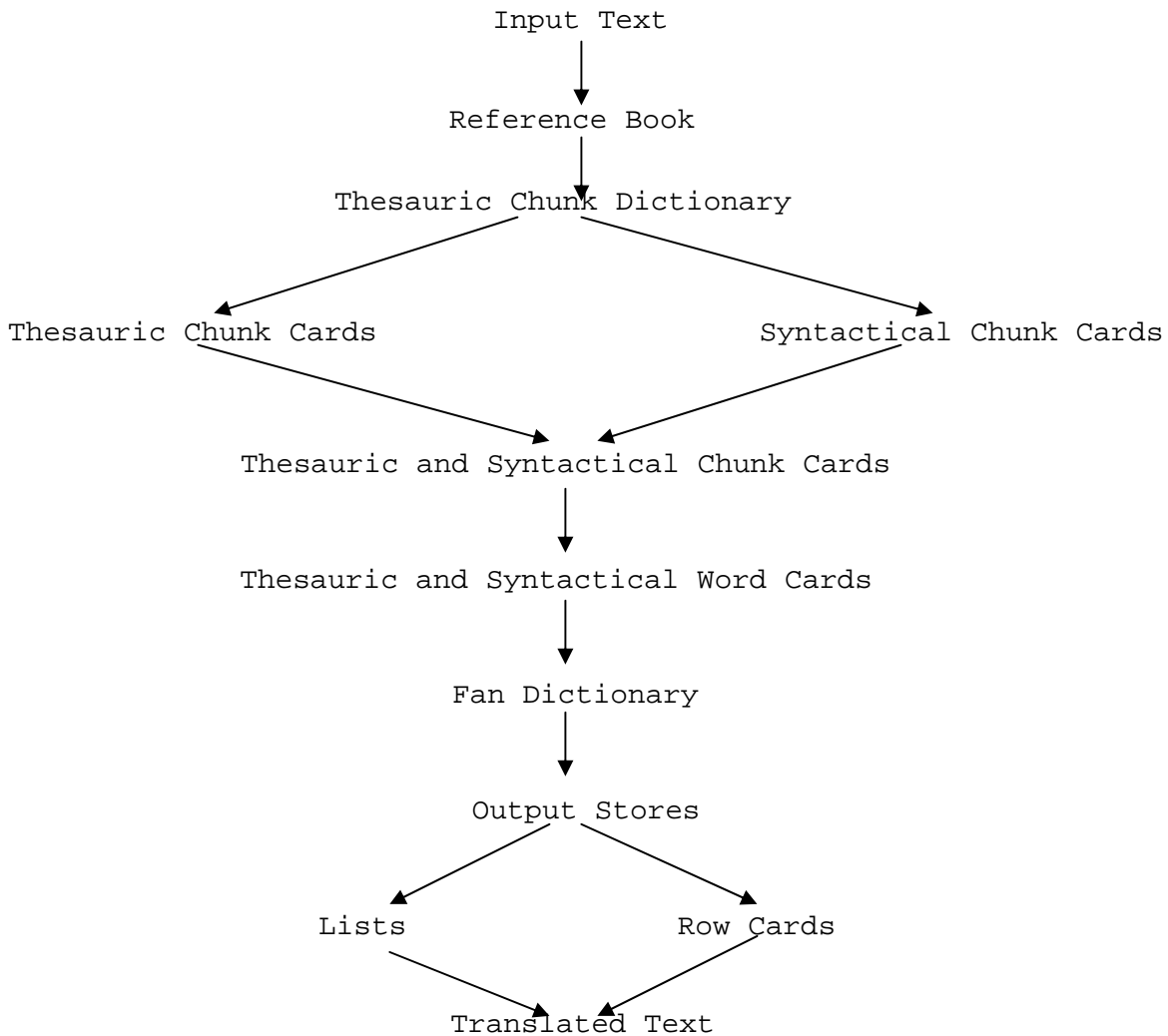GENERAL SCHEMA
OF A
THESAURIC TRANSLATION
PROGRAMME
USING A PUNCHED CARD TECHNIQUE

"A programme which presupposes an infinite space and
an infinite time."

     This schema was composed after a colloquium of punched-
card programmes, held at the premises of the Cambridge Language
Research Unit, June 19th - 24th, 1958. The Colloquium was
attended by members of C.L.R.U.; also by Stephen Whelan, R.R.E.,
Malvern: by T.M. Staniforth, late of R.R.E. Malvern: and by
R.A. Fairthorne, R.A.E., Mathematical Services Division,
Farnborough.

Joy M. Staniforth.
20/6/58

A Guide to the Order of the Programme.

Input Text

↓

Reference Book

↓

Thesauric Chunk Dictionary

Thesauric Chunk Cards   Syntactical Chunk Cards

Thesauric and Syntactical Chunk Cards

↓

Thesauric and Syntactical Word Cards

↓

Fan Dictionary

↓

Output Stores

Lists   Row Cards

Translated Text

<u>Reminders</u>

1)   When sorting on certain rows only, remember to cut out the
  other rows with the drum provided for the purpose.
2)   At the moment, when reproducing part of a card only on the
  duplicator, a grid card must be used.

———————————————

    We assume the existence of a chunking reference book, which
indicates the method *of* chunking each word. [At the moment this
exists in the form of a card index.]
    Chunk your input text according to this reference book.
    Number your input text as follows:-
Paragraph I

Sentence I

1.1	1.2	2	3.1	3.2	4.1	4.2	4.3	5	6.1	6.2
Galli	a	est	omn	is	di	vis	a	in	part	is
1	2	3	4	5	6	7	8	9	10	11

Any punctuation sign is numbered as if it were a whole word.   The
numbers which appear above the text are punched on to the cards in

the S Pack. (The pack which contains the cards bearing the syntactical, monolingual and arrangement information only.) This form of numbering will probably be a temporary procedure, which as tests proceed can be abandoned.

The numbers which appear below the text are stamped on to the cards.

We assume the existence of a dictionary pack made as at present but minus the unambiguous-category specifications and Lists (i.e. minus the information in columns 31-78 of the top row). Thus more room is left for monolingual and syntactical information.

This pack can also be minus the alphabetical code. This, however, is optional at the moment for whilst this code serves no useful purpose this part of the card is as yet not required for anything else, and the code does serve to match the S Pack with the D Pack.

Pull out of this dictionary (i.e. the Thesauric chunk dictionary) the cards required for the text under consideration.

As the chunks of each word are got out an intersection is made between them enabling certain unwanted endings to be dis-carded, this is called the pun removing procedure and is as follows:-

The card containing the stem chunk or, where there is more than one chunk in the stem, the meet of the stem chunks (this is obtained by superimposing the cards on which the separate stem chunks are punched and using them as one card), is matched against the cards containing the endings in turn. Only those which intersect twice in the section of the card which contains mono-lingual information are retained, the others are replaced in the dictionary. It is hoped that this section of the card will eventually be coloured so as to make the intersections easier to detect.

At present the cards obtained from the dictionary which in future we shall call pack D must be reproduced. It is hoped eventually that these dictionary cards will be mass-produced so that the actual cards taken from the dictionary can be used for translation, the dictionary being replenished from a store of dictionary cards.

Feed pack D into the reproducer in the order of the input text. Stamp numbers on the reproduced pack in accordance with numbers you have written below the chunks on your input text, duplicates being given the same number. Replace the cards you took from the dictionary back in the dictionary.

Now reproduce rows X and Y only from the cards in pack D plus generating a pack which contains Arrangement, Monolingual

and syntactical information and list numbers only. We will
call this pack, pack S.

   Add to the cards in the S pack, numbering of the form that
is written above the text. Thus the card which contains the
information about the chunk - vis - will be numbered paragraph 1,
sentence 1, word 4, chunk 2.   The paragraph nos are to be
punched in columns 11 and 12 the sentence nos in columns 13 and
14, the word nos in columns 15 and 16 and the chunk number in
column 17.

Card
-vis→

Row	11	12	13	14	15	16	17		
X									
Y									
0									
1		▨		▨					
2							▨		
3									
4						▨			
5									
6									
7									
8									
9									

   Now perform your "syntactical" tests on the S pack until
you have only one card left for each chunk. [Noting as a
temporary expedient that rows 0-9 in columns 1-10 and 18-80 can
be used to carry additional information provided that you do not
want to carry this information forward.]

   Now stamp these cards with the appropriate numbers according
to the input text, then reproduce their X and Y rows only on to
new cards SJ, stamping the new cards with the appropriate numbers.

   Now take pack D and make join cards for each individual
chunk in the first sentence as follows:-

   Take the card of the 1st chunk in the sentence and super-
impose it on the card of the 2nd chunk. If there is any inter-
section whatever reproduce this intersection on a new card J_1,
then intersect the 1st chunk card with the 3rd chunk card and so
on to the end of the sentence making all reproductions of inter-
sections on to the same card J_1. Then replace the 1st chunk
card by the 2nd chunk card and intersect the 2nd chunk card with
the 1st, 3rd and subsequent chunk cards to the end of the sentence,
reproducing all intersections on a card J_2. Repeat this perfor-
mance until a 'J' card has been made to correspond to every chunk
in the sentence. As each 'J' card is produced stamp it with the
number of the generating chunk.

   Now reproduce these J cards minus their X and Y rows stamping
the new J cards with the same number as the one from which they
were produced.

— 3 —

Now make new cards of the join of the J cards and SJ cards which bear the same number. Thus producing the final join card for each chunk. These cards must be stamped with the number of the J and SJ card which generated them.

Now check the stamped numbers on the final join cards of the chunks against the numbers on the input text [nos below text]. Make join cards for each word from the final join cards for each chunk. Stamp each card as it is made with the numbers of the chunks it contains. We will in future call these cards T cards.

We will assume that the fan Dictionary is made up of cards stamped as at present but minus the alphabetical code.

In row 79 of the fan card will be put the number which specifies the type of list which this card tells you to refer to, if it tells you to refer to one at all.

This number is coded in binary, but is written along a column not a row. Thus no 63 in binary is   and is stamped thus:-

Lists will be on 'Peek-a-Boo' (Taylor) cards. The perpendicular axis contains list numbers which will come from columns 21-30 inclusive row X (Top row) on a T card.

The horizontal axis consists of the specification number of the list which will come from column 79 of the fan cards. Should these specification numbers overflow a card, then a 2nd and even a third card can be taken. (No more than 3 should be necessary.) Thus, eventually, there will be 1, 2 or 3 peek-a-boo cards for each language.

Column 80 of your fan card can be numbered in a similar way to bring you to your row cards. For if a fan card and the row card to which it refers bear the same number in column 80, a sorter could find the row card required out of a pack of row cards.

Now take the T card of your first word and having set your sorter accordingly, sort the fan dictionary by the 1st thesauric head hole that appears in this T card (card T_1). Giving us a pack T_{11}.

[When syntax is sufficiently reliable the fan dictionary will first be sorted by the most general syntax operator (i.e. the one punched furthest to the left on the card) thus reducing the size of the dictionary needed for this and the following sorts.]

Then resort pack T_{11} by the 2nd Thesauric head hole of card T_1 and reproduce the resulting card or cards being careful to write on them the word or words written on the originals.

Then having replaced the originals of these cards in the pack T_{11} resort it according to the 3rd thesauric head on card T_1 again reproducing and then replacing the originals of the resulting cards.

Proceed like this until the pack T_{11} has been searched by all the thesauric head holes on card T_1. (Except the 1st hole which generated the pack.)

The resulting cards are then stored in a safe place and the output pack T_{11} is returned to the Fan Dictionary.

The Fan Dictionary is then sorted by the 2nd thesauric head hole in the card T_1 thus generating pack T_{12}.

This pack T_{12} is then sorted by the 1st, 3rd etc., thesauric head holes on card T_1, the resulting cards being added to those previously stored and the pack T_{12} being replaced in the Fan Dictionary.

This procedure is continued until each Thesauric head hole on card T_1 has generated a pack T_{1r} from the fan dictionary and all resulting cards are now in one store T_1. Place card T_1 in the same store in case the list number on it is needed.

Then take the next word in the sentence which is on card T_2 and repeat the above procedure; putting the resulting cards in a store T_2 together with card T_2.

Continue thus until all words in the sentence have been dealt with.

Now by comparing the cards in the stores and where necessary referring to the lists or row cards indicated on them, it should be possible to derive a translation.