

Appendix II

ESTABLISHING THE CONTEXTUAL MEANING OF WORDS

by  
C. Wordley

May, 1959

The purpose of this paper is to establish the meaning of the word "pen" in the two following contexts:

- 1) The pen is in the inkstand.
- 2) The inkstand is in the pen.

At a preliminary glance at these two sentences, one would think that the thesaurus method would render the meaning of "pen" in each of these sentences as "plume". Before making this assertion, one should be familiar with the translation programme of the Cambridge Language Research Unit. Using this programme, I will describe a test that will establish the contextual meaning of "pen" in each of these sentences,

The part of the programme with which we are concerned is the limiting of the "fan" uses<sup>(1)</sup> of the words in a sentence by intersections. In the programme, this algorithm is used for finding the meaning of the input language "chunk" that is applicable to the context in which it occurs. This algorithm may be stated as follows:

The punched-cards representing words of the input language are firstly intersected, on their Thesauric Dictionary entries, over a large piece of input text (a paragraph). The most frequently occurring thesaurus heads are found (these heads have the greatest relevance to this particular text).

These heads are then punched onto a "grid" card<sup>(2)</sup>, and this "grid" card is intersected with each of the word cards. For each intersection a card is produced on which is punched the "meet" of these two cards. This produced pack of cards is stored.

Each of the sentences is then bracketted<sup>(3,4)</sup>. The word cards of each sentence are then intersected only within their immediate bracketting constituents. The intersections of each word card with each of its bracketting constituents are noted on one card. This card is then compared with the equivalent card produced in the first series of intersections. If

these cards do not give the same meaning for the word being considered, the second card is preferred; if they give the same meaning, but one of them has a few more head specifications, the "join" of the two cards is taken. If one of the cards gives more than one meaning of the word, the "meet" of the two cards is taken.

The final specification is then compared with the output language dictionary, and the equivalent card, or the nearest equivalent card, is found and this will give the output word, or words.

A test on the above indicated lines was carried out for each of the sentences. The English "Fan" Dictionary (i) was searched and the cards representing "pen", "in" and "ink-stand" were drawn out. (It is not necessary to consider the other words in the sentences as they do not affect the meaning of them.) The entries for the above words are as follows:

1) PEN

<u>ROGET HEAD</u>	<u>NUMBER</u>	<u>CODING NUMBER</u> <sup>(ii)</sup>
Region	181	161
Inclosure	232	204
Limit	233	205
Writing	590	445
Book	593	445 (List)
Hinderance	706	532
Restraint	751	566
Prison	752	566

2) IN

Existence	1	11
Intrinsically	5	15
Completeness	52	58
Component	56	60
Inclusion	76	74
Contents	190	168
Receptacle	191	168
Nearness	197	174
Contiguity	199	176
Interiority	221	196

i) The English Fan Dictionary consists of a pack of punched-cards, which represent words in the English Language. The information on these cards define the words in terms of thesaurus heads. Each hole punched in the cards represent a thesaurus head.

ii) The coding number which represents the heads when punched on the cards is not the same as the head number in Roget. This is due to the fact that with our present scheme, there are only 780 bits on the punched-cards that may represent the 1,000 heads in Roget. It was necessary therefore to compact Roget (5,6,7).

Centrality	222	197
Investment	225	200
Interjacency	228	205
Inclosure	232	204
Concavity	252	223
Direction	278	244
Approach	286	252
Convergence	290	256
Arrival	292	258
Ingress	294	260
Reception	296	262
Insertion	300	266
Qualification	469	370
Meaning	516	400
Method	627	473
Importance	642	484
Conduct	692	522
Restraint	751	566
Prison	752	566

3) INKSTAND

Permanence	141	121
Receptacle	191	168
Writing	590	445

As the largest piece of text for both parts of the text is the sentence, no long range intersections can be carried out. (Although an intersection over the sentence was carried out, mainly through curiosity).

Results of intersections are listed below:

<u>Intersection</u>	<u>Result (Coded No.)</u>	<u>Thesaurus Head</u>
PEN $\cap$ IN	204	Inclosure
	566	Restraint, Prison
PEN $\cap$ INKSTAND	445	Writing, book
IN $\cap$ INKSTAND	168	Receptacle

From the above results it may be observed that the thesaurus technique gives both the "plume" meaning and "enclosure" meaning of pen when intersections are carried out over the whole of each of the sentences.

With the introduction of a simple bracketting procedure <sup>(3,4)</sup>, the following results are obtained. The sentences are bracketted as follows:

1. ((THE PEN)/(IS (IN (THE INKSTAND))))
2. ((THE INKSTAND)/(IS (IN (THE PEN))))

In the first sentence, it is noted that from the bracketting, "pen" would not be intersected with any of the other meaningful chunks in the sentence. Therefore no resolving of the meanings of this word may be obtained; the original specification is kept (which give all the meanings of pen). As the sentence is the only context that is given, no long range

card is available for comparison\*. It is most certain that any context in which this sentence may occur would have the general theme of "writing" and "book", therefore these would be amongst the most frequent thesaurus heads. (This means that from the "meet" of the long range intersections card and the immediate bracketting constituents intersection card, which is the original specification, the "plume" meaning of "pen" would be obtained.

In the second sentence, "pen" would be intersected with "in", as "in" is one of "pens" immediate bracketting constituents. This intersection would restrict the meaning of "pen" to "inclosure", "restraint", and "prison". If the long range intersection differed from this meaning, this second would be preferred. It shows therefore that the correct meaning of "pen" would be obtained in this intersection process using the thesaurus method of translation.

This short test shows that the thesaurus method can distinguish the different meanings of the same words when they appear in very similar contexts.

---

\* It is a rule with the thesaurus method that intersections take place over long ranges of text. As only one sentence was given as the example, the full value of these intersections cannot be obtained. Therefore the method used had to do the best it could over this short sentence.

C. Wordley.  
Cambridge Language  
Research Unit.

References:

- 1 Fans and Heads. M. Masterman.
2. General Schema of a Thesaurus Translation programme using a punched-card technique. by Joy M. Staniforth.
3. The A.F. Parker-Rhodes Translation Procedure. A. F. Parker-Rhodes.
4. A Procedure for analysing form-class sequences using punched-card machinery. M. Kay.
5. Coding system (Compacted Roget. A New Proposal).
6. Can some Thesaurus Heads be Superimposed on Others? M. Shaw.
7. Compacting Roget's Thesaurus. M, Shaw.

INK-STAND

FAN CARD

1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8

HOLLERITH 4-252 THE BRITISH TABULATING MACHINE CO., LTD. 17 PARK LANE LONDON W.1. PRINTED IN ENGLAND

IN

FAN CARD

1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8

HOLLERITH 4-252 THE BRITISH TABULATING MACHINE CO., LTD. 17 PARK LANE LONDON W.1. PRINTED IN ENGLAND

IN

FAN CARD

1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8

HOLLERITH 4-252 THE BRITISH TABULATING MACHINE CO., LTD. 17 PARK LANE LONDON W.1. PRINTED IN ENGLAND