Quarterly Progress Report of the Research Laboratory of Electronics, M.I.T., July, 1957

(XV. MECHANICAL TRANSLATION)

## B. GENERATION OF NOMINAL COMPOUNDS IN ENGLISH

In any sentence-generating grammar, such as would be needed for the synthesis of target-language sentences or would form the basis of any reasonable heuristic device for analyzing source-language sentences, it will be necessary to provide for automatic generation of strongly recursive constructions that have a multiplicity of kernel-sentence sources (1). An excellent example of such constructions is the nominal compound in English, characterized by the concatenation of two or more words under a stress superfix of a primary stress (/) followed by tertiary (\) or secondary (^); e.g., flý wheèl, tíme bòmb, táx collèctor, Secûrities and Exchánge Commìssion, búrble pòint, eléctron deficiency, Mersénne nùmber, scìntillátion côunter, fócal pòint, and so on.

If we choose a compound of any one type, say adjective plus noun, as in hót bòx, bláckhèad, dárkròom, we note that not all members of the first category may occur together with each member of the second, but that those which do occur together are just those that also co-occur in sentences of the form The Noun is Adjective:

The man is mad 

Madman

The season is rainy 

Rainy season

but not:

\*The tone is hungry ..... \*Hungry tone

\*The duchess is interstitial ...... \*Interstitial duchess.

(Note: The asterisks indicate that the forms do not occur in English.)

Since these sentences are all in the kernel, and since we wish to avoid having to state more than once the particular selections of one constituent by the other, it is reasonable to attempt to derive the compounds from the kernel sentences by means of grammatical transformations. The immediate difficulty is that not all compounds can be derived from the same type of kernel sentence, and a great variety of kernel sentences will be required for generating all types of compound.

The nominal compounds of English have been studied with a view to generating them by transformations from kernel sentences. Forty-five distinct types have been isolated, ranging in complexity from simple transformations like

The rain falls — Rainfall

to more complicated and perhaps more dubious transformations like

The butterfly has a tail

The tail is like a swallow

Swallowtail.

Many compounds seem to be derivable by several different transformations without exhibiting any corresponding ambiguity. Thus, while the compound "ether extraction"

## (XV. MECHANICAL TRANSLATION)

might be generated in two different ways — corresponding to its two different meanings:

They extract the fat with ether — Ether extraction (like "smallpox vaccination")

This apparatus extracts the ether — Ether extraction (like "arms shipment"),
the compound "ion trap" may also be generated in two different ways but seems to have only one meaning:

This coil traps ions — — — — — Ion trap (like "doorstop")

This trap is for ions — — — — Ion trap (like "flour sack").

Another difficulty, at present, is that in several cases a given type of source sentence appears to yield two or more kinds of compound by means of different transformations, but not all of these source sentences appear to undergo all of the transformations involved. Thus, for example, the source sentence X + V - es + N may, in some cases, be transformed into the compound V - ing + N, as in:

He drinks the water ———— Drinking water

She wears this apparel 

Wearing apparel

and many other cases into N + V-ing, as in:

He chews tobacco — Tobacco chewing

but not all sentences underlying the second type of compound yield compounds of the first type, as in:

He dodged the draft — > Draft dodging but not \* Dodging draft.

Finally, exact specification of all of the required grammatical transformations necessarily awaits the preparation of a complete grammar of English.

R. B. Lees

## References

1. For explanation of linguistic terms see Quarterly Progress Report, April 15, 1957, p. 133, and references cited therein.