

TOWARDS A UNIVERSAL DEFINITION OF "SUBJECT"*

by

Edward L. Keenan

*Research for this paper was supported by a grant (#2994) from the Wenner-Gren Foundation and a grant from the Social Science Research Council in Britain.

I The Problem

In this paper I will attempt to provide a definition of the notion "subject of" which will enable us to identify the subject phrase(s), if any, of any sentence in any language.

Such a definition is needed in universal grammar in order for the many universal generalizations which use this notion to be well defined. For example:

- 1 Accessibility Hierarchy (Keenan and Comrie, 1972): NPs on the upper end of the AH, given below, are universally easier to relativize than those on the lower end. Thus some languages (Ls) have relative clause forming strategies which apply only to subjects; other Ls have strategies which apply only to subjects and direct objects, others have ones which apply only to the top three positions on the AH, etc.

Subj > Dir Obj > Ind Obj > Obl > Gen > Obj of Comp

- 2 Functional Succession Principle (Perlmutter and Postal, 1974): If one NP can be raised out of another then it assumes the grammatical relation (subject of, direct object of, etc.) previously borne by the other.
- 3 Advancement Continuity Principle (Johnson, 1974c; Trithart, 1975; Keenan, 1975a): If a L can advance (e.g., via operations like Passive) NPs low on the AH to subject then it can advance all intermediate positions to subject. Thus if a L has a locative voice (e.g., the school was seen Mary at by John) then it necessarily has a direct object (= Passive) voice.

Clearly generalizations like 1-3 determine constraints on the form, and substance, of possible human languages. But to verify them and determine their universality it is necessary to be able to identify subjects, direct objects, etc., in a principled way across Ls. If we use different criteria to identify subjects in different Ls then "subject" is simply not a universal category and apparently universal generalizations stated in terms of that notion are not generalizations at all. In addition, absence of identifying criteria for subjects, etc. makes verification of putative universals like 1-3 difficult. Counter examples can be rationalized away by merely saying that an offending NP is not "really" a subject, etc. Finally the claim that we have primitive intuitions concerning which NPs are subjects founders, like all arguments from intuition, when the intuitions of different individuals do not agree. For example, it would appear that George (1974) and Keenan and Comrie (to appear) have differ-

ent intuitions concerning which NPs are subjects of simple transitive sentences in Dyirbal (Dixon, 1972).

II.0 Defining "Subject of"

We are not free to define a notion like "subject" in any way that suits our purposes. There is a large body of lore concerning the notion, and any proposed definition must at least largely agree with the traditional, and to some extent, pretheoretical usage of the term. Our approach then will be to collect a large and diverse set of cases from different Ls in which our pretheoretical judgments of subjecthood are clear. Then we shall attempt to abstract from this set a set of properties which are characteristic of subject NPs and then try to determine some combination of the characteristic properties which will be jointly necessary and sufficient to pick out the subject of an arbitrary sentence in an arbitrary L in a way that is in conformity, of course, with our pretheoretical intuitions in the clear cases.

However, even a cursory examination of subjects across Ls reveals that in many Ls subject NPs are characterized by properties which are not only not universally valid, they are peculiar to the particular L in question. For example, in Latin, subject NPs carry a characteristic case marking (the nominative). But that particular ending probably occurs as a nominative marker in no other L; there is no universal nominative case marker. And in Malagasy (Malayo-Polynesian; Madagascar) subjects characteristically occur clause finally, whereas in most Ls subject NPs precede the other major NPs within clauses. Nonetheless, we do want to say that the prima facie evidence that an NP is the subject of some sentence in Latin is that it is nominative, and in Malagasy that it is clause final. Consequently we want to phrase a universal definition of "subject" in such a way as to allow that different Ls may use language specific means to mark subject NPs. This we shall do in the following way:

First, universal means of distinguishing a privileged subset of sentences in any L will be given. These sentences will be called the semantically basic sentences (henceforth b-sentences) and their subjects will be called basic subjects (henceforth b-subjects). Then we shall attempt to provide universally valid criteria for identifying subjects of b-sentences in any L. Once the b-subjects have been identified in any particular L then the full set of properties characteristic of b-subjects in that L can be determined. So in some Ls b-subjects may have a certain case marker, or position, or they may exhibit a very specific type of transforma-

tional behavior, or even have semantically specific properties. Finally, once the full complement of b-subject properties has been determined for a given L, subjects of non-basic sentences will be defined to be those NPs, if any, which present a clear preponderance of the properties characteristic of b-subjects. Thus in any given L, subjects of non-basic sentences may present very few of the universal properties of subjects, but still be clearly identifiable as subjects in that L since they possess very many of the language specific properties of b-subjects in that L.

Note further, that on this type of definition, subjects of certain sentences, and more generally of certain sentence types, will be more subject-like than the subjects of others. The reason is that they will exhibit more of the complement of properties which characterize b-subjects in general. Thus the subjecthood of an NP (in a sentence) is a matter of degree. ✓

In addition, it seems to me that subjects in some Ls will be more subject-like than those of other Ls in the sense that they will in general, present a fuller complement of the properties which universally characterize b-subjects. Very possibly, for example, European Ls are more subject oriented than those Sino-Tibetan Ls discussed by Li and Thompson (this volume).

II.1 The Definition of Basic-Sentence in a Language

- 4 For any Language L,
 - a. a syntactic structure x is semantically more basic than a syntactic structure y if, and only if, the meaning of y depends on that of x. That is, to understand the meaning of y it is necessary to understand the meaning of x.
 - b. a sentence in L is a basic sentence (in L) if, and only if, no (other) complete sentence in L is more basic than it.

Concerning 4a, there is no simple way of determining whether some sentence e.g. is more basic than another since such a determination requires that we understand the meaning of the two sentences. So some cases will surely be problematic. But many cases we feel are quite clear. E.g., John is a linguist is clearly semantically more basic than Fred thinks that John is a linguist since we cannot understand the meaning of the latter without understanding that of the former. If we didn't know what John is a linguist meant, we wouldn't know what Fred is thinking. Further, each of the

following structures is semantically more basic than all of those which follow it: John sang, John sang off-key, John didn't sing off-key, the fact that John didn't sing off-key, some newspaper reported the fact that John didn't sing off-key, some newspaper which reported the fact that John didn't sing off-key, etc. The relation more basic than (henceforth MBT) is transitive. If x MBT y, and y MBT z, then x MBT z. For if we need to know the meaning of x in order to know that of y, and that of y is required in order to know that of z, then clearly we can't understand the meaning of z without understanding that of x. Note however that the meanings of two different structures may simply be independent, so neither need be more basic than the other (1). E.g., the meaning of John's leaving surprised everyone neither depends on nor is dependent upon the meaning of if anyone screams Mary will faint.

Concerning 4b, the b-sentences in any L are defined to be the maximally basic structures having the category "sentence," although we do require that a sentence be in the set of b-sentences if the only other sentences more basic than it are "too context dependent" for their meaning (that is, they do not express a "complete thought"). The intent here is to rule out cases where the only NPs occurring in b-sentences would be pronouns. Thus in many Ls, if not in English, it is arguably the case (Keenan, 1972a) that pronominal sentences like he hit him are more basic than ones like John hit Bill. Compare e.g. 5a and 5b from Swahili.

5 a. a- li- m- piga
3sg past-3sg hit
human human
subj obj
"he hit him"

b. Juma a- li- m- piga Faru
Juma he-past-him-hit Faru
"Juma hit Faru"

Sentence 5b differs in meaning from 5a solely in that the reference of the pronominal elements in 5a is made more specific. To determine e.g. the truth of 5b we must first determine the reference of the pronominal elements as Juma and Faru, and then determine that 5a is true. So we can't really understand the meaning of 5b without being able to understand that of 5a, so 5a is plausibly more basic than 5b. Nonetheless 5a is "incomplete" in that it does not really tell us who is being talked about. Plausibly then no complete sentence in Swahili is more basic than 5b, or 5a, so both of these are among the b-sentences in Swahili (although the

basicness of the tense marking in the example would have to be further considered).

II.2 Some General Properties of Basic Sentences

While there is no mechanical procedure for identifying the set of b-sentences in a L, the set will generally have several characteristic properties which makes the identification of a fairly large set of relatively basic sentences reasonably easy.

Thus, in general, if the meaning of one structure depends on that of another, then the form of that structure also depends on the form of the other. In general then we expect that b-sentences will exhibit the greatest morphological and syntactic potential of the sentences in a L. Thus they will present the greatest range of tense, mood, aspect, mode, and voice distinctions. They will have the greatest privileges of occurrence: they will accept the greatest range of verbal and sentential modifiers; they will be the easiest to embed and adjoin to other sentences, the easiest to nominalize and internally reorder, the easiest to relativize, question and topicalize out of, the easiest to pronominalize and delete into, etc. In other words, the b-sentences are roughly the "simplest" sentences syntactically. For example, 6a below is clearly more basic than 6b.

6 a. Mary doesn't like John anymore

b. As for John, Mary doesn't like him anymore

For clearly, to understand 6b we need to know what Mary doesn't like him anymore means, where him is understood to refer to John. But this is exactly what we need to know to understand 6a. The syntactic test for b-sentencehood then predicts that 6a has more potential than 6b, which is clearly correct. E.g., we can nominalize 6a, Mary's not liking John anymore, but not 6b, we can cleft on Mary in 6a, It is Mary who doesn't like John anymore, but not in 6b, etc.

Semantically speaking as well, b-sentences can be expected to have certain characteristic properties. They will e.g. generally be structurally unambiguous. The reason is that in general the operations which form less basic structures from more basic ones affect only part of the meaning of the less basic one. And since b-sentences in general have fewer "parts" than non-basic ones they present fewer possibilities for the interpretation of which parts are understood to be affected by the operation.

Consider for example passives in English. Although the case is less clear than the ones considered earlier, we con-

sider passives to be, in general, less basic than the corresponding actives. For what we must understand about the relations between the participants and the action in a passive sentence is basically the same as that of the active. But in addition, in the passive we require that the referent of the object NP be identifiable independently of that of the agent (see 3.0 in Section II.3 for more discussion), that is, loosely, the patient is more of a topic in the passive than the active sentence. Assuming then that actives are more basic than passives then we would expect from our argument above that at least some ways of forming complex structures from simpler ones would engender more ambiguities when applied to passives than to actives. And this seems to be the case. Consider e.g. the addition of adverbials like willingly in 7. (See Lasnik and Fiengo, 1974, for some discussion.)

- 7 a. The police arrested John willingly
b. John was arrested by the police willingly

In 7a no indication that John was a willing participant in the act is given. But 7b is ambiguous according as it was John or the police who acted willingly. A further instance, more problematic in some respects, concerns pairs like

- 8 a. Every boy kissed a girl
b. A girl was kissed by every boy

Sentence 8b seems to me more clearly ambiguous according as it was necessarily the same girl or not. But 8a seems to me to have only the reading on which it is not necessarily the same girl (though it might be). Not everyone agrees with these judgments, however, and in the (remarkable) absence of a clear pretheoretical intuition concerning what it means to say that a structure is ambiguous the case is hard to argue.

In addition to the general absence of structural ambiguity b-sentences may be expected to have other semantic properties which more directly depend on the fact that they are the semantically primitive sentences. Thus they will in general be declarative and affirmative. Further in many cases non-basic sentences seem semantically richer than the more basic ones whose meaning they depend on. It is tempting to think that the b-sentences of a L will always be among the least informative sentences of the L (Talmy Givón, personal communication). But while this is often so, I think it not always the case. Thus there is a class of non-basic sentence forming operators whose semantic effect is precisely to attenuate to some degree the more specific meaning of the more basic sentences. Thus I would want to say that John left is

more basic than John didn't leave, John might have left, and It is possible that John left. Yet all the latter sentences, in some intuitive sense, tell us less about how the world is than does the more basic John left.

Another measure of the semantic "simplicity" of b-sentences concerns their relative freedom from presupposition. Thus many of the clear examples of presuppositional structures (see e.g. Keenan, 1970) involve embedded sentences (e.g. relative clauses, factive complements of verbs like realize, surprise, etc.). And generally if a sentence contains an embedded sentence then its meaning depends on that of the embedded one and so it is not basic. Unfortunately, however, b-sentences are not presupposition-free. The use of proper names and demonstrative NPs in b-sentences normally carries the presupposition that the NP has a referent. It seems to me likely however that aside from lexically specific presuppositions of basic predicates the presuppositions of b-sentences can be limited to existence claims.

A final characteristic of b-sentences: It is not necessarily the case that a b-sentence of one L will translate as a b-sentence of another L. The basic predicates of different Ls codify somewhat different concepts. So e.g. the president resigned receives no b-sentence translation in Ls spoken only by peoples which do not have the sort of political institutions referenced by terms like president and resign. A more interesting question perhaps is whether the syntactic types of b-sentences are the same across Ls. The question is an empirical one and the answer is not obvious. E.g. while passive sentences may be non-basic in English they are at least much more basic (see Note 1 for this more generalized use of "basic") in many Malayo-Polynesian Ls such as Maori, Tagalog, Malagasy. In particular the privileges of occurrence of passives in those Ls are much greater than in English. Further, syntactic properties which are obligatory in some Ls may only be optional in others, and possibly then not present in the b-sentences of those Ls. E.g. in English all sentences, and therefore the b-sentences, are marked for tense/aspect. But in Indonesian such marking is optional. Sentences lacking such marking occur frequently in discourse (and in isolation, in translation). Plausibly then the b-sentences in Indonesian express "naked propositions" whose meaning is made more specific by the addition of tense/aspect adverbs in non b-sentences.

II.3 Characteristic Properties of Basic Subjects

Below we present a list of 30 odd properties which subjects characteristically possess (if any NP in the L does).

The properties may be pragmatic, semantic, or syntactic. And of the syntactic ones, some concern properties internal to a single sentence and others concern the relation between a b-sentence and some modification of it. The properties are organized into groups and sub-groups according to their relationships. Plausibly in many cases properties in the same sub-group are not independent (although they can be verified independently) but we cannot at the moment show this, in general.

Furthermore, we have not been able to isolate any combination of the b-subject properties which is both necessary and sufficient for an NP in any sentence in any L to be the subject of that sentence. Certainly no one of the properties both necessary and sufficient, and in our statement we point out counter examples both ways for properties which might have been thought fully general.

Consequently we must have recourse to a somewhat weaker notion of definition. We shall say that an NP in a b-sentence (in any L) is a subject of that sentence to the extent that it has the properties in the properties list below. If one NP in the sentence has a clear preponderance of the subject properties then it will be called the subject of the sentence. On this type of definition then subjects of some b-sentences can be more subject-like than the subjects of others in the sense that they present a fuller complement of the subject properties.

Note further that on this type of definition "subject" does not represent a single dimension of linguistic reality. It is rather a cluster concept, or as we shall say, a multi-factor concept. Many basic concepts in social science are multi-factor concepts. Thus one's intelligence is a combination of abilities (Thurstone, 1956): verbal comprehension, immediate recall, numerical manipulation, visualizing flat objects in three dimensions, and deductive reasoning. So one can be intelligent in different ways and to different degrees. And being a subject is, we claim, more like being intelligent than, for instance, like being a prime number. The factors which compose the concept of subject might coincide with our groupings of properties, though in the worst of cases each of the 30 odd properties would be an independent factor.

II.3.1 The Subject Properties List (SPL)

We present four major categories of b-subject properties. The first, Autonomy Properties, is by far the largest.

A. Autonomy Properties

1. Independent Existence. The entity that a b-subject

refers to (if any) exists independently of the action or property expressed by the predicate. This is less true for non-subjects. Thus in a student wrote a poem the existence of the poem is not independent of the act of writing, whereas the existence of the student is. Other examples: someone committed a booboo, defined a term, proved a theorem, etc.

2. Indispensability. A non-subject may often simply be eliminated from a sentence with the result still being a complete sentence. But this is usually not true of b-subjects. E.g. John hunts lions (for a living), John hunts (for a living), *hunts lions (for a living).

Several ergative Ls however do appear to permit unspecified subject deletion. Notably Tongan, Eskimo, and Tibetan.
Tongan (Churchward, 1953)

- 9 a. Na'e tamate'i 'e Tevita 'a Koliath
killed subj David obj Goliath
"David killed Goliath"
- b. Na'e tamate'i 'a Koliath
killed obj Goliath
"Goliath was killed"

(NB: 'a marks both transitive objects and intransitive subjects.)

3. Autonomous Reference. The reference of a b-subject must be determinable by the addressee at the moment of utterance. It cannot be made to depend on the reference of other NPs which follow it. Thus if two NPs in a b-sentence are to be stipulated as being the same in reference it will either be the non-subject which gets marked (perhaps deleted) or the rightmost NP. Thus in English we could never say He-self admires John_i for John_i admires himself, for in the first sentence the reference of the subject cannot be determined independently of that of a following NP, so the subject would not be autonomous in reference. Note however that in Ls in which b-subjects can follow other NPs it is sometimes (but not always) possible for a leftmost NP to control the reference of the subject. The subject's reference is still determinable at the moment of utterance since it is stipulated as being coreferential to an NP whose reference has already been established.

Tagalog

- 10 a. sinampal ng lalake ang babae
+pass
hit-by Agt man subj woman
"the woman was hit by a/the man"

- b. sinampal ng lalake ang kaniyang sarili
hit-by Agt man subj his self
"the man hit himself"

Here subject NPs i.e. ang phrases (see Schachter, this volume, for some problems with the notion of "subject" in Philippine languages) normally may precede or follow objects, as long as both follow the verb. But when objects are definite (as they are if specified as being coreferential to a definite subject--and subjects are always definite in Tagalog) they must be presented as surface subjects. And in this case, contrary to the usual pattern, the subject must follow the agent phrase. Thus its reference is determinable at the moment of utterance.

Samoan (ex. from Chapin, 1970)

- 11 a. sa sogi e ioane ia lava
past cut Agt John he emph
"John cut himself"
b. sa sogi ioane ie ia lava
past cut John Agt he emph
"John cut himself"

Here again the NP whose reference is dependent on that of another occurs second. In 11a that NP is plausibly a subject if that sentence is passive. If Samoan is ergative in this case then John is the subject and it is the object that is pronominalized. But then in 11b it would be the ergative subject whose reference was not independent. But on either analysis the reference of the b-subject is always determinable at the moment of utterance. In fact, we know of no clear counter examples to this property. So autonomous reference is plausibly a universal necessary condition on b-subjecthood.

But this property does seem to understate the facts a bit. Thus in many Ls in which the subject may follow the objects, such as Malagasy, Gilbertese, and Tzeltal (see Keenan 1974), the reference of the subject cannot be made to depend on that of the object, even though the object precedes and so the property of autonomous reference would not be violated. E.g.

Malagasy

- 12 a. manaja tena Rabe
respect self Rabe
"Rabe respects himself"

- b. *manaja an-dRabe tena/ ny tena-ny
respect acc-Rabe self the self-his
"He-self respects Rabe"

There are several other properties of subjects that plausibly are related, more or less closely, to the property of autonomous reference.

3.1 b-subjects are always (in general, not necessarily in every sentence) among the possible controllers of stipulated coreference, either positive or negative (see Keenan, 1975a for discussion). Thus,

3.1.1 b-subjects in general can control reflexive pronouns. And in some Ls control of reflexives within clauses is largely restricted to b-subjects. E.g. in Malagasy, Japanese, and German.

3.1.2 b-subjects are among the possible controllers of coreferential deletions and pronominalizations. Note:

- 13 a. John_i talked to Bill_j for awhile and then he_{i,j} left
b. John_i talked to Bill_j for awhile and then $\emptyset_{i,*j}$ left

3.1.3 The possible controllers of backwards pronominalization and deletions include b-subjects. E.g.

- 14 a. When he_{i,?j} got home, John_i talked to Bill_j
b. On $\emptyset_{i,*j}$ arriving home, John_i talked with Bill_j

3.2 The NPs which control "switch reference" indicators include b-subjects. Thus in the Eastern Highlands Ls of New Guinea as well as among many groups of American Indian Ls, subordinate verbs will carry either one of two affixes according as their subject is coreferential or not coreferential with the subject of some other clause. E.g.

Hopi (ex. from Pam Munro, personal communication)

- 15 pam navoti:ta pam mo:titanif^{-q}
-qa?e
he thinks he win { -diff. subj.
-same subj.
"He thinks that he will win"

Thus if the suffix -q is chosen on the subordinate verb the person who is being thought about is necessarily different from the one doing the thinking, whereas they are the same if the suffix qa?e is chosen. For further discussion see Keenan, 1975a; Jacobsen, 1967, Munro, 1974a. Our point here is only that control of switch reference is largely limited to subjects. We note however that switch location markers rather

than switch subject markers are reported for Angaataha (Huisman, 1973) in New Guinea.

3.3 The NPs which control verb agreement, if any, include b-subjects. (For an argument that verb agreement is, at least historically, pronominalization, see Givón, this volume). We further note that verb agreement across clauses appears restricted to subjects, though this is only attested to our knowledge for the Ls of the E. New Guinea Highlands. See Keenan, 1975a for examples.

Since verb agreement is one of the properties which people have considered definitional of subjects, it is worth noting that it fails to be a necessary condition on b-subjecthood since in very many Ls verbs agree with no NP. E.g. Swedish, Sinhalese, Afrikaans, Thai, Vietnamese, Chinese, Japanese, Maori, Malagasy, etc. Similarly verb agreement fails to be a sufficient condition on b-subjecthood since in many Ls verbs agree with NPs in addition to subjects, e.g. Basque, Chinook, Arosi (Melanesian), Jacaltec (Mayan), Kapampangan (Philippine), Hungarian, Georgian, Blackfoot, Machiguenga (Arawak; Peru), etc. Note further that in a very few cases verbs may agree with objects but not with subjects. E.g. Avar (Caucasian), Mabuiag (Australia), and, very partially, Hindi.

3.4 b-subjects are the easiest NPs to stipulate the coreference of across clause boundaries.

3.4.1 If reflexive (i.e. essentially anaphoric) pronouns in sentence complements of verbs of thinking can be bound by NPs in the matrix clause then these pronouns can always occur in subject position in the complement clause.

Yoruba

- 15 Ojo_i ro po on_i / ò_j mu sasa
 Ojo_i thinks that he_i / he_j is clever
 "Ojo thinks that he is clever"

3.4.2 NPs which can be coreferentially deleted in sentence complements when coreferential with matrix NPs always include subjects.

Malagasy

- 16 Nihevitra Rabe_i fa notadiavin- dRasoà ø_i / izy_j
 thought Rabe_i that was looked-for-by Rasoà ø_i / he_j
 "Rabe thought that (he) was being looked for by Rasoà"

Similarly the most likely NPs to under Equi-NP deletion include b-subjects.

- 17 a. John_i wants ø_i to help Fred
 b. *John_i wants Fred_j to help ø_i

3.4.3 The NPs which can be coreferentially deleted across coordinate conjunctions include b-subjects.

- 18 a. John_i went up to Fred_j and ø_i insulted him_j
 b. *John_i went up to Fred_j and he_j insulted ø_i
 c. *John_i went up to Fred_j and he_i insulted ø_j

3.4.4 The NPs which can be coreferentially deleted under verb serialization generally include b-subjects.

Akan (ex. from Schachter, 1974a)

- 19 Kofi de aburew_i ø_i gu nsum
 Kofi takes corn flows water-in
 "Kofe pours corn into the water"

3.5 Absolute Reference. In the overwhelming majority of cases, if a b-sentence is true then we understand that there is an entity (concrete or abstract) which is referred to, or has the property expressed by, the b-subject. Thus John worships the Sun Goddess, if true, requires that there be someone referenced by "John" but does not require that there exist a Sun Goddess, in distinction to The Sun Goddess worships John, where Sun Goddess is a subject. Other examples of this sort: John bought a present for the prime minister, cursed Santa Claus, is talking about the perfect woman, etc.

Even if the subject of a basic (or at least fairly basic) sentence is "indefinite" we still normally understand that there must be an object with that property, whereas this is often not the case for indefinite object phrases. Thus a student owes John a report, if true, does imply the existence of a student but not of a report. Other examples of not necessarily referential objects: John ordered a beer, painted a pony, resembles an elephant, imitates alchemists, etc.

However if weather expressions are basic, and they appear to be, then absolute reference is not a necessary condition on b-subjecthood, since weather expressions may have "dummy" i.e. non-referential subjects. E.g. it is raining, etc. Most b-sentences distinguish between an object spoken about and some property it has or some activity it is involved in. But in simple statements about the weather there appears to be little distinction between the activity (the raining) and the object involved (the rain). So in general, if a

weather sentence a subject-predicate form (they may consist of just a single verb) then either the subject will be semantically weak or the predicate will (e.g. Russian rain goes).

3.6 Presupposed Reference. Certain operations like negation, questioning, and conditionalization (below) have the effect of suspending the reference of normally referential NPs. The reference of a b-subject however is harder to suspend under these operations. Thus (from Donellan, 1966) De Gaulle was the king of France implies the existence of a king, but De Gaulle wasn't the king of France need not have this implication. It can be used simply to deny that De Gaulle had a certain title or office. On the other hand, The king of France wasn't de Gaulle, where the king of France is subject, still does imply the existence of a king. Analogous claims hold for the pair Was de Gaulle the king of France? and Was the king of France de Gaulle?. Conditionalization of the following sort (Larry Horn, personal communication) seems to work the same way. Thus if the coup had succeeded de Gaulle would have been the king of France does not imply the existence of a king of France, whereas if the coup had succeeded the king of France would have been de Gaulle does, and the difference between the two sentences is merely that in one the king of France is a subject and in the other it isn't.

Note of course that if b-sentences may have indefinite subjects, e.g. a student attacked John, then under the most natural form of denial, no student attacked John, the existence implication of the subject is not preserved. It is not clear to me at the moment whether we want to consider such sentences as basic in English or not (see 3.8). Arguably understanding such a sentence requires that we understand there exists a student, or perhaps students exist, and so sentences with indefinite subjects are perhaps not basic.

3.7 Metaphoric Idioms. These often suspend the reference or existence implication of NPs. And again, b-subjects are the most reluctant of the major NPs to abandon their reference. Thus in the man took the bull by the horns, let the cat out of the bag, has an ace up his sleeve, is looking for a needle in a haystack, etc. only man has its literal referent. Normally if the reference of a b-subject is suspended in an idiom then so also is the reference of the other major NPs. E.g. the fat's in the fire, the early bird gets the worm, etc.

3.8 Topic. b-subjects are normally the topic of the b-sentence, i.e. they identify what the speaker is talking about. The object they refer to is normally known to both

speaker and addressee, and so is, in that sense, old information. If a L has special topic or old information markers (Japanese, Korean) they will most naturally be used on subjects.

3.9 "Highly Referential" Nps, e.g. personal pronouns, proper nouns, and demonstratives can always occur as subjects. In some Ls, e.g. Malagasy, Tagalog and Philippine Ls generally, Kinyarwanda (by and large) and probably much of Bantu generally, subjects of b-sentences must be definite. And in Tagalog, direct objects must be indefinite. So if an NP position cannot be filled by definites that is evidence that it is not a subject, and if it cannot be filled by indefinites that is evidence that it is a subject.

3.10 Subjects are the most natural targets of "advancement" (Perlmutter and Postal, 1974) transformations. That is, roughly, if a L can assign to one NP in a clause the position, case marking and verb agreements appropriate to another NP in the clause then it can assign the position, case marking, and verb agreements of subject NPs to non-subjects (and we say that that NP has been advanced to subject). E.g. Passive in English advances direct object to subject. Many Ls, e.g. Bantu generally (see Keenan 1975a), can advance NPs to object. But all such Ls can also advance NPs to subject, whereas the converse fails. So subjects are the most accessible targets of advancement processes. (See Johnson, 1974c for discussion.)

3.11 Basic, or relatively basic, subjects have wider scope, logically speaking, than non-subjects. (See Keenan, 1974 for justification of why this is related to the autonomous reference property of subjects.) Thus suppose we are given a sentence in some L containing the main verb kiss and two quantified NPs, every man and a woman. If the truth of the sentence most naturally allows that the choice of woman can vary with the choice of man, as in every man kissed a woman, then that is evidence that every man occurs as a subject. But if, on preference, the choice of woman must be made independently of that of the man, as in a woman was kissed by every man, then that is evidence that a woman is subject.

3.12 b-subjects are normally the leftmost occurring NP in b-sentences. Note, however, that in a few cases Ls have fairly fixed word order in which the subject follows one or more objects. E.g. Malagasy, Tzeltal (Mayan), Mezquital Otomi (Oto-Manguan; Hess, 1968), Gilbertese (Micronesian). Note also Ls like Walbiri (Hale, 1967) in which basic word

order appears totally free, and Ls like Tagalog in which NPs in b-sentences occur in any order as long as they are all after the verb.

3.13 The NPs which can be relativized, questioned, and cleft include b-subjects. In some Ls, e.g. Malagasy, only subjects can be relativized.

3.14 The NPs whose possessors can be relativized, questioned, and cleft include b-subjects. E.g. in Tagalog, it appears, possessors of objects cannot be questioned, but possessors of subjects can.

3.15 A personal pronoun is rarely present in a position relativized if that position is a b-subject one. So even if a L, like Arabic, Fijian, or Welsh, normally presents such pronouns, as in the girl that John gave the book to her it will not normally say the girl that she gave the book to John but only the girl that gave the book to John.

3.16 b-subjects are always among the NPs in a L which can undergo raising. E.g.,

- 20 a. John believed Fred to have struck the gatekeeper
b. *John believed the gatekeeper Fred to have struck

3.17 Subjects can always be expressed by morphologically independent, possibly emphatic, pronouns. These pronouns can be conjoined with full NPs.

3.18 NPs which "launch" floating quantifiers (e.g. all the boys left/the boys all left) include subjects (Perlmutter and Postal, 1974).

B. Case Marking Properties

1. b-subjects of intransitive sentences are usually not case marked if any of the NPs in the L are not case marked (Greenberg, 1966). Exceptions: in Motu (Malayo-Polynesian, New Guinea; Capell, 1969) both transitive and intransitive subjects are marked (different markers) but transitive objects are not marked. Similarly in several Yuman Ls, e.g. Mojave (Munro, 1974) direct objects are unmarked but transitive and intransitive subjects are marked (some marker).

2. The NPs which change their case marking under causativization include b-subjects.

Malagasy

- 21 a. manasa lamba Raso
+acc +nom
wash clothes Raso
"Raso is washing clothes"

- b. mampa-nasa lamba an-dRasia Rabe
+acc +acc +nom
cause-wash clothes Raso Rabe
"Rabe is making Raso wash clothes"

3. The NPs which change their case marking action nominalizations include b-subjects. Usually a b-subject changes to a possessor case or the non-subject agent case. E.g.,

- 22 a. John swept the floor
b. John's sweeping (of) the floor
c. the sweeping (of) the floor by John

C. Semantic Role

1. The semantic role (Agent, Experiencer, etc.) of the referent of a b-subject is predictable from the form of the main verb (Li and Thompson, this volume). Some semantic category information, e.g. animacy, is usually also predictable, but semantic restrictions on objects are usually more specific than those on subjects (Edith Moravcsik, personal communication).

2.1 b-subjects normally express the agent of the action, if there is one. Note that this property cannot be used to identify subjects of sentences in which there is no agent, and sentences of that sort will be numerous among the b-sentences in a L. E.g. John is tall, is in Chicago, is a plumber, etc. Note further that in Dyirbal (Australia; Dixon, 1972) the NP in b-sentences which has most of the Reference properties cited in B.3 above does not express the agent. See Keenan and Comrie (1972) for justification of this point. Thus expressing the agent, if there is one, does not seem even a sufficient condition on b-subjecthood.

2.2 Subjects normally express the addressee phrase of imperatives. But note that in many Malayo-Polynesian Ls, e.g. Maori and Malagasy, imperatives are frequently in non-active forms, and the addressee phrase, if present, appears as a passive (or other type of non-active) agent phrase. E.g.,

Maori (ex. from Hale, 1968)

- 23 tua- ina te raakau raa (ke te toki)
fell-passive the tree yonder (with this ax)
"be chopped down (by you) the tree there (with this ax)"
="chop down the tree there (with this ax)"

2.3 b-subjects normally exhibit the same position, case marking, and verb agreements as does the causer NP in the most basic type of causative sentence. Again we note

however that Maori and Malagasy causatives are very easily passivized, so the agent there is not a surface subject.

Malagasy

- 24 a. mampianatra angilisy an-dRabe aho
cause-learn English acc-Rabe I
"I am teaching Rabe English"
- b. ampianara-ko angilisy Rabe
cause-learn-by me English Rabe
"Rabe is taught English by me"

D. Immediate Dominance

The b-subject is immediately dominated by the root node S. This is the type of definition given in Aspects (Chomsky, 1965) and may represent a necessary condition on b-subjecthood. It is difficult to tell, since, as far as I know, there is no simple test to determine whether or not a subpart of a sentence is a constituent.

For example, in many Ls person/number particles or tense/aspect particles form a higher level constituent with the main verb. By parity of reasoning then they should form higher level constituents with the subject in those Ls in which they are bound to the subject. If so, then the subject would not be immediately dominated by the root node S. Thus in Fred's the one, or Fred'll go, only Fred is the subject, not Fred'll or Fred's. In the English cases, however, such sentences are arguably not basic. But in Luiseño (Uto-Aztecan) or Walbiri they arguably are.

Luiseño (ex. from Hyde, 1971)

- 25 Xwaan-po wiiwish naachaxan-an
Juan -2 or .3 sg acorn mush eat +future
+future
"Juan will eat acorn mush"

Further, it is quite clear that being immediately dominated by the root S is not a sufficient condition for b-subjecthood. We note three categories of cases where more than one NP is arguably immediately dominated by the root S.

1. Ls whose unmarked word order is VSO. While these are a minority type across Ls, they probably constitute between 5 to 10 percent of the worlds Ls. Thus, Malayo-Polynesian: Maori, Samoan, Tongan. Semitic: Classical Arabic. Celtic: Welsh, Breton, Scots Gaelic. Nilo Saharan: Maasai. American Indian: Chinook (Penutian), Jacaltec, (Mayan), Zapotec (Oto-Manguenan).

2. In some Ls the relative position of subject and object is completely free. E.g. Walbiri (Australia), Tagalog (Philippines).

3. In many Ls cited as SOV there is little evidence for a VP constituent. E.g. in Turkish, Hindi, and Persian wh-interrogative words are naturally placed in the object slot, even if they question the subject NP. In Tibetan this placement seems obligatory. For arguments that Japanese lacks a VP constituent see Hinds (1974). For arguments that VP constituency is limited to SVO Ls see Schwartz (1972).

III.0 The Utility of a Multi-Factor Concept of "Subject"

Using the Subjects Properties List (SPL) one can, we claim, identify the b-subjects in any L. Then the full set of properties characteristic of b-subjects in any given L can be determined, and then subject NPs of non basic sentences can be identified. While complex, this means of identifying subjects does permit us to verify the universal generalizations stated in terms of "subject" at the beginning of this paper.

This concept of subject has, however, another use: It suggests that we look for cross-L generalizations which express relations between the properties in the SPL. One such generalization, which seems to us to be valid is the following: In general, non-basic subjects are never more subject-like than basic subjects. In other words, in any given L, subjects of non-basic sentences frequently do not have quite as full a complement of the subject properties as do subjects of b-sentences. One reason for this is that syntactically derived subjects are, by our tests, usually somewhat less subject-like than b-subjects. To consider just the case of passive sentences in English, note that the subject does not express the agent, need not have the property of autonomous existence (e.g. a beer was ordered by John does not imply the existence of a beer), and controls reflexives less easily than b-subjects (*?John was insulted by himself). This suggests that operations which create derived subjects may do so to a greater or lesser extent.

Thus perhaps some properties of subjects are harder to pass on to underlying non-subjects than are others, and conversely some properties will be harder for NPs which have lost their subject status to lose. If the subject properties can be ordered in terms of how hard they are to pass on to other NPs then we would have another universal generalization. Namely, if an operation which derives a complex sentence from a simpler one passes on one of the subject properties to another NP in the derived sentence then it necessarily passes

on all the properties higher on the ordering than the given property. Further, if certain transformations were actually defined as subject creating ones (e.g. Perlmutter and Postal, 1974 would define Passive as an operation which converts an object to a subject) then each such transformation could simply be marked according to how far down the ordering of subject properties it could extend in assigning subject properties to another NP. So some operations would be more subjectivizing than others.

An attempt to empirically support any particular ordering of the 30 odd subject properties, however, would go beyond the bounds of this paper. So we shall here simply suggest a hypothesis concerning a partial ordering of the properties, and then present some evidence in support of that hypothesis, acknowledging that more evidence would be needed for our conclusions to be definitive.

III.1 The Promotion to Subject Hierarchy (PSH)

<u>Coding Properties</u>	>	<u>Behavior and Control Properties</u>	>	<u>Semantic Properties</u>
position > case marking > verb agreement		deletion, movement, case changing properties, control of cross-reference properties, etc.		Agency, autonomous existence, selectional restrictions, etc.

The claim made by the PSH is that if an NP in a derived sentence is assigned any of the three categories of subject properties then it is assigned all the higher categories. And within the category of coding properties, if an NP acquires the verb agreements characteristic of subjects then it must also acquire the case marking and position; and if it acquires the case marking then it must acquire the position. So the characteristic position of subjects is the easiest property to assign to a derived subject. Further, the PSH also claims that the subject properties assigned to a derived subject may be any initial sequence of those on the PSH. So a derived subject may e.g. present the coding and behavioral properties of b-subjects but not the semantic properties (which is very frequently the case. We know of no clear cases in which derived subjects become e.g. agents. However, the direct and inverse theme markers in Algonkian need to be further investigated in this regard. See Frantz, 1971 for some discussion.) Or a derived subject may present only the coding

properties of b-subjects. Thus derived subjects may look like subjects without behaving like them, but if they behave like subjects then they look like subjects.

We should stress here that the relation we postulate between coding and behavioral properties represents merely a hypothesis. It states that while a derived subject may have the position and morphological characteristics of b-subjects, it is possible that they do not raise, or delete, or relativize, or control co-reference, etc. in the way characteristic of b-subjects. E.g. Maori has a transformation which moves certain types of possessives, namely pronouns, preverbally. But it applies only to active subjects, not to subjects of passive sentences. In Jacaltec (Craig, 1975) certain types of Equi apply only to active (i.e. basic) subjects, and do not apply to subjects of passive sentences. And so on. But our evidence in support of the hypothesis is only impressionistic, and not systematic. (See Timberlake, this volume, for more systematic support for this claim from N. Russian.)

We have, however, more evidence in support of the ordering of the coding properties, which we shall refer to as the Coding Hierarchy (CH).

III.2 The Coding Hierarchy

The first prediction made by the CH is that some derived "subjects" in some Ls would take on the characteristic position of b-subjects in those Ls but not acquire the characteristic case marking and verb agreements.

Some support for this claim comes from Biblical Hebrew in which subjects characteristically occur after the verb, are not case marked, and trigger verb agreement. In passive sentences, however, two major patterns appear (Richard Steiner, personal communication). Either the derived subject has the full complement of coding properties, as in 26a, or it has only the position, retaining its former case marking, and the verb going into a 3rd sg, masc. form, so not agreeing with anything, as illustrated in 26b.

- 26 a. $l\bar{o}-'elle\ t\bar{e}h\bar{o}leq$ $h\bar{o}-'or\bar{e}\bar{s}$ (Numbers 26:53)
to-these shall-be-divided the-land
fem. 3sg, pass. fem, 3sg.
- b. $b\bar{e}-\bar{y}or\bar{o}l\ y\bar{e}h\bar{o}leq$ $'e\theta\ h\bar{o}-'or\bar{e}\bar{s}$
by-lottery shall-be-divided acc the-land
masc. 3sg pass. fem. 3sg.
(Numbers 26:55)

A second piece of supporting evidence comes from Kimbundu (Bantu, spoken in Angola. Data from Talmy Givón. See also

Chatelain, 1888). The basic word order in Kimbundu is SVO. NP subjects and objects are not case marked, and verbs agree with subjects only. In the passive, however, the derived subject moves to subject position, but does not trigger the expected subject agreement. Rather that is filled by the 3pl human prefix (regardless of the person and number of the derived subject). The verb does, however, take the direct object pronoun which does agree with the derived subject. The clear inference here is that this type of passive developed from an "impersonal" active of the sort they saw John, then a topicalization of the object yielding, John, they saw him, and finally the possibility of reintroducing the agent, John they saw him by me. Sentences 27a-27d illustrate the relevant data.

- 27 a. nga-mono Nzua
I -saw John
- b. nga-mu- mono
I -him-saw
"I saw him"
- c. Peter na Dick a- mono Nzua
Peter and Dick they-saw John
"Peter and Dick saw John"
- d. Nzua a- mu- mono kwa meme
John they-him-saw by me
"John was seen by me"

This type of passive may be characteristic of a group of West Bantu Ls (it occurs as well in Luvale; see Horton, 1949) and is distinct from the suffixal passive characteristic of Swahili, Shona, Luganda, etc.

A third piece of very interesting support for the prediction comes from Maasai (Nilo-Saharan. All data from Tucker and Mpaayei, 1955). The basic word order is VSO, and in passives the old object occurs in the immediate postverbal position, as predicted.

- 28 e-nyer Tinkoi nkishu
3-love Tinkoi cattle
"Tinkoi loves cattle"
- 29 e-rik-i nkishu aainei lmurran
3-led-pass cattle by young-men
"my cows will be led by the young men"

Further, full NPs are case marked, by tone!

- 30 a. e-dol embártá
3-see horse (acc)
"he sees the horse"
- b. e-dol embartá
3-see horse (nom)
"the horse sees him"

And in passive sentences the derived subject retains the accusative tone. (We mark tone only where it is relevant to our discussion.)

- 31 e-isis- i Sirónkà
3-praise-pass Sironka (acc)
"Sironka is praised"

Compare:

- 32 a- dol Sirónkà
lsg-see Sironka
"I see Sironka"

Note as well that passive sentences in Maasai are intransitive in that no NP other than the derived subject is needed for the sentence to be complete.

As regards agreement, verbs normally only agree with subjects. (Note that the prefixal e- indicates either 3sg or 3pl subject agreement.) However, when the direct object is first or second person singular, the agreement prefix changes. Compare:

- 33 a. áa- dol (nánú)
3sg subj (acc) lsg
lsg obj
"he sees me"
- b. á- dol (nanú)
lsg subj see (lsg, nom)
"I see him"

Thus we can tell by the verb marking whether a first or second person singular functions as a subject or an object. And in passives, when the derived subject is first or second person singular, it triggers the verb prefix appropriate to 3 person subject and lsg or 2sg object!

- 34 áa- rik- i
3sg subj nauseate-pass
lsg obj
"I am nauseated"

Thus it appears that in Maasai the passive verb always has a 3 person subject agreement marker, reminiscent of the Kimbundu

passive in this respect, and that where the passive verb agrees with the derived subject, namely when it is 1sg or 2sg, it agrees with it as though it were an object. Thus the derived subject retains its object case marking and verb agreements, acquiring only the position of subject.

Final, but in some ways less convincing, evidence for our claim comes from Ls like Latin and German. Here when Passive advances accusative direct objects to subjects they acquire the full complement of coding properties. However, a few verbs in both Ls take their direct objects in the dative case. And in these cases one could argue that when the verbs are passivized the dative NP takes on the subject position but retains its dative case marker and does not trigger verb agreement, the verb reverting to a 3sg form. (Exs. from Gildersleeve and Lodge, 1913.)

35 mihi invidetur (ab aliquo)
 1sg envied (by someone)
 dat +pass
 "I am envied"

The problem with this data for our analysis, as Paul Postal, (personal communication) pointed out, is whether there is any real sense in which the underlying object has acquired the subject position (it very clearly has not taken subject case or subject verb agreement). This is perhaps better analyzed as an impersonal passive in which the old subject gets demoted to oblique status and the verb becomes passive, but nothing gets promoted to subject position, the other NPs merely remaining where they were. (See Keenan, 1975b for some discussion of impersonal passives.)

The PSH might also be interpreted to mean that we could expect to find Ls in which subjects demoted by operations like Passive lost only their characteristic position but not their case marking or verb agreements. That is the subjects properties that are the hardest to acquire are also the hardest to give up. At the moment we have relatively little evidence for this, although the Maasai data, if further analyzed, might support it. Further, the following example from Luiseño (Munro, 1974a) is suggestive.

36 ?ivi no-naawu-ki no-yo po-lo?xa
 this my-dress-poss my-mother her-make
 "this dress was made by my mother"

Now Luiseño is dominantly SOV, with objects case marked, usually. Subjects trigger a kind of agreement which may tack onto the end of the first word in the sentence or, in more complex sentence types, show up as a possessive-type affix on

the verb. Now, as the passive above illustrates, the demoted subject loses its sentence initial position, but remains nominative and appears to trigger a kind of possessive agreement on the verb. Historically speaking this is natural, since it would appear that passives in Luiseño, at least this type, are similar to those in Chemehuevi (Munro, 1974b), which look like this dress is what my mother made, where mother would be expected to remain nominative and trigger verb agreement.

III.3 A second prediction made by the Coding Hierarchy is that derived subjects in some Ls would acquire the position and case of b-subjects, but not trigger the verb agreements characteristic of b-subjects. We have less evidence for this, but some nonetheless which is suggestive.

Thus in Welsh, active (=basic) sentences present VSO order. Verbs agree only in person with subjects (although they also agree in number with pronominal subjects). Neither full NP subject nor object are case marked but pronominal constructions for subjects and objects are different. In passives, the old object now occurs in the expected subject position, and uses the nominative pronominal construction, and so to that extent, acquires the case of b-subjects. But the derived subject does not trigger verb agreement. The verb becomes morphologically invariable. (Examples below from Bowen and Jones, 1960).

- 37 a. gwelir fi
 is-seen I
 "I am seen"
- b. gwelir di
 is-seen you
 "you are seen"
- c. dysgir Cymraeg gan yr athro
 is-taught Welsh by the teacher
 "Welsh is taught by the teacher"

A second example which supports this point in the CH concerns cases where a demoted subject loses its position and case marking, but does not lose its ability to trigger verb agreement. Kapampangan (Philippines; all data from Mirikitani, 1972) is illustrative here. Kapampangan is verb initial in basic word order and seems to permit a fair degree of freedom of word order of full NPs after the verb, although this cannot be judged with certainty on the basis of the data in Mirikitani, op. cit. Surface subjects take the preposition ing and other NPs have other prepositions. Verbs

agree with subjects in b-sentence types.

- 38 sumulat ya ng poesia ing lalaki
write he obj poetry subj boy
"the boy will write a poem"

In passive sentences the former object gets assigned the subject pre-position and appears to take the subject position (but position may not really be criterial of subjects), and it acquires the capacity to trigger verb agreement. The old subject, while losing its position and most importantly its pre-position, still retains the possibility of triggering verb agreement. Thus the derived verb now agrees with two NPs, the derived subject and the former subject.

- 39 isulat na ya (=ne) ning lalaki ing poesia
be-written it he agent boy subj poem
"the poem was written by the boy"

Another example of a L in which demoted subjects still trigger verb agreement is Achenese (Lawler, 1975).

III.4 Some Possible Counter Examples

One interesting class of possible counter example to the claims in the CH is presented by the possibility of promoting to subject various types of locatives in Bantu Ls. Consider 40a and 40b from Chicewa. (See Trithart, 1975 for discussion.)

- 40 a. John a- nathamang-ir- a ku sukulu
John he- ran dir-indic to school
"John ran to school"
- b. ku sukulu ku- nathamang-idw- ir- a- ko
to school loc- ran -pass-dir-indic loc
ndi John
by John
"School was run to by John"

It appears in 40b that the derived subject triggers verb agreement and acquires the subject position but retains its locative case marking. If so, this clearly violates the claim of the CH. Note, however, that the subject agreement it triggers is not one of the normal noun class markers, as is the case for b-subjects. Rather, it is the locative marker itself. This suggests that the original locative phrase ku sukulu "to school" has been reanalyzed as a mere NP, the old locative ku now being interpreted as the noun class marker. Then the verb agreement is regular, and the subject no longer

carries an oblique case marker, and the CH is not violated. Further work is needed to determine whether this analysis receives any independent motivation.

A second counter example comes from Biblical Hebrew (again supplied by Richard Steiner, personal communication). While the general pattern for passives is as cited in Section 3.1 above (either the derived subject takes the position, case, and agreements of b-subjects, or else it only takes the position, retaining its case and the verb becoming 3sg), there are a very few cases in which a derived subject remains accusative but does trigger verb agreement.

- 41 wē- 'eθ hō- 'oh lēfōnw mēvo'oreθ
and-acc the-brazier in front of him was-kindled
fem 3sg fem 3sg pass

It would appear then that Hebrew, at this stage, was in the midst of analyzing subjects of passive sentences as real objects. Sometimes all the surface, i.e. coding, properties are assigned to subjects, and sometimes not. In a large majority of cases where not all the coding properties are assigned to derived subjects the properties that are assigned are in accordance with the CH, but in a few cases this is not so. (Note incidentally, that the accusative NP which triggers verb agreement in 41 also occurs clause initially, plausibly not the subject position in Biblical Hebrew.)

This type of historical shift may also serve to rationalize a final counter example to the CH pointed out by Butler (personal communication) in which the evolution of impersonal constructions in Old and Middle English is discussed. Such constructions had a major NP which was dative/accusative and did not trigger verb agreement--the verb being fixed as 3sg. Such constructions survive in frozen expressions until somewhat later, e.g. me-thinks, if you please, etc. Now one of the ways the impersonal constructions were lost is that the surface dative/accusative was reanalyzed as a subject, becomes nominative and triggers verb agreement. Butler points out, however, that during the reanalysis "Constructions like the following occasionally turn up with think, seem, and ail: Me-seem my head doth swim. (1571. Damon and Pythias. 79.) ..." It appears then that during an historical transition in the reanalysis of a subject NP we can have NPs with subject position and verb agreements but which do not have subject case marking. Judging from Butler's use of "occasionally," however, we may infer that, like Hebrew, this violation of the CH was not the norm, but merely the reflection of an instability a surface pattern. Nonetheless the instability does lead to "occasional" violations of the CH and cannot be dismissed.

IV.0 Conclusion

We have attempted to provide a definition of the notion "subject of" which would be universally valid in the sense that it would allow us to identify subjects of arbitrary sentences from arbitrary languages. The definition we proposed, while cumbersome, does nonetheless allow us to verify many universal generalizations stated in terms of that notion. And in addition it has suggested further generalizations concerning universal properties of subjectivizing transformations.

Postscript: The information on the evolution of impersonal constructions in Old and Middle English communicated to me by Butler, cited on p. 331, is contained in an unpublished paper by Milton C. Butler, titled, "The re-analysis of impersonal constructions in Middle English."

Note

1. We might note, however, that in general the meaning of a non-basic structure will depend on the meaning of only finitely many other structures. We might take this number to determine the degree of basicness of the structure. Then any two structures can be compared with regard to basicness using the normal "greater than or equal to" relation. Thus the basicness of a structure x would be greater than or equal to that of y if and only if the degree of basicness of x is greater than or equal to that of y . In this way the relation more basic than can be extended to apply to all the structures in a L .