
TM-3304

# Dictionary Circles

10 January 1967





35

# TECHNICAL MEMORANDUM

(TM Series)

This document was produced in connection with a research project sponsored by SDC's independent research program.

## Dictionary Circles

Karen Sparck Jones

10 January 1967

SYSTEM
DEVELOPMENT
CORPORATION
2500 COLORADO AVE.
SANTA MONICA
CALIFORNIA
90406

( page 2 blank)


## ABSTRACT

If we say that two words are linked if one defines the other in a standard dictionary entry, it follows that we may find chains of word-senses which are successively linked in this way, and also circles or closed chains of linked senses. It is evident that these connections between different word senses in a vocabulary as a whole will constitute a network of variable density representing more or less closely associated sets of words. However, given a specific set of definitions of 'link', 'chain' 'circle' and so on, and a data base consisting of the relevant selections from a dictionary, the problem is to find procedures for discovering chains and circles from the linked word-senses given by dictionary material, which are both theoretically and practically satisfactory.* If we take the rather general definitions of these notions given by Olney* as a starting point. it turns out that several alternative search procedures can be associated with them, because they may be interpreted in different ways, which may give quite different results for the same body of data, depending on the restrictions on the starting point for a search, the mode of extension of a chain, and so on. Basically, a distinction can be made between an unrestricted attempt to find all the chains arising from each separate word in a set, and a more restricted attempt based on searching for all the chains linking word pairs in the set: the chains obtained for a given set of words may not be at all alike in the two cases, with corresponding effects on the circles. The first method is in fact very unrewarding, so that the second is the only real possibility: even here, however, there are choices to be made in the presentation of the initial lists of items which are immediately linked to each given word-sense, and in the modes of searching through successive lists to find a chain, if a suitable compromise between maintaining the source information intact, identifying all the (shorter) chains and (smaller) circles in it, and doing so in a finite time, is to be reached.


---

*Olney, J. C. A research plan for investigating the structure of an English lexicon. SDC document TM-(L)-3331 (in preparation).

(page 4 blank)

CONTENTS



APPENDIX

BLANK PAGE

## 1. INTRODUCTION

We can say that the semantic structure of a dictionary is the network of relations between word-senses given by the occurrence of one word in the definition of another. A special interest attaches to those points in the network where words are more closely related than usual, as is the case, for example, where word "a" defines word "b," and vice versa; and more generally, if we say that two words are linked if one defines the other, we can have chains of words which are so linked, because the first word is defined by the second, and the sense in question of the second by a third, and so on; and we can have a dictionary circle if a pair of words are linked by two different chains. Presumably relatively dense areas of the network, or lexical fields, consist of sets of words which are mutually connected in this way.

There is no need to say why we should be interested in these notions of link, chain, circle and field: anyone interested in semantics ought to be. If any attempt to analyze text semantically depends, as it appears to do, on the use of a semantic classification, we have to investigate the relational and classificatory structure of our vocabulary. There are various ways of doing this: one is to look at the actual behaviour of words, in terms of their occurrences, in actual text. Another is to study a good dictionary, since this can be regarded as a summary of large quantities of text-based information. The suggestion that this last might be attempted is put forward by Olney, where an investigation of dictionary circles in English using the information given by Webster's Seventh New Collegiate Dictionary (W7), is proposed. It is well known that there are many difficulties about constructing semantic classifications in general, especially if this is to be done automatically on a large scale. But it might appear that processing dictionary definitions to find links, chains and circles, might be more straightforward. It turns out, however, that this is not the case.

That dictionary processing is not so straightforward becomes clear if we take some plausible set of definitions of these concepts of link, chain, circle and so on, and see how they work out when we try to establish clear-cut procedures for finding circles in an actual dictionary. The definitions given by Olney, for instance, present a number of problems which have to be resolved if the attempt to find groups of related words in W7 is to amount to more than low-level sorting on an ad hoc basis. Of course, the practical difficulties which tend to arise when anything ambitious is attempted with so much data may mean that first experiments have to be fairly limited, though these may well produce a large number of interesting facts about English. The point is that the problems which arise when we think about looking for these groups of words in any other than a very restricted way are interesting ones, and so should be investigated.

In what follows, therefore, the notions of link, chain and circle as they may represent the relations between words in a dictionary are studied from the starting point presented by the proposals in Olney's paper. Much of the discussion will necessarily be very detailed, and may indeed appear excessively so, but my contention is that this is inevitable if the cash value of the general suggestion that actual dictionary material should be examined is to be obtained; and since this general suggestion is an attractive and interesting one for anyone concerned with vocabulary structure, we want to know what its cash value is, so that no apology for the detail which follows is needed.

The general questions of interest which arise from the approach to the analysis of dictionary material presented by Olney are:

(1) What are the consequences of the formal definitions of 'link', 'chain', and 'circle' given there?

(2) How do these formal definitions work out for someone who wants to obtain concrete lexical facts about a particular language from a given dictionary?

(3) How do they fit a realistic program for obtaining such facts?

The emphasis in Olney's report is on dictionary circles, so that the relationship between two words which are associated by a circle is regarded as particularly important: the other connections by links and chains are of interest mainly because circles depend on them. So that the problems just mentioned will be discussed in what follows in terms of methods of obtaining dictionary circles.

Since the discussion follows from Olney's, his original definitions are given for convenience here:

Let a be a sense of word A and b be a sense of word B.

Def: a is linked to b if either: (1) B is a content word and B is used in the description of a in W7 or A is a content word and A is used in sense a in the description of b in W7;

or (2) there is a content word C which is used in the same* sense in the W7 descriptions of both a and b.

*I.e., is used in a sense that falls under the same sense description in W7.

Def: $\alpha$ is a chain if $\alpha$ is an ordered n-tuple such that every element of $\alpha$ is a word-sense and the ith element of $\alpha$ is linked only to the (i-1)th and (i+1)th elements of $\alpha$, where $1<i<n$.

Def: $\alpha$ and $\beta$ are independent chains if $\alpha$, $\beta$ are chains having n and m elements respectively and it is not the case that there are elements $e_i$, $e_{i+1}$ in $\alpha$ and $e'_j$, $e'_{j+1}$ where $1\leq i<n$ and $1\leq j<m$, unless $e_1$ is linked to $e_{i+1}$ by virtue of condition (1) above and $e'_j$ is linked to $e'_{j+1}$ by virtue of condition (2) above or vice versa.

Def: a and b form a dictionary circle with respect to $\alpha$, $\beta$ if $\alpha$, $\beta$ are independent chains having n and m elements respectively, and element $e_1$ in $\alpha$ = a = element $e'_1$ in $\beta$, and element $e_n$ in $\alpha$ = b = element $e'_m$ in $\beta$.

These definitions do not obviously suggest any search procedures for chains and circles. In what follows I shall consider what sort of procedures for finding circles might follow from them in a straightforward way. It may be that what we get looks unsatisfactory, but if so we have to revise our definitions accordingly, or at least pin them down by associating them with specific search procedures.

Notes: (a) The definition of a two-step link in (2) above can be disregarded as redundant and confusing.

(b) The initial definitions in Olney's report have been given a special twist by Ziff; but he is primarily interested in the special case where two words define each other; and while philosophers and logicians can legitimately concentrate on this case, we should look at others as well.

## 2. POSSIBLE APPROACHES TO FINDING DICTIONARY CIRCLES

There are two of these:

- The exhaustive approach:

  Given a certain set of words, we can identify all the links, chains and circles arising from any members of the set; we can then select those circles in which some specific words that we are particularly interested in appear.

- The selective approach:

  We can take specific words we are particularly interested in, and find as many chains and circles connecting them as possible.

It is worth noticing that if circles are formed from chains, it is possible that we might find some chains which were subsequently eliminated through not forming circles satisfactorily. The questions of interest then are:

(a) whether we would get the same results from these two approaches with respect to particular words; and

(b) whether the definitions of link, chain and so on can be interpreted in the same way in each case.

## 3. OBTAINING DICTIONARY CIRCLES EXHAUSTIVELY

### 3.1 LINKS

The starting point for everything is the link between two word senses. Obviously, if the development of chains rests on the fact that if a * is linked to b, and b to c, we have the same sense of b each time. I shall assume in what follows that this is always the case, so that when I write "word" I mean word-sense, so that if a word is linked to two other words, we have one sense of it in each case, and one sense of the other words, and so on, consistently.

In TM-3331 two words are linked as noted above, that is either if a occurs in the definition of b (a defines b for short), or if b occurs in the definition of a (b defines a). We thus have asymmetrical links, though when we say simply that a and b are linked we appear to be dealing with a symmetrical relationship, and this ambiguity may cause trouble when we consider chains and circles. We can represent this definition of a link thus:

a and b are linked either (1) if a defines b = a - / b**

or (2) if b defines a = a / - b .

Then a - / b = b / - a

a / - b = b - / a .

These equivalences appear trivial, but it is easy to get in a mess thinking about chains and circles if we write something in one way rather than the other.

---

*Quotation marks are omitted for convenience.

**"- /" and "/ -" are used as typographically convenient equivalents for the more obvious "—>" and "<— ".

However, since a and b may be linked if either (1) or (2) holds, we can simply represent a link between a and b thus:

$$a - b = b - a.$$

3.2 CHAINS

The definition of a chain of words (Olney) follows naturally from that of a linked pair of words: let a two-word chain be either

$$a - / b$$

or $a / - b$ (with equivalences as above).

a three-word chain will be:

$$a / - b / - c = c - / b - / a$$
$$a / - b - / c = \text{etc.},$$
$$a - / b / - c$$
$$a - / b / - c .$$

a four-word chain will be:

$$a / - b / - c / - d = d - / c - / b - / a$$
$$a / - b / - c - / d = \text{etc.},$$
$$a / - b - / c / - d$$
$$a / - b - / c - / d$$
$$\text{etc, total } 16 .$$

etc.

This kind of thing is clearly going to get very complicated. In principle, moreover, we may have not merely directional links, but different types of link, each of which may hold in two directions: we could, for example, have a occurring as a synonym in the definition of b, or as in antonym, or b as a synonym of a in the definition of a, and so on. Such relationships are clearly very interesting, but become inordinately difficult to manipulate for more than two- or three-item chains. And since I am primarily interested in the general consequences of the notions of links, chains and circles, I shall exclude this possibility here. I shall simply say that a and b are linked, say, if b occurs in the definition of a, without saying how it occurs.

It may in fact be convenient, since directional links can become very awkward, to disregard them as well and consider only a link with direction unspecified, between two words. Each of the chains given above can therefore be abbreviated thus:

two-word a - b = b - a

three-word a - b - c = c - b - a

four-word a - b - c - d = d - c - b - a

etc. It must be emphasized, however, that though the direction of the links is suppressed, the order of the items in the chain must be preserved: in constructing it we work from one end, so that if we start with a, we attach b to it if b is linked to a, and then add c to the chain if c is linked with b. Thus though the words in the two following chains are the same, and two of them are similarly linked in each, the chains as wholes are not the same:

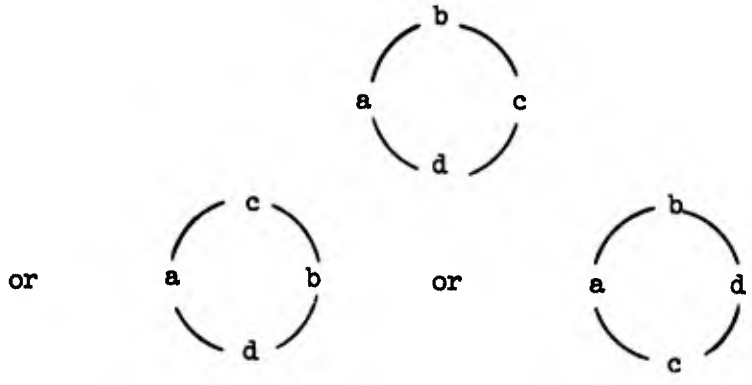
a - b - c - d

a - b - d - c .

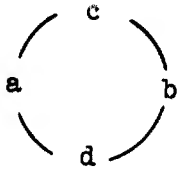
## 3.3 CIRCLES

Now consider the construction of circles. There are two ways of looking at a circle:

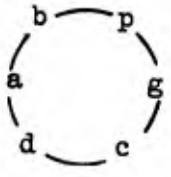
(1) As simply containing some specified item or items, where, even if we are interested in several items, these may or may not be all the items in the circle; thus a circle containing only a, b, c, d can be either



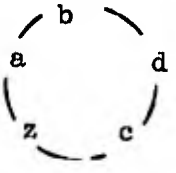
If we are interested in a specific order of the items, say, a, c, b, d, we may only have



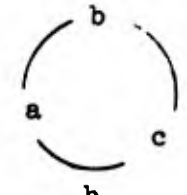

If other words are allowed as well, we may have, for example:




or

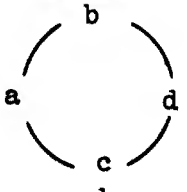



(2) As consisting of two combined chains whose ends correspond. In this case we are interested in two specific items, the end ones for the two chains: this is the approach adopted in Olney's paper. Thus for the two words a and b, we may have the following circles, derived from the indicated pairs of chains.



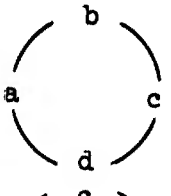

a - b

a - c - b



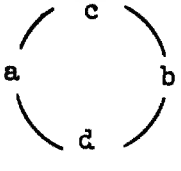

a - b

a - c - d - b



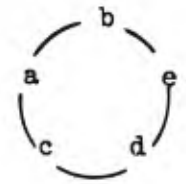

a - b

a - d - c - b



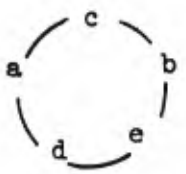

a - c - b

a - d - b




a - b

a - c - d - e - b



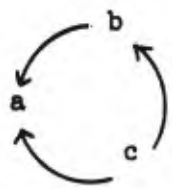

a - c - b

a - d - e - b

(The three four-point circles are the same as the first three given under (1), but this is incidental, since we are interested specifically in a, b, c, and d in the first case, but only in a and b in the second, and the fact that the circles obtained for them include c and d is immaterial.)

The difference between (1) and (2) is that looking at a circle as a pair of chains means that it must be constructed in a specific way: for (1) we could construct circles by using a single loop chain, or by combining two chains, or more than two: what is more important, we may get a circle containing specific words, and even two of them, without these having been the end words of the component chains. Thus if we are interested in a and b, we may get a circle containing a and b by method (1) from combining two chains like d - a - c and d - b - c. But since Olney's paper deals only with circles of type (2), where we have two words of interest and these are the end words of two combined chains, we can confine ourselves to these in what follows. It should however be noted that this difference is important chiefly from the point of view of procedures for finding circles: if we had found every circle in our dictionary by some method or other, we could look for circles containing a specified pair of words tout court since a circle is a circle once it is formed.

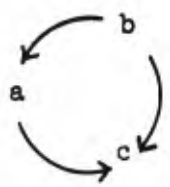
It must, of course, be remembered that simplifying the link reduces the number of circles: if we have directional links, we can have different circles containing the same words, like




a / - b

a / - c - b

and




a / - b

a - / c / - b

## 3.4 PAIRED-CHAIN CIRCLES

Confining ourselves now to circles formed from paired chains, the first point is that the smallest possible such circle, which consists of two two-word chains, is special since we can only obtain it by specifically taking directional links into account. To combine the two chains

a - b

and

a - b

meaningfully, we must in fact have

$$a - / b$$

and

$$a / - b .$$

The two-point circle, in other words, depends on our explicitly knowing not simply that a and b are linked, but that a defines b and b defines a.

I shall not say more about this special circle: though it may be of philosophical interest in connection with definition, from the point of view of dictionary circles in general, it is both quirky and uninteresting. The interest and problems of dictionary circles come with the larger circles.

Suppose, then, that we look for two-chain circles using a realistic example,* to make the whole thing more concrete, as follows:

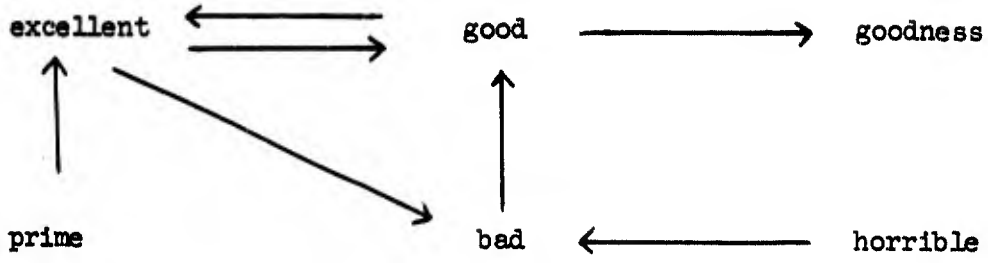
We have some information about the relations between a number of words which are given by their occurrences in one another's dictionary definitions, which we have obtained simply by picking out the 'content words' in the sense descriptions in question: thus if we have the entry "a: a kind of b," for instance, we say that a and b are related. The actual definitions may involve different kinds of relation, say if one word is said to be the opposite of another, and so on, but these details can be suppressed here to avoid complexity, though for a really refined vocabulary description one would have to take them into account. Our illustrative selections are as follows:

| | | |
|---|---|---|
| good | : | excellent, bad |
| goodness | : | good |
| excellent | : | good, prime |
| bad | : | excellent, horrible . |

That is, we have a sense of "good" which is defined by both "excellent" and "bad," a sense of "excellent" which is defined by "good" and "prime," and one of "bad" which is defined by "excellent" and "horrible." I am assuming, moreover, that each word has only one sense thoughout, so that it is the same sense of "bad" which defines "good" and is defined by "horrible." We can represent the relations between these words as given by their occurrences in the various definitions in a diagrammatic way as follows, x —> y being interpreted as x defines y since it occurs in the sense description of y. In principle, such a diagram is what we want to finish up with, to show the circles we have found, but it serves a useful expository purpose here.

---

* It was taken from an actual dictionary, but I cannot remember which; the fact that I may not have an accurate description of the specific English words in question is irrelevant from the present point of view.




Looking at this diagram, we would intuitively say that we ought to find a two-member circle containing "good" and "excellent," and two three-member circles containing "good," "excellent" and "bad." (They differ in the direction of the link between "good" and "excellent.") If our procedure for finding circles does not turn up these, there is something wrong.

With our definition of a link between two words, we can now obtain word pairs from this information as follows:

| | | |
|---|---|---|
| good | /- | excellent |
| good | /- | bad |
| goodness | /- | good |
| excellent | /- | good |
| excellent | /- | prime |
| bad | /- | excellent |
| bad | /- | horrible. |

Since chains and circles depend on linked words, this list of word pairs provides the basic information from which we hope to obtain circles; or to put it differently, it constitutes the vocabulary description in which we hope to find circles. And the whole process of identifying circles is a matter of operating on the various pairs in the list.

So when we now try to set up chains, what do our original definitions imply as to the way we do this? In fact, all that we are told is that a chain consists of conjoined linked pairs, the only restriction being that no word should occur twice in one chain. In other words, if we have word a with b linked to it, and we have another pair consisting of b with c linked to it, we can, if we have the same sense of b in each case, form the chain a - b - c; and presumably we should pursue any chain as far as we can on this basis. Suppose, then, that we set out to make the longest possible chains by stringing together <u>any</u> linked pairs in our set which have a common member. We may, for example, given that each word is used in the same sense throughout our list, start with "goodness" and proceed to "good," and then since

"good" is defined by "excellent" add the latter, and since "excellent" defines "bad," add "bad," and then since "bad" is defined by "horrible" add this, after which we can proceed no further. This gives us the chain

goodness / - good / - excellent / - bad / - horrible .

Or if we start with "excellent," which defines "good," we can add the latter, then take "bad," which defines "good," and then take "horrible," which defines "bad," to get

excellent - / good / - bad / - horrible .

The important point is that the definitions of link and chain as they stand permit us to start with any word in our set and take any successive pairs with common items to form our chain. It is possible that the idea of looking for the longest chains is a mistake, but there is no reason for saying so yet.

On this basis, then, and trying every word in our list as a starting element, we can obtain 39 chains altogether, including three with only two members, though we can reduce the number to 30 since several are simply reversed duplicates. The reduced list is given as Fig. 1 (see Appendix).

Now when we look at these chains, there are none with corresponding end words which do not also have common subchains, i.e., strings of two or more words linked in the same way, as appears in Fig. 2; but two chains with corresponding ends cannot form a circle unless they are independent, that is have no common subchains, and indeed no common words other than their end ones. From these chains as they stand, therefore, we cannot obtain any circles, though this is intuitively very unsatisfactory when we think of the interconnected character of the diagram illustrating the ways in which the words in our set are linked.

Could one, therefore, nevertheless get some chains by either (1) processing these chains in some way; or (2) revising the method of obtaining them, within the general framework of our exhaustive approach?

(1) In principle, we might process these chains either by (a) eliminating some of them; or (b) using only parts of each chain.

. Elimination

Given two or more chains with corresponding ends and also common subchains, we might eliminate all but one member of such sets; but the possible criteria for doing this, like retaining the longest, do not seem workable in practice, or leave us with no chains at all.

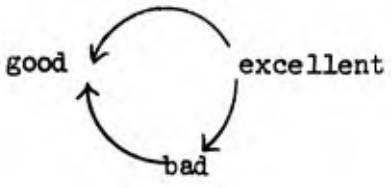
. Selection

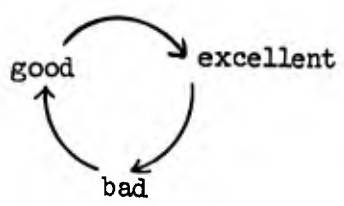
Given some chain which has a common subchain with another, we might delete this common subchain, and retain only the remainder as a chain; but again the attempt to do this leaves us with no chains at all. Alternatively, we might select any common subchains and treat them as chains. If we do this in this case, we are at least left with some chains, but none of them will form a circle. (See Fig. 3.)

So that the only conclusion we can come to is that applying the formal definitions in the most straightforward and unrestricted way is liable not to give us any circles.

(2) Now consider restricting the formation of chains. Suppose that we start not simply with a list of all our pairs, without notes of their source definitions, as we have done so far, but with a list which records the source of each pair, so that when we derive chains, we have an additional restriction to the effect that we cannot combine pairs from the same entry, which is after all a very reasonable restriction. Thus if we start with "excellent," which defines "good," we cannot add "bad," since this also defines "good" in the same entry, but we can add "goodness," We cannot, that is, use more than one pair from any entry in a given chain. If we do this for our example, and we again look for the longest chains, we get 13 different ones. But the attempt to form circles from them fails because there are few chains with corresponding ends, and none of these are independent, as is shown in Fig. 4. And the alternative of taking only chains formed within an entry is even more unrewarding.

A further alternative is to abandon the search for the longest chains only and to look instead for all the chains up to and including a specified length (which is justified by some independent argument). Suppose, for example, that we say that we will take all the linked word pairs as two-word chains, and also all the three-word chains, subject, say, to the sensible restriction just mentioned that only one link in a three-word chain can be obtained from a particular entry. We are regarding a two-word chain as separate from a three-word one of which it is a subchain; and we are disregarding the fact that a three-word chain might in principle be extended. If we do this for our "good" example we get 7 two-word chains, and 10 three-word ones (see Fig. 5) which give us three circles, 1 two-word one for "good" and "excellent," and two, with differently directed links, for "good," "bad" and "excellent," thus:



Note that these are the circles we earlier said intuitively that we should get. On the other hand, the restriction to chains of a certain length is very artificial, and requires solid justification, as does the treatment of sub-chains of longer ones as independent entities. The actual search procedure is also rather uneconomic, since the same circle may be found with different pairs of chains for its component words.

## 4. CONCLUSION ON THE EXHAUSTIVE APPROACH

It is wholly unprofitable if pursued in an undiscriminating way: we failed, for instance, to get even the two-word "good"-"excellent" circle earlier; this method will only give results if it is restricted in a somewhat artificial way, and even then it is rather effortful.

We now have to ask, therefore, what the alternative selective approach will do for us. It seems that it should avoid the difficulties we have just encountered: if we are looking for a chain between two specified words, a and b, we will not meet the particular problem raised by searching for the longest chains, namely that we can have two chains a - b - c - b and a - e - c - f, say, which ought to give us a circle from a - b - c and a - e - c which we, in fact, miss through extending our chains beyond c. But we should equally avoid arbitrary upper limits on our chains: if we look for all the chains between all the pairs of words in our set, this should naturally include those whose end words happen to be given by looking for all the chains of a specified length from each word. So, hopefully, the selective approach will work better than the one we have just considered; in particular, though we might not do this in practice, if we were to look for all the chains between all the pairs of words, up to chains of a certain preassigned length to avoid the extremes we would get with a large collection of words, we would have a kind of exhaustive approach, though a more sensible one than that just considered because both ends of each chain are specified and not only one.

How, then, might we proceed to find circles on this basis?

## 5. OBTAINING DICTIONARY CIRCLES SELECTIVELY

The contrast between approaches in Sec. 3. and Sec. 5., is between getting all the chains we can, and especially the longest possible ones, from some initial dictionary information, and then selecting the ones which happen to have ends that we like; and starting from end points that we like and finding chains between them, which may well not give us the longest chains we could generate from our initial word pairs. In each case, though, what we have available are sets of word pairs, and forming chains is in principle a matter of scanning the set and combining appropriate items. But it is nevertheless convenient to think of the second approach as being much more directly concerned with dictionary entry list manipulations than the first. So in what follows the whole business of constructing circles will be reconsidered.

## 6. LISTS

We will say that a dictionary list contains a given word, the entry word, and those other words which define it; we will imagine for simplicity's sake that a word has only one sense, so that the list contains items defining this sense. Thus, if we refer back to our previous example, we have two lists for "good" and "bad":

| | | | |
|---|---|---|---|
| good | excellent | bad | good |
| | bad | | horrible |

In principle, as we have seen, two words a and b are linked either if a defines b or vice versa, i.e., in this case either if b occurs in the list for a, or if a occurs in the list for b. However, correlating this notion of a link with an exhaustive but comparatively rational procedure for operating on lists to find circles may present many problems unless a list is interpreted not as containing only the words defining a given word, but as containing all the words which are linked to a given word because the latter appears in their lists: thus if x defines y, x and y are linked, and y should then appear in x's list, as well as x appearing in y's; and in general, if x and y are linked either because x defines y or y defines x, each word should appear in the other's list. If this is not done, our defining lists are asymmetrical, and must be used as such consistently; and the difficulty is that a simple procedure for using such lists will almost certainly give incomplete results, while any attempt to construct circles in a thorough way from them is liable to involve appalling complications. On the other hand, duplicating all the original dictionary information may be objected to both because it is an effort, and because there is a danger that the initial facts about the relations between words which were actually found in the source dictionary will be obscured. In these circumstances, therefore, we have to choose the lesser of two evils, and it seems that it is probably preferable to restrict the use of lists by permitting only asymmetrical ones, and to accept that only a subset of the links which actually hold between the words concerned will actually be picked up. In this way, at least the original dictionary information will be preserved and the search procedure kept fairly simple. These remarks, of course, only apply if we want to find long (or long-ish) chains and large (or large-ish) circles in a fairly thoroughgoing way.

However, given that we form a circle by combining two chains, it is clear that the single-link chain must have a special status, since it seems reasonable to say that we have a single-link chain either if a is in the list for b or vice versa: the preceding remarks are concerned with longer chains, and the point about single-link chains between two specified words a and b is that they may be handled differently from the linked word pairs involving any two words which are simply components of longer chains. It is, of course, the case that the two-word circle must also be special, as before; and I shall, in fact, disregard it in this context, for the same reasons as before.

What we are interested in now, therefore, is obtaining circles consisting of two chains whose end points are specific words, from these lists; so that what we want are procedures for starting at one end and reaching the other, via a series of words which are linked by the occurrence of one in the list for another. In principle, we are interested in any such chains (up to some defined upper limit of length); in any case, even if we are primarily concerned only with the shorter ones, we still have to have a procedure which will find a number of chains from which we select the shorter ones, unless we deliberately confine ourselves to quite short chains, like one- or two-link ones, to start with.

Suppose, then, that we have the following lists, where the left-hand item in each case is the entry word, and the right-hand ones occur in its definition, "-" being used for words we are not interested in here:

| a | b | c | d |
|---|---|---|---|
| - | - | - | - |
| b | c | b | a |
| - | - | | |
| c | d | | |
| - | | | |

The definition of a link is as before, namely that x and y are linked if either occurs in the other's list. Thus a and b and a and c are linked, and b and c are linked, in fact both because c occurs in b's list and b in c's.

But when we come to constructing two-chain circles from these lists, there are two alternative ways of doing it.

Before we consider these, however, it should be noted that there are consequences of the use of lists for finding compatible chains which apply in both cases, as follows: If a circle is defined as consisting of two independent chains with corresponding ends, then with respect to all circles containing more than three elements, pairs of chains of different lengths are possible. For a three-element circle, the only possibility is that one chain will be a two-word, i.e., one-link, chain, and the other a three-word, i.e., two-link, one, thus:

a - b giving

a - c - b

But for all larger circles, alternative combinations are possible: thus for a four-element circle we may have either a one-link chain and a three-link one, or two two-link ones; and, in general, we will have as many pairs of chains for a circle with a given number n of elements as it is possible to divide n:

thus

for n = 5, we have 1 and 4
or 2 and 3 ;

for n = 6, we have 1 and 5
2 and 4
3 and 3 ,

etc.

This point is obviously of practical significance in considering search programs.

## 7. CHAINS AND CIRCLES

Turning now to the formation of circles, the two methods of obtaining chains depend on one or the other of the following ways of using two lists to link three elements together. Suppose that we wish to find a chain between a and b, when there is no single link between them. The alternatives are:

(1) Take the list for a and the list for b, and see if they contain a common element. For example:

| | | | | | |
|---|---|---|---|---|---|
| a | - | b | - | gives | a b |
| | c | | c | | \ / |
| | - | | - | | c |

.

We can call this the V method.

(2) Take the list for one element, say a, and taking the list for one of its members, see if b occurs in the latter. For example:

| | | | | | |
|---|---|---|---|---|---|
| a | - | c | - | gives | a — c |
| | c | | b | | —b |
| | - | | - | | |

.

We can call this the L method.

In these two cases we get the same chain between a and b, namely a - c - b. But note that the two results have been obtained with different sets of lists: in the first case we used the lists for a and b, in the second those for a and c; and note further that the two sets of lists given for (1) and (2) respectively will not give the results obtained under (2) and (1) respectively, at least if these are the only lists relevant to a and b that we have to work from. There may obviously be cases where we would get the same result either way from the same initial information, as would be the case if we had the three lists:

| | | |
|---|---|---|
| a - | b - | c - |
| c | c | b |
| - | - | - |

though even here we are using different parts of the whole in the two procedures. It is, however, clear that we cannot rely on this happening, and that we are therefore liable to get different results from the same set of lists according to which method we use to form our chains.

Further still, note that (1) is a symmetrical method, since we start at both ends of the chain at once, while (2) is asymmetrical since we start at one end and follow a chain to see if we meet what we want at the other; and in this case we may get different results according to which end we start. Thus, if we have the three lists:

| | | |
|---|---|---|
| a - | b - | c - |
| c | d | b |
| - | - | - |

we get our chain a - c - b if we start from a, but not if we start from b.

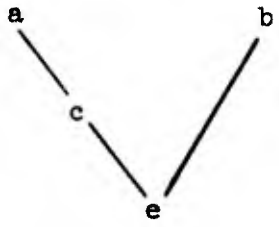
It should also be emphasized that the symmetrical method (1) has to be carried through consistently for longer chains (and this is more than a simple matter of programming convenience).

Thus given the lists:

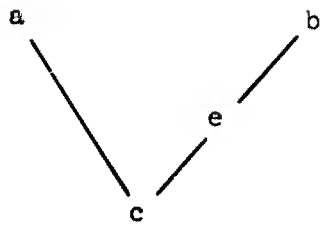
| | | | |
|---|---|---|---|
| a - | b - | c - | d - |
| c | d | e | e |
| - | - | - | - |

we proceed from a and b simultaneously to comparing the lists for c and d, to get the chain a - c - e - d - b. But note that for an even number of words in a chain we necessarily have an asymmetrical last step, which we can illustrate

thus




or




I have not given the direction of the links in these examples; but it may well be necessary with this kind of list manipulation to preserve the direction of the links in a chain since this is so closely associated with its derivation: though the direction in principle follows from the method of derivation it may sometimes be unclear in detail, as would be the case with chains of the kind just illustrated. If the direction is not noted, this information might be lost, even if the method of derivation was known, say if a chain were reversed to close a circle.

Given these alternative methods of obtaining chains, we can now consider their effect on an attempt to find circles. Suppose, for instance, that we try to find circles with our earlier example using "good," "goodness" and so on. The dictionary lists in this case are

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| good | excellent | goodness | good | excellent | good | bad | excellent |
| | bad | | | | prime | | horrible |

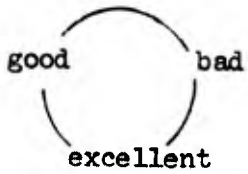
1. Suppose that we specify "good" and "bad" as the words we are interested in:

   (a) if we use the V method we get the two chains

   good - bad

   good - excellent - bad

   and therefore have the circle



(b) If we use the L method we get nothing.

2. Now suppose that we specify "good" and "excellent" as our words. We can indeed get a two-point circle for them, but we can disregard this to see if we can find anything else more interesting:

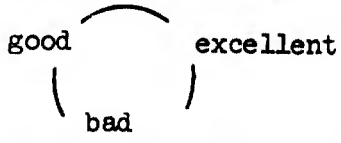
(a) If we use the V method, this time we get nothing,

(b) but if we use the L method we get

good - excellent

good - bad - excellent

and therefore have the circle

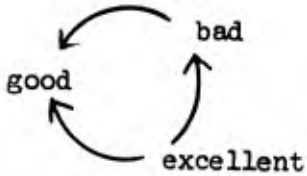



In fact, we have two circles, because "good" and "excellent" are linked two ways, so that we have two two-word chains for them, giving the circles



Note, however, that we would not get anything if we chose to start with "excellent": though perhaps this means that we should take special steps to see if there is a two-word chain starting from either end, whatever word we start with for longer chains.

A point of interest about this example is that the two circles where the link direction are suppressed is the same. But if directed links are used, though we get two circles under 2(b), we get only one of them under 1(a), namely




None of the other possible end words in this example will give us anything.

The disparity between the results obtainable from the same initial data by the two methods is better illustrated by a larger example, as follows:

Suppose that we have dictionary lists for some imaginary words a, b, c, d, e, f, g, h, as follows:

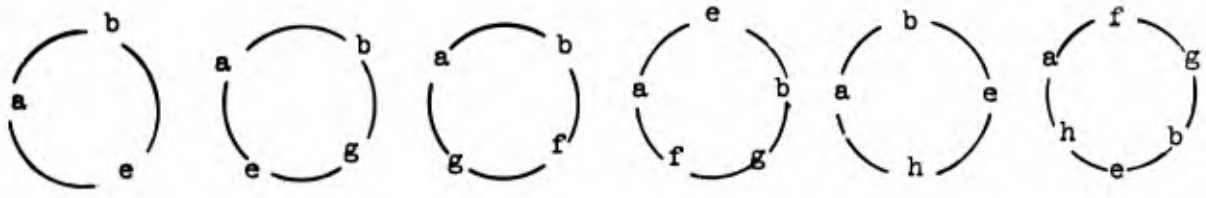
| a | b | b | - | c | d | d | f | e | b | f | g | g | b | h | e |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | d | | | | e | | | | c | | | | | | |
| | e | | | | | | | | f | | | | | | |
| | f | | | | | | | | g | | | | | | |
| | h | | | | | | | | | | | | | | |

1. Take a and b as interest words.

   V method: we get a - b only, therefore no circles

   L method, from a: we get a - b
   e - e - b
   a - e - g - b
   a - f - g - b
   a - h - e - b ;

   the independent pairs among these chains give the circles

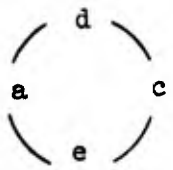



   from b: we get nothing (except that we can use a - b as b - a), and so cannot obtain any circles.

2. Take a and c as interest words.

V method: we get a - d - c
a - e - c

giving the circle




L method from a: we get a - e - c
a - d - e - c

which are not independent, so we get no circles.

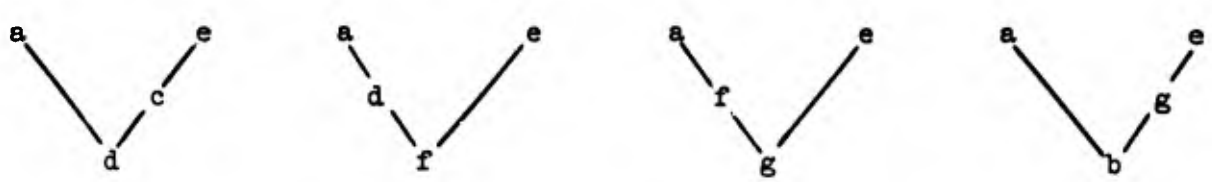
from c: we get nothing.

3. Take a and e as interest words.

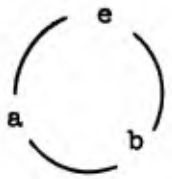
V method: we get a - e
a - b - e
a - f - e
a - d - c - e
a - d - f - e
a - f - g - e
a - b - g - e *

---

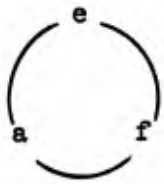
*The actual formation of these four-member chains is:



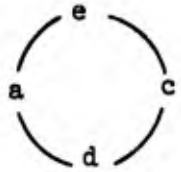
a - e
a - b - e



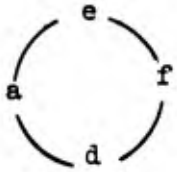

a - e
a - f - e



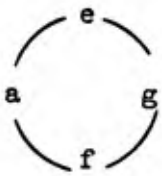

a - e
a - d - c - e



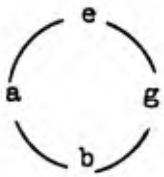

a - e
a - d - f - e



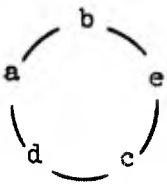

a - e
a - f - g - e



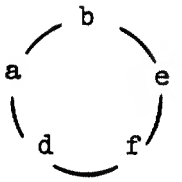

a - e
a - b - g - e



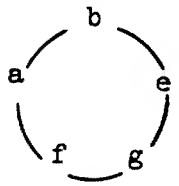

a - b - e
a - d - c - e



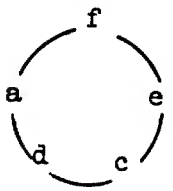
a - b - e
a - d - f - e



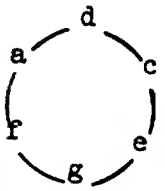

a - b - e
a - f - g - e



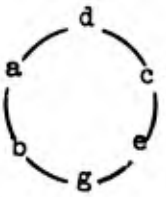

a - f - e
a - d - c - e



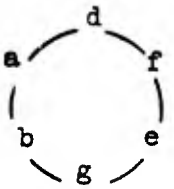

a - d - c - e
a - f - g - e




a - d - c - e
a - b - g - e

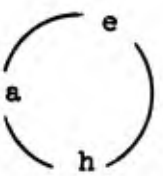



a - d - f - g
a - b - g - e




L method, from a: we get a - e
a - h - e

giving the circle




from e: we get nothing (except e - a = a - e).

The fact that these two methods of obtaining chains, and so of forming circles, can give quite different results is somewhat distressing; and we may therefore feel that there is some point to the earlier suggestion that all the links in an initial set of dictionary entries be duplicated, since this will presumably enable us to obtain the complete set of circles which are jointly obtainable by both methods, with either one of them. Suppose, therefore, that we duplicate all the links, but mark the duplicates in some way to distinguish the parts of each list which are obtained in this way from those given by the original source dictionary definitions. Thus for our last example, we would have a revised set of lists as follows:

| a | b | b | a' | c | d | d | a' | e | a' | g | b | h | a' |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | d' | | e' | | e | | c' | | b | | c' | | e |
| | e | | g' | | | | f | | c | | e' | | |
| | f | | | | | | | | f | | f' | | |
| | h | | | | | | | | g | | | | |
| | | | | | | | | | h' | | | | |

If we now take a and b as the words we are interested in, we find that we can obtain all the chains we obtained by the L method but missed with the V method, by the V method; and for a and e, we get those we obtained by the V method by the L method as well. But note that it is now possible to obtain further chains by either method which were not obtainable before, such as a - d - c - e - b. This is disagreeable, since one feels that such chains, which could not be obtained from the original lists as straightforwardly derived from the source dictionary, must be ersatz ones. Of course, if we restrict ourselves only to short chains, say of two or three words, this duplication would not produce these complications, and would not involve much work. If one is interested only in short chains, the duplication is an obvious way of ensuring that one gets all such chains.

The alternative is to search by both methods together. Let 'Def. x' be the list for x, and 'Def. Def. x' the list formed by combining all the lists for members of the list for x. Then the combined procedure for finding chains between a and b will be:

Take 'Def. a' and look for b; and vice versa; a positive result gives the one-link chain for a and b. If the result is negative, take 'Def. a' and look for 'Def. b'; this is the V method for a two-link (i.e., even-numbered chain). If no result is obtained take 'Def. Def. a' and look for b; this is the L method for a two-link chain from a; taking 'Def. Def. b' and looking for a is the reverse. If no result is obtained take 'Def. Def. a' and look for 'Def. b,' and 'Def. Def. b' and look for 'Def. a', for a three-link chain by the V method. (Both searches are required because the number of links is odd.) If no

results are obtained take 'Def. Def. Def. a' and look for b, or the reverse, for a three-link L chain from a, or b; and so on. Obviously this is going to involve a very large number of list comparisons if longer chains are accepted, though if we are interested only in shorter ones it is the natural solution to the problem of ensuring that we have got everything. Though note that in this case it will almost certainly be more economical to duplicate the original information and search by one method better than search by both.

## 8. CONCLUSION ON THE SELECTIVE APPROACH

It is considerably better than the exhaustive approach, but is not without its problems.

## 9. OVERALL CONCLUSION

Some form of the second, that is selective, method is clearly the only genuine possibility, both from the lexicographic and computing point of view; but it is evident that any such procedure must be a restricted one in some way, and since this is the case it is probably desirable to incorporate the restrictions in the initial definitions in some way. How the whole is worked out, however, depends considerably on strictly independent considerations concerning the interest or utility of longer chains and larger circles. If one believes that from a semantic point of view the relationship connecting the end points of longer chains is rather remote, or that the relationships between a variety of words connected only by longer chains or larger circles is fairly loose, this may or may not be a justification for concentrating only on limited chains and circles; but the point has to be argued on its own merits, and this is outside the scope of the present note. Such considerations are naturally especially relevant when we look at the further suggestion that sets of words or word senses which are mutually linked by chains and circles for lexical or semantic fields: clearly there is something to the argument that the members of such fields should be fairly close to one another. But it may also be the case that there are significant relationships between items which are not so close, and that these have to be taken into account in one's picture of the vocabulary. The way in which chains and circles form fields, however, is so large a topic that I shall not go into it here.

## APPENDIX

good - / goodness
excellent / - prime
bad / - horrible
good - / excellent / - prime
good / - excellent / - prime
good - / excellent - / bad / - horrible
good / - excellent - / bad / - horrible
good / - bad / - excellent / - prime
good / - bad / - horrible
excellent - / good / - bad / - horrible
excellent / - good / - bad / - horrible
excellent - / good - / goodness
excellent / - good - / goodness
excellent - / bad / - horrible
excellent - / bad - / good - / goodness
bad - / good - / excellent / - prime
bad - / good / - excellent / - prime
bad - / good - / goodness
bad / - excellent / - good - / goodness
bad / - excellent - / good - / goodness
bad / - excellent / - prime
goodness / - good / - excellent / - prime
goodness / - good - / excellent / - prime
goodness / - good - / excellent - / bad / - horrible
goodness / - good / - excellent - / bad / - horrible
goodness / - good / - bad / - excellent / - prime
goodness / - good / - bad / - horrible
prime - / excellent - / good - / goodness
prime - / excellent / - good - / goodness
prime - / excellent - / good / - bad / - horrible
prime - / excellent / - good / - bad / - horrible
prime - / excellent - / bad / - horrible
prime - / excellent - / bad - / good - / goodness
horrible - / bad - / good - / excellent / - prime
horrible - / bad - / good / - excellent / - prime
horrible - / bad - / good - / goodness
horrible - / bad / - excellent - / good - / goodness
horrible - / bad / - excellent / - good - / goodness
horrible - / bad / - excellent / - prime

Figure 1. Longest Chains Obtained by the Exhaustive Method

| | |
|---|---|
| good : prime | good - / excellent / - prime |
| | good / - excellent / - prime |
| | good / - bad / - excellent / - prime |
| good : horrible | good - / excellent - / bad / - horrible |
| | good / - excellent - / bad / - horrible |
| | good / - bad / - horrible |
| excellent : horrible | excellent - / good / - bad / - horrible |
| | excellent / - good / - bad / - horrible |
| | good / - bad / - horrible |
| excellent : goodness | excellent - / good - / goodness |
| | excellent / - good - / goodness |
| | excellent - / bad - / good - / goodness |
| bad : prime | bad - / good - / excellent / - prime |
| | bad - / good / - excellent / - prime |
| | bad / - excellent / - prime |
| bad : goodness | bad - / good - / goodness |
| | bad / - excellent / - good - / goodness |
| | bad / - excellent - / good - / goodness |
| goodness : prime | goodness / - good / - excellent / - prime |
| | goodness / - good - / excellent / - prime |
| | goodness / - good / - bad / - excellent / - prime |
| goodness : horrible | goodness / - good / - excellent - / bad / - horrible |
| | goodness / - good - / excellent - / bad / - horrible |
| | goodness / - good / - bad / - horrible |
| prime : horrible | prime - / excellent - / good / - bad / - horrible |
| | prime - / excellent / - good / - bad / - horrible |
| | prime - / excellent - / bad / - horrible |
| good : goodness | good - / goodness |
| excellent : prime | excellent / - prime |
| bad : horrible | bad / - horrible |

Figure 2. Longest Chains With Corresponding End Words

good - / goodness
excellent / - prime
bad / - horrible
good / - excellent - / bad
good - / excellent - / bad
good / - bad / - excellent

Figure 3. Selected Common Subchains of the Longest Chains

good - / goodness
excellent / - prime
bad / - horrible
good / - excellent / - prime
good / - excellent - / bad
excellent - / good - / goodness
bad - / good - / goodness
bad - / good - / excellent
excellent / - good - / goodness
good - / excellent - / bad
prime - / excellent - / bad
good / - bad / - excellent
good / - bad / - horrible

Figure 4. Chains Formed Under Restrictions on the Use of Entries

Chains:

good / - excellent
good / - bad
goodness / - good
excellent / - good
excellent / - prime
bad / - excellent
bad / - horrible

good / - excellent / - prime
good / - excellent - / bad
excellent - / good - / goodness
good / - bad / - excellent
good / - bad / - horrible
bad - / good / - goodness
bad - / good - / excellent
goodness / - good - / excellent
good - / excellent - / bad
prime - / excellent - / bad

Matching pairs:

good / - excellent - / bad
good / - bad

good - / excellent - / bad
good / - bad

good / - bad / - excellent
good / - excellent

good / - bad / - excellent
good - / excellent

bad - / good - / excellent
bad / - excellent

Figure 5. Two- and Three-Word Chains

BLANK PAGE

Unclassified
Security Classification

## DOCUMENT CONTROL DATA - R&D

*(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)*

| | |
|---|---|
| 1. ORIGINATING ACTIVITY (Corporate author)<br>System Development Corporation<br>Santa Monica, California | 2a. REPORT SECURITY CLASSIFICATION<br>Unclassified<br>2b. GROUP |
| 3. REPORT TITLE<br>Dictionary Circles | |
| 4. DESCRIPTIVE NOTES (Type of report and inclusive dates) | |
| 5. AUTHOR(S) (Last name, first name, initial)<br>Jones, Karen S. | |
| 6. REPORT DATE<br>10 January 1967 | 7a. TOTAL NO. OF PAGES<br>34 / 7b. NO. OF REFS |
| 8a. CONTRACT OR GRANT NO.<br>Independent Research for Scientific Community<br>b. PROJECT NO.<br>c.<br>d. | 9a. ORIGINATOR'S REPORT NUMBER(S)<br>TM-3304<br>9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) |
| 10. AVAILABILITY/LIMITATION NOTICES<br>Distribution of this document is unlimited | |
| 11. SUPPLEMENTARY NOTES | 12. SPONSORING MILITARY ACTIVITY |

13. ABSTRACT

If we say that two words are linked if one defines the other in a standard dictionary entry, it follows that we may find chains of word-senses which are successively linked in this way, and also circles or closed chains of linked senses. It is evident that these connections between different word senses in a vocabulary as a whole will constitute a network of variable density representing more or less closely associated sets of words. However, given a specific set of definitions of 'link', 'chain' 'circle' and so on, and a data base consisting of the relevant selections from a dictionary, the problem is to find procedures for discovering chains and circles from the linked word-senses given by dictionary material, which are both theoretically and practically satisfactory. If we take the rather general definitions of these notions given by Olney as a starting point. It turns out that several alternative search procedures can be associated with them, because thay may be interpreted in different ways, which may give quite different results for the same body of data, depending on the restrictions on the starting point for a search, the mode of extension of a chain, and so on. Basically, a distinction can be made between an unrestricted attempt to find all the chains arising from each separate word in a set, and a more restricted attempt based on searching for all the chains linking word pairs in the set: the chains obtained for a given set of words may not be at all alike in the two cases, with corresponding effects on the circles.

DD FORM 1 JAN 64 1473

Unclassified
Security Classification

| 14. KEY WORDS | LINK A | | LINK B | | LINK C | |
|---|---|---|---|---|---|---|
| | ROLE | WT | ROLE | WT | ROLE | WT |
| Word Senses<br>Chains of Word-Senses<br>Vocabularies<br>Dictionary Circles | | | | | | |

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through _____."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through _____."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through _____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS), (S), (C), or (U)*

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.