Reference Restricting Operators in Universal Grammar

Edward L. Keenan
UCLA

In a standard logic (SL) the logical structure we assign to a sentence like he hit himself is roughly xHx, in which the identity of reference of the two referential positions is guaranteed by using the same variable in both positions. I will propose in this paper however an Extended Logic (EL) in which the two referential positions in logical structure (LS) are represented by distinct variables and an overt identity operator is used to force their coreference. Thus the LS for he hit himself will be (Ix,y)xHy, read as "identifying x and y, x hit y."

We shall support our claims for EL over SL using the five criteria below in terms of which the adequacy of a proposed LS is to be evaluated.

1. Primary Logical Adequacy. LSs for natural language sentences (Ss) must correctly determine the logical properties of such Ss. Roughly this is done as follows: The notions true and false in a state of affairs are formally defined on the LSs for Ss. Then one LS is said to be a logical consequence of another just in case it is true in every state of affairs in which the first is true. So to satisfy the criterion of Primary Logical Adequacy it must be the case that whenever one sentence is judged to follow from another then its LS must be true in every case in which that of the other is true. If this fails then the LSs we are using are not adequate. E.g., if the LS we assigned to John is dumb were not true whenever the one we assigned to John is rich and dumb was, then this assignment of LSs is not adequate, since, pretheoretically, John is dumb must be true if John is rich and dumb is true.

As regards logical presupposition,[1] we say that a LS Q is a presupposition of a LS P just in case it is true whenever P or its logical negation, -P, are true. Thus if Q is not true then neither P nor its negation are true, and P is then neither true nor false but has a third value (or perhaps is left undefined in this case). Intuitively, the presuppositions of a sentence are that part of the meaning of a sentence which is not affected by denying (or questioning) it. E.g. the student who cried was sad presupposes some student cried since the latter is true if either the former, or its natural denial, the student who cried wasn't sad, are true. In formulating the EL I am not concerned with how presuppositions are to be represented, it is sufficient that we distinguish in some way between what part of the information in a sentence (or LS) can be affected by negation and what part can't.

Clearly then the basic logical properties of a S are determined by the truth and falsehood conditions of the LS we assign it. Let's compare then these conditions for the two LSs proposed for he hit himself. In SL, xHx is true just in case the thing 'x' denotes is in the H relation to itself. It is false if it isn't. In EL, (Ix, y)xHy is true just in case x and y

denote the same object and that thing is in the H relation to itself. The LS is false if 'x' and 'y' denote the same object, but that object is not in the H relation to itself. Otherwise, i.e. if 'x' and 'y' do not denote the same object, the LS is assigned the third value (or left valueless).

The principle difference in the two LSs is that that of EL allows for the possibility that the two referential positions do not corefer, and in that case treats the LS in the same way it treats others whose presuppositions fail.

In SL however there is no way that the two positions can fail to corefer, since they are filled by the same referential expression (RE). Nonetheless, in both LSs coreference of the two REs is presupposed since in each case if the LS is either true or false the two REs must corefer. EL has only the very slight advantage over SL in that it makes the statement of the presupposition more explicit. Furthermore, in addition to representing the major presupposition of he hit himself both proposals would seem to represent the same consequences, since if either LS is true the two REs must corefer and the object referred to must be in the H relation to itself. We will assume then that there are no consequences or presuppositions of he hit himself which can be shown to obtain using one of the LSs proposed but not the other.

The two LSs do not so much differ with regard to basic logical adequacy then as they do with regard to the means used to express the presupposed coreference of the two referential positions. SL uses repetition of REs, and EL uses distinct REs and an explicit operator which identifies their reference.

We consider now criteria to use in choosing between LSs of the same degree of logical adequacy. First however we should justify why we want to make such a choice. We do, because we are interested in making generalizations concerning the logical expressive power of natural languages. Several such generalizations have been offered in Keenan [1973 and 1975]. E.g. it was shown that languages which present personal pronouns in positions relativized (e.g. the man that Mary saw him, rather than the man that Mary saw) permit the formation of a larger class of relative clauses than languages which do not present such pronouns. And we explained this on the grounds that pronoun retaining languages presented in surface more of their logical structure than pronoun deleting languages. But it is obvious that to make such generalizations rigorously we need a well defined set of logical structures expressible in any given language.

2. Criteria for Choosing Between LSs of the Same Degree of Logical Adequacy.

2.1 Simplicity. While this criterion is notoriously hard to apply in many cases, both in logic and in syntax, it does seem intuitively clear that the SL representation of he hit himself, xHx, is simpler than the EL one, (Ix,y)xHy since the former needs only a binary predicate symbol and two REs, whereas the latter needs all of this plus a new (relative to SL) category of sentence

operator--one which forms sentences from sentences by constraining the reference of REs which otherwise might have their reference interpreted independently. Clearly we must justify using a new category of logical operator to do something we can do without it.

And the justification is basically that, independently of representing the coreference in he hit himself, we need a class of operators of which the identity one is but a special case. Adding the identity operator then does not increase the types of logical categories we need.

Consider the LS need for sentences like (1) from Mojave (Pam Munro, pers. comm.):

(1) a. n^{y}a- iva:- {k / m} yaamo:m- k
b. when-arrive+3sg- {ss / ds} drunk+3sg-past
'When he_i arrived {he_i / he_j} was drunk'

In (1a) the use of the 'same subject' marker -k in the subordinate clause forces the interpretation that the person who arrived and the one who was drunk are the same. And in (1b) the 'different subject' -m marker forces the interpretation that they are different. (The marking remains the same when either clause takes full NP subjects as well.)

The problem for SL here is how to represent the LS of (1b) where the two REs incorporated into the verb must have different reference. Merely using distinct variables in the relevant referential positions in LS does not, and cannot in principle, force non-coreference. Distinct variables simply refer independently, and may, by accident so to speak, be assigned to the same individual. In fact, in SL there are LSs in which necessarily distinct variables are interpreted as the same individual. E.g. $(\forall x)(\forall y)$ ($x = y$ and zHw).

We might consider using a LS with distinct variables, conjoined with the LS which states that the variables differ in reference. Then the LS of (1b) above would be roughly:

(2) $(x \neq y)$ and (when x arrived y was drunk)

But (2) is not logically adequate because it explicitly asserts the non-coreference of x and y, which is presupposed in (1b). If it were merely asserted in (1b) then its natural denial, (3), could be true merely if x and y referred to the same object and when x arrived y was drunk was true. But in fact the non-identity of reference is not affected by negation. (3) still requires that the one who arrived and the one who was drunk be different, and states merely that y was not drunk when x arrived.

(3) n^{y}a- iva:- m, yaamo:m- mpotc
when-arrive+3sg-ds, drunk+3sg-neg
'When he arrived he wasn't drunk'

In EL however it is easy to represent the LSs of sentences like (1b). We merely use a reference restricting operator (RRO) which stipulates that two REs must be assigned different referents in order for the sentence operated on to be either true or false. Somewhat more formally then, (Nx,y)S will be true just in case x and y refer to different things and S is true; it will be false just in case x and y refer to different things and S is false. Otherwise it is third valued (or valueless). In EL then the LS of (1b) will be roughly (Nx, y)(when x arrive, y was drunk).

Now since SL seems to provide no logically adequate way to represent presupposed difference in reference, EL is to be preferred to it on the grounds of primary logical adequacy. Thus we have independent motivation for a class of RROs, and using simply another one to stipulate positive (rather than negative) co-reference does not increase the basic complexity of the logic.

We should further insist that marking negative co-reference is by no means an isolated phenomena across languages. It is common in Yuman languages like Mojave [Munro, 1974; Jacobsen, 1967; Winter, nd] as well as certain Uto-Aztecan languages like Hopi [Keenan, 1975]. It is further a typological trait of the Eastern New Guinea Highlands languages e.g. Fore (4), Scott [1973].

(4) a. kana-{ogá- / nta-} na wa-tá́- y- e
b. come-{ds 3sg past / ss 3sg past}- 3sg go-past-3sg-indic
'He$_i$ came and he$_j$ / he$_i$ went'

Note further that many languages, e.g. Turkish, Swedish, Finnish, distinguish reflexive from non-reflexive possession. Thus in Turkish (5a) [Eser Erguvanli, pc] the person hit must be someone's friend other than Ali's, whereas in (5b) it is Ali's friend who was hit.

(5) a. Ali o- nun arkadaş-i- na vur-du
Ali 3sg-3sg gen friend- 3sg poss-dat hit-past
'Ali hit his (≠ Ali's) friend'
b. Ali (kendi) arkadaş-i- na vur-du
Ali (self) friend- 3sg poss-dat hit-past
'Ali hit his (own) friend'

Finally note the many languages e.g. Yoruba (6) mark both positive and negative coreference between subjects of verbs of thinking and saying and subjects of their sentential complements.

(6) Ojo$_i$ ro pe {on$_i$ / o$_j$} mu sasa
Ojo thinks that {he (=Ojo) / he (≠Ojo)} is clever

2.2 Maximizing Generalizations Concerning Logical Expressive Power of Natural Languages. Other things being equal we should adopt those LSs which permit a natural statement of generalizations concerning the relation between LSs and the syntactic means used to express them. Now by representing both positive and negative coreference as special cases of RROs we can naturally state the following generalization: 'If two referential positions are in the domain of a negative RRO then they are also in the domain of a positive RRO'. That is, if a language can stipulate negative coreference between two positions then it can always stipulate their positive coreference as well. On our approach then we are stating that there is a conditional dependency between two members of a given class of LSs. This type of dependency is already known to be natural (e.g. if verbs agree with subjects in gender then they agree in number, etc.). On the other hand, if the LSs of positive and negative coreference are unrelated it seems much more arbitrary that there should exist any dependency relation between them.

The generalization relating positive and negative coreference suggests a simplification of our notation in which positive coreference is treated as the unmarked case. Henceforth, instead of writing (I$\underline{x}$,$\underline{y}$) we shall simply write ($\underline{x}$,$\underline{y}$), it being understood that when no overt RRO symbol is present then identity of reference is intended. Negative coreference of course will still be noted as (N$\underline{x}$,$\underline{y}$).

There is a second generalization we can make on RROs: Namely, any RRO always includes main clause subjects among the REs with respect to which the reference of other REs can be restricted. That is, subjects are always among the controllers of any type of reference restriction.[2]

2.3 Generality. If distinct Ss have a logical property in common then, other things being equal, this property should be represented in the same way in logical structure.

Here we would like to exhibit several other cases of positive coreference which can all be expressed by RROs of the sort we are proposing but which, in SL, would either be represented differently or not at all.

Case 1. Consider the positive coreference which obtains between full NPs and pronominal markers on verbs, as in Kinyarwanda (7) [Alexandre Kimenyi, pc].

(7) abanyeshuuri ba- rasoma igitabo
the students they-read book
'the students are reading a book'

In traditional terms such pronominal forms are considered to be agreements with the subject, and transformationally are thought to be introduced by a rule which copies certain features of the (possibly derived) subject onto the verb. On our approach however we consider such "agreements" to be independently generated pronominal forms, and the full NPs they corefer to to be operators which

place restrictions on their reference. Namely, they force the pronoun to refer to a specific person in the case where the full NP is e.g. a proper noun, or they constrain the pronoun to take its reference within a certain class in the case of a common noun. Thus the LS we propose for John drinks (ignoring tense) is (J,x) (x drink), read as "identifying x with John, x drinks." Similarly a (specific) man drinks will be represented as (M man, x)(x drink), which will be read as "restricting x to the set of men, x drinks." That is, the LS is true just in case x drink is true, where x refers to some object in the set of men. All men drink on the other hand will be represented as (All men, x)(x drink), which is true just in case x drink is true, where x is an arbitrarily chosen man. Analogously, some men drink is represented as (some men, x)(x drink).

If our analysis of agreements as pronominal forms is adopted we can explain several otherwise unexplained facts. Notably:

1) In languages in which the "agreements" are explicit (i.e. morphologically segmentable, and their form varies with the noun they corefer to or with the inherent properties of the referent, rather than with the verb subclass) they can function alone in main clauses as independently referring elements. Thus while the subject and object agreements in (8), from Swahili [Jean Tremaine, pc] might be argued to be copied from the full NPs, they cannot be so argued in (9) since there are no full NPs. But (9) can easily be used in a situation in which the hitter and the hittee are visibly present to speaker and hearer even though they have not been previously mentioned in context.

(8) Juma a- li- m- piga Ali
Juma he-past-him-hit Ali
'Juma hit Ali'

(9) a- li- m- piga
he-past-him-hit
'he hit him'

2) Verb "agreements" may code semantic features about their referent not present in the full NP and hence not obtainable by copying. A common case is where verbs "agree" in gender with a full NP which is not marked for gender either overtly or covertly (like proper names in English). Thus from Russian, Sasha pila = Sasha drank (fem), whereas Sasha pil = Sasha drank (masc). Hebrew and Avar provide further examples of this sort. More striking perhaps are cases where the verb "agreement" differs in number from the NP it corefers to. This is illustrated in (10) from Walbiri [Hale, 1973],(11) from Daga [Murane, 1974], and (12) from Spanish [Alfredo Hurtado, pc].

(10) ngarka-Ø ka- na pula- mi
man- abs pres-I shout-pres
'I (a man) am shouting'

(11) oaenapan war-apen ta-inton...
people get-inf do-1pl past
'We people were trying to catch (the pig)...'

(12) a. los mujeres protestamos pero...
the women complain-1pl but...
'We women complain but...'

b. los lingüistos tenéis una terminología muy poco
the linguists have-2pl a terminology very little
elegante
elegant
'You linguists have a very inelegant terminology'

A third type of case concerns the semantic relations which NPs bear to their verbs. Thus in Hadza (13) [Keenan, 1972] is ambiguous according as the lion killed the buffalo or the buffalo killed the lion. But the pronominal affixes on the verb are not ambiguous. The final "agreement" unequivocally refers to the agent, the preceding one the patient.

(13) seseme-ko //o-ta- kwa nak'oma-ko
lion- female kill-her-she+past buffalo-female
'the lioness killed the female buffalo' or
'the female buffalo killed the lioness'

Adopting our analysis of verb agreements as pronominal forms, then many simple sentences in many languages evidence positive coreference between full NPs and pronominal forms. This coreference is naturally represented in our logic by using further instances of RROs, as indicated above. In SL this coreference is not generally represented at all. Our logic appears more general than SL then, since coreference in a diversity of structure types is expressive in the same way.

Case 2. In SL the device of indicating positive coreference by repeating REs, even in simple reflexive cases, does not extend to the full range of such structures. Thus while we might represent John hit himself as jHj, this repetition will not work directly for sentences like Everyone hit himself, since the LS of Everyone hit everyone does not express the right truth conditions. In our system however the same identifying operators we need for simple cases apply without modification to these more complex ones. Thus the LS for John hit himself will be (John, x)(x,y)xHy, read as "identifying John with x, identifying x with y, x hit y." Similarly Everyone hit himself will be (All person,x)(x,y)xHy, read as "restricting x to people, identifying x with y, x hit y."

Case 3. Clearly the coreference expressed in a sentence like he hit only himself and only he hit himself is basically the same as that of he hit himself. But this is difficult to represent in SL since the two REs, the subject and object of hit, are the same RE and hence it is difficult to have an operator like only operate on one of them independently. In our system however the minimal

meaning of only is easy to represent. (only x)S, read as "only x is such that S" is true just in case S holds of x and is false of anything different from x.[4] Otherwise it is valueless (or third valued). Assuming this logical analysis of only we can represent only he hit himself as (only x)(x,y)xHy, read as "only x is such that identifying x with y, x hit y." And he hit only himself will be (x,y)(only y)xHy, read as "identifying x with y only y is such that x hit y." Similarly we can naturally distinguish the LSs of Nixon likes only himself, (n,x)(x,y)(only y)(xLy), and Nixon likes only Nixon, (n,x)(n,y)(only y)xLy, as well as Only Nixon likes only himself, (n,x)(only x)(x,y)(only y)xLy, and Only Nixon likes only Nixon, (n,x)(n,y)(only x)(only y)xLy. We note that logically adequate structures for these sentences can be provided by SL, but they are highly unnatural by our criterion in 2.4.

Case 4. Since RROs may restrict the reference of two REs, it is easy to represent the double binding in Bach-Peters sentences like the man who slapped her really loved the woman who insulted him, where the reference of the two NPs the man who slapped her and the woman who loved him is not independent one from the other. Further, this type of double binding is more prevalent across languages than has previously been realized. Note the 'Janus construction' in Turkish (14) [Eser Erguvanli, pc). This construction is also present in Luiseño [Pam Munro, pc] and Persian [Galust Mardirussian, pc].

(14) $\text{vali}_i\text{-si}_j$ $\text{köy}_j\text{-ü-nü}_i$ metetti
mayor_i-3sg poss_j village_j-acc-3sg poss_i praised
'Its mayor praised his village'

In (14) the reference of its must be to village, and the reference of his must be to mayor. Thus the reference of its mayor and his village cannot be determined independently of each other. Note further that the pattern of coreference is presupposed. If we deny (14) we obtain (15) in which the pattern of coreference is not changed.

(15) vali_i- si_j $\text{köy}_j\text{-ü-nü}_i$ met et-me- di
mayor_i-its_j village_j-acc-his_i praise-neg-past
'Its mayor didn't praise his village'

Note further that this type of double binding may extend across clause boundaries (at least in underlying structure).

(16) $\text{vali}_i\text{-si}_j$ $\text{köy}_j\text{-ü-nün}_i$ güzul
mayor_i-its_j village_j-acc-his(gen)_i beautiful
ol-dug-u- nu soyle-di
be-nom-acc-poss say- past
'Its mayor said that his village was beautiful'

The LSs we provide for such sentences are as follows: (14) = (w,y)(x,z)(w's mayor, x)(z's village, y)(x praise y). The truth conditions of (14) are, informally: "x praised y, where x is identified with w's mayor and w is identified with z's village, and z in turn is identified with x, the mayor."

Since representing double binding has been problematic in SL but requires no additional operators in EL we conclude that EL is more general than SL.

2.4 Correspondence Principle (Naturalness). We would like to argue that certain LSs are more natural than others in that their formal properties correspond more closely to the surface syntactic properties of the natural languages Ss which express them. What constitutes a closer correspondence will doubtless be problematic in many cases, but some cases seem to us clear enough to constitute an argument in favor of EL over SL. The following correspondence principle (CP) covers many of these cases: "LSs are more natural according as distinct elements of the LS correspond to distinct elements in surface, and identical elements in LS correspond to identical elements in surface."

The reason why CP is natural is that speakers do make inferences i.e. assess logical consequences, on the basis of information provided in surface structure. If the LSs we use to represent the surface forms have many formal properties in common with the surface form we have at least some hope that we have represented properties of surface forms which speakers actually use in coding their logical properties.

Consider e.g. a trivial instance of CP. Suppose that even within SL one were to propose (17c) as a LS for (17a) rather than the more usual (17b).

(17) a. Socrates is a man
b. man(s)
c. (man(s) and ((∃x)Horse(x) or not (∃x)Horse(x))

Since (17c) is the conjunction of (17b) with a tautology it follows that (17b) and (17c) always have the same truth value and hence have the same degree of primary logical adequacy as representations for (17a). But CP states clearly that (17b) is the more natural representation of the two since (17c) contains many distinct elements e.g. Horse, or, etc. which are not present as distinct elements in (17a).

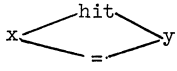
Consider now our two proposals for representing positive coreference. The SL one expresses the coreference in he hit himself or John hit himself by repeating the subject RE in object position. If this were natural by CP we would expect to find across languages that reflexives sentences were expressed literally as he hit he or John hit John. But in fact this never happens. No language expresses positive coreference in this environment by using identical NPs in the subject and object slots. Even if the subject and object are both proforms, they are probably always distinct in some

way. Hence SL is unnatural by CP since its LSs present identical elements in places where natural languages use distinct elements. By the same token, our logic is in this respect more natural since it uses distinct elements in the referential positions.

Furthermore many languages (but far from all) distinguish in surface the coreferential expression from the identity binding element. Thus in the switch reference cases cited earlier the marker of same reference or different reference is distinct from the REs whose reference is restricted. And even in many simple reflexive constructions we can often distinguish the referential element from the identity element. Thus in many languages, e.g. Basque (18), Georgian, Berber (Tamazight), Tera, Hebrew, etc. John hit himself is literally something like John hit his head, his bone, his self, etc. where head, etc. represent the referential element, and the possessive adjective or pronoun represents the identity operator.

(18) Gizona-k bere burua jo zuan
man-erg his head hit he-had

On the other hand our LSs appear unnatural in that no language normally renders He hit himself as "identifying he with him, he hit him." That is languages do not have repeated occurrences of the subject and object proforms. This suggests that we modify our logical notation so that two REs be directly related by two or more relational symbols--the main one, and the 'logical' one of identity or non-identity. Thus we might use something like



which would indicate that x bears both the hit and the identity relation to y. This seems to us a real possibility, but the notation would have to be formally worked out before it could be taken seriously.

We also should mention that many languages do not present two surface REs in simple reflexive constructions--rather the verb takes a reflexivity marker and no object proform appears. By CP this type of reflexive should have a different LS than the ones we have been considering, and this seems to us correct, although space prevents us from arguing the point.

Finally, consider the naturalness of the LSs we propose for simple non-reflexive sentences like John hit Bill, namely (j,x)(b,y)xHy, We have already indicated that in many languages the variables attached to the predicate show up as 'pronominal agreements' on the verb. It further happens, to a surprising extent, that full NP subjects and objects also carry pronominal indices which match those on the verb making them look extremely similar to the LSs we propose. E.g. (19) from Swahili [Alex. Kimenyi, pc].

(19) wa- naume wa- li- m- piga m- ke
they-man they-past-her-hit she-woman
'the men hit the woman'

From Avar [Anderson, nd], who notes,pc, that this matching is not synchronically productive).

(20) v-as v-eker-ula
-boy -run-pres
'the boy runs'

From Genoese [B. Vattuone, 1975?]

(21) A Katayni a vende i pesi
3sg fem Catherine 3sg fem sells 3pl fish
'Catherine sells fish'

It appears then, that at least for certain simple structures in many languages, EL appears more natural than SL. And overall then, the criteria in 2.1-2.4 largely argue in favor of our analysis of RROs over the analysis of SL.

Footnotes

[1]For some discussion of logical presupposition see Keenan [1972], Kartunnen [1973] and Van Fraasen [1969].

[2]A possible counterexample: Angaataha [Huisman, 1973] is cited as having switch location markers, both positive and negative, but not switch subject markers. That is, if the location of the action of an early clause is changed in the next clause the verb takes a certain marker; if no change then the verb marking is different.

[3]In fact, we would be inclined to propose as underlying structures simply a set of REs expressions and a non-logical predicate, together with the various relations which the REs bear to the predicate (e.g. agent of, patient of, etc. as well as to each other (e.g. identity, non-identity).

4. **It is false if S holds of x and something different from x.**

APPENDIX

FORMAL SEMANTICS OF REFERENCE RESTRICTING OPERATORS

We assume a definition of interpretation similar to that in Keenan [1972]. Essentially an interpretation of the formal language specifies a universe of discourse, provides a unique discourse name for each object in it and allows these names to occur in the same positions as free variables, and gives an interpreting function f, which interprets sentences as truth values, variables and proper nouns as members of the universe of discourse, and common noun phrases as subsets of the U of D. We shall refer to the