XVI. MECHANICAL TRANSLATION[*]

V. H. Yngve
Carol M. Bosche
Elinor K. Charney
J. L. Darlington
D. A. Dinneen
B. J. Fraser

M. Gross[†]
Renate I. Heil
Muriel S. Kannel
E. S. Klima
K. C. Knowlton

J. D. McCawley
T. More, Jr.
W. K. Percival
A. C. Satterthwait
M. R. Quillian
J. J. Viertel

A. RELATEDNESS BETWEEN GRAMMATICAL SYSTEMS

Since the seventeenth century, grammars of the English language have contained judgments on the grammatical status of historical developments in the syntax of the inflected forms of the interrogative and personal pronouns. At present, the use of WHO for WHOM as the objective form of the relative and interrogative pronoun is generally accepted, at least in the United States, as colloquial speech. The defense of 'It is me' against 'It is I' also has a lengthy history. Finally, constructions like 'Him and me are staying', although not accepted as Standard English, have been acknowledged as forming as much a part of a natural and consistent linguistic system ("Vulgar" English) as their socially more acceptable relatives. The changing syntax of the pronouns is, of course, basically a historical phenomenon. However, because of the normative approach and conservatism of formal education, as well as the strong formative influence of literary tradition, many educated speakers of American English have (as more or less distinct styles of speech) the different syntactic subsystems characterizing all of these usages. From the point of view of descriptive linguistics, phenomena like these raise two questions: What differences in syntactic structure are represented by such variations in the shape of sentences? and What is the relationship between the different systems so characterized? In the present report, these questions will be approached in the following way: The syntactic structure of each system will be considered to be revealed by the set of rules which most economically generates the sentences of the system.[1] That set of rules is its grammar (G). The relationship between one system (L_1) and another (L_2) will be thought of in terms of the rules (E_{1-2}) that must be added to the grammar (G_1) of L_1 in order to account for the sentences of L_2. A fundamental structural difference, varying in nature and degree, will be considered to exist between systems L_1 and L_2 when the set of rules G_2 for most economically generating the sentences of L_2 is not equivalent to G_1 plus its extension E_{1-2}. From the point of view of comparing systems, the particular pairing and direction that I chose for extending the grammar G_1 of one system to account for the sentences belonging to another system are

Table XVI-1. Representative sentences of the four related systems.

	L_1	L_2	L_3	L_4
α	1) She could see him near me. 2) He and I left. 3) We two left. 4) We all left. 5) It was I.		It was me.	Him and me left. Us two left.
β	6) Could she see him? 7) Who could see him? 8) Whom could she see? 9) With whom did he speak? 10) Whom did he speak with? 11) Who was it? 12) Who was the leader?	Who could she see? Who did he speak with?		
γ	13) He knew whether she could see him. 14) He knew who it was. 15) He knew who was the leader. 16) He knew whom he spoke with.	He knew who he spoke with.		
δ	17) The leader who could see him left. 18) The leader whom I saw left. 19) The leader with whom he spoke left. 20) The leader whom he spoke with left.	The leader who I saw left. The leader who he spoke with left.		

admittedly arbitrary. However, except when so noted, the order in which the styles are considered recapitulates comparable aspects in the historical development of the pronouns. Of course, my presentation deviates in an essential way from true historical perspective: The sequence of pronominal systems abstracted from consecutive stages of the language is collapsed, as it were, and each system treated as part of an otherwise identical total system.

Table XVI-1 contains sentences that are representative of the four systems to be related. A blank space under any L signifies that the sentence is the same in its final shape as that one on its immediate left. It is assumed that, aside from the differences in final shape presented by the sentences in Table XVI-1 and by the other sentences of the systems like them, the rest of the sentences have the same shape.

Consider, first, only L_1, which is typical of a style that might be referred to as "elegant literary." Classifying representative sentences in the particular four types given in Table XVI-1 (a, β, γ, δ) follows traditional grammatical usage: a, simple declaratives; β, interrogative structures corresponding to the latter; γ, structures corresponding to interrogatives but contained as constituents of complex declaratives (and related structures) in the form of indirect questions; δ, sentences themselves containing structures, relative clauses, which are closely related to independent sentences. The notions — common to intuitively based traditional classification — of "correspondence" and "relatedness" between types of structures would not be without analogues in the rigorous formal description of the language. It will be found that accompanying such impression of "correspondence" and "relatedness" are formal facts like the recurrence of certain basic grammatical features between representatives of different structural types. For example, the same sorts of subjects and objects (to use traditional terms for major functional units in the syntactic system) occur with the same verbs. Also, although the phenomenon may be of lower order with respect to basic grammatical features, the same agreement holds between subjects and verbs, regardless of the type of structure. But, although the functions of the major elements remain the same, the final shapes of sentences differ according to the type of structure. In particular, the order of the elements differs. In a the subject is before the finite verb; the finite verb forms the head of a chain of verb forms, interrupted only by certain adverbs, and the object occurs after that verb chain. With respect to the order of such elements, group β is more complicated. Often the particular word order of one sentence as compared with that of another can be attributed to clear grammatical differences elsewhere in the sentence (e.g., inversion in β when accompanied by an interrogative word, as contrasted to the word order in a; direct order in a subordinate clause as in γ, but inversion elsewhere as in β). The basic order of elements will be that (or those) from which the ultimate particular word orders, as well as other grammatical features, are most simply specified. Often, while major functional elements like subject and object cannot be

177

specified in a simple way on the basis of their position in the total set of sentences representing all possible structural types, their position in terms of the basic order of elements is significant for their specification. Clearly, in English the relationship between the order of elements and their grammatical function is simpler in α than in β; and γ and δ are more complicated since they contain instances of α and β. For convenience, in fact, we can refer to the word order of α, with only a few modifications, as basic.

The position of an element is thus relative to the level of syntactic structure. Among grammatical features describable by position at the level characterized by the presence of basic word order is the occurrence of the objective form of pronouns in L_1. The occurrence of HIM rather than HE, and WHOM rather than WHO in the various structures of L_1 exemplified in Table XVI-1 follows from a principle of syntactic function that may be stated in terms of position in the simple declarative sentence; i.e., that pronouns following transitive verbs and prepositions occur in their objective form. Thus, 1) in L_1 is the ultimate form of a sentence like She – could – see – he + CASE – near – I + CASE. The specification of function at the level of the simple sentence in terms of the occurrence of inflectional forms corresponds to the traditional notion, whereby HIM and WHOM are the forms taken by the pronouns as "objects" of verbs and prepositions, and whereby WHOM in 8)-10), 16), 19), etc. is still considered to be the "object" of a transitive verb or preposition, even though the "object" is separated from its governor and may even precede it.

Consider, first, the direct questions β. The direct question can be thought of essentially as a single sentence, specified by an interrogative marker (Wh) introducing the sentence (e.g., 6) in L_1 would have the form Wh – she – could – see – he + CASE). The interrogative specifier Wh- can remain unattached or can have attached (indicated by +) to it various elements of the sentence, including instances of the pronoun marked with CASE; e.g., 8) in L_1: Wh + he + CASE – she – could – see, and 7) in L_1: Wh + she – could – see – he + CASE. The special word order of direct questions can be explained as the attraction of the finite verb form to Wh and its attachments; i.e., Wh – she – could – see – he + CASE $\Rightarrow$ Wh – could – she – see – he + CASE; Wh + he + CASE – she – could – see $\Rightarrow$ Wh + he + CASE – could – she – see; Wh + she – could – see – he + CASE already has the finite verb beside it. In direct questions, unattached Wh- is not represented phonologically. In indirect questions, which constitute questions embedded as subordinate clauses in other sentences, unattached Wh- appears as WHETHER. Without further justification, I symbolize subordination by THAT, which blocks the Wh-attraction resulting in the inverted word order of questions; i.e., Wh + he + CASE – she – could – see ultimately yields 'Whom could she see?' but under subordination, we have

21) he – knew $\begin{cases} \text{THAT – she – could – see –} \\ \quad \text{he + CASE} \\ \text{Wh + she + THAT – could –} \\ \quad \text{see – he + CASE} \\ \text{Wh + he + CASE + THAT –} \\ \quad \text{she – could – see} \\ \text{Wh + THAT – she – could –} \\ \quad \text{see – he + CASE} \end{cases} \Rightarrow$ he – knew $\begin{cases} \text{that – she – could – see –} \\ \quad \text{him} \\ \text{who – could – see – him} \\ \\ \text{whom – she – could – see} \\ \\ \text{whether – she – could –} \\ \quad \text{see – him} \end{cases}$

The relative clause also represents embedding with subordination, with the condition that the two sentences involved have an identical constituent. Sentence 16) in L_1 represents the subordination of 'I saw the leader' to the identical noun in 'The leader left'. It is simplest to consider the subordinated sentence after pronominalization; i.e., I – saw – he. After case-marking, the latter has the form: I – saw – he + CASE, which, when attached to Wh and embedded, with subordination, into the Nominal of 'The leader left' yields: The – leader – Wh + he + CASE + THAT – I – saw – left. Marked case is again determined in terms of position with respect to basic word order. Conjoined forms like 2) can be thought of as the replacement of a plural subject by any number of other compatible subjects: Given he – left and I – left, we can embed the subjects in They – both – left to yield: He – and – I – both – left. The marked or unmarked quality of the pronouns is carried over into conjunction in L_1.

In the grammar comprising the rules for generating the sentence of L_1, the rules of the phrase-structure level, abbreviated to the special demands of the present report, would have the following general appearance[2]:

22) S → (Wh) Nominal – Tense $(V^m) \begin{cases} \begin{Bmatrix} V^{be} \\ Verb \end{Bmatrix} \text{Predicative} \\ Verb \ \text{(Nominal)} \end{cases}$

23a) Verb – Nominal → $\begin{Bmatrix} V^t \\ V^p - prep \end{Bmatrix}$ Nominal

23b) Verb – Predicative → V^c – Predicative

23c) Verb → V^I

24) V^t → $\begin{cases} \text{know} \\ \text{see} \end{cases}$

25) V^{be} → be

26) V^c → become

27) V^p → speak

28) $V^I \rightarrow$ leave

29) speak – prep $\rightarrow$ speak – with

30) $V^m \rightarrow$ can

31) tense $\rightarrow \left\{ \begin{array}{l} \text{Past} \\ \text{Present} \end{array} \right.$

32) Nominal $\rightarrow \left\{ \begin{array}{l} \text{Pronoun} \\ \text{John} \\ \text{the leader} \end{array} \right\}$ Number

33) Number $\rightarrow \left\{ \begin{array}{l} \text{Sg} \\ \text{Pl} \end{array} \right.$

34) Pronoun $\rightarrow \left\{ \begin{array}{l} \text{I} \\ \text{he} \\ \text{she} \end{array} \right.$

Other obligatory rules, of transformational nature, carry into sentences the strings emerging from the phrase-structural level, along with operations accumulated in the transformational level.[3]

35) Tense attachment: Tense – V $\Rightarrow$ V + Tense

36) DO – proclysis: – Tense $\Rightarrow$ do + Tense

37a) Wh + Pronoun – Number $\Rightarrow$ who

37b) Wh + Pronoun + CASE – Number $\Rightarrow$ whom

38a) I + CASE – Sg $\Rightarrow$ me; he + CASE – Sg $\Rightarrow$ him

38b) do + Past $\Rightarrow$ did; can + Past $\Rightarrow$ could; speak + Past $\Rightarrow$ spoke;
   be + Past $\Rightarrow$ was; leave + Past $\Rightarrow$ left; see + Past $\Rightarrow$ saw

The transformational rules, in their appropriate order, which account for the grammatical features associated with case would be:

39) I. Case-marking: $\left\{ \begin{array}{l} V^t \\ \text{prep} \end{array} \right\}$ Pronoun $\Rightarrow \left\{ \begin{array}{l} V^t \\ \text{prep} \end{array} \right\}$ Pronoun + CASE

   II. Conjoining: S': X' – Pronoun + CASE – Pl – Y'
            S": X" – Nominal" – Y"
            S'": X'" – Nominal'" – Y'"

            $\Rightarrow$ X' – Nominal" – and – Nominal'" – Pl – Y'

III. Wh-attachment, including subordination where appropriate[4]:

$$\text{Wh-}\ldots \left\{ \begin{array}{l} \text{Prep - Pronoun + CASE} \\ \text{Pronoun (+ CASE)} \end{array} \right\} \text{Number} \Rightarrow$$

$$\text{Wh} + \left\{ \begin{array}{l} \text{Prep - Pronoun + CASE} \\ \text{Pronoun (+ CASE)} \end{array} \right\} \text{Number (+ THAT)} \ldots$$

IV. Preposition placement[5]: Wh + Prep - Pronoun $\Rightarrow$ Prep - Wh + Pronoun

V. Wh-attraction:

Wh (+ Pronoun + CASE - Number) - Nominal - Tense (V^m) Verb $\Rightarrow$

Wh (+ Pronoun + CASE - Number) - Tense (V^m) Nominal - Verb

Consider now L_2, one variety of Colloquial English. To account for the sentences of L_2 proceeding from the grammar of L_1, a rule (E_{1-2}) must be added which reduces clause initial WHOM to WHO. The case form, however, is retained when WHOM is preceded by a preposition, cf. 9) versus 10) and 18) versus 19) in L_2. The extension E_{1-2} would occur after Wh-attachment, but before preposition placement, and thus leave ... WITH WHOM .... Thus $G_1 + E_{1-2}$, the G_1-based grammar of L_2, would have the following difference in appearance.

40) $G_1 + E_{1-2}$: ...

      III. Wh-attachment

      E_{1-2}. Wh + Pronoun + CASE - Number $\Rightarrow$ Wh + Pronoun - Number

      IV. Preposition placement

      ...

The extended set of rules with G_1 as a basis accounts for the sentences of L_2. From the point of view of L_2 independently, however, the sentences are most economically described by a grammar (G_2) that differs from that of L_1 in the order of the transformational rules. In G_2, the rule of case-marking is dependent on the position of the elements of the sentence as they occur after Wh-attachment. The difference in order of rules reflects the fact that while in L_1 case-marking is dependent on function (namely whether or not the element is a grammatical object), in L_2 CASE, or more properly lack of CASE, is a concomitant of position as reckoned after certain major distortions of the basic order of elements. Thus in L_2, CASE is not associated with functional elements if they do not occupy a position after the motivating factors. Other things being equal, the grammar G_2 would consist of the same phrase-structure rules as G_1, but the order of the transformational rules would be different.

41) G_2: ...

      I. Wh-attachment

     II. Case-marking

    III. Conjoining

    IV. Preposition placement

     V. Wh-attraction

L_3 differs from L_2 in containing a further extension of the positional determination of CASE to cover also pronouns representing the Predicative; i.e., those following V^{be} and V^c. Position in this case is reckoned, as elsewhere in G_2, after Wh-attachment. Thus 5) is marked in L_4 but 11) and 14) are not. Extension E_{2-3} would have the following effect:

$$\left\{\begin{matrix} V^{be} \\ V^c \end{matrix}\right\} \text{Pronoun} - \text{Number} \Rightarrow \left\{\begin{matrix} V^{be} \\ V^c \end{matrix}\right\} \text{Pronoun} + \text{CASE} - \text{Number},$$

or that which amounts to the same thing,

$$V\,[\text{Pronoun} - \text{Number}]_{\text{Predicative}} \Rightarrow V\,[\text{Pronoun} + \text{CASE} - \text{Number}]_{\text{Predicative}}.$$

42) $G_2 + E_{2-3}$: ...

      I. Wh-attachment

     II. Case-marking: $\left\{\begin{matrix} V^t \\ \text{prep} \end{matrix}\right\}$ Pronoun $-$ Number $\Rightarrow$

$$\left\{\begin{matrix} V^t \\ \text{prep} \end{matrix}\right\} \text{Pronoun} + \text{CASE} - \text{Number}$$

$$E_{2-3}. \quad \left\{\begin{matrix} V^{be} \\ V^c \end{matrix}\right\} \text{Pronoun} - \text{Number} \Rightarrow V - \text{Pronoun} + \text{CASE} - \text{Number}$$

    III. Conjoining

    ...

In G_3 the difference attributable to E_{2-3} is represented by the extension of case-marking to all pronouns following a verbal form V, so that CASE becomes entirely a positional feature without functional reference. The phrase structure of G_3 remains the same, as does the ordering of transformational rules. From the point of view of units represented in the phrase structure, case-marking is generalized in G_3, in that it applies to a unit of a higher level of the phrase structure: Verb rather than V^t and, as extended by E_{2-3}, V^c.

43) G_3: ...

II. Case-marking: $\begin{Bmatrix} Verb \\ V^{be} \end{Bmatrix}$ Pronoun – Number $\Rightarrow$

$\begin{Bmatrix} Verb \\ V^{be} \end{Bmatrix}$ Pronoun + CASE – Number

...

In terms of the class of units V, the generalization can be expressed simply as:

44) G_3: ...

II. Case-marking: V – Pronoun – Number $\Rightarrow$

V – Pronoun + CASE – Number

...

It should be noted that the ordering of Wh-attraction after case-marking accounts for the fact that in the final form taken by sentences, the unmarked form can also occur regularly following a V; e.g., 'Has he_ left ?', 'Who is he_ ?'. When case-marking applies to these sentences, however, they have the form Wh – he_ – has – left and Wh + Pronoun – he_ – is.

From the point of view of G_3, the sentences in L_4 manifest a further extension of case-marking to all pronominal forms before Tense, except for interrogative-relative WHO and unconjoined pronouns directly adjacent to Tense:

45) E_{3-4}: X – Pronoun – Number – Y $\Rightarrow$ X – Pronoun + CASE – Number – Y

provided that X $\neq$ Wh+ and Y $\neq$ Tense

Grammar G_4 differs from G_3 + E_{3-4} with respect to its rules in both the transformational and the phrase-structure levels. The basic forms of the personal pronouns emerging from phrase structure are those previously resulting from combination with the case markers. The order of the transformational rules is different; case-marking now occurs even later in the grammar than conjoining and is limited in application, in the case of the personal pronouns, to a position that is immediately before the finite verb. The ultimate positional nature of case-marking in L_4 thus does not even observe the weakly functional principle of similarity involved in conjoining. If sentences like 9) and 19) are admitted in L_4, then the marking of the relative-interrogative form remains essentially the same as in G_3. Thus, while the marked forms of the personal pronouns yield I, HE, THEY, etc., that of the interrogative-relative would yield WHOM. It would probably lead to greater homogeneity not to consider the sequence of a preposition followed by WHOM in '... with whom he spoke ...', as in 9) and 19), as part of L_4. Then in G_4, Wh-attachment would analyze the string: Wh – he – spoke – with –

183

Pronoun only as: Wh – ... Pronoun ... and not also as Wh – ... Prep – Pronoun ...,
as was the case in the other grammars. Consider now constructions like 'John,
according to whom we were at fault, would not condone such action', as well as related
ones, not discussed here, like 'The concert during which he slept was a great success',
i.e., constructions of L_1, etc., which do not permit the separation of the preposition
from its object in relative clauses. If this style of construction, too, is not considered
to be a part of L_4, then Wh-attachment is simplified and preposition placement is omit-
ted from the grammar, as are operations involved in realizing the form WHOM. Rules
that are unnecessary when L_4 is considered in this way are bracketed [ ]. In L_4 as in
the other systems, case-marking applies before the inverted-order characteristic of
direct questions. This accounts for the presence, in certain sentences, of I, HE, THEY,
etc. after the finite verb. Apparent exceptions to the rule of case-marking in G_4 such
as 4) in L_4 or 'They each left' versus 'Us two left' are best explained on the basis of
the special structural position of words like EACH and ALL in these sentences: namely,
that, although grammatically restricted by the nature of the nominal, these forms are
constituents rather within the predicate of the sentence, as is clear when the predicate
is in expanded form; e.g., 'They have all left'. The structure of G_4 would have the
following difference:

46) G_4: 33) ...

34) Pronoun → $\begin{cases} me \\ him \\ her \end{cases}$

35) ...

36a) $\begin{cases} \text{same as in} \\ G_1, \text{ etc.} \end{cases}$: Wh + Pronoun – Number $\Rightarrow$ who

36b) [Wh + Pronoun + CASE – Number $\Rightarrow$ whom]

37a) me + CASE – Sg $\Rightarrow$ I; him + CASE – Sg $\Rightarrow$ he

The transformations accounting for the introduction of CASE in 36b) and the ultimate
word order of the sentences would now have the following form and order:

47) I. Wh-attachment

II. Conjoining

III. [Preposition placement with case-marking: Wh + preposition – Pronoun –
Number $\Rightarrow$ preposition – Wh + Pronoun + CASE – Number]

IV. Case-marking: – Pronoun – Number – Tense $\Rightarrow$
– Pronoun + CASE – Number – Tense
(the sequence "– Pronoun ..." differentiated from "+ Pronoun ...")

Table XVI-2. Summary of differences in ordering of main transformations

L_1	L_2	L_3	L_4
Pronoun → I, he, she			Pronoun → me, him, her
Case-marking	Wh-attachment		Wh-attachment
Conjoining	Case-marking	Case-marking expanded	Conjoining
Wh-attachment	Conjoining		[Preposition placement, with concomitant CASE]
Preposition placement	Preposition placement		Case-marking
Wh-attraction	Wh-attraction		Wh-attraction

V. Wh-attraction

It should be noted that in the structure resulting from conjoining, the constituent 'Pl', which represents the number of the constituent replaced by the conjoined nominals, leaves a string to which 47) IV does not apply. In the derivation of 2) in L_4, for example, at the level of case-marking in G_4 the string has the following shape: him – Sg – and – me – Sg – Pl – Tense – leave.

We have examined in sequence four systems. A given system and that following it in the sequence were compared to see how the differences between them, in terms of the sentences each contains, are reflected in their individual grammars. An operation common to the four systems — the attachment, motivated by position, of a marker (case) to a certain class of words (pronouns) — is found to manifest a progressively weaker relationship to function, as illustrated in Table XVI-2.

E. S. Klima

References

1. This notion of grammar is developed by N. Chomsky, Syntactic Structures (Mouton and Company, s'Gravenhage, 1957). A detailed treatment of specific grammatical problems can be found in R. Lees, A grammar of English nominalizations, Int. J. Am. Ling. 26, 3 (1960).

2. The use of the unnecessary symbol V^{be}, rather than simply BE, is part of a convention for abbreviating the class of elements formed by V^c, V^t, V^I, and V^x with the otherwise nonoccurring symbol without superscript V.

3. Similar rules would attach Pl as 's' to an immediately preceding noun and ultimately remove all other number markers. Thus the dangling 'Pl' (plural) resulting from conjoining in 34) is removed with the same operation that removes the empty 'Sg' (singular) after nouns in English, and to this extent is no more arbitrary.

4. For a more precise analysis, I would propose deriving questions, but not relative clauses, from sentences in which the Pronoun that is to be attached to Wh is one of the indefinites: SOMEONE, ANYONE, or SOMEBODY. In this way, certain characteristics of questions (direct and indirect), but not of relative clauses, would be accounted for; e.g., 'Who else left?' is like 'Someone else left'; 'Whom did he see of importance?' is like 'He saw someone of importance'. If we consider CASE to be applicable also to the indefinites, then the rest of the analysis need not be changed essentially.

5. The operation, designated here as preposition placement, which is necessitated by the sequence Wh – prep – pronoun ... left by Wh-attachment is not an indication of the arbitrariness of Wh-attachment, for the latter serves also to prevent the reduction in relative clauses of '... with whom ...' to '... with that ...' on a par with the permissible reduction of '... whom ...' to '... that ...'; e.g., 'The man whom I spoke with', 'The man that I spoke with' but 'The man with whom I spoke', and not 'The man with that I spoke'. Thus between 39) III and 39) IV there is a rule that is applicable only to relative clauses: Wh + Pronoun + CASE – Number – THAT $\Rightarrow$ THAT. Further, the subordinator THAT may itself be deleted except in certain specifiable contexts, and in others must be deleted: 'He knows that they left' ' He knows they left', 'The man that I spoke with left' 'The man I spoke with left'.