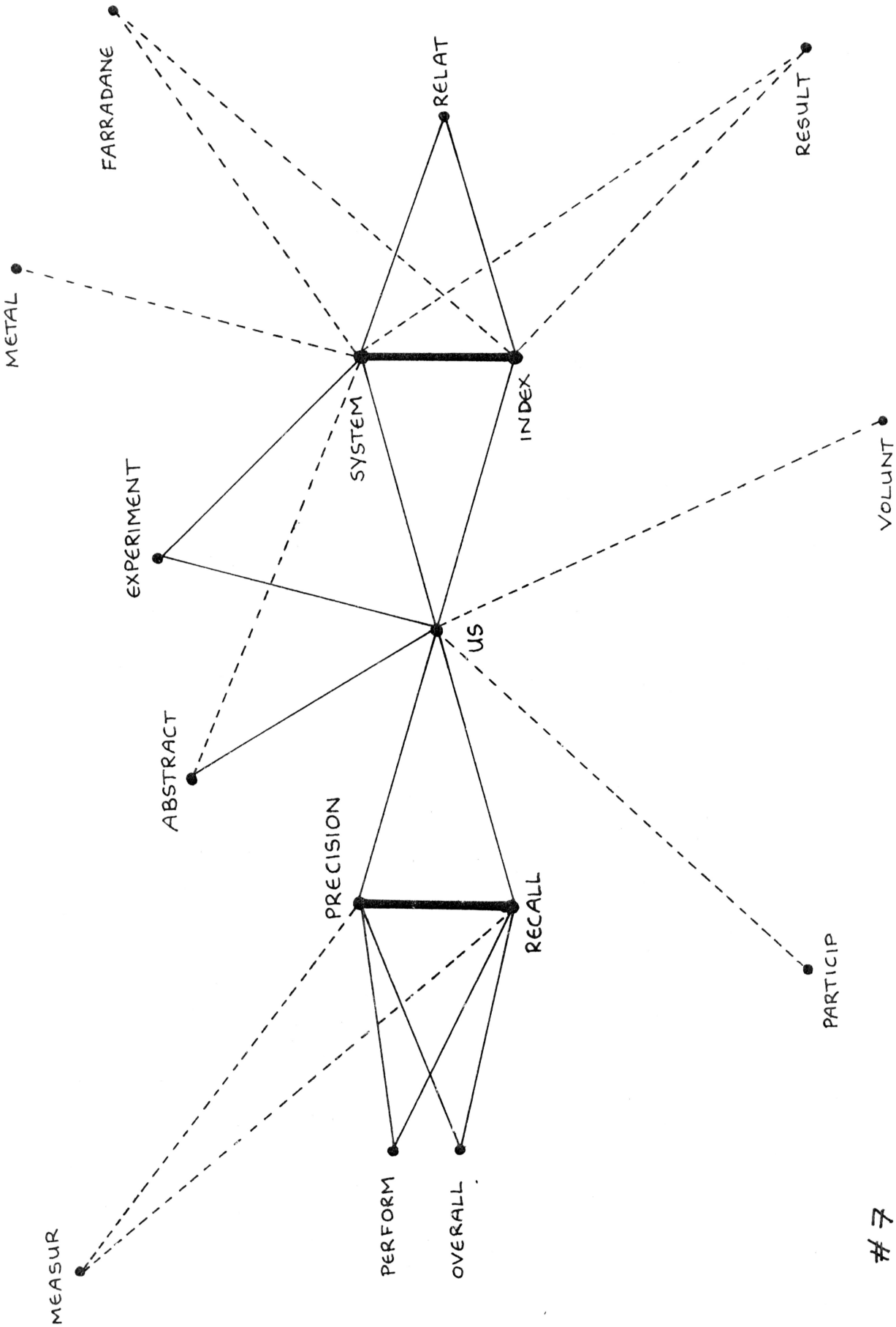
#4


PREMIS
FUNDAMENT
DEDUC
HUMAN
SPECTR
CONCEPT
CHANG
INFORMAT
PAPER
SCIEN
PHENOMEN
PARTIT
TEXT
STRUCT
RECIPI
SEND
TERM
DEFIN

#5

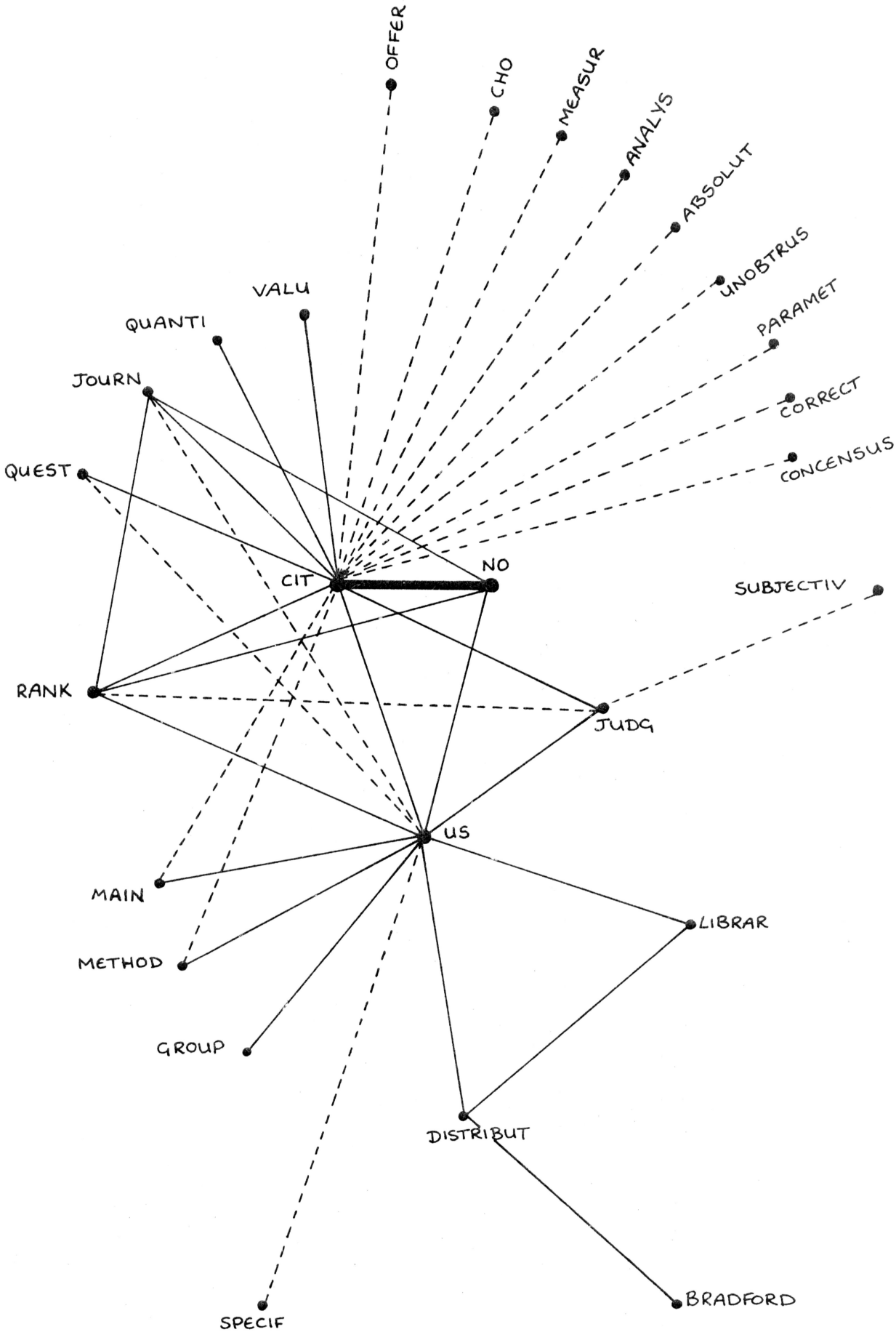

YULES
MODEL
INCREAS
SIZ
CHARACTER
REPRESENT
NUM
CLASS
SAMPLE
DISTRIBUT
BIBLIOMETR
ITEM
ACTUAL
POPUL
BRADFORD

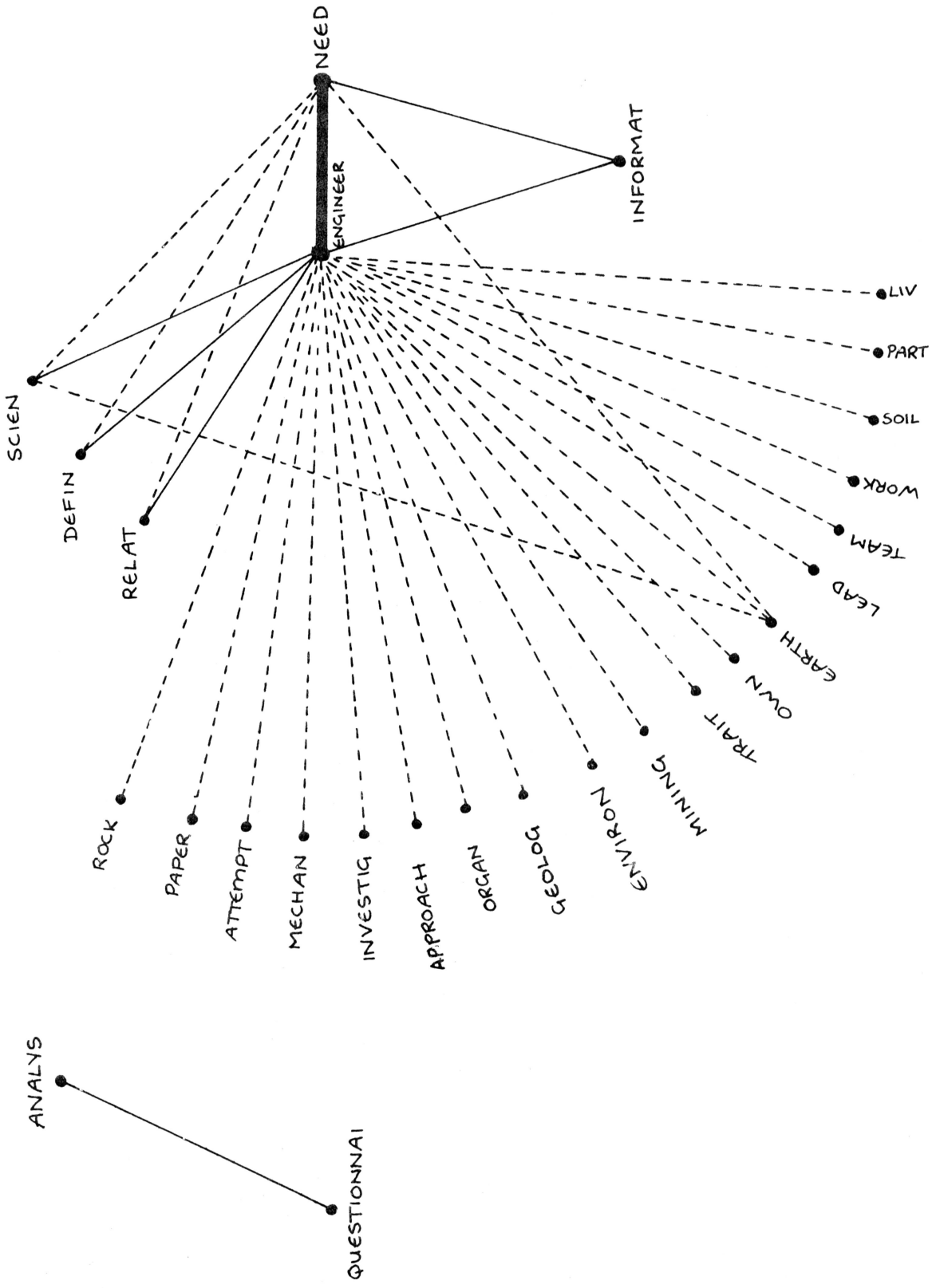
#6


FARRADANE
RELAT
RESULT
METAL
SYSTEM
INDEX
EXPERIMENT
VOLUNT
US
ABSTRACT
PRECISION
RECALL
PARTICIP
MEASUR
PERFORM
OVERALL

#7

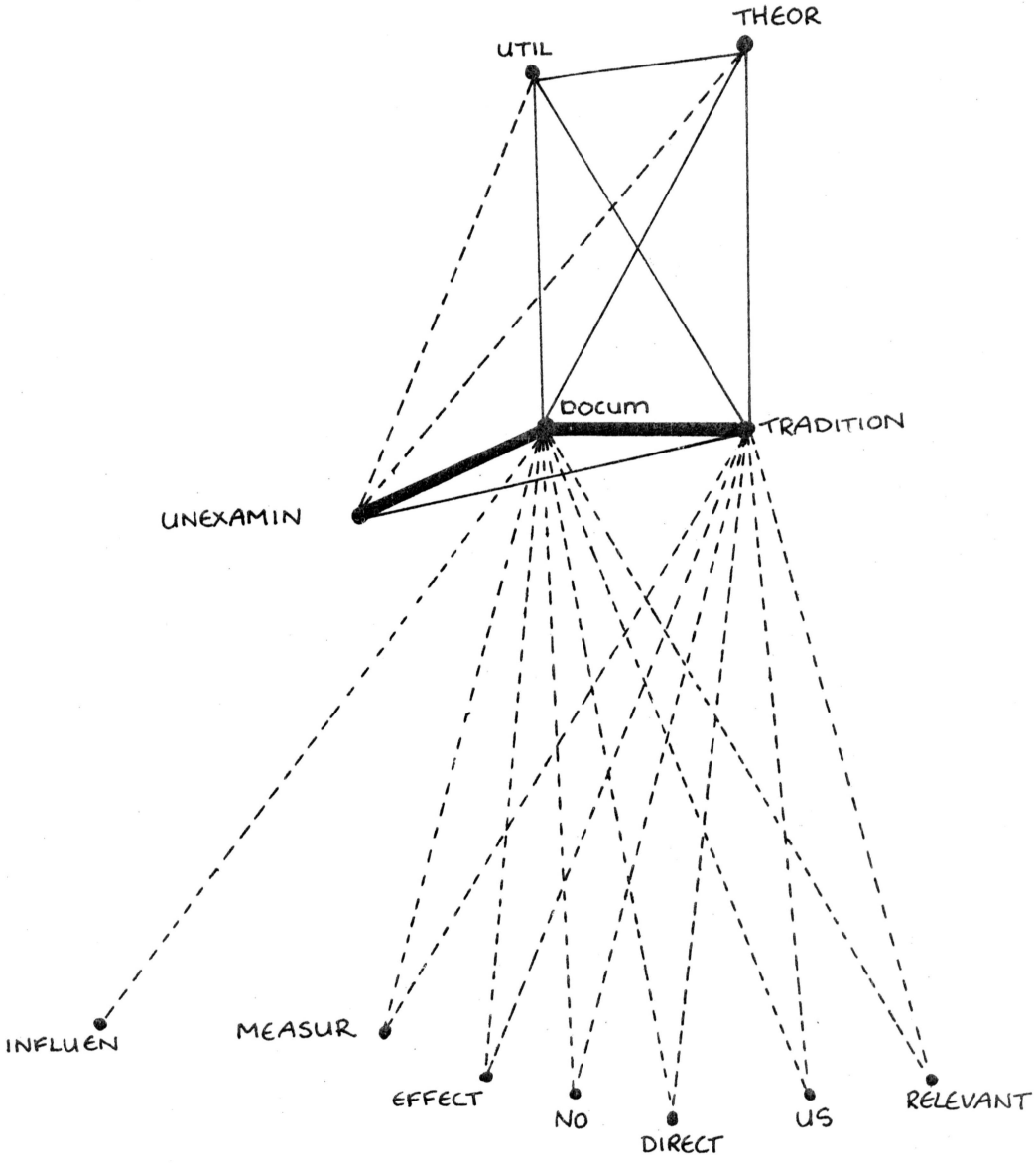
#8

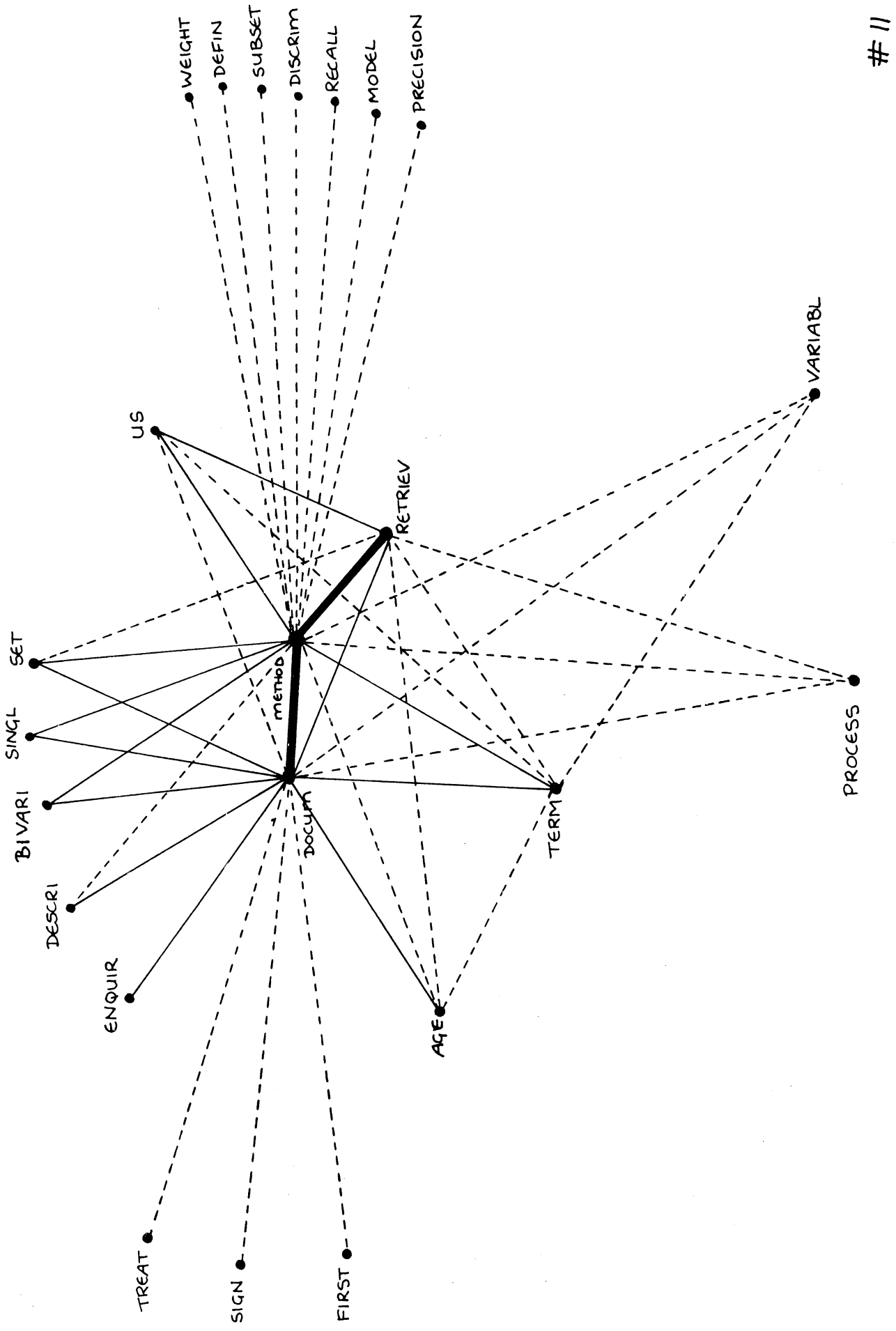

OFFER
CHO
MEASUR
ANALYS
ABSOLUT
UNOBTRUS
PARAMET
CORRECT
CONCENSUS
VALU
QUANTI
JOURN
QUEST
CIT
NO
SUBJECTIV
RANK
JUDG
US
MAIN
LIBRAR
METHOD
GROUP
DISTRIBUT
BRADFORD
SPECIF


NEED
ENGINEER
INFORMAT
SCIEN
DEFIN
RELAT
ROCK
PAPER
ATTEMPT
MECHAN
INVESTIG
APPROACH
ORGAN
GEOLOG
ENVIRONM
MINING
TRAIT
OWN
EARTH
LEAD
TEAM
WORK
SOIL
PART
LIV
ANALYS
QUESTIONNAI

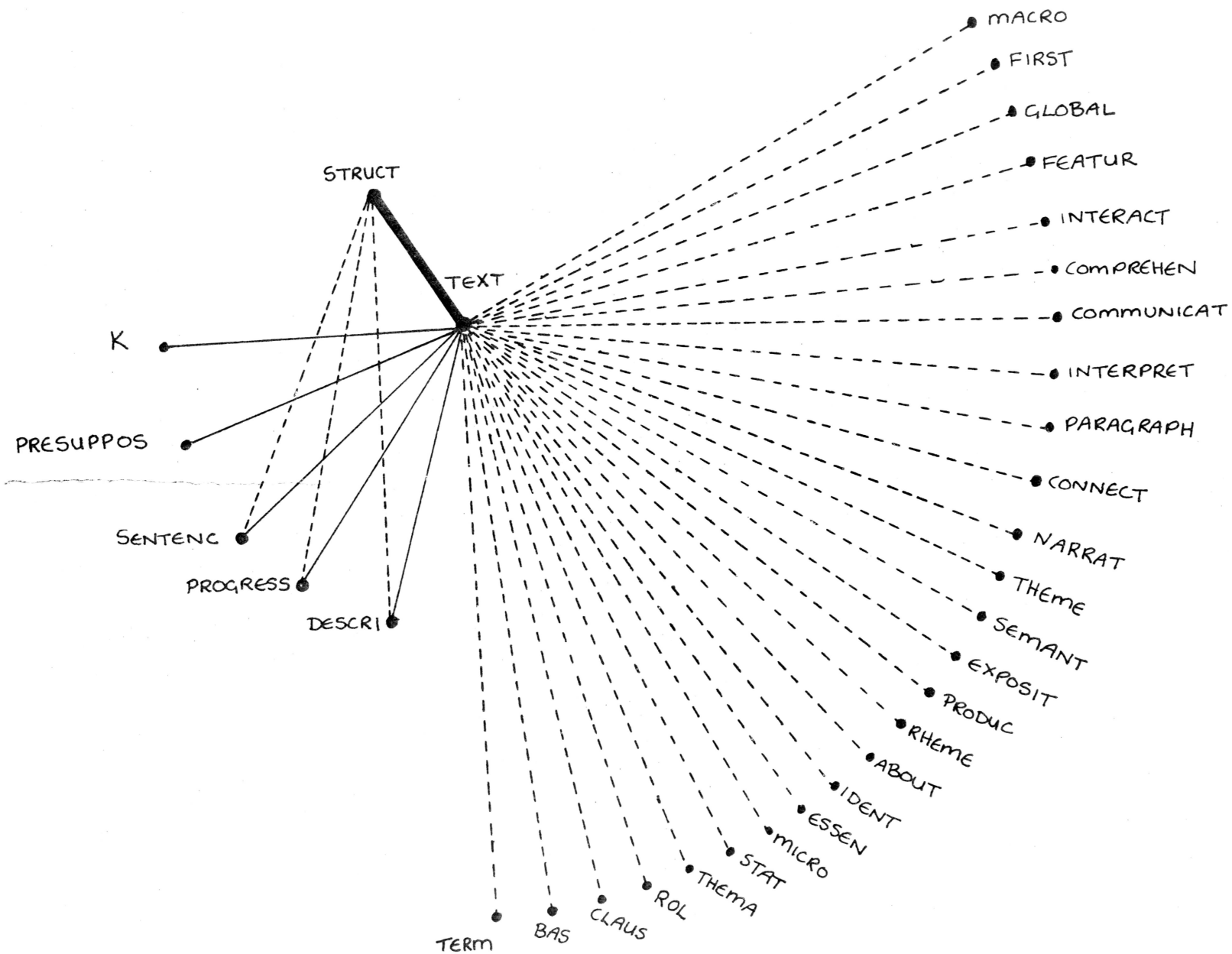
#9

#10

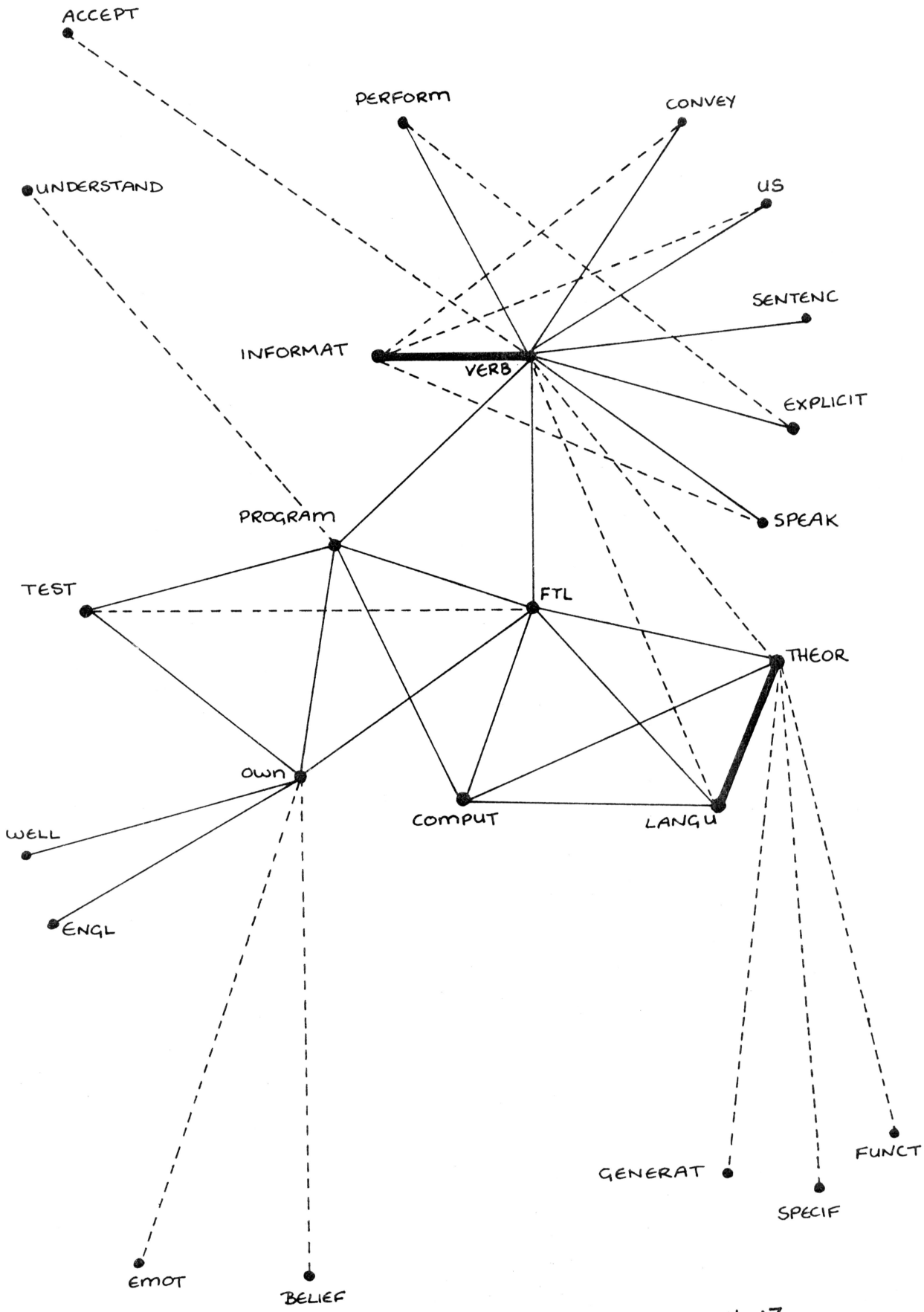

THEOR
UTIL
DOCUM
TRADITION
UNEXAMIN
INFLUEN
MEASUR
EFFECT
NO
DIRECT
US
RELEVANT


WEIGHT
DEFIN
SUBSET
DISCRIM
RECALL
MODEL
PRECISION
US
RETRIEV
VARIABL
SET
SINGL
METHOD
BIVARI
DESCRI
ENQUIR
DOCUM
TERM
PROCESS
AGE
TREAT
SIGN
FIRST

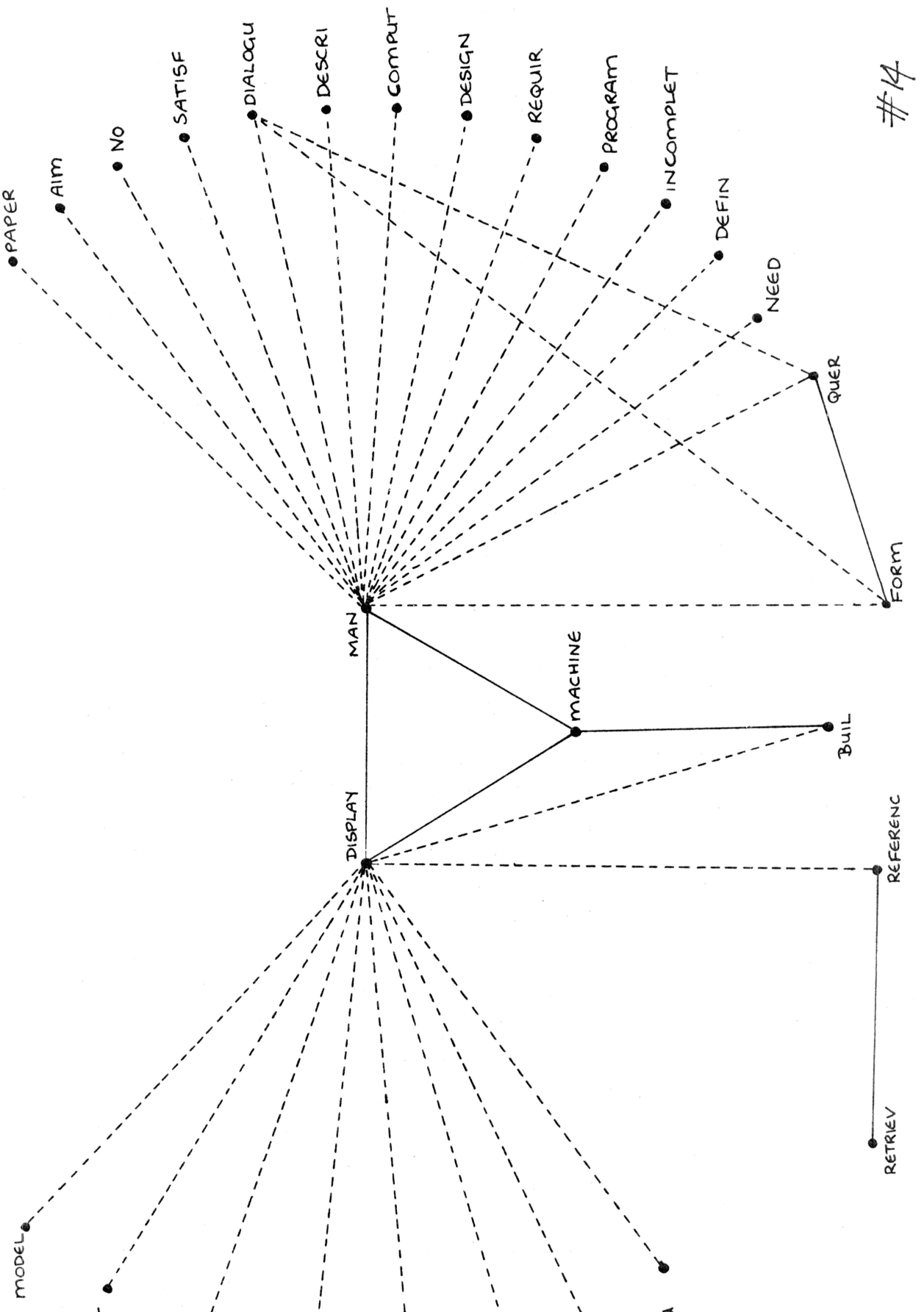
#11


MACRO
FIRST
GLOBAL
FEATUR
INTERACT
COMPREHEN
COMMUNICAT
INTERPRET
PARAGRAPH
CONNECT
NARRAT
THEME
SEMANT
EXPOSIT
PRODUC
RHEME
ABOUT
IDENT
ESSEN
MICRO
STAT
THEMA
ROL
CLAUS
BAS
TERM
STRUCT
TEXT
K
PRESUPPOS
SENTENC
PROGRESS
DESCRI

#12


ACCEPT
PERFORM
CONVEY
UNDERSTAND
US
SENTENC
INFORMAT
VERB
EXPLICIT
SPEAK
PROGRAM
TEST
FTL
THEOR
own
COMPUT
LANGU
WELL
ENGL
GENERAT
FUNCT
SPECIF
EMOT
BELIEF

13


PAPER
AIM
NO
SATISF
DIALOGU
DESCRI
COMPUT
DESIGN
REQUIR
PROGRAM
INCOMPLET
DEFIN
NEED
QUER
FORM
MAN
MACHINE
BUIL
DISPLAY
REFERENC
RETRIEV
MODEL
REACT
CHO
US
STAT
VARI
INTEREST
CONSEQU

#14

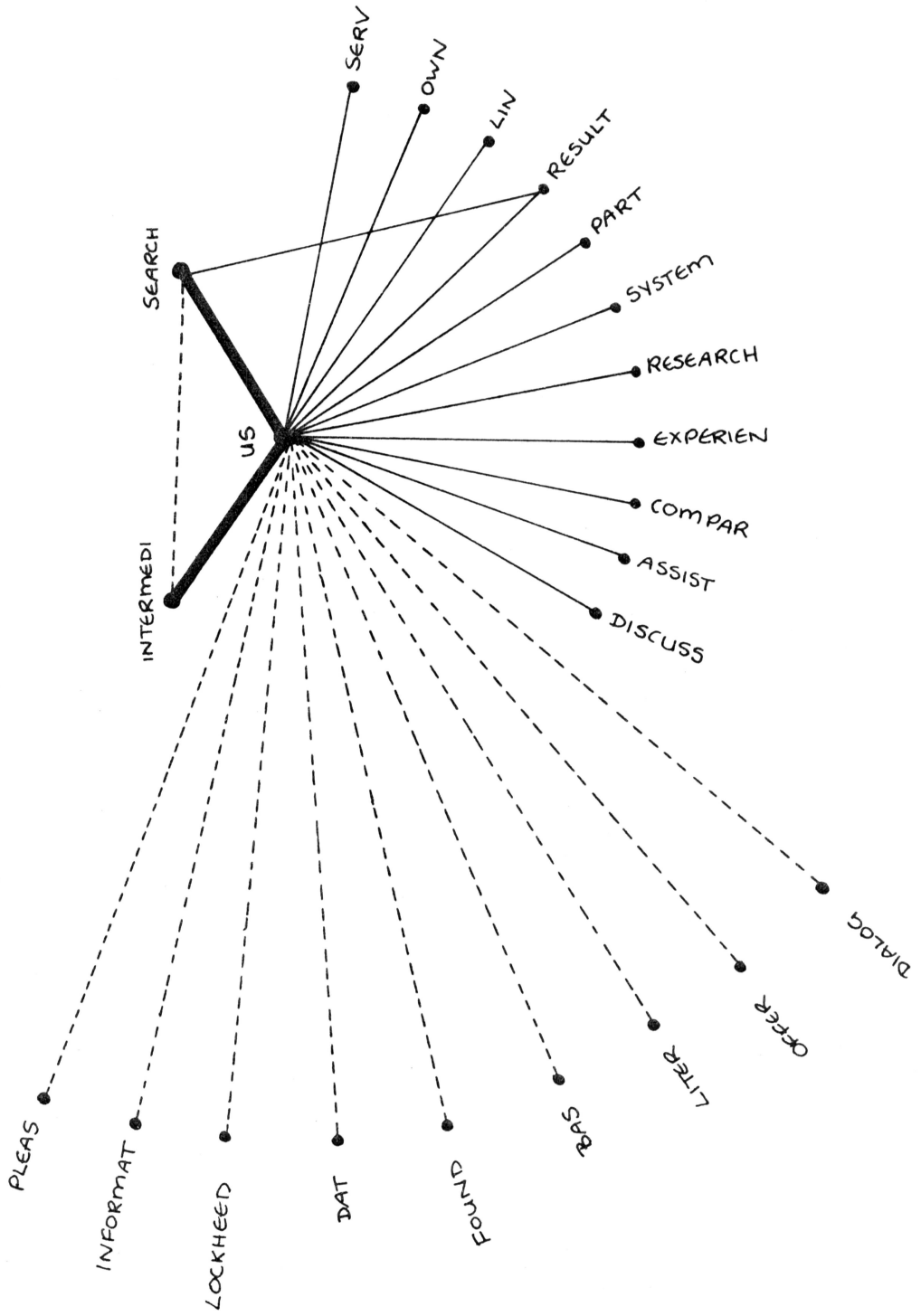
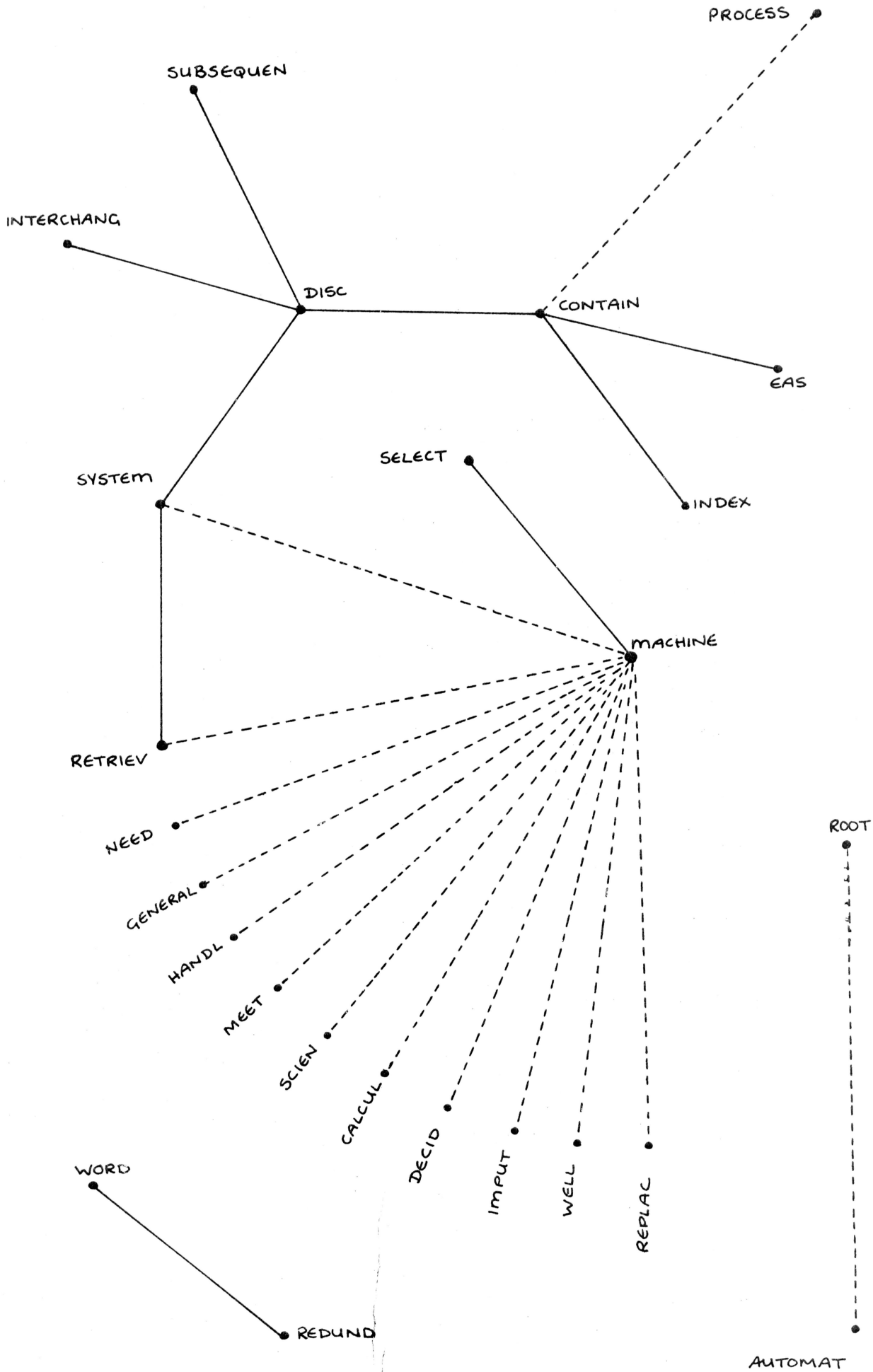
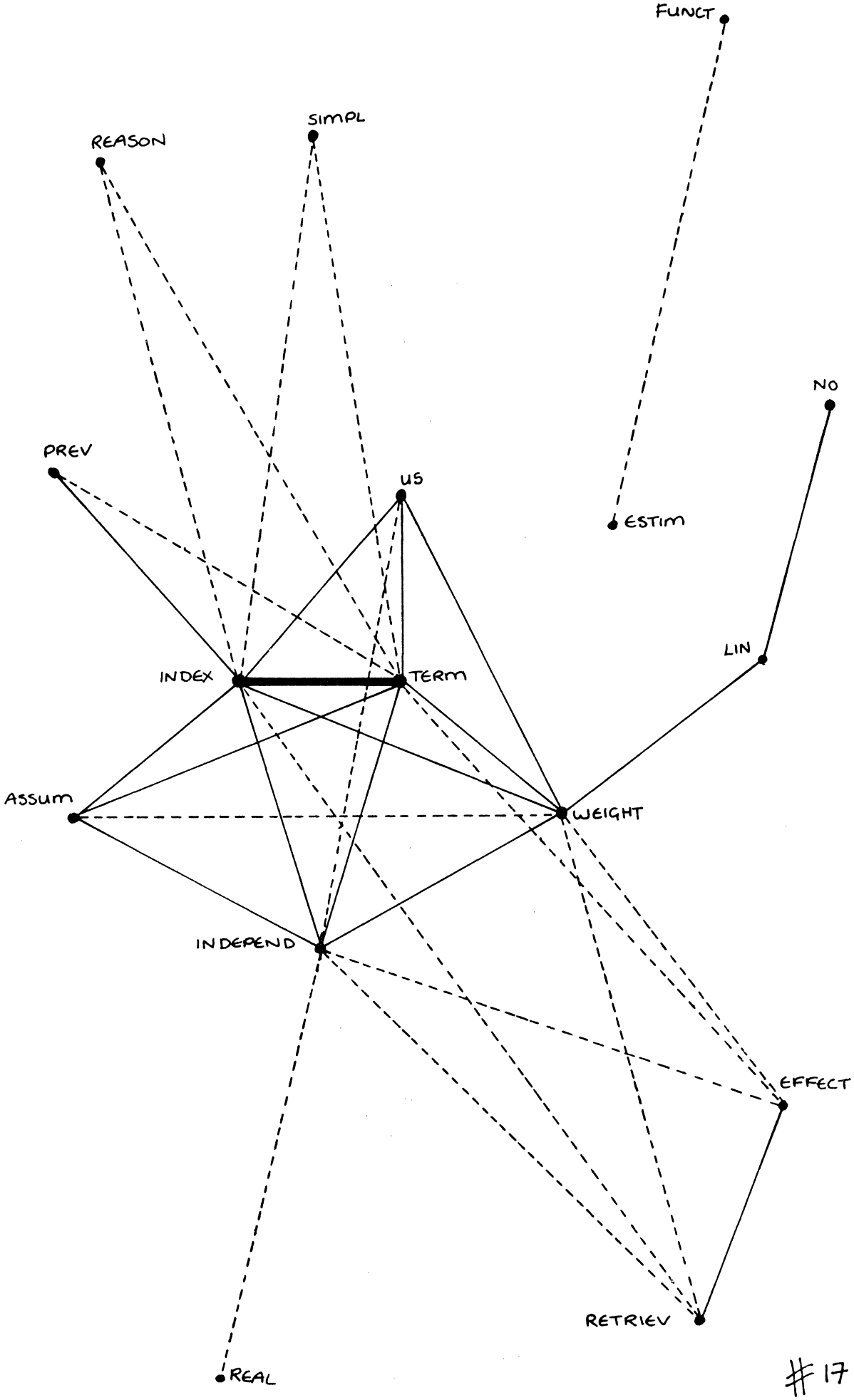

SERV
OWN
LIN
RESULT
PART
SYSTEM
RESEARCH
EXPERIEN
COMPAR
ASSIST
DISCUSS
SEARCH
US
INTERMEDI
PLEAS
INFORMAT
LOCKHEED
DAT
FOUND
BAS
LITER
OFFER
DIALOG

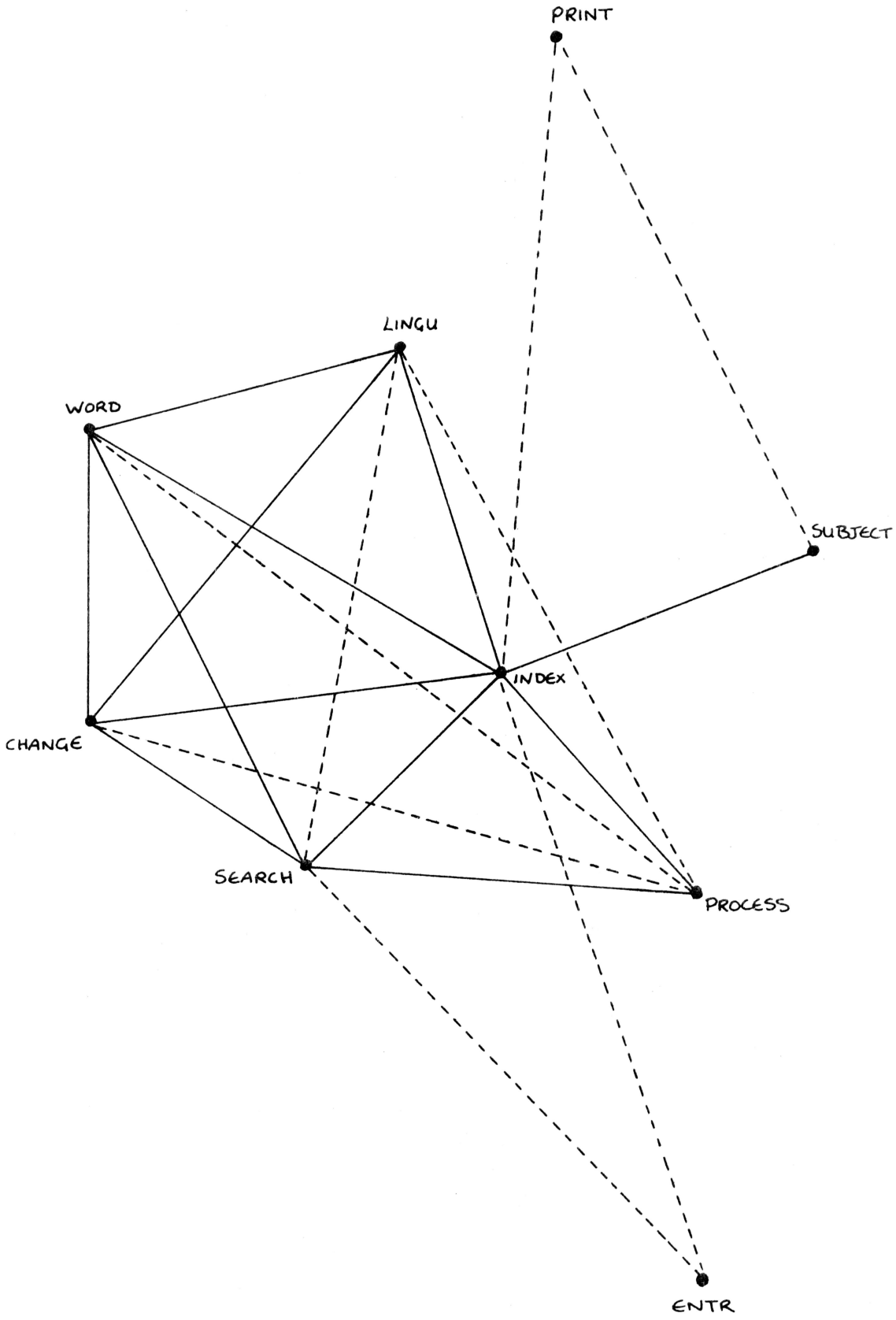
FIG #5


PROCESS
SUBSEQUEN
INTERCHANG
DISC
CONTAIN
EAS
SELECT
SYSTEM
INDEX
MACHINE
RETRIEV
NEED
GENERAL
HANDL
MEET
SCIEN
CALCUL
DECID
IMPUT
WELL
REPLAC
ROOT
WORD
REDUND
AUTOMAT

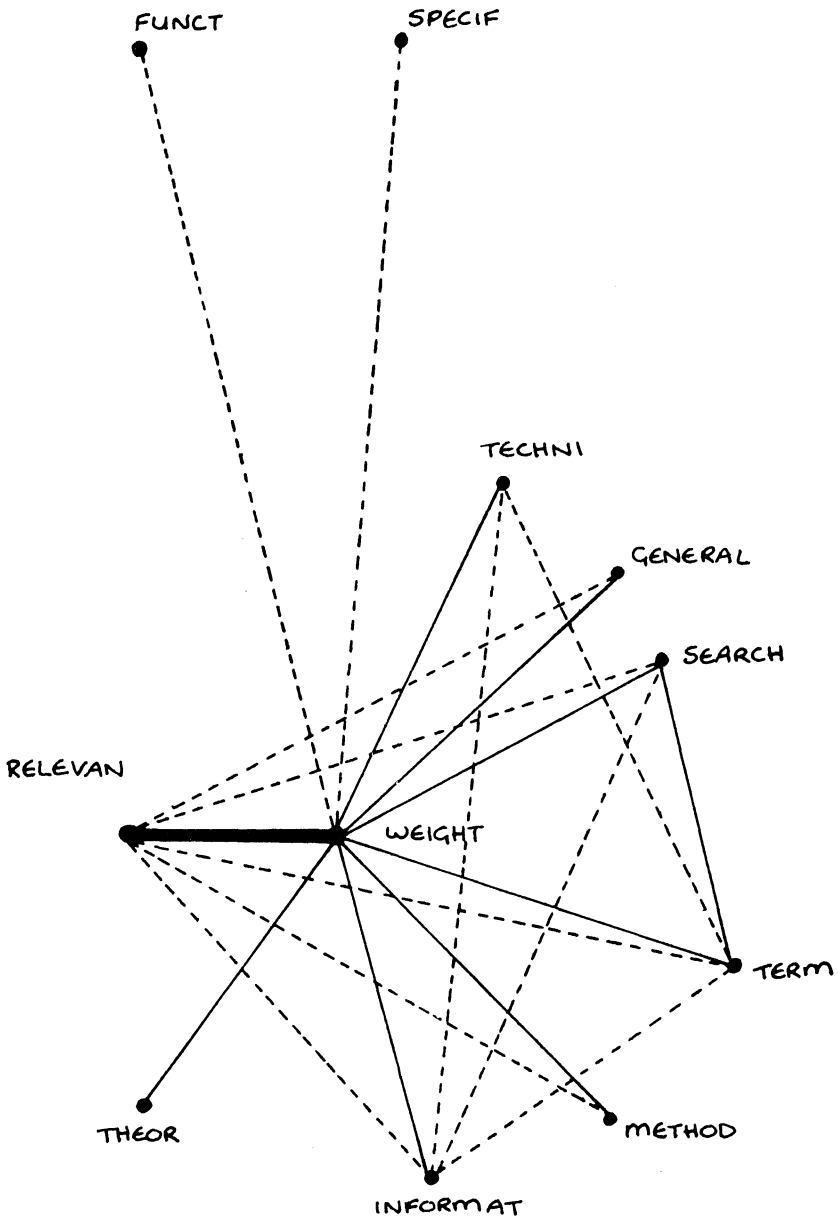
#16


FUNCT
SIMPL
REASON
NO
PREV
US
ESTIM
LIN
INDEX
TERM
ASSUM
WEIGHT
INDEPEND
EFFECT
RETRIEV
REAL

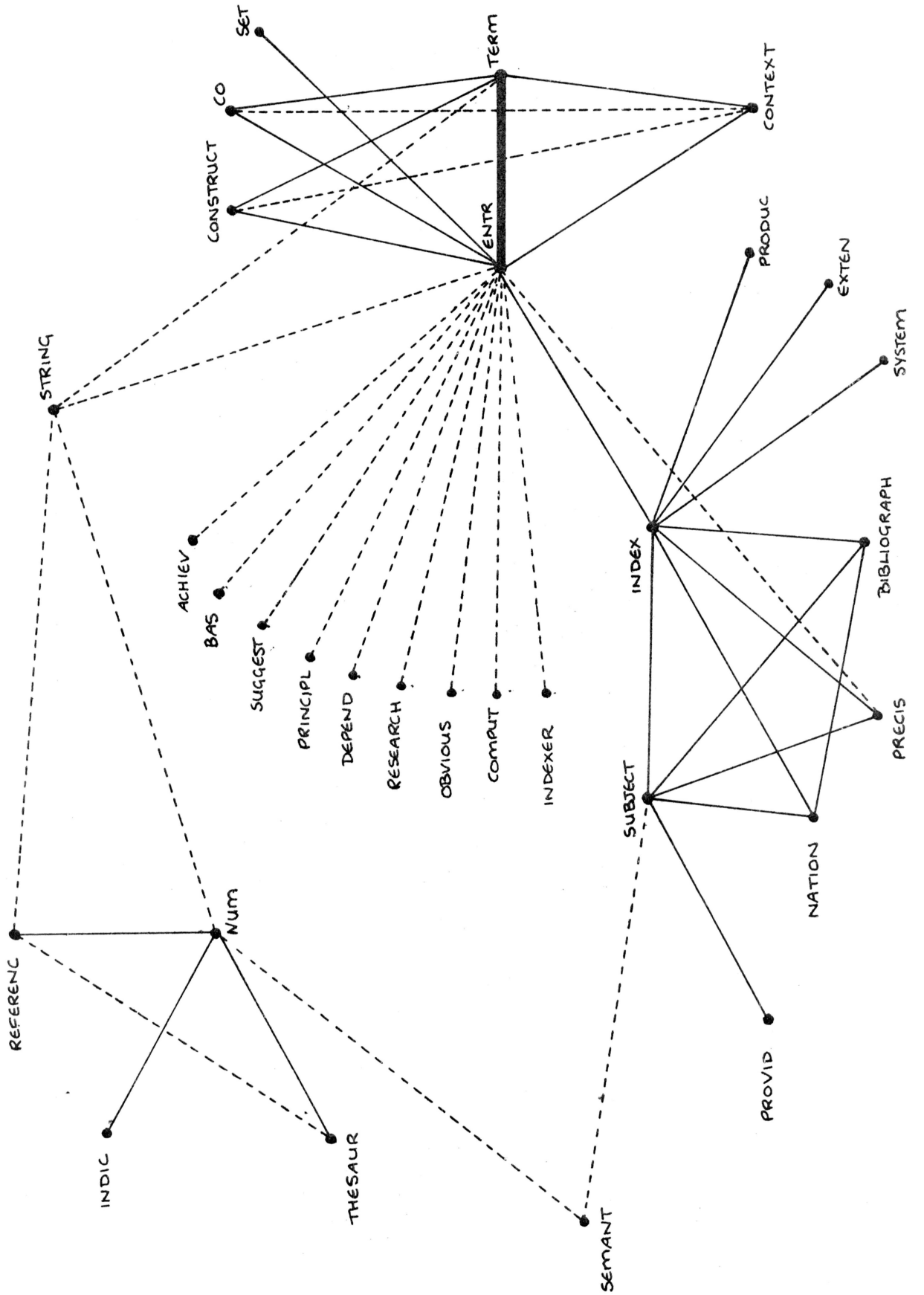
#17

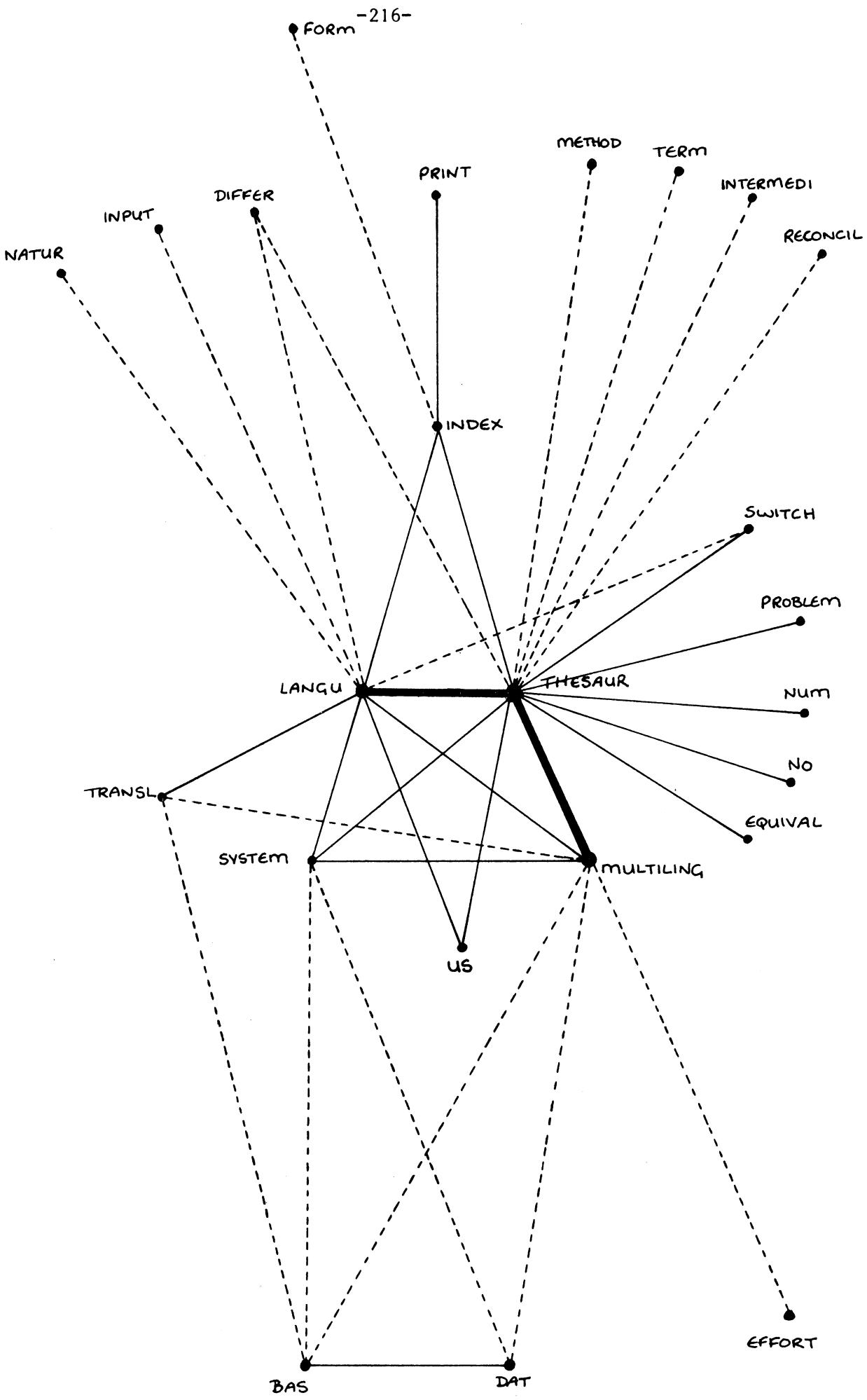


#19

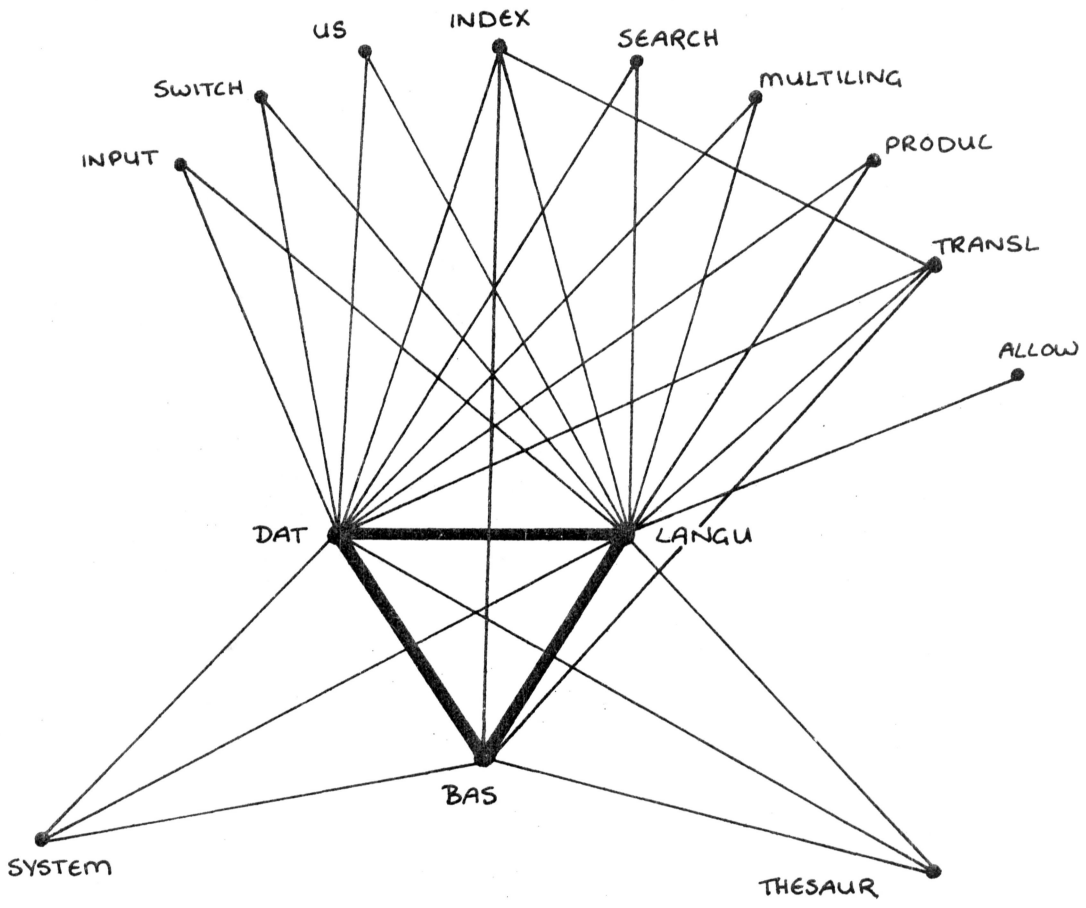

FUNCT
SPECIF
TECHNI
GENERAL
SEARCH
RELEVAN
WEIGHT
TERM
THEOR
METHOD
INFORMAT

※20


#21
SET
TERM
CO
CONTEXT
CONSTRUCT
ENTR
STRING
PRODUC
EXTEN
SYSTEM
ACHIEV
BAS
SUGGEST
PRINCIPL
DEPEND
RESEARCH
OBVIOUS
COMPUT
INDEXER
INDEX
BIBLIOGRAPH
PRECIS
SUBJECT
NATION
PROVID
REFERENC
NUM
INDIC
THESAUR
SEMANT

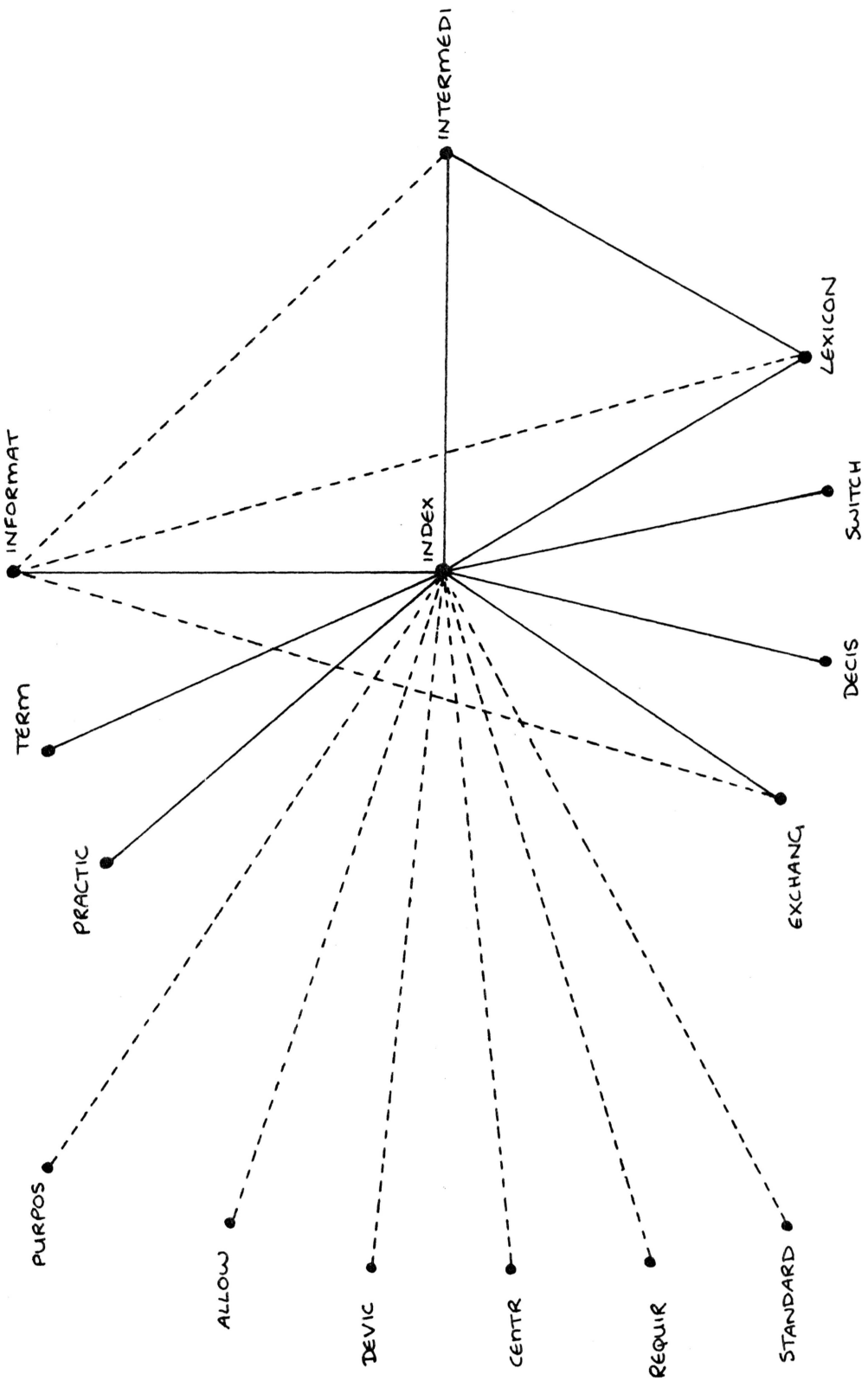

FORM
METHOD
TERM
PRINT
INTERMEDI
DIFFER
INPUT
RECONCIL
NATUR
INDEX
SWITCH
PROBLEM
LANGU
THESAUR
NUM
NO
TRANSL
EQUIVAL
SYSTEM
MULTILING
US
EFFORT
BAS
DAT

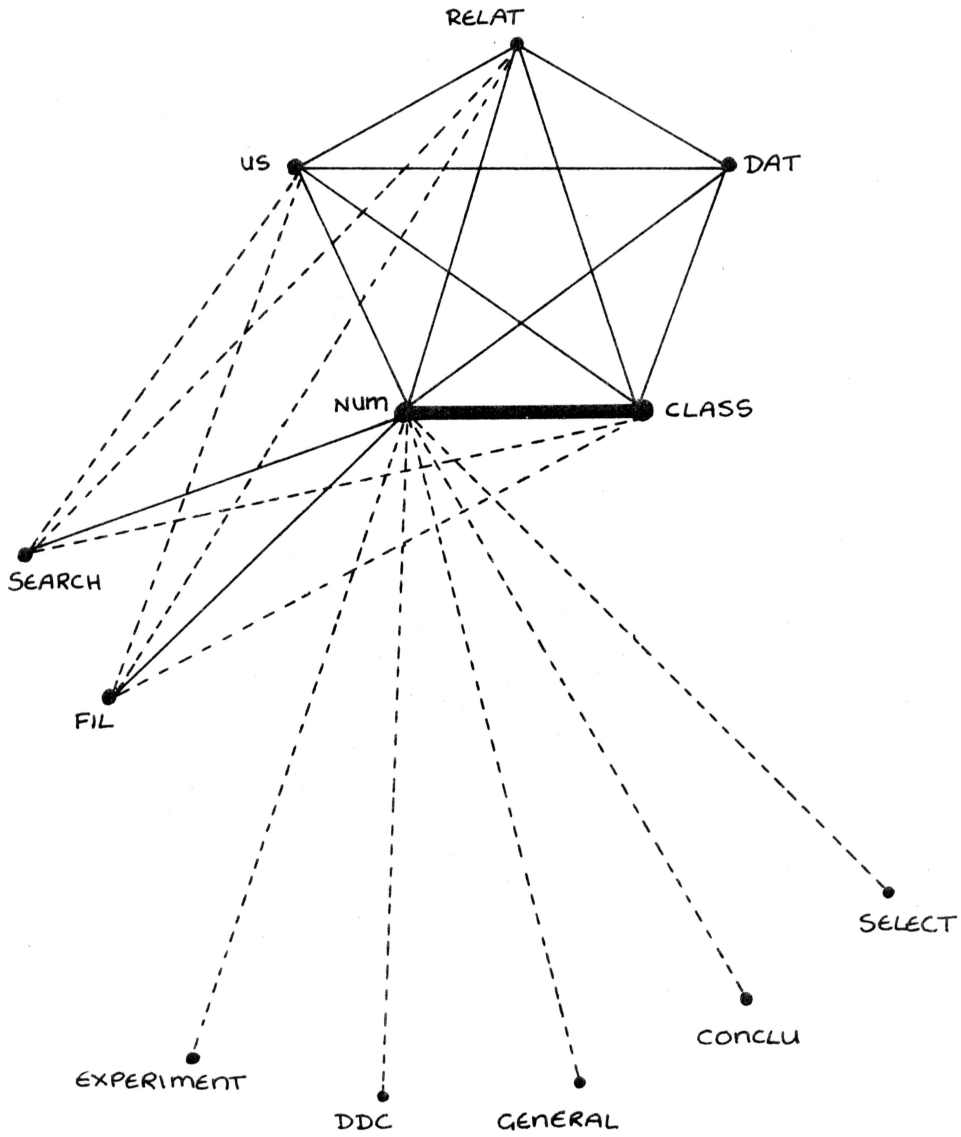
22


US
INDEX
SEARCH
SWITCH
MULTILING
INPUT
PRODUC
TRANSL
ALLOW
DAT
LANGU
BAS
SYSTEM
THESAUR

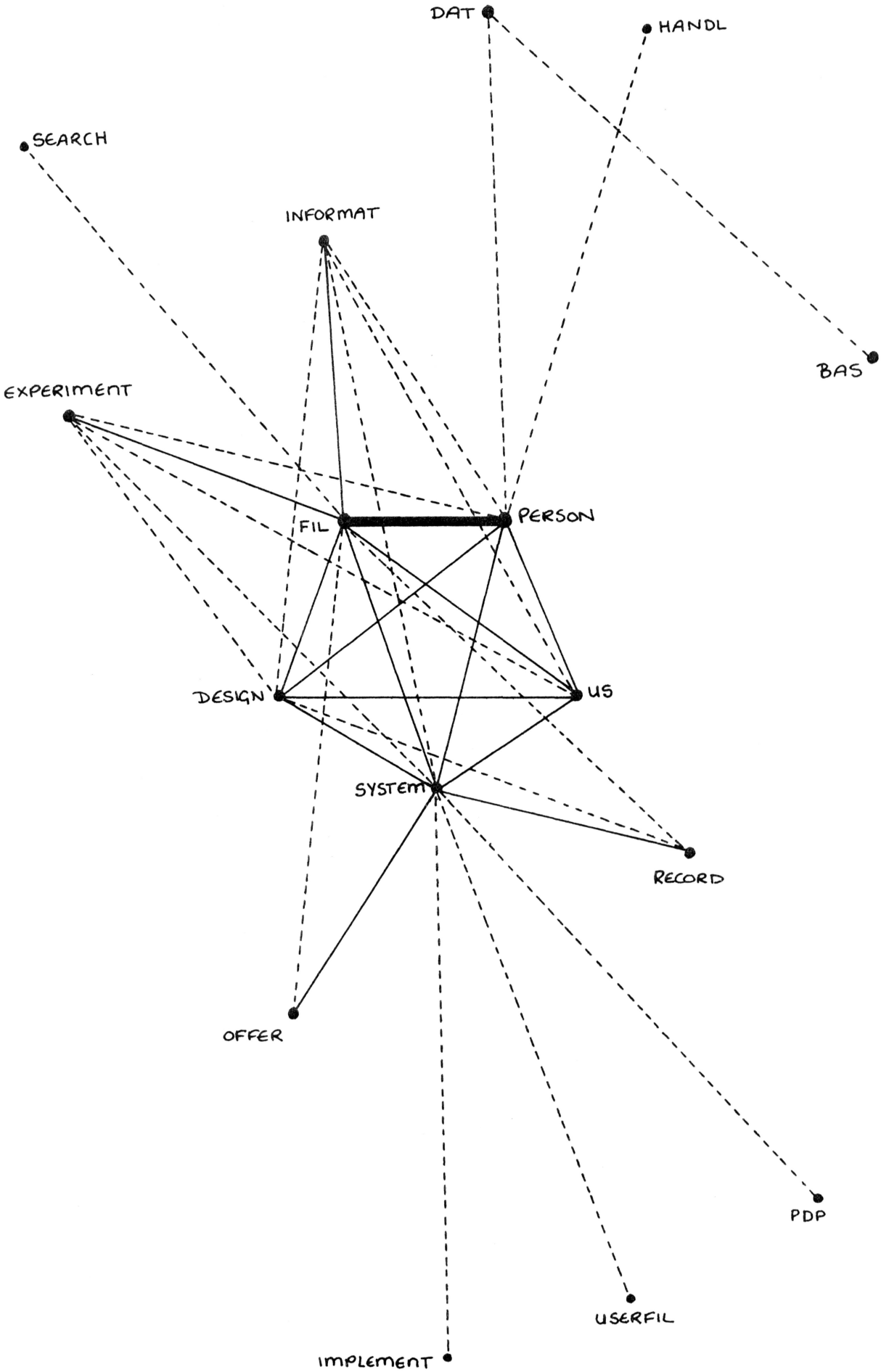
23

#24

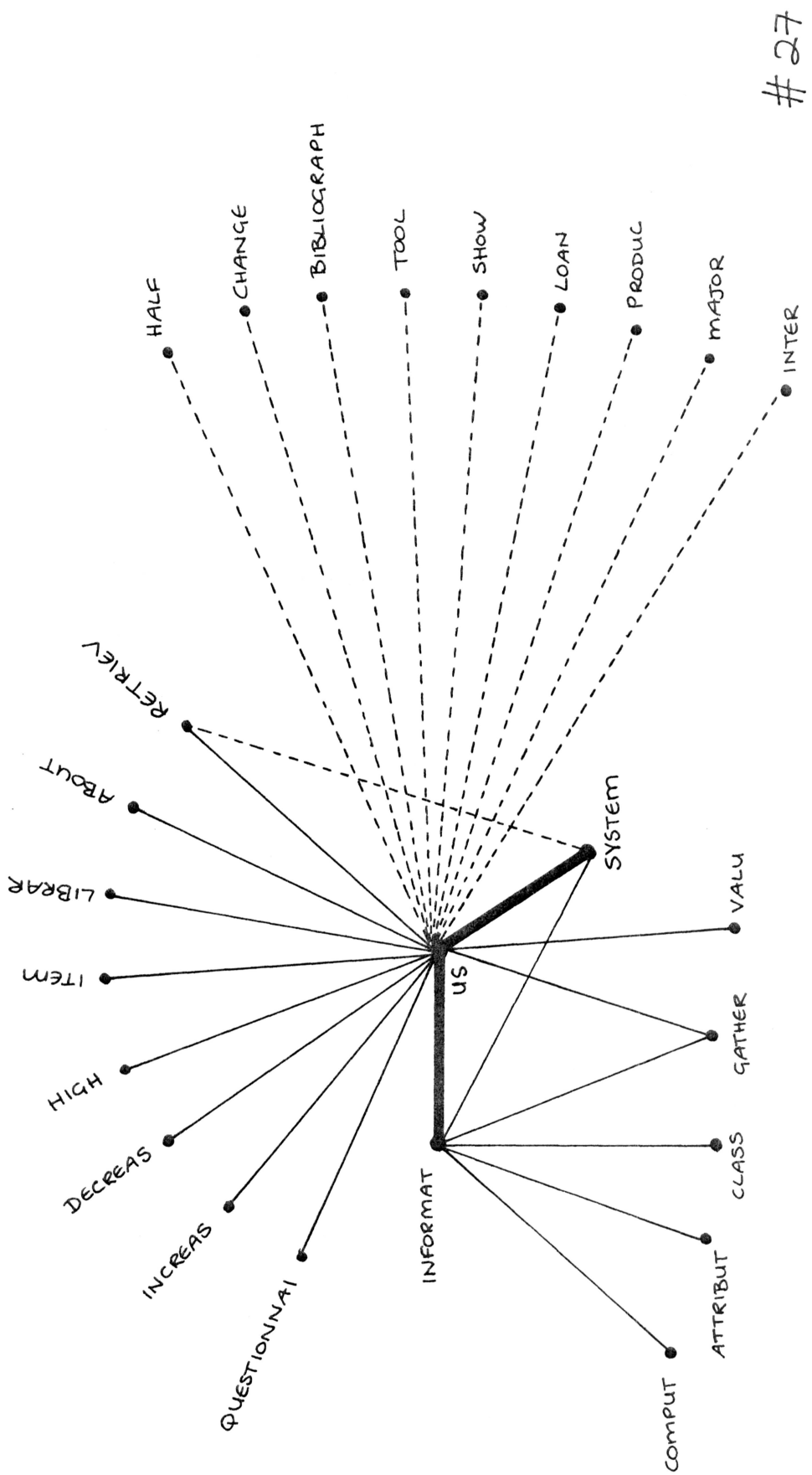

INTERMEDI
LEXICON
INFORMAT
INDEX
SWITCH
TERM
DECIS
PRACTIC
EXCHANG
PURPOS
ALLOW
DEVIC
CENTR
REQUIR
STANDARD

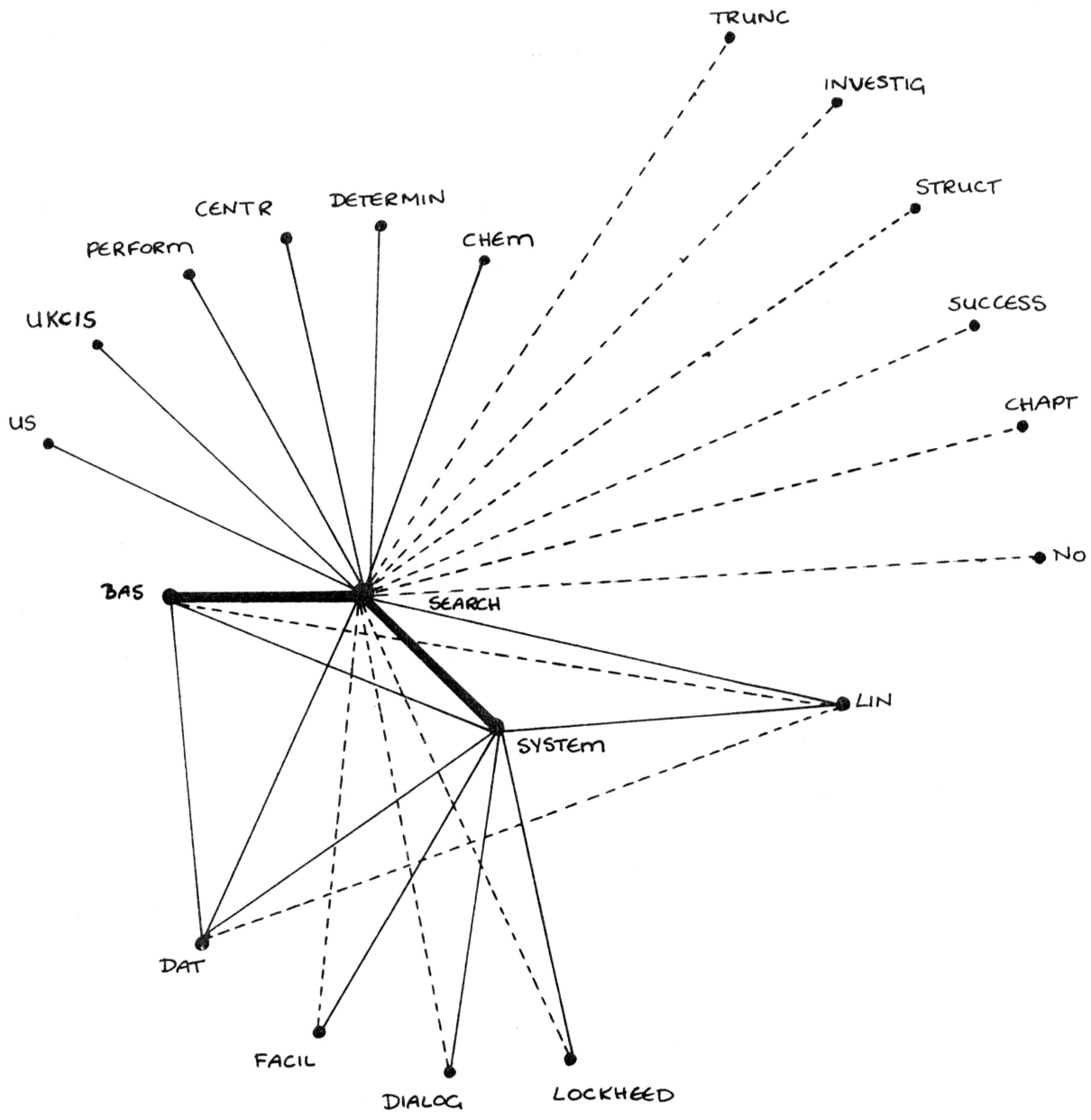

RELAT
US
DAT
Num
CLASS
SEARCH
FIL
SELECT
CONCLU
EXPERIMENT
DDC
GENERAL

#25

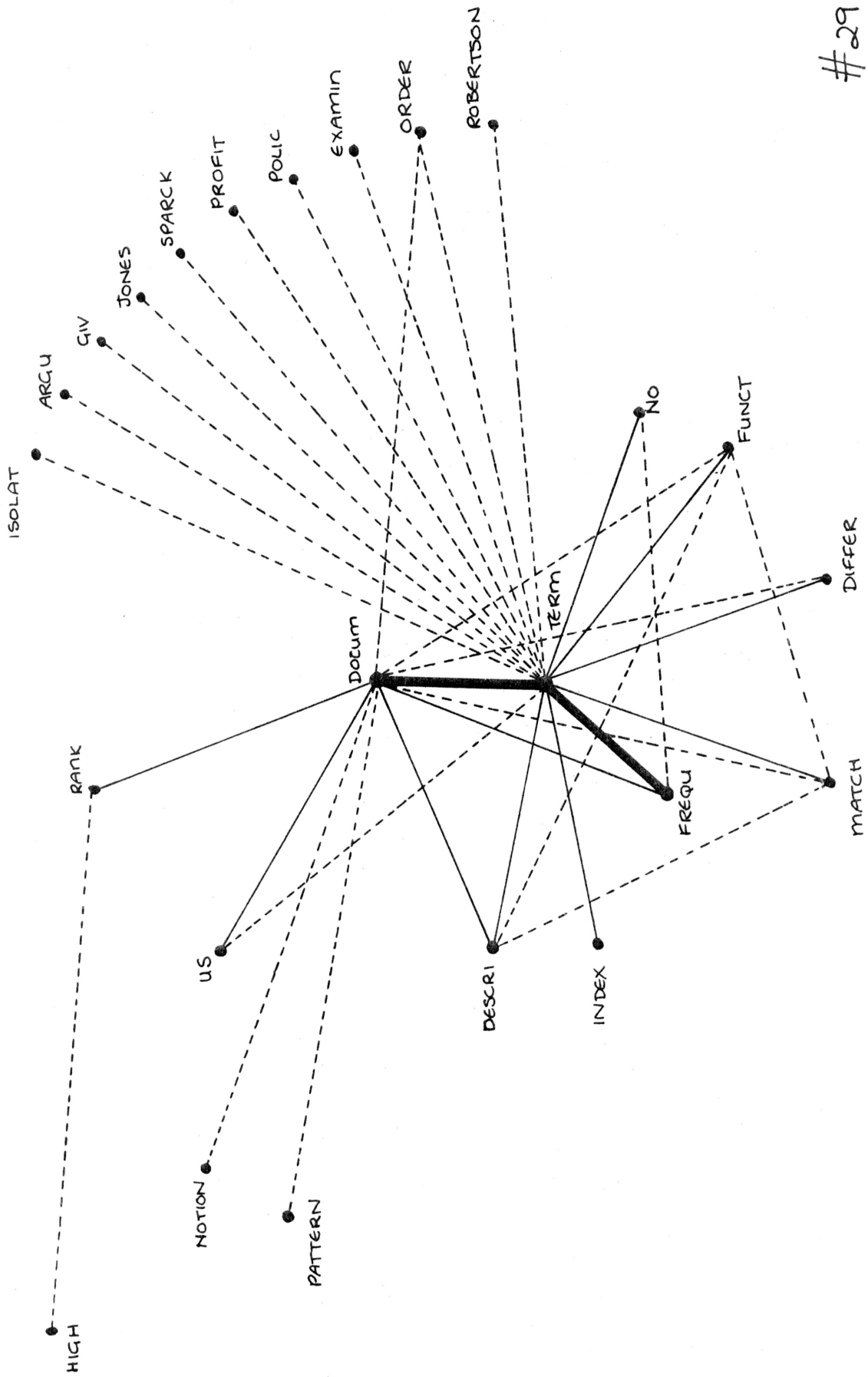

DAT
HANDL
SEARCH
INFORMAT
BAS
EXPERIMENT
FIL
PERSON
DESIGN
US
SYSTEM
RECORD
OFFER
PDP
USERFIL
IMPLEMENT

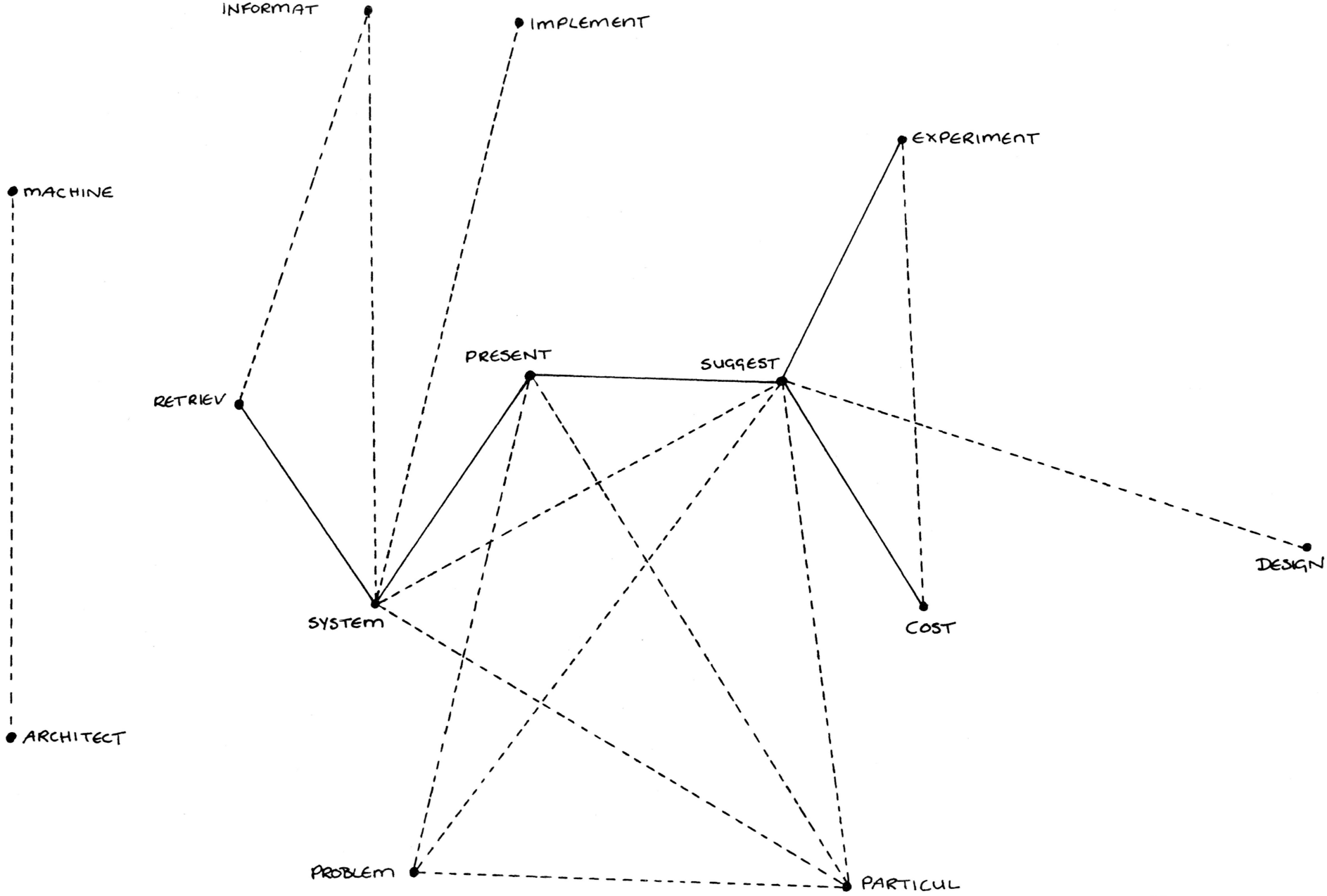
#26


#27
HALF
CHANGE
BIBLIOGRAPH
TOOL
SHOW
LOAN
PRODUC
MAJOR
INTER
RETRIEV
ABOUT
LIBRAR
ITEM
HIGH
DECREAS
INCREAS
QUESTIONNAI
SYSTEM
US
INFORMAT
VALU
GATHER
CLASS
ATTRIBUT
COMPUT

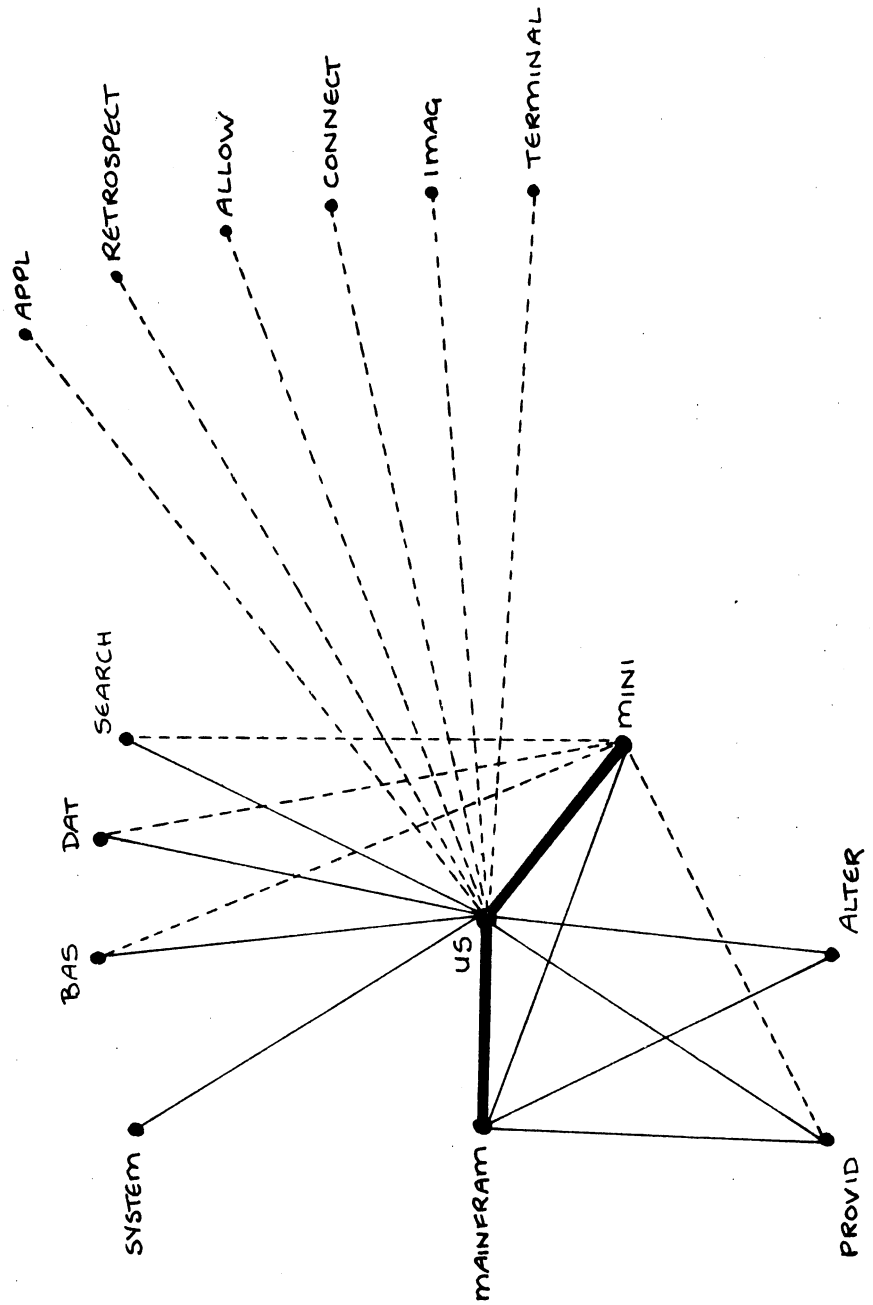

TRUNC
INVESTIG
STRUCT
SUCCESS
CHAPT
No
DETERMIN
CENTR
PERFORM
CHEM
UKCIS
US
BAS
SEARCH
LIN
SYSTEM
DAT
FACIL
DIALOG
LOCKHEED

※28

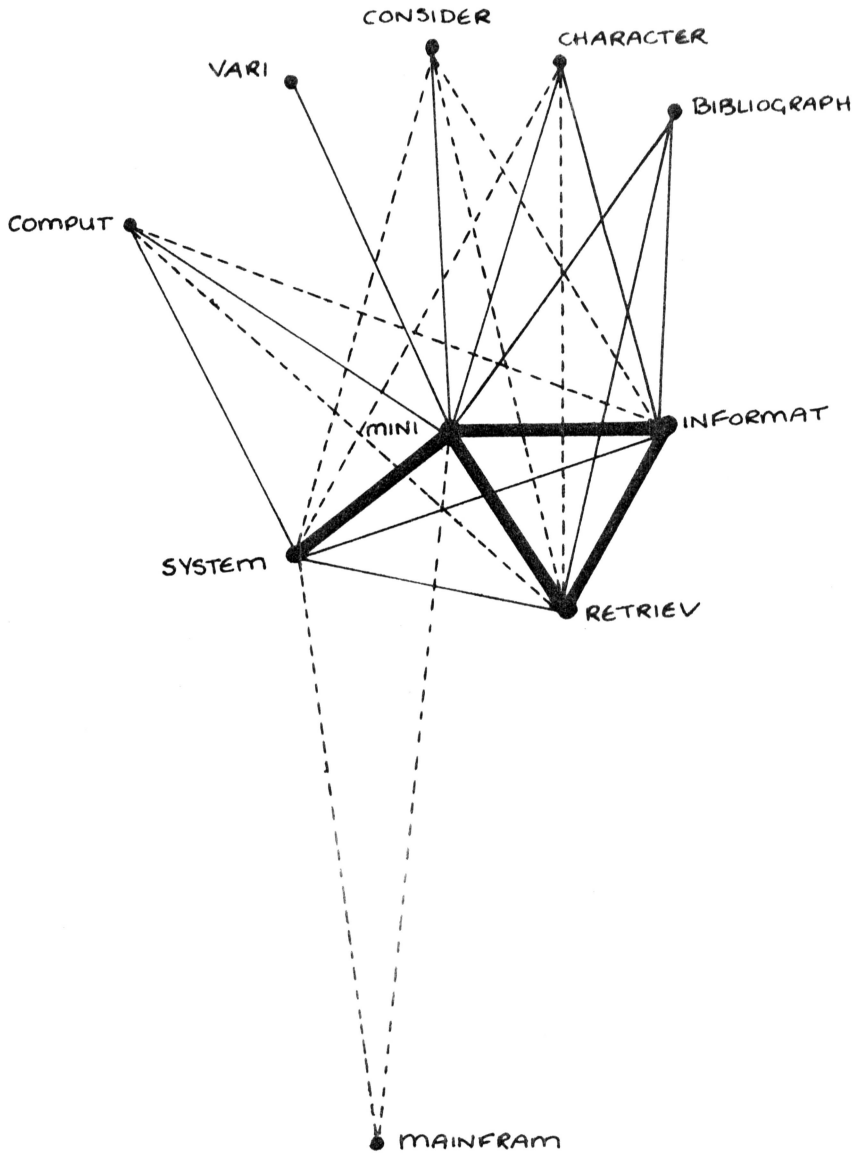

#29
ROBERTSON
ORDER
EXAMIN
POLIC
PROFIT
SPARCK
JONES
CIV
ARGU
ISOLAT
NO
FUNCT
DIFFER
TERM
DOCUM
RANK
FREQU
MATCH
US
DESCRI
INDEX
NOTION
PATTERN
HIGH


INFORMAT
IMPLEMENT
EXPERIMENT
MACHINE
PRESENT
SUGGEST
RETRIEV
DESIGN
SYSTEM
COST
ARCHITECT
PROBLEM
PARTICUL

#30


APPL
RETROSPECT
ALLOW
CONNECT
IMAG
TERMINAL
SEARCH
DAT
BAS
SYSTEM
US
MINI
ALTER
MAINFRAM
PROVID

#31


CONSIDER
CHARACTER
VARI
BIBLIOGRAPH
COMPUT
MINI
INFORMAT
SYSTEM
RETRIEV
MAINFRAM

#32