

Logical expressive power and syntactic variation in natural language

EDWARD L. KEENAN

I will argue here¹ that the task of formally defining logical structures (LS) for natural languages (NL) has a linguistic interest beyond the immediate one inherent in representing logical notions like *entailment*, *presupposition*, *true answer to*, etc. The reason is that LS can be used as a basis for describing, and in some cases explaining, certain kinds of syntactic variation across NL. Below we consider three types of comparison between LS and the surface structures (SS) which can be used to express them in different NL.

In the first comparison we will demonstrate that NL differ significantly in their capacity to form restrictive relative clauses. We explain this variation in terms of the *Principle of Conservation of Logical Structure* (Keenan (1972b)).

In the second comparison we show that NL differ in their capacity to stipulate the co-reference of NP positions, but we provide no explanation for this variation.

And in the third comparison we show that the expression of indirect questions (indQ) varies in restricted ways across NL and propose a *Principle of Logical Variants* which explains this on the basis of the LS we propose for indQ.

1. Type 1 comparison

Here we compare LS with the SS which can be used to express them in various NL. Obviously if a LS can be naturally expressed in some NL but

¹ Many of the generalizations made in this paper concerning cross language variation required data from a large variety of languages. I should here like to express my thanks to the following linguists and students of language for their help in obtaining the data from the indicated languages: R. Beynon – Welsh; M. Butar-Butar – Batak; B. Comrie – Persian, Turkish, Polish, and Russian; R. P. G. de Rijk – Basque; K. Ebert – Fering (N. Frisian) and especially for making available her field notes on Kera; G. Gembries – Old and Middle English; J. Hawkins – German; R. Hull – Korean; A. Janhunen – Finnish; J. Payne – Welsh; A. Salmond – Maori; P. Sgall – colloquial Czech; A. Sinha – Hindi; H. van Riemsdijk – Zurich German; and B. Vattuone – Genoese.

not others then the former are logically more expressive *in that respect* than the latter.

For example the LS expressed by the relative clauses (RC) in (1)–(3) below are not naturally expressible in English, but they are in the NL indicated:

(1) *Arabic*

alrajul allathi *hua* wa ibnahu thahabu ille New York
the man that *he* and son-his went to New York
*‘the man that he and his son went to New York’

(2) *Welsh*

het y gwn y dyn a’ *i* gadewodd ar y ford
the hat that I-know the man who *it* left on the table
*‘the hat that I know the man who left (it) on the table’

(3) *Zurich German*

də baəb wo də hanz z kryyt glæwbt daz d maria
the boy that the Hans rumour believes that the Mary
ən aakfalə heb
him attacked has
*‘the boy that Hans believed the rumour that Mary attacked (him)’

Nor are these possibilities merely isolated facts about Arabic, Welsh, and Zurich German. As indicated by our glosses, these NL differ from English in that they characteristically present a personal pronoun in the NP position relativized. (We henceforth refer to such NL as +PRO, and NL like English which do not present such pronouns as –PRO.) To verify that the greater expressive power of the +PRO languages was in fact due to the presence of the pronoun in the RC we tested RC formation in a variety of ‘difficult’ environments in several +PRO languages and compared the results with the corresponding environments in –PRO NL. The results are summarized in Table I.¹

¹ Concerning the less well known NL in Table I we note the following: Aoba-1 is one of 12 ‘melanesian’ languages (whose names our informant did not know) spoken on the island of Aoba, the New Hebrides. We arbitrarily designated our informant’s language as Aoba-1. It is spoken in the village of Nagea, Longana District.

Batak (Toba dialect) is a Malayo-Polynesian language spoken in Sumatra. The Czech entry refers to colloquial Czech, not literary Czech which uses relative pronouns rather than a constant *rel*-marker with personal pronouns retained in the positions relativized. Gilbertese is a ‘micronesian’ language spoken in the Gilbert Islands. Kera is a ‘Chadic’ language spoken in the south of Chad near Fianga. Urhobo is a Kwa language spoken in the delta region of Nigeria; and Malagasy is a Malayo-Polynesian language spoken throughout Madagascar.

TABLE I

Relativization is possible into:

	co-NP	VP-S	NP-S	IndQ	RC	Conj	Oblique	Poss-NP
+PRO								
Aoba-1	yes	yes	yes	yes	no?	yes	yes	yes
Arabic	yes	yes	no	yes	yes	yes	yes	yes
Batak	yes	yes	no	no	no	yes	yes	yes
Czech	no	yes	no	yes	yes	yes	yes	yes
Genoese	yes	yes	yes	yes	yes	yes	yes	yes
Gilbertese	no	yes	no	yes?	no	yes	yes	yes
Hebrew	yes	yes	yes	yes	yes	yes	yes	yes
Kera	**	yes	no	yes	yes	**	yes	yes
Persian	yes	yes	yes	yes	yes	yes	yes	yes
Urhobo	yes	yes	**	no	no	**	yes	yes
Welsh	no	no	yes	yes	yes	yes?	yes	yes
Zurich German	no?	yes	yes	yes	no	yes	yes	yes
-PRO								
English	no	yes	no	yes?	no	yes	yes	yes
Russian	no	no	no	no	no	yes	yes	yes
Finnish	no	no	no	no	no	yes	yes	yes
Hindi	no	no	no	no	no	yes	yes	yes
Malagasy	no	yes	no	no	no	no	no	no
Tagalog	no	no	no	no	no	no	no	no

KEY. *Rows*: the entries are largely self explanatory, save ** which means that the context specified by the column heading does not arise in a natural way.

Columns: *co-NP*: a single member of coordinate NP, as in 'the man and the woman live in Chicago'; *VP-S*: sentence complements of verbs, as in 'John thinks that Mary kissed the man'; *NP-S*: sentence complements of N, as in 'John believes the rumour that Mary kissed the man'; *indQ*: indirect question, as in 'John knew which man kissed the woman'; *RC*: relative clauses, as in 'I saw the man who kissed the woman'; *Conj*: both sides of an overt coordinate conjunction in different grammatical cases, as in 'The man entered the room and Mary kissed the man'; *Oblique*: an object of a pre- or post-position, as in 'John hid the money under the bed'; *Poss-NP*: a genitive, as in 'The woman's cow died'.

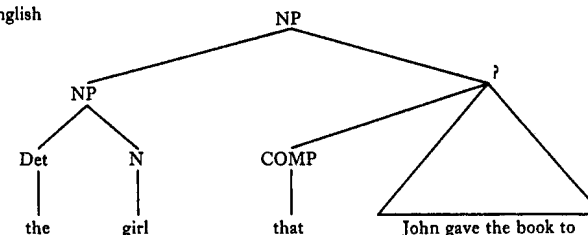
Clearly the +PRO languages generally permit the formation of RC in a greater variety of environments than do the -PRO languages. Why? We feel that it is because the +PRO languages present in surface more of the logical structure of the RC than do the -PRO languages and that in general the logically more perspicuous expression can be used in otherwise more difficult positions (where an NP position is 'difficult' if there are many NL that do not permit it to be relativized at all).

To see this, consider that, logically speaking, a RC determines (\approx refers to) an individual (or set of individuals) by first specifying a larger set – that referred to by the head NP – and then restricting that set to those member(s) of which a particular logical sentence – hereafter *the restricting sentence* – is true. Thus for an individual to be correctly referred to by (4),

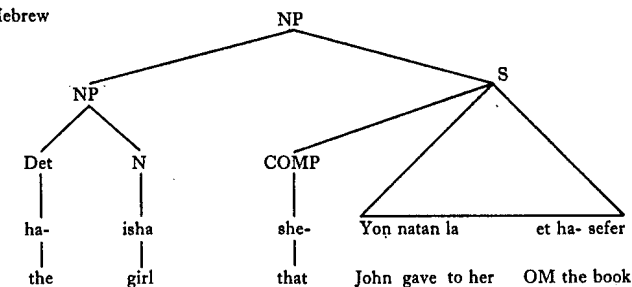
(4) the girl that John gave the book to

the individual must, in the first place, be a girl, and in the second place, be such that the sentence *John gave the book to her* is true of her. But now compare (4a), the SS of (4), with that of (4b), its natural rendition in Hebrew, a +PRO language.¹

(4) a. English



(4) b. Hebrew



Clearly in Hebrew the restricting sentence of LS is fully presented in surface, since the presence of the pronoun *la* means that the subordinate clause in (4b) has the full SS of a sentence. But in (4a) the subordinate clause in English lacks an NP in an essential argument position of the main verb *give* and consequently is not a SS sentence. Clearly then the +PRO languages present more of the LS of RC in surface than do the -PRO languages.

¹ We are ignoring here the logical properties of the definite determiner since, for purposes of our examples, they do not distinguish the Hebrew and the English RC.

It seems natural to attempt to generalize this point to the following explanatory principle:

Conservation of Logical Structure (CLS)

Let x be a logical structure and let x_i and x_j be distinct surface structures (from possibly distinct NL) which express x . Then if x_i presents more of x than does x_j , the syntactic transformations which generate x_i apply in a greater range of cases (or at least in more 'difficult' cases) than those which generate x_j .

Admittedly our inability to measure how much logical structure has been presented by some SS in a particular case, and our inability to measure the 'range' and 'difficulty' of cases to which a sequence of transformations apply, means that the application of CLS will be problematic in a great many cases. Nonetheless, in a few cases, such as the RC comparison we have been considering, its application will be relatively unproblematic and its explanatory force reasonably clear.

2. Type 2 comparison

Here we compare two or more semantically distinct LS with their SS expression in various NL. (By *semantically distinct*, as opposed to *semantically* (or *logically*) *equivalent*, we mean that the LS are interpreted differently in at least one model (state of affairs). E.g. such LS for sentences must have different truth values in at least one model; LS for singular definite descriptions must refer to different individuals in at least one model, and so on.)

Clearly if semantically distinct LS are naturally expressed by syntactically distinct SS in one NL but by the same SS in another, then the former is logically more expressive than the latter, *in that respect*.

The varying possibilities of stipulating co-reference in various NL determine an interesting class of type-2 differences. English (modern), and many NL, stipulate co-reference by contrasting marked and unmarked forms of pronouns as in (5).

- (5) a. John hit himself
b. John hit him

Clearly the marked pronoun in (5a) forces us to understand that the person hit was the same as the one doing the hitting, whereas this is not the case in (5b). Thus the situations in which (5a) is true need not coincide with those in which (5b) is true, and consequently the LS of (5a) is distinct from that of (5b).

Note that the stipulation of co-reference above need not be expressed in surface by pronoun marking. In a few NL at least, a contrast between a pronominal form and a deletion is used in such cases. (Where helpful in the examples below we use matching subscripts to mark NPs whose co-reference is being stipulated; otherwise different subscripts will be used.):

(6) *Japanese*

Taro_i -wa \emptyset_j /zibun_i-o tataita
Taro_i him_j/himself_i struck

(7) *Arosi* (Solomon Islands)

a. E urao_i na rubuia \emptyset_j
the woman_i hit her_j
b. E urao_i si a rubuia haria_i
the woman_i hit self_i

In light of these examples one might feel that universally NL could stipulate co-reference between subjects and objects of verbs. But this is not the case. In (8)–(12) the NL cited characteristically do not stipulate co-reference between subjects and objects, and are thus logically less expressive than e.g. Modern English

(8) *Middle English*¹ (Chaucer, *The Knight's Tale*)

- a. He leyde hym, bare the visage, on the beere (v. 2877)
(*He* = Theseus, *hym* = Arcite, as is clear from context.)
b. At Thebes, in his contree, as I sayde,
Upon a nyght in sleep as he hym leyde (v. 1384)
(*he* = Arcite, *hym* = Arcite.)

(9) *Fering* (A north Frisian dialect)

John hee ham ferreet
John has him/himself betrayed

(10) *Maori* (New Zealand)

ka patu ia ki a 'ia
hit he OM person him/himself
(mkr)
'he hit him/himself'

Gilbertese

E taraia Teerau
he-hit him/himself Teerau

¹ My especial thanks go to Genevieve Gembries, Universität Stuttgart, who worked through many Old and Middle English texts and found several 'minimal pairs' of the sort cited in (8).

- (12) *Tahitian* (Tryon (1970))
 'Ua ha'apohe 'oia 'iana
 was-kill he him/himself
 'He killed him/himself'

Furthermore, as the 'distance' between NPs increases it becomes more difficult to stipulate co-reference. An interesting pair in this regard are subjects of verbs like *think* and the subjects of their sentence complements. English for example does not generally stipulate co-reference in these positions. Thus the pronoun in (13),

- (13) John thinks that he is clever

does not change in form according as it is stipulated to refer to John or not, even though we might conceivably expect something like (14):

- (14) *John thinks that he-self is clever

Some languages however, e.g. Japanese and Korean, can stipulate co-reference in these positions by contrasting reflexive and non-reflexive pronouns:

- (15) *Korean*
- | | | | | |
|------------------------|---|----------------------------|---|-----------------------------|
| John _i -eun | { | keu _j -ka | } | yeonglihatago sankaghabnita |
| | | keu-chacin _i -i | | |
| John _i top | { | he _j subj | } | clever thinks |
| | | he _i self subj | | |
- 'John_i thinks that he_j/he_i is clever'

A few other languages use an emphatic form of the pronoun (e.g. *òn* vs. *ó* in Yoruba, *to* vs. *wi* in Kera) to stipulate co-reference in such positions, leaving the unemphatic form to be interpreted as not being co-referential with the matrix subject:

- (16) *Yoruba*
 Ojo ro pe (*òn* / *ó*) mu sasa
 Ojo_i thinks that (he_j/he_i) is clever
- (17) *Kera*
 Golsala dig minti (to /wi) bi cuuru
 Golsala thinks that (he_i/he_j) is intelligent

Furthermore, use of the pronoun/deletion contrast appears a more frequent way to mark co-reference in these cases than for subjects and objects of the same verb. (Note that in the examples below we only consider pairs of sentences differing solely by the presence or absence of the pronoun: in particular the *that*-complementizer is retained in the cases without pronouns.)

noun: in particular the *that*-complementizer is retained in the cases without pronouns.)

- (18) *Finnish*¹
 John_i luuli että (\emptyset _i /hän_j) oli sairas
 John_i thought that (he_i/he_j) was sick
- (19) *Malagasy*
 Nihevitra Rabe_i fa notadiavin- dRaso (\emptyset _i /izy_j)
 thought Rabe_i that looked for by Raso (he_i/he_j)
 'Rabe_i thought that (he_i/he_j) was being looked for by Raso'
- (20) *Polish*
 Jan_i powiedział, że (\emptyset _i /on_j) przyjdzie
 John_i said that (he_i/he_j) would come

NL then are logically different in that they vary with regard to the NP positions that are naturally stipulated as being co-referential. But for the moment we can offer no general explanation for this variation.

3. Type 3 comparison

Here we compare the SS expression of LS which are distinct but semantically equivalent (that is, they are always interpreted in the same way in a given model). Clearly distinguishing the SS expressions of such LS does *not* increase the logical expressive power of a NL. Consequently we might expect that (other things being equal) NL would vary according as they more naturally expressed the one or the other in surface. The *Principle of Logical Variants* below makes this intuition more precise. To state it we shall first make explicit a notational convention tacitly used earlier. Namely, for any natural language L_i and any logical structure x , x_i will denote any surface structure in L_i which expresses x . Secondly, we define:

Definition 1: Logical structures x and y are *logical variants in a logical context F* just in case Fx and Fy are logically equivalent. That is, just in case putting either x or y in the context F yields logically equivalent results.

We can now propose the following explanatory principle:

Principle of Logical Variants (PLV)

If x and y are logical variants in a context F then languages vary naturally according as they express x_i or y_i in the context F_i .

¹ Finnish can also stipulate co-reference in these positions by deleting the *että*-complementizer and using a non-finite verb in the subordinate clause.

Clearly the more interesting applications of PLV would be in the cases where the variants x and y are themselves semantically distinct but have their differences neutralized in the context F . For then logically expressive NL would have different SS for x and y and this difference would signal a difference in meaning; hence such NL would not in general be indifferent as to which surface structure x_i or y_i was used in an arbitrary context. But PLV precisely specifies one context where we do expect indifference – namely, those in which, for independent logical reasons, the meaning differences between the SS is neutralized.

Here we shall claim that the LS of singular definite RC (21a) and that of wh-questions (21b) are logical variants in the contexts determined by indirect question frames, e.g. 'John knows—', 'John remembers—', etc., as they appear in (21c).

- (21) a. the student that left early
 b. Which student left early?
 c. John knows which student left early

Clearly (21a) and (21b) are not logically equivalent so their LS must be distinct. Now if we can establish that their LS are variants then PLV will predict that NL will vary according as the object clauses of indQ have the form of a RC or of a wh-question (Q).

Let us note first that this prediction appears to be correct to a large extent. Thus in many NL – e.g. Finnish, Tagalog, Korean – the form of a simple Q does not change when it is embedded in indQ contexts:

- (22) *Finnish*
 a. Millä veitsellä John tappoi kanan?
 Which knife-with did John kill (the) chicken?
 b. Mary ei tiennyt millä veitsellä John tappoi kanan
 Mary not know which knife-with John killed chicken
 'Mary didn't know with which knife John killed the chicken'

And, as is well known, in many other NL (English, French, and to a lesser extent, German), the form of a simple Q changes only slightly when embedded in indQ frames.

What is perhaps less well known is that there are also many NL in which the object clauses in indQ are identical with (or in a few cases merely very similar to) singular definite relative clauses. E.g.

- (23) *Yoruba*
 a. *Relative Clause*
 Okunrin (na) ti obinrin na lu

man (the) that woman the hit
 'the man that the woman hit'

- b. *Wh-question*
 Okunrin wo- ni obinrin na lu?
 man which foc woman the hit
 'Which man did the woman hit?'

- c. *Indirect question*
 Tale mo okunrin ti obinrin na lu
 Tale knows man that woman the hit
 'Tale knows which man the woman hit'

Note that in (23c), if the interrogative form *woni* is used instead of the relative form *ti* the result is clearly ungrammatical:

- (23) d. *Tale mo okunrin woni obinrin na lu

Other NL in which such object clauses are like RC are: Shona, Malagasy, Urhobo, Roviana (Solomon Islands), Sa (New Hebrides), and Kera (Chad). In fact in English we often get both RC and Q structures in indQ. Thus the (a) and (b) sentences below are paraphrases (at least on one reading).

- (24) a. John knows the route the plane will take
 b. John knows which route the plane will take
 (25) a. John knows the year that Mary was born in
 b. John knows which year Mary was born in
 (26) a. John knew in advance the horse that would win
 b. John knew in advance which horse would win

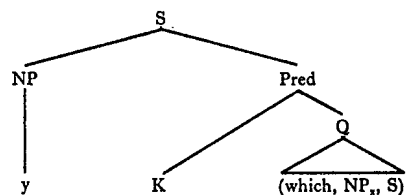
Clearly then if we can show that the LS of Q and of singular definite RC are variants in indQ frames then PLV will explain the variation we have observed in those positions. Below we present informally the main lines of such a demonstration. The analysis of direct questions is based on that in Keenan and Hull (1973a).

First, within the formal logic we provide distinct LS for such RC and Q. The former have, roughly, the form: (*the*, NP_x , S) which can be read as *the NP which is such that S* (where S is a logical sentence (formula) having a free occurrence of the variable x). Semantically, such structures are interpreted as the unique member of the universe of discourse which is in the set of things that the noun phrase NP_x is interpreted as and of which the sentence S is true. Logical questions on the other hand have the form (*which*, NP_x , S) which can be read as *Which NP is such that S?* Semantically they are interpreted as operators which combine with answer phrases (= definite NPs) to form a logical sentence. That is, question-answer pairs are treated

as propositions, things that determine truth values in a model. And in particular, restricting ourselves for purposes of exposition to singular questions, a definite NP will be a *true* answer to (*which*, NP_x , S) just in case it is interpreted as the unique NP_x of which S is true. A true answer to such a question then will always refer to (be interpreted as) the same object that (*the*, NP_x , S) refers to.

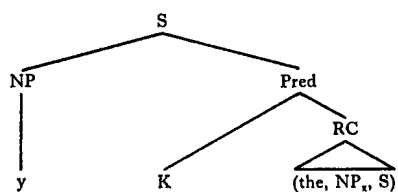
Second, we add to the logic a class of transitive predicates K , K_1 , K_2 , ... (= *know*, *remember*, etc.) which form intransitive predicates from either NPs of the form (*the*, NP_x , S) or from questions. Thus within the logic we have two formally distinct types of sentence, represented roughly as in (27) and (28) below. Now, to show that the LS of RC and Q are variants in indQ frames, we must specify the truth and falsehood conditions for (27) and (28) in a way which both accounts correctly for their consequences, presuppositions, etc., and which shows that they are logically equivalent – that is, that (27) and (28) have, in effect, the same truth and falsehood conditions.

(27)



'y knows (remembers, etc.) which NP is such that S'

(28)



'y knows (remembers, etc.) the NP which is such that S'

We shall informally suggest such a definition. But notice first why it is plausible that the definitions should be such as to determine the logical equivalence of (27) and (28). (28) says in effect that John can identify the referent of *the NP which is such that S*. And (27) says that John knows an answer to the question *which NP is such that S?* But to know an answer to such a question is precisely to be able to identify its true answer set, and this is just the referent of *the NP which is such that S*.

Consider now LS having the form (27). A formal statement of their semantics is given in Hull (this volume). The intuition he formalized is this: to say, e.g. that *John knows which student Mary invited* is to say that John knows an informative answer to the question *Which student did Mary invite?* (An answer is informative in a model if it is true but not necessarily so, i.e. there are other models in which it is a false answer.) Such an answer is given by a noun phrase NP' which *identifies* the student that Mary invited. That is, the LS of NP' is *the student that Mary invited* is both true and informative. So to say that John knows which student Mary invited is to say that, for some such NP' , *John knows that NP' is the student Mary invited*. More generally then:

Definition 2: (y , K , (*which*, NP_x , S)) is true in a model \mathcal{M} iff there is an NP' such that:

- (i) NP' is a true informative answer to (*which*, NP_x , S) in \mathcal{M} and
- (ii) (y , K that ($NP' = (\text{the}, NP_x, S)$) is true in \mathcal{M}

(We are assuming here the formal semantics of factives, e.g. *know that*, etc., given in Keenan (1972a).)

Concerning (28), the intuition that we want to formalize in sentences like *John knows the route the plane will take* is that John can identify the referent of the RC *the route the plane will take*. (Note that it is not the personal sense of *know*, as in *to know John* that we are trying to capture. Rather it is the more sophisticated use, as in *to know who John is*. In several NL of course this ambiguity does not arise since these meanings are expressed by different verbs, e.g. *connaître* vs. *savoir* in French, *kennen* vs. *wissen* in German, etc.) Now to be able to identify the referent of a RC is, linguistically, to be able to refer to it (using of course some noun phrase NP') in an alternative, that is, informative, way. There must then be some such noun phrase NP' equal in reference to the RC and such that John knows that the equality holds. Somewhat more generally then,

Definition 3: (y , K , (*the*, NP_x , S)) is true in a model \mathcal{M} iff there is an NP' such that:

- (i) ($NP' = (\text{the}, NP_x, S)$) is true and informative in \mathcal{M} and
- (ii) (y , K that ($NP' = (\text{the}, NP_x, S)$) is true in \mathcal{M}

Clearly condition (ii) here is identical to the corresponding condition in Definition 2. And condition (i) here is equivalent to condition (i) in Definition 2 since it says in effect that NP' informatively identifies the true answer set of the question (*which*, NP_x , S). Consequently the truth conditions of LS of the form (27) and (28) are the same. Similarly we could develop the

falsehood conditions of (27) and (28) in a natural way which would show them to be the same. Thus (27) and (28) would be shown to be logically equivalent, and so the LS of relative clauses and wh-questions would be logical variants in contexts determined by indirect questions.¹

4. A transformational solution to type 3 problems?

The type of variation we have explained in terms of (PLV) might appear to some to be more naturally represented within a transformational framework. That type of solution (hereafter the *T-solution*) would be roughly as follows: posit a single LS for indirect questions and then transformationally derive either RC or Q in these contexts depending on the NL. This solution would appear simpler than ours in that it needs only one type of LS for indQ where we have two. But our solution is to be preferred over the T-solution for two quite general reasons:

First, the T-solution offers no *explanation* as to why it is natural to get both RC and Q structures in indQ. Why for example do we not get focused (= cleft, in English) sentences in these positions? They are syntactically similar in many ways to RC and Q, as argued in Schachter (1973) and Keenan and Hull (1973b). Furthermore, why are we indifferent to the difference between RC and Q only in indQ? Why not generally? PLV provides an explanation for this (assuming that something like our LS for RC and Q are correct). Note that even if transformations deriving RC and Q from the same source in indQ contexts could be independently motivated we would still want an explanation as to why NL permit the existence of transformations which derive SS which are both syntactically and, generally, semantically distinct, from the same source in certain contexts.

And second, the T-solution creates a great many syntactic problems not created by our approach:

In the first place, it is clear that RC and Q will each have to have their own syntactic source in general (since they are syntactically distinct structures having in general a different distribution). So the common source for RC and Q in indQ will have to be different from at least one of these independently needed underlying structures, which means that surface RC or Q (or possibly both) will be derived from two distinct sources. This is inelegant and counter-intuitive (since there is in general no reason to think that either RC or Q have two distinct syntactic structures).

¹ The logical properties of indirect questions are not well known, and any proposal for representing their LS at this point will not be definitive.

And in the second place, the more syntactically distinct two types of SS are the less natural it is to derive them from the same source. Now although RC and Q are syntactically similar in many respects they are never systematically identical and the differences run deeper than the surface distribution of determiners (e.g. *the* vs. *which*). Thus quite generally, in any given NL the NP positions that can be relativized are not exactly the same as those that can be questioned. Some of the differences are language specific, and some are rather more general.

In the former category we might cite the following facts from German and English. In German many speakers find it difficult or impossible to relativize into sentence complements of verbs like *glauben* (= *think*) but it is often possible to question these positions:

(29) German

- a. Du glaubst, daß Fritz das Mädchen liebt
You think that Fritz the girl loves
'You think that Fritz loves the girl'
- b. Wen glaubst du, daß Fritz liebt?
'Who do you think that Fritz loves?'
- c. *Das Mädchen, das du glaubst daß Fritz liebt
'the girl that you think that Fritz loves'

On the other hand, in English we can 'simultaneously' relativize into NPs occurring across coordinate conjunctions in different grammatical cases, but the corresponding question is ill-formed:

- (30) a. the man who came early and who(m) Mary attacked
- b. *Who came early and \emptyset /who Mary attacked?

(Note that the question *Who came early and who did Mary attack?*, while well-formed, does not simultaneously question two NP positions. That is, it does not request the identification of one individual with two properties; rather it requests two identifications of individuals, each of which has one property.)

Finally, concerning the more language general differences in the formation of RC and Q, we feel that in general RC formation is a more esoteric process than Q formation. It is likely that all NL allow the major and near-major NP positions of main verbs to be questioned. (Even in very elementary grammars of languages we usually find a list of question words like *who?*, *what?*, *when?*, *where?*, *whose?*, etc.) On the other hand, many NL have only very restricted RC forming means. (See Keenan and Comrie

(1972) for a detailed justification of this claim.) For example many Malayo-Polynesian languages can only relativize on subject NPs.

A particularly nice differentiating position in this respect is genitives (e.g. 'the woman's cow died'). This position is often difficult to relativize, and in several languages – e.g. Tagalog, Basque, Malagasy, Fering – it is not naturally relativizable at all. Yet genitives can usually be questioned:

(31) *Basque*¹

- a. Apaiz-aren liburuak irakurri ditut
 priest-the-of books read have-them-I
 'I have read the priest's books'
- b. Noren liburuak irakurri dituzu?
 who-of books read have-them-you
 'Whose books have you read?'
- c. *liburuak irakurri ditudan apaiza
 books read have-them-I-rel priest-the
 'the priest whose books I have read'

Similarly in Fering we have an interrogative *whose* (= *Hokkers?*) and we do as well in Tagalog (*Kaninong?*).

In conclusion, I have attempted to represent three types of syntactic variation across languages in terms of the logical structures needed to represent the consequences, presuppositions, etc. of NL structures. In two of these cases we were able to propose explanatory principles, that of *Conservation of Logical Structure* and the *Principle of Logical Variants*, which enabled us to explain the observed variation.

REFERENCES

- Hull, R. (1973). 'A semantics for superficial and embedded questions in natural language'. This volume.
- Keenan, E. (1972a). 'On semantically based grammar', *Linguistic Inquiry*, 3, No. 4, 413–61.
- (1972b). 'The logical status of deep structures'. To appear in the *Proceedings of the 11th International Congress of Linguists*. Bologna, Italy.
- Keenan, E. and Comrie, B. (1972). 'Noun phrase accessibility and universal grammar'. Paper read at the winter meetings of the Linguistic Society of America.

¹ The examples here are due to R. P. G. de Rijk and represent the indigenous way of forming RC (e.g. head noun to the right of the subordinate clause, as is normal for SOV languages). Some dialects, under the influence of Romance, can in addition form RC with the head noun to the left of the subordinate clause, and this type does admit of relativization on genitives.

- Keenan, E. and Hull, R. (1973a). 'The logical presuppositions of questions and answers'. In J. Petöfi and D. Franck, eds., *Präsuppositionen in der Linguistik und der Philosophie* (1973). Frankfurt: Athenäum.
- (1973b). 'The logical syntax of direct and indirect questions'. To appear in *Parasession of the Ninth Regional Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society.
- Schachter, P. (1973). 'Focus and relativization', *Language*, 49, No. 1, 19–46.
- Tyron, D. (1970). *Conversational Tahitian*. Canberra: Australian National University Press.