

**PAPERS FROM THE
FOURTH REGIONAL MEETING
CHICAGO LINGUISTIC SOCIETY**

APRIL 19-20, 1968

**Department of Linguistics
University of Chicago
Chicago, Illinois**

PAPERS FROM THE
FOURTH REGIONAL MEETING
CHICAGO LINGUISTIC SOCIETY

APRIL 19-20, 1968

Edited by

Bill J. Darden
Charles-James N. Bailey
Alice Davison

Department of Linguistics
University of Chicago
Chicago, Illinois

Copyright © 1968
by
Department of Linguistics
University of Chicago

PREFACE

In the spring of 1965, through the effort and initiative of Chairman Doris Bartholemew, the Chicago Linguistic Society held its first regional meeting. Since then, the spring meeting has been an annual event. The quality of the papers has steadily improved. This year, it was decided that they were worthy of publication. We would like to express our gratitude to the University of Chicago for providing a subsidy for the printing of this volume. Particular thanks go to Professor Eric P. Hamp, Chairman of the Department of Linguistics, and Dean Robert E. Streeter, Division of the Humanities, for their personal efforts in obtaining the money.

The primary criteria for the preparation of this volume were speed and economy. In order to be able to sell such a small edition at a reasonable price, we asked that the speakers provide us with self-edited manuscripts that were ready for photographing. The speakers were provided with instructions as to margins and style, but, as will be obvious to the reader, not all of them followed our instructions.

Four papers which were read at the meeting were not submitted for publication:

Lloyd B. Anderson: Jungles of Semantic Graphs.

Hans G. Guterbock: The Old Hittite Ductus of Cuneiform.

David E. Pingree: Epigraphy and the Chronology of Northern India during the Scythian period.

David L. Stampe: Yes, Virginia...

CONTENTS

SYNTAX

Robert Binnick: On the Nature of the "Lexical Item" . . .	1
Bill J. Darden: Is the English Perfect an Embedded Past? A Statement from the Devil's Advocate . . .	14
Georgia M. Green: On <u>Too</u> and <u>Either</u> , and Not Just <u>Too</u> and <u>Either</u> , Either	22
Carolyn G. Killean: Interesting Features of Gender-Number Concord in Modern Literary Arabic	40
Stanley E. Legum: The Verb-Particle Construction in English, Basic or Derived?	50
Larry W. Martin: Some Relations of Embedding in English Nominals	63
James D. McCawley: Lexical Insertion in a Transformational Grammar without Deep Structure	71
Jerry L. Morgan: Some Strange Aspects of <u>It</u>	81
Arnold M. Zwicky, Jr.: Naturalness Arguments in Syntax	94

PHONOLOGY

Eric P. Hamp: Acculturation as a Late Rule.	103
Earl M. Herrick: A Stratification Restatement of a Problem in Manyika Phonology	111
Robert Howren: Stem Phonology and Affix Phonology in Dogrib (Northern Athapaskan)	120
Kostas Kazazis: Sunday Greek.	130
Dale E. Woolley: Some Observations on Stridency in English	141

OTHER PAPERS FROM THE REGULAR SESSIONS

Joseph C. Beaver: Progress and Problems in Generative Metrics.	146
Wilbur A. Hass: Perception Versus Cognition in Linguistic Theory and Developmental Psychology .	156
Keith Percival: The Notion of Usage in Vaugelas and in the Port Royal Grammar.	165

SPECIAL SESSION ON WRITING SYSTEMS

✓ Charles-James N. Bailey: The Pronunciation of <u>Zeta</u> in Ancient Greek	177
✓ I. J. Gelb: Grammatology and Graphemics	194
✓ Peter H. Salus: On the Evolution of the Runic Alphabet	202
✓ Michael C. Shapiro: On the Original Direction of the Brāhmi Script	208
✓ Arlene R. K. Zide: A Brief Survey of Work to Date on the Indus Valley Script.	225
✓ Norman H. Zide: Graphemic System in the Ol Cemed Script.	238

On the Nature of the "Lexical Item"

Robert Binnick
(University of Chicago)

At the outset two initial assumptions should be made which are controversial, and, were it not for limitations on this paper, deserving of discussion. The first is that the level of deep structure in transformational syntax is the same level of linguistic description as that of semantic representation and that which provides the input to the transformational component. That is, the structures directly generated by the base component of a transformational grammar are semantic, not syntactic. The second is that synonymous sentences have identical representations on this level. These points have been discussed by Gruber (1965) and Lakoff (1967) and their defense can be seen in those papers. The relation of these assumptions to my topic, the nature of lexical items, will become clear shortly.

Transformational semantics has interested itself with the lexicon without really explicating its nature beyond showing that lexical items lead a tripartite existence: phonological, syntactic, and semantic. The phonological existence plays no role in the base, as it is irrelevant on that level whether the item meaning 'cow', for example, is cow, vache, or even *zilchtraum, for it still means 'cow'; nor does it play any rule in the operation of the syntactic transformations, since one could not conceive of a rule which affected only, for example, verbs beginning with a stop. The phonological representations of the items can therefore be mapped onto them at the surface syntactic level, and need not concern us further in discussing the basic nature of lexical items.

At this point consider the class of transitive motion verbs in English, such as cross, enter, follow, and so on. These are systematically ambiguous, as each member of the class can serve either as an active, motive verb, as in

1. Caesar crossed the English Channel.
2. The announcer entered the studio.
3. James Bond followed the enemy spy.

or a stative, locative verb, as in

4. Route 80 crosses Iowa.
5. The railroad enters the town near the post office.
6. 'B' follows 'A' in the alphabet.

This distinction is not merely semantic but is amply supported by syntactic differences as well, for the co-occurrence restrictions in each case are different. Thus

7. Caesar is crossing the English Channel.

is the present tense of

8. Caesar crossed the English Channel.

but

9. The railroad is entering the town near the post office.

is an inchoative usage of the verb enter, not comparable to sentence (5) or to (10):

10. The railroad entered the town near the post office.

which is ambiguous. (This marginal, predictable inchoative usage is not possible with the active, motive sense.)

The manifold ambiguity of these "gestalt" verbs, as I call them, is shared by only two other verbs, go and come, which are not gestalt verbs because they are not transitive and because they do not incorporate in their meaning a preposition, as do the gestalt verbs. Nonetheless, with the proper prepositional complements, go and come can replace any gestalt verb. That is, for every sentence containing a non-idiomatic usage of a gestalt verb, there is a synonymous sentence with go and another with come; for example, sentences (1-6) can be paraphrased as follows:

11. Caesar went (came) across the English Channel.

12. The announcer went (came) into the studio.
13. James Bond went (came) after the enemy spy.

14. Route 80 goes (comes) across Iowa.
15. The railroad goes (comes) into the town
near the post office.
16. 'B' goes (comes) after 'A' in the alphabet.

The undoubted stylistic differences between go and come need not concern us here. Go and come even paraphrase the inchoative usage of (9), thus:

17. The railroad is going (coming) into the town near the post office.

This strange parallelism is easily explained in terms of one of our assumptions made earlier. The sentences with go and come, and those with the gestalt verbs, are derived from the same underlying structure. This should not be surprising, for the glosses of gestalt verbs are invariably phrases containing go or come and a preposition, to which the gestalt verb in question is often morphologically related, as in the cases of across and cross, around and surround, in and enter, and others. Furthermore, the class of objects of a gestalt verb is identical to that of the objects of the preposition occurring in its paraphrase, whether the two are morphologically related or not. Thus the objects of follow are those of after.

Jeffrey Gruber, in the work mentioned above, has stated that the sentences with go and come are closer to the deep structure than those with the gestalt verbs, and that these were derived transformationally by a process called "incorporation" in which a string of one or more underlying lexical items are mapped into a single derived lexical item. An important piece of evidence he presented concerns the use of non-gestalt motion verbs intransitively with prepositional-phrase complements, as in

- 18a. John climbed down the ladder.
- b. up the mountain.
- c. along the ledge.

Climb, like gestalt verbs, incorporates in its gloss a preposition, but syntactically it is not transitive, but pseudo-intransitive, since it does not require an object:

19. John climbed all day.

and semantically it is more complex than the gestalt verbs, so that it is quite different from them. Nonetheless, the fact that when climb takes an object, the direction indicated is invariably up, as in

20. John climbed the mountain.

which is synonymous to (18b), shows that climb incorporates a preposition, namely up, although climb, unlike ascend, is not primarily a direction-of-motion verb, but rather a manner-of-motion verb, as shown by 18a). Clearly climb incorporates a preposition, so there is no reason to doubt that the gestalt verbs do as well.

The question arises whether this process of incorporation is not in fact a kind of deletion. Suppose cross and across were allomorphs of the same morpheme, and that sentence (1) for example is derived from (11) by simply deleting the main verb. On the surface this is not such a bad idea, since go and come can be paraphrases of the verb be, as shown by sentences (21, 22).

21. 'B' is after 'A' in the alphabet.
22. James Bond is after the enemy spy's secrets.

and be is notorious for its tendency to be deleted. Furthermore, if we take incorporation to mean declaring two adjacent lexical items in some string by fiat to be a single lexical item, then incorporation seems not only trivial but counter-intuitive, and deletion seems a much more likely explanation.

However, if, as I shall show, items which are incorporated into a gestalt verb, for example, start out in the deep structure at wildly different places, then incorporation is neither trivial nor indefensible, and has wide implications for the nature of surface words and their derivation from more abstract items in the lexicon.

A similarity of go and come to be is that the causative of all three is put. It is clear that some verbs have related causative verbs completely unrelated to them morphologically. Thus send and bring are the

causatives of go and come respectively, as seat is the causative of sit and fell of fall.¹ But in the locative sense, the causative of both go and come is not send or bring but put, as in

- 23. The engineers put the highway across Canada.
- 24. The Phoenicians put 'B' after 'A' in the alphabet.

Gestalt verbs, however, have no causative counterparts in sentences like

- 25. M. had James Bond follow the spy.
(motive)
- 26. The Phoenicians made 'B' follow 'A' in the alphabet. (locative)

Nonetheless, there is a peculiar usage of many gestalt verbs, particularly some, like surround, which are never motive. The causative of

- 27. The stones surrounded the megalith.

is probably not (28) (which some speakers claim is a good sentence),

- 28. *The Druids had (made) the stones surround the megalith.

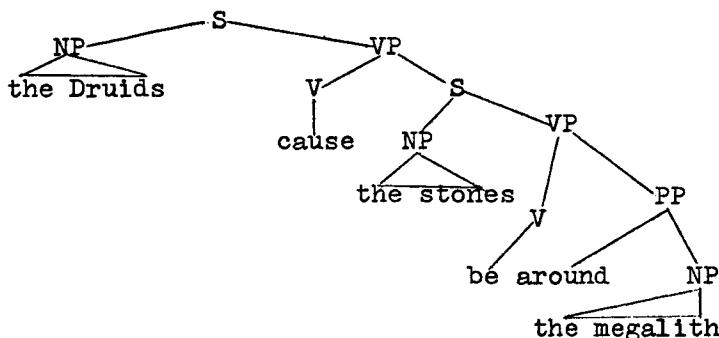
but rather

- 29. The Druids surrounded the megalith with stones.

The underlying structure of this last would be something like the very simplified tree in (30).

¹ Charles Pasley has suggested to me that take may be the causative of one of these verbs. Its status is not clear.

30.



There is in English a class of derived transitive verbs closely related to these causative gestalt verbs. These verbs, such as jail, imprison, cage, shelter, etc., are all derived from nouns which denote objects which are also places. Each of these verbs can be paraphrased by put plus a complement phrase containing the root noun as the object of a motive preposition such as into or unto, thus

31. The sheriff jailed Robin Hood.

is paraphrased by

32. The sheriff put Robin Hood into jail.

If we modify the verb in (32) the adverb usually goes at the end:

33. The sheriff put Robin Hood into jail for four years.

Such adverbs clearly modify the Robin Hood into jail phrase rather than the main verb. Thus in (34a)

34a. The sheriff, who was in Nottingham,
jailed Robin Hood in Reading.

34b. The Queen was crowned in Reading.

the phrase in Reading clearly does not modify the entire sentence, as it does in (34b). In sentence (34a) there is no single constituent for the adverbial phrase to modify, but in its paraphrase there is the phrase mentioned above, namely, Robin Hood into jail.

Now the prepositions occurring with put are typically motive. Their objects are those which serve as objects of motive verbs, such as climb or enter. These prepositions are systematically related to locative prepositions, as into to in and unto to on; most motive prepositions, however, double as locatives, as through does, and even into and onto have largely been replaced by their locative counterparts in colloquial English. If the motive prepositions are derived from their locative counterparts, as seems likely, this may serve as evidence that motion verbs are ultimately derived from inchoative locatives, i.e., from come to be.

Consider the objects of the verb make. These are identical for the most part with those of be, suggesting that make is the causative of be just as send is the causative of go. A conspicuous absence, however, is that make never takes any of the locative complements that occur with be. Thus the passive causative of John is king, is John was made king, but that of John is in Rome, is not *John was made in Rome, but John was put in Rome.

Put can be considered, therefore, to be the locative causative of be. Sentence 31 is therefore truly paraphrased by

35. The sheriff caused Robin Hood to be in jail.

If the deep structure of a sentence with a verb like jail contains a causative pro-verb with a locative complement containing the verb be, then deletion cannot be involved in the derivation of the surface verb jail. If the meaning of jail is cause to be in jail, then the derivation ultimately consists of shifting the included main verb into the matrix sentence, along with other material.

The same process of incorporation is clearly at play in the case of the gestalt verbs. The structure underlying

36. John went across the river in a punt.

also underlies

37. John punted across the river.

and

38. John crossed the river in a punt.

There are some problems with an analysis like this which treats punt as an incorporation of the object of a locative preposition into the main verb and across as a similar incorporation of a preposition. First, why can the preposition of across the river and not that of in a punt, or the noun of in a punt, but not that of across the river be so incorporated? There is no obvious answer unless we are to accept Charles Fillmore's case for case. Furthermore, if the noun is to be incorporated, it can have no modifiers, so that the structure underlying

39. John's friends went across the river
in twenty-two gaily decorated orange-
colored punts.

will underly not

40. *John's friends twenty-two gaily
decorated orange-colored punted
across the river.

but rather (41).

41. John's friends punted across the river
in twenty-two . . . punts.

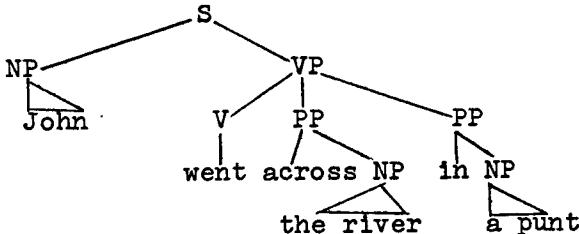
Furthermore, additional adverbs can appear. You can very well say

42. The sheriff jailed Robin Hood in a dank,
medieval dungeon.

Another problem is that only certain kinds of nouns and prepositions can be incorporated. This fact has the virtue of establishing natural semantic classes on a principled basis, but the vice of leaving one to wonder whether in some language there might not be motion verbs incorporating not only nouns like punt but nouns like yesterday, river, and so on, as well.

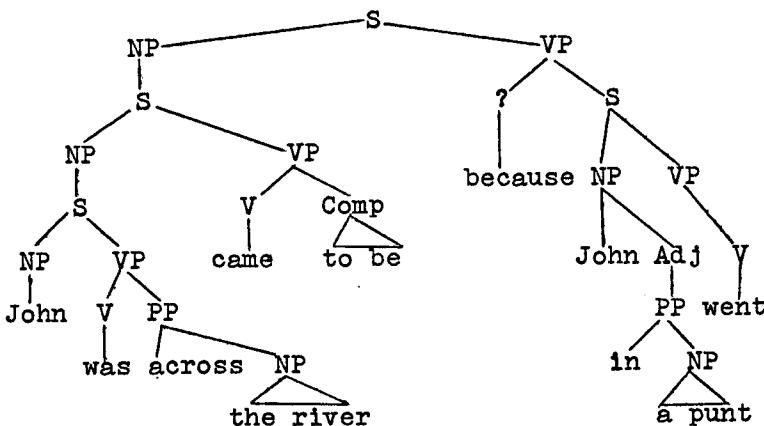
In any case, these problems suggest rather strongly that an underlying structure like (43),

43.



or any structure in which the two prepositional phrases are treated as coordinate modifiers of the verb, is inadequate. It may be that the basic structure is more like (44). The argument for such a treatment is unfortunately too lengthy to present here.

44.



In any case, given the proper transformational apparatus, the lexicon will need list only a handful of basic, non-derived motion verbs, probably only be, go, and come.² All other locative and motive verbs,

²It may be that go and come are semantically complementary, in which case there will be only one motion verb, go. This might in turn be derived from be, suggested above.

including all the complex motion verbs like run and climb, the gestalt verbs like surround and enter,³ and the instrumental motion verbs like punt, boat, and so on, as well as all the causatives of these various verbs, can be derived by transformations. In short, I am taking a position quite opposed to the recent position of some conservative transformationalists who want to stuff the lexicon à outrance.

Nonetheless, I am unwilling to go so far as to completely accept the opposing position. Although cross is derived ultimately from go, the former is a transitive, the latter an intransitive verb. Whereas you can say

45. The river was crossed by Lewis and Clark.
you cannot say

46. *The river was gone across by Lewis and Clark.⁴

Therefore, at the point at which the passive transformation operates, cross must exist as a real verb. Another rule which will demand that cross be a real verb at that point is the auxiliary shift transformation.

Clearly, then, at some point cross must be not only a verb, i.e., a constituent with the label verb, but also a transitive verb. Although most syntactic properties are reflexes of semantic ones, there are some syntactic properties which undoubtedly do not originate in the base, such as grammatical gender, or the ability to take certain endings which permit certain endings which permit certain transformations (e.g., that which derives verbs in -en from adjectives). The latter type in general are closely

³A distinction should be made here between locative-motive gestalt verbs which derive from go and/or come, like enter, and locative ones, which derive from be, like surround.

⁴I know of at least one native speaker who can say (46), and three who understand it out of context, although they claim it is "ungrammatical".

related to morphophonemics in that features like [\pm French], needed anyway in phonology, play a role in their formulation. For this reason, lexical items at some point must be syntactic units with individual complexes of syntactic idiosyncracies. If certain verbs, like cross, are derived, then they cannot bear such complexes in the base, any more than they can have phonological identities in the base. But once cross is declared a verbal constituent, its syntactic properties must be noted. There will be many transformations that will apply before cross becomes a verb, and many that apply after. The ones applying before do not take syntactic information into account at all, but merely shove constituents around. Those that apply after are perfectly ordinary transformations that do take syntactic information into account.

What I propose, then, is to have all morphemic units declared at some single point. That is, go into would be constituentized and labeled as a verb. It is at this point that syntactic lexical items come into being. A split-level theory like this has the advantages of giving a role in transformational analysis to traditionally observed units which are neither surface words nor underlying lexical items, but nonetheless have linguistic reality to the extent that people have talked about them for two thousand years.

APPENDIX I

Notes on Latin

The use of prepositions like in in Latin with accusative vs. ablative follows the motive-locative distinction. Note that Latin permits certain derivations impossible in English which nonetheless seem to have not dissimilar bases from English, thus Caesar Germanos flumen traicit. (Civil Wars 1.83) 'Caesar throws the Germans across the river.' = 'Caesar crosses the Germans the river.' (note tra- prefix.)

APPENDIX II

Classes of Verb Mentioned in the Paper

Pure locative: be.

Pure motive: go, come. (can serve as locative as well)

Motive-locative gestalt verbs: enter, exit, precede,
follow, approach, ascend, descend, surpass,
reach . . .

Locative gestalt verbs: surround, subtend, transcend,
enclose . . .

Causatives of be: make, put (locative)

Causatives of simple motive verbs: send (go), bring
 (come).

Complex motion verbs: climb, scurry, hasten, walk . . .

Instrumental (complex) motion verbs: boat, punt, jet,
 . . .

BIBLIOGRAPHY

- Fillmore, Charles. 1968. The case for case. Proceedings of the 1967 Texas conference on language universals, ed. by Emmon Bach and Robert Harms.
- Gruber, Jeffrey S. 1965. Studies in lexical relations. M.I.T. dissertation.
- Lakoff, George. 1967. Instrumental adverbs and the concept of deep structure. PEGS (duplicated).

IS THE ENGLISH PERFECT AN EMBEDDED PAST?
A STATEMENT FROM THE DEVIL'S ADVOCATE

Bill J. Darden

(University of Chicago)

Recently two generative grammarians, Emmon Bach (1967) of the University of Texas and James McCawley of the University of Chicago, working along different lines, have come to the conclusion that the perfect in English is best analysed as an embedded past tense.

McCawley's as yet unwritten analysis is based in part on papers by T. R. Hofmann (1966) and Susan Fisher (1967). He argues that past-tense sentences, when embedded after modals, in infinitive constructions, or in gerundives, have the past tense show up as "have", while the present is realized as zero. The evidence that these constructions are really embedded pasts is provided by the adverbs with which they may appear. For instance, the time adverb "yesterday" requires past tense, not perfect, while "by now" requires perfect, yet either of these adverbs can follow the sentence: "John is believed to have left for Philadelphia." Likewise, "John is believed to have already met Charles when he met Sarah" is an underlying past perfect, paraphrased by "It is believed that John had already met Charles when he met Sarah." "John is believed to be there" has present realized as zero.

The rules necessary to account for the above phenomena are as follows: first, past goes to "have" and present goes to zero after "to", "ing" or "modal", then "have" goes to zero before another "have" dominated by "Aux". The second rule is necessary in order to prevent the occurrence of two "have's" in the case of embedded past perfects.

What is proposed is to add "tense" to the environment for this rule, treating present and past perfect as past embedded in present and past, respectively. This analysis would account for the result of tense agreement in indirect speech, when it operates. For example, "He said, 'I am happy' or 'I was there'" which yield "He said that he was happy" or "that he had been there." Here, present embedded in past gives zero, leaving a simple past, while past embedded in past yields past perfect.

This mechanism can obviously take care of the

form of the perfect, but language is not only form. It must also be demonstrated that this proposed underlying structure is not inconsistent with the meaning of the perfect. The difference between "I read that book" and "I have read that book" seems to be that in the simple past the action is localized at a past time, while in the perfect the action is treated as prior to but somehow relevant in the present (Jesperson, 1927; 243-6). The perfect not only does not specify the time in the past, it does not allow adverbs which do so, for example, **"I have read that book last night." The "prior action relevant in the present" includes the "present result of past action" meaning of the perfect; for example, "John has opened the window," implying that the window is still open or there is some other existing result, such as dirty fingerprints on the sill. The "up to now" meaning of the perfect, for example, "I have read three books" describes an action which took place within a time span extending from an indefinite point in the past up to the present. The other end point of the span can be specified by an adverb, for example, "since yesterday" but the specific time of the action may not be referred to. The addition of progressive aspect makes the action continuous throughout a time span extending up to the present: "I have been reading." Many languages allow an adverb to take care of the anteriority and express the English present perfect progressive as simple present. For example, Russian: Ja čitaju uže tri casa, "I have been reading for three hours." The literal translation, "I am reading three hours already," is easily understandable, but should be marked with a six-pointed star.

All of the meanings of the English perfect include a reference to only one specific time, the present (or the past in the case of the past perfect). The reference to the time of the action is limited to one of priorness. Thus if we add the qualification that a past, when embedded, loses the ability to refer to a specific time, we can claim that the meaning of the perfect is derivable from the fact that it is an embedded past.

Note that, with respect to information about the past, the perfect is less marked than the past. Unmarking as a result of embedding is a possible explanation (Darden, 1968). This interpretation of the perfect is helpful in understanding historical change. If something happens to break the connection of the perfect with the present, the perfect becomes the unmarked past. This change is not uncommon. In Slavic, where it occurred, the motivation for the change was probably the development of a new aspect, perfective/imperfective. In East and West Slavic the

perfect merely supplanted the aorist. In Bulgarian, in Macedonian, and in Serbo-Croatian dialects, the meaning of marked past progressed from "occurring at a certain time in the past" to "occurring at a specific time in the past as attested by the speaker's personal experience," that is, a witnessed past.

The above treatment would also account semantically for one of the side effects of the rules quoted. The rule that eliminates "have" before another "have" would allow for an indefinite number of embedded pasts. All would be changed to "have" and all but one would be eliminated. If one assumes that an embedded past can only mean priorness to a point specified by the sentence in which it is embedded, then the semantic result is the elimination of all but one perfect. Let us assume a sentence in which past_1 is embedded in past_2 , which is embedded in present. The meaning of past_1 , when embedded, is "prior to the point in time referred to by past_2 ." However past_2 , because it itself is embedded, loses (or does not have) the ability to specify a point in time, and simply means "prior to the present." The end result is that all we know of the action predicated by past_1 is that it is prior to the present.

We can now conclude that an interpretation of perfects as embedded pasts is possible, provided that one allows for the fact that embedded pasts differ semantically from surface pasts. This provision is of course crucial. In everything said above, "embedded past" could be rewritten "perfect" with no loss of validity; that is, there could be underlying perfects. Most generative grammarians would prefer this to having transformations which change meaning. One could of course claim that the two meanings are in complementary distribution. One could have contextual semantic readings for "past", depending on whether or not it was embedded. However, this type of complementary distribution is relevant only in going from form to meaning. This is impossible within the theory of grammar which McCawley advocates. In a recent paper (McCawley, 1967) he attacked the notion of a deep structure from which a semantic reading is obtained. He argued that we must start from the semantics and proceed directly to the surface by a series of transformational rules. In this theoretical framework, different semantic readings are different units. "Past" and "perfect", if they have different meanings, must be different underlying predicates. (This, of course, is imposing my definitions on McCawley's system.)*

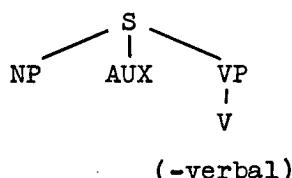
* See appendix

Note also that this complementary distribution is assumed, not proved. If one analyzes all perfects as embedded pasts, it is obviously impossible to find a perfect which is not an embedded past. However, we can look for embedded pasts which do not show up as "have". There are such examples: "I am upset at your coming in so late last night" and "It was foolish to come at six o'clock" must have embedded pasts because of their time adverbs, according to the original argument. There are also formal differences between these constructions and those which one might be tempted to call underlying perfects. Definite adverbs such as "last night" are carried along in the embedding into a gerundive or infinitive construction (see above). Such adverbs are not permitted, however, when a past is supposedly embedded to form a perfect. If this were the same process, there would be no reason not to allow "I have gone last night."

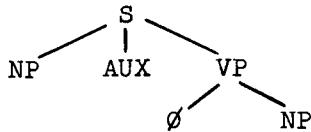
Thus the very evidence on which McCawley based the argument that "have" comes from an embedded past can be used to show that this is not the case in the perfect. This, in my opinion, indicates that McCawley's analysis is wrong.

Bach's paper on "have" and "be" has finally been published, so rather than a summary, I will present a reinterpretation of his ideas in the light of more recent work. He basically proposed that "verb" should be an optional category in phrase structure. In verbless constructions, either "have" or "be" is inserted. "Be" is inserted before everything which Chomsky (1965: 107) included in the category "predicate", for example, predicate adjective, predicate nominative. Before an NP which is not a part of "predicate", that is, an NP in the object position, "have" is inserted. Since Bach wrote his paper, everything in the category "predicate" has been reinterpreted as being of the same syntactic category as verbs. When something marked [-verbal] appears under a verb node, "be" is inserted (Bach, forthcoming; Lakoff, 1966; Darden, 1967). The difference between "I am a wife" and "I have a wife" is then in the verbal or non-verbal function of the noun phrase "a wife."

Diagrammatically, the environment for "be" insertion is:



and for "have" insertion:



Bach analyzes the perfect as a construction satisfying the above conditions for "have" insertion, but with a sentence embedded in the objective NP position. This is not a claim that perfects are transitive constructions, but that this type of construction is unmarked. In Jakobsonian terms, those constructions which take "be" could be considered marked as attributive, while those that take "have" are simply non-attributive. In this construction, the subject of the embedded sentence is deleted by the rule that deletes identical subjects after verbs such as "want": for example, "I want to go" with "I go" embedded, vs "I want him to go" with "he go" embedded. The perfect construction must be marked as necessarily satisfying the conditions for this rule.

There is superficial evidence for a similarity between perfects and transitive constructions with "have". NPs which appear in these constructions can form adjectives with the ending of the verbal participle: for example, "a three legged stool," "a stool which is three-legged," and "a stool which has three legs," which are related as are "a grown man," "a man who is grown," and "a man who has grown."

There is, however, a syntactic difficulty. Bach inserts both "have" and "be" directly into the auxiliary. He fails to account for the difference in the behavior of the perfective "have" and other "have's" in transformations. For instance, we have the option of "Have you a book?" or "Do you have a book?" but only "Has he gone?" not "Does he have gone?" This problem can be solved by inserting "have" first under the verb node. There is a later rule which shifts "have" to the auxiliary, if there is no modal (Hofmann, 1966; section 5). This rule is optional except when "have" is followed by the perfect construction. "Have" then is a zero verb, while "be" is a zero auxiliary.

In fact, once one agrees that all true verbless constructions take "have", it is no longer necessary to claim that the perfect is embedded in a noun phrase. If there is such a thing as verb phrase complementation, the perfect could be an example.

Bach, like McCawley, fails to adequately account for the constraints on co-occurrence with adverbs. He precludes all time adverbs from the embedded sentence in the perfect and claims that only those adverbs which may occur in the main sentence are allowed. That is, in the present perfect, only adverbs which may occur with the present are allowed. This allows adverbs like "by now" or "already": "He is there by now," "He has arrived by now," and prohibits adverbs such as "yesterday", *"He is there yesterday," *"He has arrived yesterday." However, it does not account for time adverbs such as "since yesterday" or "three hours" which occur with the present perfect but not with the present. "I have been here since yesterday," *"I am here since yesterday"; "I have been here three hours," *"I am here three hours" (I am here three hours every day" is a different usage). This, as well as the special constraint on the transformation which shifts "have" to the auxiliary, is an argument for a special perfect construction.

However, I find Bach's analysis preferable in that it has the information of the anteriority of the action carried in the surface by the morpheme which forms the perfect participle, while McCawley implies that this information is carried in the verb "have".

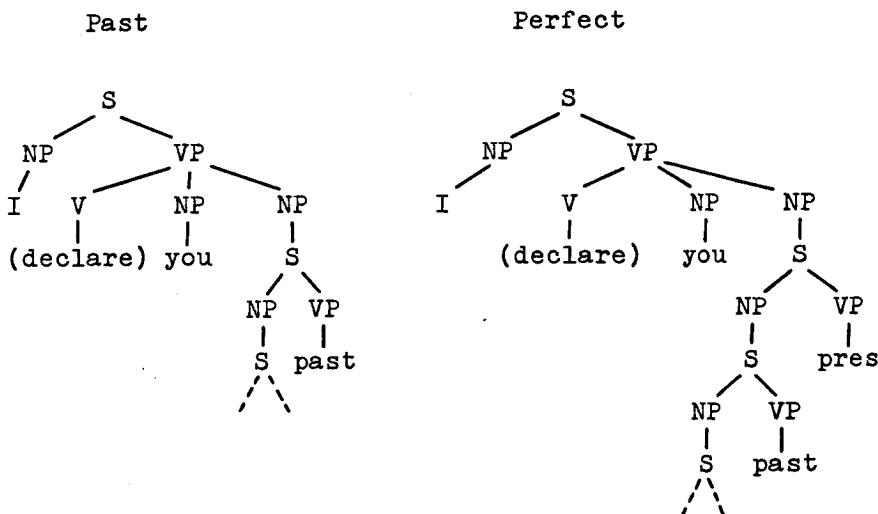
Bach's analysis is also preferable, since for his analysis it is not absolutely necessary that the embedded sentence be in the past tense. He says it is, but he does not justify his assertion. He seems to have assumed that all embedded sentences have tense, and, given the choices of tenses, past was the only reasonable possibility. Since Ross (1967) has treated tense itself as an embedded sentence, there is no longer any reason to require that embedded sentences have tense.

The motivation for eliminating an underlying perfect is one of formal economy. Formal economy is no longer the all-powerful criterion that it was a few years ago. If there is a difference in meaning, and I think that I have demonstrated that there is, then this difference should be revealed in the underlying structure. It would take the form of a special nominalizer or complementizer, which would show up on the surface as the morpheme of the perfect participle.

APPENDIX

McCawley, during the discussion of my paper, pointed out that his analysis is not so inconsistent as I had claimed. He separates the taxic from the demon-

strative function of tense, or rather, he claims that tense is only taxis. It only shows the temporal relationship between sentences. This claim is made possible by an assumption of an inherent reference point for the time of the action in all sentences. Some sentences lose this reference point in the process of embedding. The interpretation of the highest tense as a relation between sentences is made possible by the postulation of a still higher "performative" sentence, whose time reference is always the present. The partial trees of past and present perfect sentences are then:



This makes for a rather strange interpretation of the perfect. In the structure above, the verb "present" seems to indicate that the time of action of the verb "declare" is the same as the time of action of the verb "past", while the verb "past" indicates that the time of action of the verb of the main sentence (unshown) is prior to the time of action of the verb "present". This is meaningless to me, and points to the necessity of inserting a verb between the two tenses. English, in fact, does this; it inserts "have". The question is, what is there in the deep structure? There does indeed seem to be something in the relationship between the subject and the predicate which is temporally unmarked, and in English the unmarked tense is "present". The only suggestion that I can make is that there is an abstract pro-verb which indicates whether or not the following predicate is to be attributive or non-attributive.

BIBLIOGRAPHY

Bach, Emmon. 1967. "Have and Be in English Syntax," Language, 43, 462-85.

Forthcoming. "Nouns and Noun Phrases," Proceedings of the Texas Symposium on Language Universals, April 13-15, 1967. Edited by Emmon Bach and Robert Harms.

Chomsky, Noam. 1965. Aspects of the Theory of Syntax. Cambridge, Mass.

Darden, B.J. 1967. "Syntactic Features in Generative Grammar," a paper read at the Third Annual Regional Meeting of the Chicago Linguistic Society.

1968. "Phrases without Paraphrases," Chicago Journal of Linguistics, 2, no. 1, 68-71.

Fisher, Susan D. 1967. "A Late Paper on Time." Unpublished ms.

Hofmann, T.R. 1966. "Past Tense Replacement and the Modal System," Report no. NSF-17. The Computation Laboratory, Harvard Univ., VIII 1-21. Cambridge, Mass.

Jesperson, Otto. 1927. Essentials of English Grammar. New York, 1967.

Lakoff, George. 1966. "Adjectives and Verbs," On the Nature of Syntactic Irregularity. Appendix A. Report no. NSF-16. The Computation Laboratory, Harvard Univ. Cambridge, Mass.

McCawley, James D. 1967. "The Respective Downfalls of Deep Structure and Autonomous Syntax," 42nd Annual Meeting, LSA, Chicago, Illinois. Dec. 28-30.

Ross, John R. 1967. "Auxiliaries as Main Verbs." Unpublished ms.

ON TOO AND EITHER,
AND NOT JUST ON TOO AND EITHER, EITHER

Georgia M. Green
University of Chicago

Although sentences such as (1) and (2):

- (1) John is not a fink, and he is not a hero either.
- (2) John is a jerk, and he is a liar too.

with the emphatic particles either and too have been mentioned in such articles as Klima's "Negation in English" (1964) and Lila R. Gleitman's "Coordinating Conjunctions in English" (1965), little, if anything has been said outside of the fact that where either occurs, there is in the preceding clause a negative pre-verb (*i.e.* no, not, never, and in some dialects, words like scarcely, rarely, and seldom (Klima 1964:261-70)), and where too occurs, there is not. This paper will treat the construction of sentences with these emphatic particles.

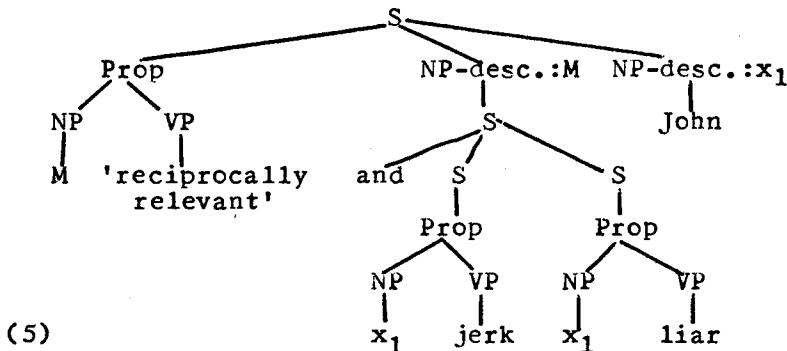
1. The most ordinary uses of the particles either and too involve an emphatic coordinate conjunction, as illustrated in (1) and (2). Like other conjunctions, this type may connect two clauses without any parts of those clauses being identical and subject to any sort of reduction, as in (3) and (4):

- (3) St. Louis is not in New York, and the Adirondacks are not in Kansas, either.
- (4) I wouldn't have any money, and you'd have to lend me your key, too.

Notice that there is always clause-final stress-pitch pattern on the last stressed word of each clause, and on the emphatic particle. This additional, separate clause intonation on the particle seems to indicate that it is itself a separate assertion. If this is the case, this assertion must be the assertion that there is a relationship or a common denominator between the assertions which are conjoined. This is not at all a far-fetched idea, since, as noted by Wierzbicka (1967:7-18) with respect to noun phrases, the presence of a common denominator is a requirement for conjoining.² Since these sentences with too and either are emphatic sentences, it makes a good deal of sense to say that they contain an assertion, realized as too or either with

clause intonation, that the conjoined elements are relevant to one another.³

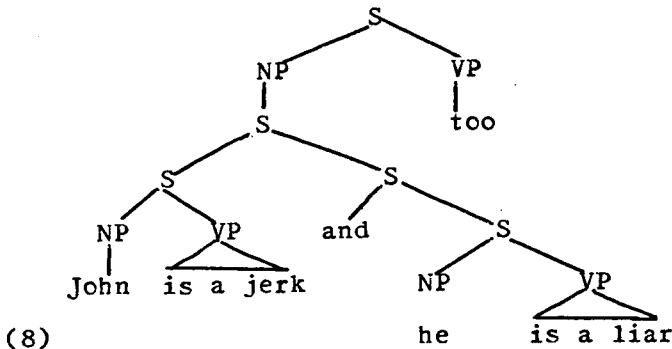
Coordinate conjunction with the emphatic particles either and too may thus be characterized as involving at least three clauses, and no semantic or syntactic constraints on the content of the conjoined clauses aside from the requirements of coordinate conjunction in general. I propose that the deep structure⁴ of (2) is roughly:



The verb phrase the semantic content of which is more fully 'members of which are relevant to each other', and which I have indicated is the predicate of a proposition of which the conjoined clauses are the subject, is predicated of a set of sentences, and is eventually realized as too or either. Too occurs in these sentences unless the highest verb in the rightmost of the sentences which compose its subject is negated in the surface structure, in which case either occurs. Thus, there is a contrast between (6) and (7):

- (6) Max didn't say that Herb was smart, and he didn't say that he was a fool, either.
- (7) Max said Herb wasn't smart, and he said that he wasn't a fool, too.

The status of too and either as predicates in the derived constituent structure:



accounts for the fact that they receive a clause intonation separate from that of the preceding clause.

2. Consider sentences (9) and (10):

(9) St. Louis is not in New York, and New Orleans is not on the East Coast, either.

(10) St. Louis is not in New York, and New Orleans is not on the East Coast either.

(10) in contradistinction to (9) implies that New York is on the East Coast. Other and...either sentences with intonation like (10) are interpreted similarly; what is explicitly predicated of the subject of the second clause is implicitly predicated of the subject of the first clause. The sentence

(11) Ho Chi Minh is responsible for a lot of deaths, and LBJ is no saint either.

implies that Ho Chi Minh is no saint. Similar behavior is exhibited by sentences with coordinate conjunction with the emphatic particles also and too with similar intonation, that is, with clause intonation for the second clause occurring before the verb phrase, as in (12):

(12) Barb is seventeen, and Wendy is old enough to have a driver's license, too.

(12) implies that seventeen, and therefore Barb, is old enough to have a driver's license.

Sentences (10) through (12) involve some kind of "pronominalization" of predicates, and make assertions involving such terms as $f(x_1)$ and $f'(x_1)$ (where f' denotes a set of semantic terms somehow derived from or related to that of f) rather than such terms as $f(x_1)$ and $g(x_1)$, which explains the fact that such

sentences as (13)

- (13) Ho Chi Minh has killed a lot of people, and
LBJ lives in Texas too.

are ungrammatical unless, for example, lives in Texas is intended (and understood) to mean something implying 'has killed a lot of people'.

It is a fact of English that non-deictic pronouns are characteristically unstressed. Likewise the intonation of nouns which are used in a manner which I will call pseudo-pronominal⁶ contrasts with that of ordinary nouns:

- (14) What's new?
The idiot called me up yesterday. (noun)
(15) Have you heard from Algernon lately?
Yes, the idiot called me up yesterday.
(pseudo-pronominal)

I do not take to be a coincidence the fact that the verb phrases which I have claimed are like pronominalizations lack stress. It is conceivable that it is a consequence of the fact that the verb phrases in the second clauses count as the second occurrences of these verb phrases, since a rule lowering the stress on non-first occurrences of given elements must be what is responsible for the fact that the "antecedents" of pronouns are unstressed when they follow the pronouns which refer to them, as in:

- (16) When he gets home, you'll get a belt from your dad.

as compared to:

- (17) Tonight you'll get a belt from your dad.

Notice that backwards pronominalization does not apply here in conjoined clauses (that is, the pronominalization may not precede its "antecedent"). The obvious explanation for this is that backwards pronominalization never applies into a clause not₈ subordinate to the one in which the 'antecedent' occurs. It is easy to see that just as with sentences like (18)

- (18) *Ho Chi Minh isn't one, and LBJ is no saint either.

there is something wrong with

- (19) Barb is old enough to have a driver's license,
and Wendy is seventeen too.

although the infelicity of the latter is logical, while that of the former is grammatical.

The use of the emphatic particles too and either in the sentences described as having pseudo-pronominalization is seen to be similar to their use in the sentences first-described, in that they contain a clause intonation for each conjoined clause, and one for the emphatic particle. One difference between the two types of sentences is that, in the second type, the verb phrase of a clause other than the first seems to be a pronominalization of some sort of the verb phrase of a preceding clause, and clause intonation for such a clause falls on the last stressed syllable preceding the verb phrase. Both types of constructions conform to the general requirements for conjoining with and, and to the requirements of what may be termed too-either suppletion; however, in contrast to the situation with ordinary too- and either-sentences, where too-either suppletion is determined by whether the main verb of the preceding derived clause was negated, here too-either suppletion is determined by the presence of a negative in the immediately preceding clause. Thus, there is a contrast between

- (20) Max said that Herb was Chinese, and he didn't say that Jóhn was oriental too.

and

- (21) Max said that Herb wasn't Chinese, and he also said that Jóhn wasn't oriental either.

A further difference between the ordinary use of either and too and their use with pseudo-pronominalization of predicates is that although sentences with the former are more emphatic with too and either they are equally good sentences without these particles, while sentences with the latter do not occur without either or too:

- (22) *Karl comes from Indiana, and Harry is a Hoosier.

We may infer from this:

- a) that these sentences have deep structures similar to that given in (5), and it is obligatory that 'reciprocally relevant' (or more simply 'too') be predicated of them since there are identical VP's or VP's which are formally semantically related in the deep structures of the conjoined sentences, or
- b) that there is no 'too' in the deep structure, but too-either is introduced by a transformational rule which is triggered by the occurrence in deep structure of VP's which are identical or formally semantically related, or

(c) that there is in the deep structure some element which is realized as too-either, but the structure is not identical to that given in (5).

I will defer my choice until later, but let me note here the fact that while alternative (a) is a rather extreme solution, the transformation mentioned in (b) would have to 1) create a considerable amount of structure fairly high in the tree, if the derived structure of sentences like (10) is to account for clause-intonation on the emphatic particle in the same way as does the derived structure of sentences like (9); or 2) tack too-either onto the end of the VP it follows, and specify somehow that it has its own clause intonation.

3. There is apparently a third kind of coordinate conjunction with emphatic particles, in which the predicate of the second clause is not a pseudo-pronominalization of the first, yet which does not seem quite like ordinary emphatic coordination. Examples:

- (23) You're a louse, and I don't think you love me, either.
- (24) Get the hell out of here, and don't shove, either.
- (25) You just got me wet, and you didn't do a very good job of wiping the window, either.

The fact that

- (26) *You're a hero, and I don't think you love me, either.
- (27) *Please leave, and please don't push, either.

are ungrammatical makes it appear that some kind of negativity is required for the first clause. The fact⁹ that in order for

- (28) You're a linguist, and I don't think you love me, either

to be grammatical, linguist must be taken as a pejorative term suggests the hypothesis that for emphatic coordination with either, both clauses must be in some way negative. Actually, as is demonstrated by

- (29) *You just got me wet, you fool, and I don't mind a bit, either.
- (30) You don't love me, and you're a louse, too.
- (31) It's good-looking, and it will wear well, too.
- (32) It's not expensive, and it tastes good, too.
- (33) It tastes good, and it's not expensive, either.

the restriction is more general: the two clauses must agree in semantic or affective negativity. When such

sentences as

- (34) He's a pacifist, and he paints well, too.

are acceptable, pacifist must be taken as connoting a positive attribute in the mind of the speaker if paints well is.

That there be an overt pre-verbal negative in the second clause remains a syntactic requirement of all conjoined sentences which have the emphatic particle either. As is clear from (29), this requirement is distinct from the semantic requirement that the two clauses agree in affective negativity. It might be concluded that the grammatical sentences of (23) through (34) are further examples of pseudo-pronominalization of predicates, and that the common semantic term which permits one verb phrase to be used as a sort of abbreviation for a preceding one is approval or disapproval on the part of the speaker. I claim not only that this connection is too vague to be called pronominalization of any sort, but also that the fact that these sentences have clause-final intonation within the verb phrase of the second clause indicates clearly that they cannot be examples of pseudo-pronominalization of predicates, because such predicates are always unstressed. In addition, sentences like (23) behave with respect to too-either suppletion like the sentences described in the first section, not like those described as having pseudo-pronominalization:

- (35) I didn't say he loved me, but I didn't say he was a louse, either.

- (36) I said he didn't love me, but I said he wasn't a louse, too.

A common denominator of positive or negative judgement is indeed required in such sentences as (23), but it is probably a consequence of the requirement for emphatic conjunction that the conjoined elements be relevant to each other.

In many respects, sentences like (23) belong to the same class as the sentences discussed in the first section. The two sub-types differ mainly in that the construction of the type just described is sensitive to a context defined by the attitudes of the speaker. This type was discussed separately in order to emphasize that while the semantic or contentive requirement of a common denominator is relatively flexible in unemphatic coordination, the semantic requirement for conjunction with emphatic too and either is more rigid. To demonstrate this difference, one may contrast a sentence which has been cited as grammatical but rather pointless

(Gleitman 1965:262):

- (37) I wrote my grandmother a letter yesterday, and six men can fit in the back seat of a Ford.

with

- (38) I wrote my grandmother a letter yesterday, and six men can fit in the back seat of a Ford, too.

which is something less than pointless. (38) is inconceivable except perhaps as an enumeration of complaints or triumphs by a mentally disturbed person, in which case it would be a token of the type of sentence just described. The establishment of a suitable context is not only more vital for the acceptance of sentences with emphatic coordination than it is for sentences with unemphatic coordination, it must be more obvious.

4. The emphatic use of either and too with the conjunction but is similar to their use with and in ordinary sentences involving identical or non-identical elements in conjoined clauses, as in

- (39) John is not a fink, but he is not a hero either.

- (40) John is fairly bright, but he is pretty absent-minded, too.

and in sentences involving pseudo-pronominalization of predicates, as in

- (41) Ho Chi Minh is a murderer, but LBJ is no saint either.

but the kind of emphatic coordination which requires an affective context is ungrammatical with but:

- (42) *You're a louse, but I don't think you love me, either.

- (43) *Get the hell out of here, but don't push, either.

- (44) *You don't love me, but you're a louse, too.

Some such sentences may occur with an unemphatic but, for example:

- (45) Get the hell out of here, but don't push.

but others may not, e.g.:

- (46) *You're a louse, but I don't think you love me.

When but conjoins clauses, its use involves a denial that the contrary of the following clause is

true, although the contrary of the following clause is to be expected in the light of preceding content.¹⁰ But may thus be considered equivalent to 'and, contrary to (your) expectation'. The addition of a parenthetical 'contrary to your expectation' in (39) is the minimal difference between the underlying (semantic) representations of (39) on the one hand, and (1) on the other.

5. The occurrence of emphatic too and either is not limited to sentences with coordinately conjoined clauses. Emphatic too and either also appear in sentences with restrictive and non-restrictive relative clauses¹¹ and in sentences with subordinate adverbial clauses, and in exactly analogous constructions. Thus there is the ordinary type:

- (47) John, who is no fool, isn't very smárt, either.
- (48) That man who is ugly isn't very smárt, either.
- (49) Although John isn't very smart, he is no fóol, either.

There is the type which involves pseudo-pronominalization of predicates:

- (50) John, who is a crook, said that Harry was no paragon of virtue either.
- (51) That man who is a crook said that Harry was no paragon of virtue either.
- (52) Because I am always late, Jóhn never gets to class on time either.

And there is the type in which there is a common affective attitude shared by the two clauses:

- (53) John, who is a louse, doesn't lóve me, either.
- (54) That guy who's so rude doesn't take good care of his péts, either.
- (55) When John is being rude, he doesn't take good care of his péts, either.

Emphatic either and too are used with complement clauses, too, if there is pseudo-pronominalization of predicates

- (56) That John says Harry is stúpid indicates that hé's not very smart either.¹²

but apparently not otherwise:

- (57) *That there are 500,000 GI's in Vietnam indicates that Johnson will not be re-elécted, either.
- (58) *The fact that you're a louse indicates that you don't lóve me, either.

Notice that backwards pronominalization cannot occur here, even if the condition of subordinate clause position is met:

- (59) *Although Harry's no paragon of virtue, Bill's a crook too.

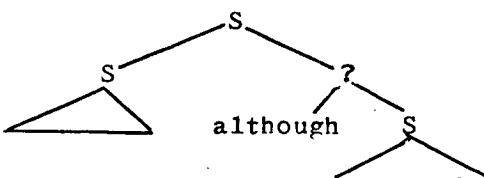
The parallel fact that pseudo-pronominalization of noun phrases does not work backwards either, as shown by (61)

- (60) Although Algernon called me up yesterday, the idiot forgot what he wanted to say.
 (61) *Although the idiot called me up yesterday, Algernon forgot what he wanted to say.

indicates that the same phenomenon is involved there, and that it is not exactly the same phenomenon as the pronominalization which results in anaphoric personal pronouns.

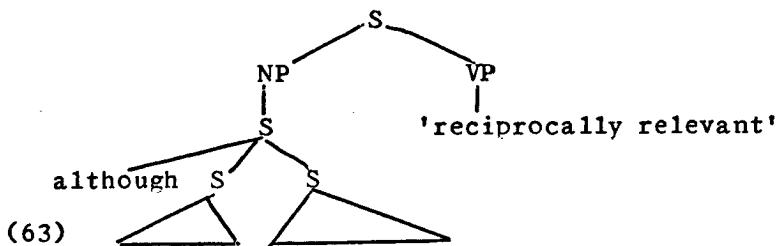
The fact that complement clauses may have the emphatic particles too and either only if pseudo-pronominalization of predicates is involved, while other subordinate clauses may have emphatic particles in both kinds of constructions suggests strongly that the emphatic particles do not have the same source in the two types, so the first alternative mentioned for their structure is ruled out.

In regard to ordinary (not pseudo-pronominal) emphatic sentences, if the proposed explanation for clause intonation on the emphatic particle in these sentences and the proposed type of underlying structure exemplified by (5) are correct, and if the sentences like (47) through (49) and (53) through (55) are derived in the same way as their coordinately conjoined counterparts, the structures of these sentences must at some fairly deep level be similar to the structures of sentences with coordinately conjoined clauses. Thus, although emphatic sentences with subordinate adverbial clauses seem to have a structure like (62)

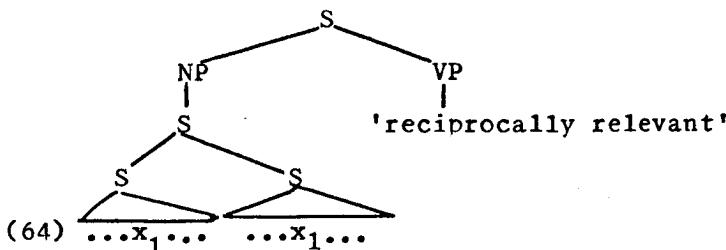


(62)

before pronominalization applies, earlier their structure would have to be more like (63)



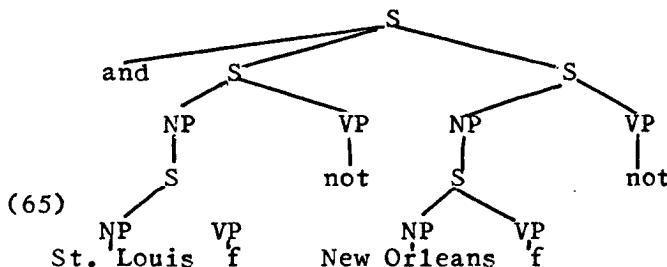
where the two clauses are members of a set of which the fact that they are relevant to each other can be predicated. Regardless of where relative clauses originate, and although they eventually end up dominated by a node which also dominates what they modify, at some intermediate stage both emphatic sentences with restrictive and those with non-restrictive relative clauses would have to have structures something like (64)



So what do the deep structures of sentences with pseudo-pronominalized predicates look like, since they are not the same as the deep structures of other sentences with too and either? Two possibilities suggest themselves. One is that in their deepest structure the related VP's of such sentences appear as identical indices which refer to a VP-description, or as identical VP-descriptions. Thus, at an early stage, the deep structure of (10)

- (10) St. Louis is not in New York, and New Orleans is not on the East Coast either.

would be



where f = 'in New York' and becomes, in its second occurrence, on the East

(65, continued) Coast, by pseudo-pronominalization.

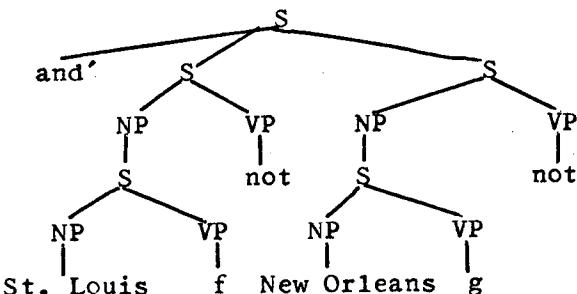
However, this means that (10) and (66) and (67)

(66) St. Louis is not in New York, and New Orleans doesn't have to put up with Rockefeller's bungling either.

(67) St. Louis is not in New York, and New Orleans isn't served by the Long Island Railroad either.

would all have the same deep structure. This amounts to putting everything the speaker knows or believes about every contentive he uses into the deep structures where those contentives occur, and has the effect of making grammar undescribable since this kind of "structure" is speaker-specific and unpredictable.

The other alternative is that the VP's in sentences like (10) and (66) and (67) are related and/or relevant to each other in the mind of the speaker, but not in some formalizable semantic way, and as suggested in the third alternative for the structure of these sentences, some element, perhaps a conjunction or an optional property of conjunctions which is, under certain conditions, eventually realized as and...too or and...either is present in deep structure and triggers the de-stressing of the second verb phrase. Conditions for the occurrence of this element would be similar if not identical to the semantic conditions for the occurrence of the predicate 'too'; the formal semantic nature of this element would be similar to that of but: it would indicate that the speaker asserts that what is true of the second (or following) subject is also true of the first (or preceding). The deep structure of (10) would then be more like:



where f = 'in New York' and g = 'on the East Coast'
(68)

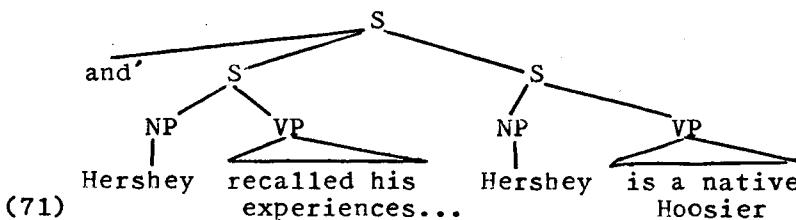
Sentences such as those in

- (69) Hershey said last week he "had a feeling that lots of engineers are going to be deferred." ...A retired Army general himself, the native Hoosier recalled his experiences as an officer on the Mexican border.¹³

clearly involve a related phenomenon: the hearer knows from the complete lack of stress on Hoosier that the native Hoosier is a description of Hershey; what is predicated of the native Hoosier is predicated of Hershey. It seems most likely that this kind of pseudo-pronominalization has as its source a non-restrictive relative clause, which presumably comes from an underlying conjoined structure. If so, sentences like (69) involve exactly the same phenomenon as sentences like (10). Thus, in context,

- (70) The native Hoosier recalled his experiences as an officer on the Mexican border.

would have the deep structure:



6. Summary. I have described two kinds of coordinate conjunction with the emphatic particles either and too, sketched their relation to conjunction with but, and shown the existence of analogous constructions in other kinds of multi-clause sentences. In this discussion it was claimed that 1) too and either in one type of sentence are the realization of an assertion by the speaker that the propositions in two clauses are relevant to each other, and 2) verb phrases in certain other sentences are pseudo-pronominalizations of other verb phrases in those sentences. The implication of these claims for the theory of generative grammar is that grammars will have to be able to refer to S's and VP's in terms which are not wholly syntactic. At the very least, the syntactic component of such grammars will have to operate in terms of semantic representations, and will have to be sensitive to the speaker's "conception of reality" and his conception of relations between his concepts to a greater extent, and for more reasons, than has been assumed up to now.

NOTES

¹I am grateful to James D. McCawley, Jerry L. Morgan, and Alice Davison for comments on earlier versions of this paper.

²Gleitman, assuming that identical common strings was the necessary condition for conjunction, claimed that all English sentences were conjoinable by virtue of the shared string S (p. 262), citing as proof the sentence:

- (a) I wrote my Grandmother a letter yesterday, and six men can sit in the back of a Ford.

which she admitted was rather pointless. Even if Gleitman's informants weren't puzzled by this sentence, I am. It requires a context to be accepted with anything but a blank stare. Gleitman does, however, mention that informants "respond oddly (they laugh; they groan)" to sentences such as:

- (b) He turned in his income tax, and he turned in his cramped compartment.

which she describes as having morphemes which are associated with different phrase-structure representations repeated in the two conjuncts (p. 263). It seems clear, however, that what these sentences lack is semantic coherence of some sort, not surface syntactic or morphological congruence. For example,

- (c) I wrote my Grandmother a letter yesterday, and six men will carry the missive to Albany tomorrow.

is perfectly acceptable. This (semantic congruence) explains both why such sentences as (d), with wholly identical conjuncts are not accepted, and why sentences like (e), with what Gleitman calls a tag, are perfectly acceptable.

- (d) *John ran home, and John ran home.

- (e) I wrote my Grandmother a letter yesterday, and I mailed it today.

³In reported speech this assertion is likewise realized as too or either. Such sentences as

- (a) John said that Harry was a louse, and that Bill was a louse too.

where it is reported that John made both statements, contrast with:

- (b) John said that Harry was a louse, and Bill is a louse, too.

where the speaker is adding a statement of his own to his report of John's statement, so it is clear that, like NP-descriptions, assertion of a relationship of relevance may occur at nodes other than the uppermost. A discussion of NP-descriptions appears in McCawley (1967).

⁴I am using the term "deep structure" to refer to a semantic representation of an utterance as an ordered and labelled tree. I use the term somewhat loosely, to refer to such a tree before many, but not necessarily before all, transformations have applied to it.

⁵If

- (a) Max said Herb wasn't smart, and he said that he wasn't a fool, either.

is a possible sentence, it is a report that Max said, "Herb isn't smart, but he isn't a fool, either." and it is more clearly phrased:

- (b) Max said Herb wasn't smart, and he also said Herb wasn't a fool, either.

or (c) Max said Herb wasn't smart, and he said Herb wasn't a fool, either, too.

(7), on the other hand, is the assertion that Max has made two statements about Herb (that he is not smart, and that he is not a fool) with no implications of temporal or other relationship between these two statements. It may be paraphrased:

- (d) Max said that Herb wasn't smart, and he also said that he wasn't a fool.

⁶Cf. Bach (1967), especially his remarks on the ambiguity of the sentence:

The idiot called me up yesterday.

⁷While it has been suggested that preceding antecedents of pronouns also lose their stress (*i.e.* that when more than one occurrence of a noun phrase appears in a sentence, all occurrences are de-stressed),

this is not the case with pseudo-pronominalization of predicates. When a verb phrase is followed by a pseudo-pronominalization of itself, it is not de-stressed unless there is contrastive stress on its subject:

In addition, of the two most ordinary intonations of the sentence:

- (c) After John saw Mary he kissed her.

that is

- (d) what happened? After John saw Mary he kissed her.
(e) What happened? After John saw Mary he kissed her.

the first seems more nearly neutral; the second seems to have contrastive stress on saw. Similarly in the pseudo-pronominalization of noun phrases discussed by Bach:

- (f) Why so sad? I wanted to talk to Algernon, but
the idiot forgot to give me his phone number.
(g) Why so sad? I wanted to talk to Algernon, but
the idiot forgot to give me his phone number.

It is similarly not obligatory to de-stress preceding antecedents of pronominalized nouns. Typical environments for both types of intonation are given below:

- (h) How come this street looks different? After vandals burned down the old library, the city built a new one.
 - (i) What happened to the old library? After vandals burned down the old library, the city built a new one.

In backwards pronominalization, however, both the pronoun and the following "antecedent" are de-stressed:

- (j) When he saw her, John kissed Mary.
(k) *When he saw her, John kissed Máry.

This would follow from a rule separate from pronominalization, and following preposing of adverbial clauses, which de-stresses second occurrences of elements. Certainly such a rule is required for sentences like:

(1) John is Irish, and Harry is Irish too.

compared to:

(m) *John is Irish, and Harry is Irish too.

⁸The constraint on backwards pronominalization has been described with reference to personal pronouns by Langacker (1966) and by Ross (1966).

⁹I am grateful to Robert Hammarberg for pointing out this fact to me.

¹⁰Cf. Reichenbach (1947:329).

¹¹I am grateful to Alice Davison for bringing these sentences to my attention.

¹²Notice that de-stressing of the pseudo-pronominalized verb phrase forces the anaphoric personal pronoun he to receive stress that it would not ordinarily have. Compare (56) with:

(a) That John says Harry is stupid indicates that he's not very smart.

although he in this sentence is ambiguous, and could refer to either John or Harry.

¹³Chicago Sun-Times, March 17, 1968, p. 72.

REFERENCES

- Bach, Emmon, 1967. "Nouns and Noun Phrases." To appear in Proceedings of the 1967 Texas Conference on Language Universals, edited by Emmon Bach and Robert T. Harms.
- Gleitman, Lila R., 1965. "Coordinating Conjunctions in English." Language 41.260-293.
- Klima, Edward S., 1964. "Negation in English." The Structure of Language: Readings in the Philosophy of Language, edited by Jerry A. Fodor and Jerrold J. Katz, pp. 246-323. Englewood Cliffs, N.J., Prentice-Hall.
- Langacker, Ronald, 1966. "On Pronominalization and the Chain of Command." Unpublished multilith, University of California at San Diego, La Jolla. To Appear in Modern Studies in English, edited by Sanford A. Schane and David A. Reibel, Englewood Cliffs, N.J., Prentice-Hall.
- McCawley, James D., 1967. "Where Do Noun Phrases Come from?" To appear in Readings in English Transformational Grammar, edited by Roderick Jacobs and Peter S. Rosenbaum, New York, Ginn.
- Reichenbach, Hans, 1947. Elements of Symbolic Logic. New York, Macmillan.
- Ross, John Robert, 1966. "On the Cyclic Nature of English Pronominalization." To appear in To Honor Roman Jakobson: Essays on the Occasion of his Seventieth Birthday, The Hague, Mouton.
- Wierzbicka, Anna, 1967. "Against Conjunction Reduction." M.I.T., duplicated.

INTERESTING FEATURES OF GENDER-NUMBER CONCORD
IN MODERN LITERARY ARABIC

Carolyn G. Killean
(The University of Chicago)

Open any textbook or descriptive grammar of modern literary Arabic and you will find statements such as the following: Literary Arabic has two genders, masculine and feminine and three numbers, singular dual and plural. These statements are not only dull; they are misleading. Arabic, in its concord system, clearly distinguishes a neuter-like gender in the plural and exhibits only two basic number categories, singular and plural. Notwithstanding the universals of Greenberg (1966b 34), the dual noun form in literary Arabic is clearly linked with the singular concord system and not the plural.¹

Literary Arabic is a heavily inflected language "concord-wise". Agreement forms exist which link verbs with their subjects, modifiers of various types including numerals and demonstrative and relative pronouns with the noun they modify and, of course, pronoun substitutes. These agreement forms can be organized into six sets which are presented in Table I. Each set represents an intersection of a traditional gender and number category in Arabic, except for Set II which is overworked. The traditional assignment of names to these six sets indicates clearly how they fit into the scheme of a statement such as the one given above. Set I is labeled masculine singular, Set II feminine singular, Set III masculine dual, Set IV feminine dual, Set V masculine plural, and Set VI feminine plural.

¹ Unless specifically designated otherwise, all further references in this paper to Literary Arabic or just Arabic should be interpreted as references to Modern Literary Arabic. I prefer this usage to that of acronyms such as MLA, MWA, etc.

TABLE I
PARTIAL LIST OF CONCORD SETS IN MODERN LITERARY ARABIC

	Set I	Set II	Set III	Set IV	Set V	Set VI
Adjectives	-∅	-a	-aani -ayni	-ataani -atayni	-uuna. -iina	-aat
Demonstrative	haadha	haadhihi	haadhaani haadhayni	haataani haatayni	haa'ulaa'i	
Verbs						
perfect	-a	-at	-aa	-ataa	-uu	-na
imperfect	y-	t-	y-...aani	t-...aani	y-..uuna	y-...na
Pronouns						
nominative	huwa	hiya	humaa	hum	hum	humma
oblique	-hu	-ha	-humaa	-hum	-hum	-humma
Numerals	-a	-∅	-aani -ayni	-ataani -atayni		

Unfortunately this labeling is not adequate to explain the behavior of these sets in agreement relationships with Arabic nouns. A closer look at the facts of Arabic concord in the plural will indicate why.

Morphologically there are two basic types of plural forms in literary Arabic with one of these types split into two sub-types. Syntactically, there are also two basic types of concord for the plural in literary Arabic with one type divided into two sub-types. However, these two systems don't match up in any neat way. The morphological plural types are traditionally termed "sound" versus "broken" plurals with the former being subdivided into a type termed masculine and another termed feminine.² Examples of these three morphological plural types are given in the 11 pairs of noun forms listed below.

Plural Forms in Literary Arabic

Broken forms: Sg.

1. 'book'	kitaab	(I)	kutub	(II)
2. 'man'	rajul	(I)	rijaal	(V)
3. 'girl'	bint	(II)	banaat	(VI)
4. 'caliph'	xaliifa ^t	(I)	xulafaa'	(V)

Sound forms:

"masculine" type

5. 'teacher'	mu ^c allim	(I)	mu ^c allimuuna	(V)
6. 'Muslim'	muslim	(I)	muslimuuna	(V)

"feminine" type

7. 'lady teacher'	mu ^c allima ^t	(II)	mu ^c allimaat	(VI)
8. 'article'	maqaala ^t	(II)	maqaalaat	(II)
9. 'meeting'	ijtimaa ^c	(I)	ijtima ^c aat	(II)
10. 'mother'	'umm	(II)	'ummahaat	(VI)
11. 'cow'	baqara ^t	(II)	baqaraat	(II)

Syntactically, the concord sets assigned to the noun forms listed under the Plural heading are predictable from form alone only in the case of the masculine sound plural - examples 5 and 6.

² Basically broken plurals in Arabic represent internal vowel changes within a stem while sound plurals refer only to suffixation of plural endings to a basic stem.

Concord Set V is identified with both, /mu^callimuuna/ "teachers" and, /muslimuuna/ "Moslems." In addition, however, concord set V is linked to example 2, /rijaal/ "men" not because of its broken plural form but because it bears a covert semantic marking, male human being.

Concord set VI, that traditionally termed feminine plural, is not assigned indiscriminately to all the examples given under the feminine sound-plural category but only to examples 7 and 10 because they refer to groups of women. In addition, the plural noun in example 3 will take set VI forms in concord because it is semantically marked, female human being.

This leaves examples 1,8,9 and 11 and no more sets of concord in Table I labeled plural by the traditionalists. I mentioned before that Set II, the traditional feminine singular concord category, was overworked and it is. All plurals in Literary Arabic except those directly marked for the semantic category of humanness (in formal terms - Person) will take Set II in all agreement situations except with numerals. This "all-others" category of nouns is clearly equivalent to a neuter gender in other languages, however, it is certainly rare to see it manifested only in the plural number, a clearly marked category. In fact, one of Greenberg's original language universals published in 1963 included the following statement: "A language never has more gender categories in non-singular members than in the singular." (74).

It is possible to avoid the label "neuter" for this "all-others" or non-Person plural category by viewing this use of Set II concord forms as a syncretization of number combined with a neutralization of gender. That is, number is completely unmarked for these plural forms and gender is restricted to the category feminine.

To justify the use of the marked gender category in a neutralization role it is possible to distinguish two Set II's. One is a marked category of forms in relation to the masculine in the number category singular and can neutralize under certain circumstances to the masculine gender. The other Set II can never neutralize to the masculine forms when it is itself the neutralization of all non-human gender assignments in the singular.

In fact, this set of so-called "feminine" forms are themselves unmarked in relation to Sets V and VI.³ Occasionally, rational or human plurals are treated as "destitute of reason", a phrase favored by one grammarian when writing of this phenomenon in 1852. In such a case, concord Set II is used in agreement with nouns formerly considered rational. The reverse situation is also possible - personification of non-human beings in Literary Arabic. When personified, animals whose plurals normally take Set II (example 11), can take Sets V and VI particularly in fairy-tale situations.

The basic concord split in the plural in Literary Arabic is thus between those nouns marked for the semantic feature, Person, and those not so marked, the former being further separated into ladies versus men. It is interesting to note that, in Greenberg's terms (1966b 30), the "dominance" of the unmarked category, male, is clearly shown by the fact that if a group of ladies are joined by one man, their sex, syntactically speaking, changes to masculine.

In order to state rules for the assignment of concord sets to plural nouns in Arabic one clearly must distinguish a dichotomy between Person nouns and non-Person nouns, whether or not one calls it a neuter/human distinction. What is both significant and interesting is that this dichotomy is not distinguished by separate concord behavior in the non-plural numbers, the singular and the dual. The rules which assign gender to an Arabic noun whether in the singular or the dual are as follows:

³R. Hetzron (1967) adopts a polarity view that gender switching indicated plural number in the original number system of Proto-Semitic. Some of his arguments for the traces of this older system in the number system in Classical Arabic are not very convincing. (185-187) However, its effect on the numerical or counting systems in Semitic are still clearly evident. He agrees with this author on the need for a restatement of Arabic gender-concord restrictions in more realistic terms (187 & 191).

All nouns are masculine, that is, take concord
Set I in agreement unless:

1. they are semantically marked for femininity,
/bint/ in example 3.

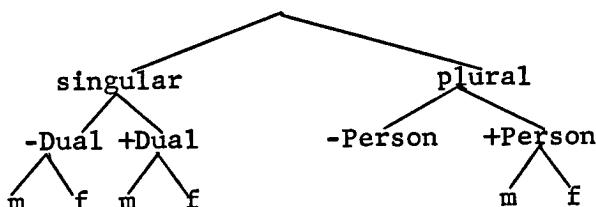
2. they are lexically marked for femininity by
convention in Arabic, for example, */ħarb/* 'war'.

3. They are morphologically marked for femininity
by one of a set of suffixes termed "feminine" in
Arabic */maqaala^t/* in example 8.

When a semantic marker is contradicted by a formal marker, the semantic marking prevails. For example, the word for caliph in Literary Arabic (example 4) is feminine in form, however, it takes Set I forms in concord because it always refers to a man, i.e. is semantically marked (-female). Similarly, a noun semantically marked as female need not always take feminine form, see example 9, *'umm/* 'mother'.

Two possible ways of visualizing the hierarchy and cross-classification of features which are necessary to account for this Arabic data are given below. The first diagram, A, considers the split of number to be the most basic in the Arabic noun system whether it is assigned from context or is lexically a part of a feature complex inherent in the noun. It is easy to see from this tree the emergence of the distinction of Person as significant in the plural and its lack of significance in the singular.

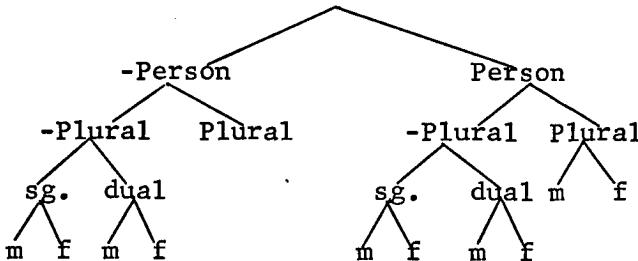
DIAGRAM A



Another possible representation of this feature hierarchy is given in diagram B. In order to account for more than just concord behavior, the basic split in the Arabic noun system is posited as that of Person versus non-Person.

Although not necessary for this discussion, the former category, Person, obviously can be further divided into the traditional pronoun categories of 1st, 2nd and 3rd person while non-Person nouns can only be considered in relation to 3rd person pronouns. There are no "I's", "we's" or "you's" in language without implications of a semantic feature of humanness involved.

DIAGRAM B



One of the strongest points Greenberg makes in his discussion of universals in 1966a (77) concerns the place of the dual in a number system which includes it. He states that the singular is unmarked in relation to the plural while the plural is unmarked in relation to the dual. The data of concord restrictions given above for Arabic tends to upset this neat generalization.

First of all, neutralization of dual concord in Literary Arabic occurs regularly in only one specific instance, in connection with the limiting quantifier "both", /kilaa/ masc. or /kiltaa/ fem. Clearly a dual form itself, this word is always followed by a dual noun or pronoun. However, for all concord related to this dual phrase, the SINGULAR sets of concord are obligatory, not the PLURAL.

Second of all, in all the modern colloquial dialects of Arabic which are related to Literary Arabic (though probably not through a direct line of descent) all specific "dual" concord forms have disappeared. (Ferguson 1959 620-621). However, noun forms are still frequently marked for duality in the dialects. These dual nouns take regular plural concord i.e. the modern reflex of Set V (reflexes of Set VI have largely disappeared also).

Superficially it appears that the dual has indeed been absorbed into the plural system and thus supports Greenberg's universal concerning its marked relation to plurality. However, one additional fact argues that this dual still retains a "special" status in Arabic which sets it outside the plural system.

The Person/non-Person distinction is still not significant in the dual. Whereas for non-Person nouns in the Arabic colloquials, optional use of the modern equivalent of Set II in plural concord situations is allowed, this optional usage is forbidden to the same nouns in the dual. Dual nouns regardless of whether or not they refer to humans must take plural concord.

There is still another argument for separating the dual from the plural in Arabic. A major distinction separating countable nouns from un-countable nouns (Count/-Count) cross classifies with the concord system set up in diagram B. Both Count and non-Count nouns can take singular and plural forms, singular and plural concord sets. However, only Count nouns in Arabic can take dual suffixes and dual concord. For example, the collective noun /šajar/ 'trees' in Arabic can be rendered countable by the addition of a suffix /-at/. This new noun formation /šajara/ 'a tree' can now be counted also in the dual /šajarataani/ 'two trees'. Collective nouns in Arabic can have relatable plural forms but no duals unless they are first rendered "countable."

The weight of evidence against considering the dual a marked extension of the plural system in Arabic prompts me to propose a number system for Arabic which excludes it. Preserving Greenberg's universal for number systems, we still have a plural which is marked in relation to a singular category. In order to take care of the dual in Arabic and the strange phenomena associated with counting in that and other Semitic languages, I propose that a separate concord system be set up, a numerical or counting system, which would include only forms from Sets I, II, III and IV. The plural and its Person/non-Person concord distinction is a part of the Arabic NUMBER system but not the Arabic NUMERICAL system of concord.

Both Count and non-Count nouns in Arabic participate in the first system. Countable Nouns alone participate in the second. The singular is a member of both systems, the dual only a member of the counting system and the plural, in Arabic at least, based on its concord relationship is a member of the NUMBER system only. This, in spite of the fact that the plural form of a noun in Arabic follows all numbers which end in the numerals 3 to 10. The gender concord set which these numerals take is assigned, despite the presence in the surface structure of a plural noun form, considering only the gender of the SINGULAR forms of the nouns as significant. In other words, as in the dual, the gender forms displayed in the entire counting system of literary Arabic have absolutely nothing to do with any distinction of Person-non-Person which has been shown to be so significant in the concord relating to the category plural.

Summary:

This paper has sought to make two points which are particularly significant considering the current emphasis on language universals in systems of concord. One is that Arabic displays a strange fondness for distinguishing more categories in its plural concord rules than it does in its simpler singular system. This proliferation of distinctions in a marked category is against the general trend cited by Greenberg. Greenberg has used classical Arabic in several of his examples but he has chosen to ignore this important fact. Perhaps he is too influenced by morphological criteria in a system that clearly depends heavily on covert semantic criteria for one of its major distinctions.

The second point is that the dual in Arabic clearly appears separate from the plural number system. Its affinity for the singular sets up either the impossible system (according to Greenberg) of a single category (non-plural) embracing one or two as opposed to three or more (plural), or one can adopt my solution - the removal of the dual from the number system altogether and the postulation of a separate "counting" or numerical system. The implications of this latter step for other languages have not yet been investigated.

Bibliography

Ferguson, Charles A. 1959. The Arabic Koine Language
35 616-630.

Greenberg, Joseph H. ed. 1963. Universals of Language.
Cambridge, Mass.

_____. 1966a Language Universals.
Current Trends in Linguistics Vol III Mouton,
The Hague 61-112.

_____. 1966b Language Universals with
special reference to feature hierarchies. Mouton,
The Hague.

Hetzron, R. 1967 Agaw Numerals and Incongruence in
Semitic. Journal of Semitic Studies XII 169-197.

THE VERB-PARTICLE CONSTRUCTION IN ENGLISH
BASIC OR DERIVED?

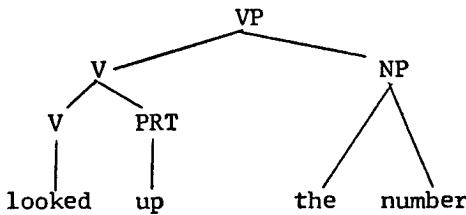
Stanley E. Legum
(The University of Texas at Austin)

By "verb particles" I shall mean such words as up and out when they are used in sentences like those in (1).

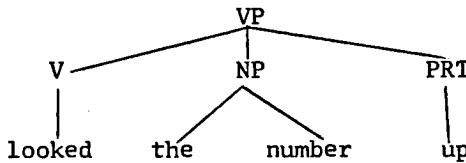
- (1) a. Susan looked up the number for her friend
- b. Susan looked the number up for her friend
- c. The chairman brought out many interesting points
- d. The chairman brought many interesting points out

As Bruce Fraser has shown in his dissertation (An Examination of the Verb-Particle Construction in English, MIT, 1965) such strings as those in (1) have tree structures such as those in (2).¹

(2) a.



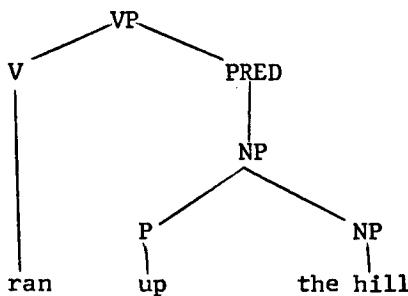
b.



¹ Although we agree as to the bracketing of particles and prepositions, Fraser and I differ widely in our labeling of nodes. Nothing essential hinges on the labeling system I have chosen as far as the arguments presented in this paper are concerned.

Structures like these are, of course, distinct from structures containing prepositional phrases. Thus the substructure (3a) of the sentence (3b) is distinct from the substructure (2a).

(3) a.



b. The boy ran up the hill

The essential difference between (2a) and (3a) is that in (2a) looked up is a constituent while in (3a) up the hill is a constituent.

Fraser has noticed a number of facts which support the above differences in bracketing and which can serve as tests for identifying recalcitrant items as either particles or prepositions. Seven of these tests are:

(4.1) PARTICLES MAY BE PERMUTED OVER DIRECT OBJECT NP'S, PREPOSITIONS MAY NOT BE

- a. The fisherman reeled in the line
- b. " " " the line in
- c. They talked about the situation
- d. * " " the situation about

(4.2) A PARTICLE CANNOT BE PREPOSED WITH A QUESTION WORD WHILE A PREPOSITION CAN

- a. *In what did the fisherman reel?
- b. About what did they talk?

(4.3) A MANNER ADVERB CAN COME BETWEEN A VERB AND A FOLLOWING PREPOSITION, BUT NOT BETWEEN A VERB AND A FOLLOWING PARTICLE

- a. *The fisherman reeled quickly in the line
- b. *The fisherman reeled quickly the line in
- c. They talked loudly about the situation

(4.4) (FOR MANY SPEAKERS) A SHORT PARENTHETICAL PHRASE CAN BE INSERTED BETWEEN A PARTICLE (WHEN IT STANDS NEXT TO THE VERB) AND A FOLLOWING NP, BUT NOT BETWEEN A PREPOSITION AND ITS OBJECT

- a. The fisherman reeled in, a broad grin on his face, the line from which a shark dangled
- b. *The fisherman reeled, a broad grin on his face, the line from which a shark dangled in
- c. *He talked about, a broad grin on his face, the situation

(4.5) IN ACTION NOMINALIZATIONS AN OF IS INSERTED AFTER A PARTICLE BUT NOT AFTER A PREPOSITION

- a. The fisherman's reeling in of the line saved his life
- b. *Their talking about of the situation pleased the dictator

(4.6) PARTICLE-NOUN PHRASE SEQUENCES CANNOT BE CONJOINED; PREPOSITION-NOUN PHRASE SEQUENCES CAN BE CONJOINED

- a. *The fisherman reeled in the line and in the fishnets
- b. They talked about the situation and {about} the issues of the day
on

(4.7) PARTICLES MAY HAVE SECONDARY STRESS; PREPOSITIONS ARE ALWAYS UNSTRESSED

- a. The fisherman reeled in the line
2 2 1
or 3 2 1
and possibly² 2 3 1
- b. They talked about the situation
2 4 1

² Fraser claims that a 2-3-1 stress pattern is acceptable in this context as well as the other patterns indicated.

Given the fact that there exist many pairs of sentences in English with substructures like (2a) and (2b), there can be no doubt that they should be transformationally related. If we rule out the possibility of a third type of structure underlying both (2a) and (2b),³ we come to the natural question: Which of these two structures is derived from the other? To derive (2b) from (2a) we would need a rule of Forward Particle Movement such as (5).

<u>(5) Forward Particle Movement</u>			
S.D.	<u>L</u> V	PRT _v	NP X] VP
	1	2	3 4
S.C.	1	0	3+2 4

CONDITION: ⁴ Obligatory if 3 > [+PRO]

Conversely, to derive (2a) from (2b), we would need a rule of Backward Particle Movement such as (6).

<u>(6) Backward Particle Movement</u>			
S.D.	L V	NP	PRT X] VP
	1	2	3 4
S.C.	1#3	2	0 4

CONDITION: ⁴ 2 > [+PRO]

CONVENTION: # denotes Chomsky adjunction

³ This decision is based on the desire to keep the mechanism involved in the description of verb-particles to a minimum. Not only is there no apparent need for any third structure different from both (2a) and (2b), but it is difficult to conceive what form such a structure might have.

⁴ The conditions on rules (5) and (6) reflect the fact that verb-particles always follow pronoun objects standing next to the verb. If Ross's Output Condition on Post-Verbal Constituents (John R. Ross, Constraints on Variables in Syntax, MIT, 1967 rule (3.41)) can be formalized, then it will probably be possible to omit the condition in both versions of the Particle Movement Rule. In any case, the conditions probably do not differ in complexity and hence should not be considered to affect the relative complexity of (5) and (6).

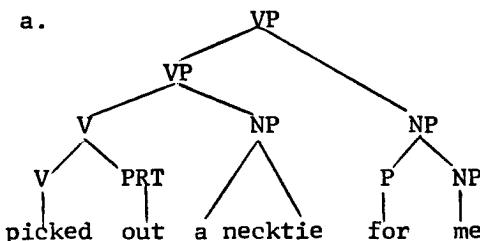
The fact that it takes more ink to state the condition on (5) than on (6) is simply a matter of our notational devices. It is not obvious a priori whether a blocking condition or an obligating condition should be a more expensive device. Instead of relying on an evaluation metric to decide between these two competing formulations, we should use the empirical evidence provided by the discussion which follows as information which should be reflected in the final form of our metric.

With these two rules in mind, we can now reformulate our basic question as: Which of the two types of Particle Movement leads to the simpler and more general description of English?

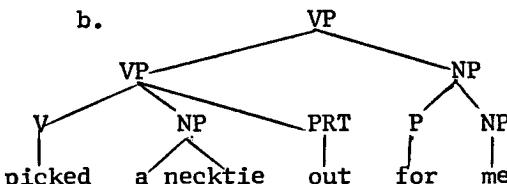
I can find no reason to prefer either (5) or (6) on the basis of a direct comparison of their formulations. There is, however, some evidence based on the interaction of the for-Dative Movement⁵ rule with Particle Movement which indicates that Forward Particle Movement provides the better analysis. Consider the sentences (7), which have the approximate structures (8).

- (7) a. Mary picked out a necktie for me
 b. " " a necktie out for me
 c. " " me out a necktie

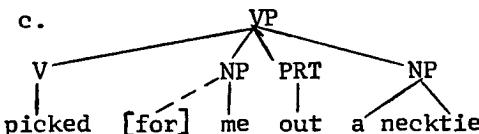
- (8) a.



- b.



- c.



With Forward Particle Movement there is no difficulty in deriving (8b) from (8a); and with Backward Particle Movement there is no difficulty in deriving (8a) from (8b). If we order for-Dative Movement and for-Deletion

⁵ For a discussion of the Dative Movement Rules and the associated preposition deletion rules see Charles J. Fillmore, Indirect Object Constructions in English and the Ordering of Transformations, Mouton and Co., The Hague, 1965.

before Forward Particle Movement, then (8c) can be derived smoothly from (8a). A serious problem arises, however, when we attempt to derive (8c) from (8b) by means of Backward Particle Movement. Either we must reformulate for-Dative Movement so as to move both the direct and indirect object NP's, or we must complicate the Backward Particle Movement rule to insure that when a particle follows two object noun phrases, as in (9) and (10a), it is obligatorily moved either next to the verb or after the first of these noun phrases.

- (9) *Mary picked me a necktie out
- (10) a. *Mary picked the boy with red hair a beautiful necktie out
- b. Mary picked out the boy with red hair a beautiful necktie
- c. Mary picked the boy out a beautiful necktie

Clearly an analysis along either of these last two lines is ad hoc and misses the generalization that verb particles interact only with the object noun phrase immediately following the verb.

There is, however, another argument available for Backward Particle Movement. According to this argument certain adjectives and prepositions which cannot conceivably be considered particles must participate in Particle Movement, but cannot be generated in the base directly next to the verb. For this reason Forward Particle Movement cannot be appropriately generalized to account for the data. Backward Particle Movement can, on the other hand, meet the test of empirical adequacy with only the smallest increment in its complexity. All that needs to be done is to replace PRT in Backward Particle Movement by a disjunction of PRT, ADJ and P1 (the specific kind of prepositional structure which is relevant for this rule). Thus (6) would become (11).

<u>(11) Revised Backward Particle Movement</u>			
S.D.	V	NP	X _{VP}
		{ PRT ADJ P1 }	
1 1#3	2	3 0	4 4

CONDITION: 2 \neq +PRO

It is my claim that the adjectives and P₁-prepositions can be generated in the base next to the appropriate verb with no more (and possibly less) mechanism than is needed to generate them elsewhere.⁶ By either giving these adjectives and prepositions the feature [+PRT] or attaching them to a PRT node, we can allow Forward Particle Movement to apply without change. Furthermore, I claim that even setting aside the for-Dative Movement data, an analysis with Forward Particle Movement comes closer to attaining the levels of descriptive and explanatory adequacy than does an analysis utilizing Backward Particle Movement. The point of the remainder of the arguments in this paper is that the adjectives and prepositions in question behave identically to particles in all relevant respects; and hence, they should be treated as "true" particles and not merely as members of the ad hoc disjunction incorporated into (11).

Let us first consider the verb-adjective sequences which behave like verb-particle combinations. Some verb-adjective pairs of this kind are: cut short, blast open, blow shut, and fling open. Notice that it is possible to say both

- (12) John cut short his conversation with Mary
and
(13) John cut his conversation with Mary short

Both the cut short cases and verb particles behave identically with respect to Rossian output constraints on the complexity of their object noun phrases. Thus we can say both

⁶ Fraser proposes two alternative formulations of the Lexical Insertion Rule which are both potentially capable of handling particles. Since the Lexical Insertion Rule does not differ from grammar to grammar within a given grammatical theory, it can have no affect on a choice between alternative analyses based on simplicity arguments.

Various formulations of the Lexical Insertion Rule can, of course, be attacked for lack of universality. Thus it seems unlikely that the rule needed for Backward Particle Movement is formulable in a universally valid form, but I have not shown it to be so.

- (14) John {cut short } his lecture which was
 {chopped up}

purported to challenge the authority of
 the President to wage undeclared war

but find both versions of (15) awkward.

- (15) ?*John { cut } his lecture which was
 1{chopped}1
 purported to challenge the authority of
 the President to wage undeclared war
 {short}
 1{ up }1

We do, of course, find (16) completely acceptable
 once we have simplified the direct object noun phrase.

- (16) John { cut } his lecture {short}
 1{chopped}1 1{ up }1

That the permissible verb-adjective combinations must be listed in the lexicon in much the same manner as the permissible verb-particle combinations can be seen from the fact that not all semantically reasonable combinations can occur. (See Fraser's dissertation for a detailed argument on this point.) Thus we do not find either (17a) or (17b):

- (17) a. *John cut brief his lecture
 b. * " " his lecture brief

That adjectives in verb-adjective combinations remain adjectives in function can be seen from the fact that they take degree adverbs and comparatives. Thus we find

- and (18) John cut his lecture very short
 (19) John cut his talk on the intricacies of
 the United States' pattern of overseas'
 investment much shorter than we had expected

Notice that these modifications of the adjectives alter their effective complexity with respect to the Output Constraints; for we find (20) and (21) somewhat uncomfortable:

- (20) *John cut very short his lecture

- (21) ?John cut much shorter than we had expected
 his talk on the intricacies of the United
 States' pattern of overseas' investment

Examples (12) through (21) lead to two related conclusions. First, that at least at some point in a derivation it is possible for an item to be simultaneously an adjective and a particle; and second, that whatever modification of the lexical insertion rule is needed to insert certain prepositions (namely those normally thought of as "particles") in the particle position, that same modification can be formulated so as to allow certain adjectives to be inserted in that position as well. Hence the cut short type combination of verb plus adjective can be handled equally well by an analysis using Forward Particle Movement and one using Backward Particle Movement.

Let us now consider that class of compound (i.e., polymorphemic) prepositions including among others into, out of, away from, off of and inside of which have been used to argue for the necessity of Backward Particle Movement. Modifying Fraser's notation slightly, I shall label the first morpheme in any such compound P_1 and shall label the remainder of the compound P_2 .

Whenever we find a grammatical sentence with a compound preposition introducing a prepositional phrase, we also find that the corresponding sentence with P_2 and its object NP omitted is also grammatical. Furthermore, if the sentence contains a direct object, there is also a second grammatical reduced sentence with the P_1 permuted over the direct object noun phrase. Thus along side of (22a) and (22d) we find (22 b,c,e,f).

- (22) a. The butler brought the dinner into the room
 b. " " " " " in
 c. " " " in " "
 d. She took her book out of her purse
 e. " " " " out
 f. " " out " "

Another fact which needs to be explained is that we do not find both a particle and a P_1 (whether compounded with a P_2 or not) associated with the same verb simultaneously. Thus although sentences (23) are all grammatical, sentences (24) are all ungrammatical.

- (23) a. John chased away the cat
 b. John chased the cat away from the house
 c. John chased out the cat
 d. John chased the cat out of the house
- (24) a. *John chased away out the cat
 b. *John chased out away the cat
 c. *John chased out the cat away from the house
 d. *John chased away the cat out of the house

Fraser has suggested that these facts can be accounted for by adopting a rule (25)--Fraser's rule (5-39)--of Prepositional Phrase Deletion and Revised Backward Particle Movement which is written so as to be inoperative for all but the first particle or P₁ which appears after the verb.

(25) PP-Deletion					
S.D.	[+V]	(NP)	P ₁	P ₂	NP
		1 2 3 4			
S.C.		1 2 3		0	

Besides the fact that it requires the adoption of the more complex of the Particle Movement Rules, such an analysis is vulnerable to two other serious objections. It will allow the base to generate the sentences (26) when away and out are considered particles as opposed to prepositions, and has no automatic way of marking them as unacceptable.

- (26) a. *John chased the cat away out of the house
 b. *John chased the cat out away from the house

More importantly, the PP-Deletion transformation deletes unrecoverable information, since more than one P₂ can appear following a P₁. Thus according to an analysis adopting rule (25), (27a) could be derived from either the structure underlying (27b) or the structure underlying (27c).

- (27) a. Billy ran away
 b. Billy ran away to Europe
 c. Billy ran away from home

An alternative way to account for the same set of facts is to adopt a Particle Incorporation rule which adjoins particles which are also prepositions to directional prepositional phrases. The ungrammaticality

of (24) and (26) now becomes an automatic consequence of the fact that only one particle can be inserted by the lexical insertion rule with each verb. Two possible formulations of Particle Incorporation suggest themselves:

(28) a.	<u>PRT Incorporation: Version 1</u>
S.D.	$\llbracket \text{I} + \text{V} \rrbracket \llbracket \text{+PRT} \rrbracket \text{V}_{\text{NP}} \llbracket \text{NP} \rrbracket \llbracket \text{+PREP} \rrbracket \text{NP} \rrbracket_{\text{VP}}$
S.C.	1 2 3 4 5
	1 0 3 2+4 5

b.	<u>PRT Incorporation: Version 2</u>
S.D.	$\llbracket \text{I} + \text{V} \rrbracket \llbracket \text{NP} \rrbracket \llbracket \text{+PRT} \rrbracket \llbracket \text{+PREP} \rrbracket \text{NP} \rrbracket_{\text{VP}}$
S.C.	1 2 3 4 5
	1 2 0 3+4 5

If we allow Particle-Incorporation to follow Forward Particle Movement and to precede Complex NP Movement (see John Ross, Constraints on Variables in Syntax, MIT, 1967, rule (5.57)), then by marking the P_2 -preposition [+directional] we can gain a possible simplification in our rules by adopting a modified version of (28b).

(29)	<u>PRT Incorporation: Version 3</u> (probably obligatory)
S.D.	$\llbracket \text{+PRT} \rrbracket \llbracket \text{+PREP} \rrbracket \llbracket \text{+PREP} \rrbracket \llbracket \text{+DIR} \rrbracket$
S.C.	1 2
	0 1+2

CONDITION: $V \neq 1$

The Particle Incorporation analysis is lent support by the fact that all of the relevant tests in (4) indicate that P_1 's are particles. Two apparent dissimilarities are revealed by the tests (4) but can be easily explained. The first of these dissimilarities is the behavior of manner adverbs (see 4.3). Although (30a) and (30b) are ungrammatical, (30c) is grammatical.

- (30) a. * Mabel carried quickly in the newspaper
 b. * Mabel carried quickly the newspaper in
 c. Mabel carried the newspaper quickly into the kitchen

These facts can be handled quite reasonably by requiring that the Manner Adverb Placement Rule(s) follow Particle Incorporation. However Manner Adverb Placement is formulated, it will have to be made sensitive to the

difference in structure created by Particle Incorporation. Notice that this is the case whether or not Particle Incorporation is preferred over PP-Deletion.

The remaining objection to an analysis with Particle Incorporation instead of one with PP-Deletion (and hence with Backward Particle Movement) is that P_1 's conjoin and can therefore not be particles (see 4.6). This argument would be devastating to a Particle Incorporation analysis except for the fact that the only examples of conjunction of P_1 's which exist are severely limited in distribution and quite clearly idiomatic. In all, only three pairs of conjoined (apparently reduced) P_1 's exist. These are in and out, on and off and off and on. I have not been able to find any examples with out and in or any of the other possible pairs of P_1 's. For this reason, I propose that the three occurring pairs of apparently conjoined and reduced P_1 's be treated in the lexicon as units. If such an analysis is indeed correct (31c) will be explained as a case of conjunction reduction while (31d) will be a case of in and out treated as a lexical unit.

- (31) a. The man let the dogs in and out
- b. ? " " " " " into the house and
 out of the house
- c. The man let the dogs into and out of
 the house
- d. The man let the dogs in and out

The non-acceptability of (32) implies that in and out is more complex than a simple particle and hence probably has some internal structure.

- (32) *The man let in and out the dogs

This fact is, of course, completely compatible with the claim that in and out is a lexical unit. Presumably the same mechanism which is used to insert idioms can also be used to insert complex particles in the deep structure. Thus the arguments based on the facts about compound prepositions, lead us to conclude that Particle Incorporation is to be preferred over PP-Deletion, and that of the Particle Incorporation analyses the one assuming Forward Particle Movement is probably more highly valued by the evaluation metric than the analysis assuming Backward Particle Movement.

The above arguments imply that particles are introduced in the deep structure next to the verb and that the deep structures in which they occur are operated upon by the following ordered rules:

- (33) a. for-Dative Movement
- b. for-Deletion
- c. Forward Particle Movement
- d. Particle Incorporation
- e. Complex NP Movement
- f. Manner Adverb Placement (may precede (e))

SOME RELATIONS OF EMBEDDING IN ENGLISH NOMINALS

Larry W. Martin
(University of Iowa)

This paper¹ argues two main points: first, that embedded relatives show properties so similar to those of conjunct sentences that any reasonable set of rules must contain generalizations covering both phenomena, and second, that accounting for relative clauses by simple embedding rules is implausible.

Consider first the general fact that if there exists a sentence containing a relative clause, there exists a corresponding conjoined sentence. The converse does not hold. That is, if we have (1a), we have guaranteed (1b).

- (1a) The president recalled the general who lied to the soldiers.
- (1b) The general lied to the soldiers and the president recalled him.

But just in case the conjuncts share no nounphrase, there is no corresponding relative. Further, whenever a conjunct is anomalous, so is the corresponding relative:

- (2a) The bomb exploded yesterday and I'll defuse it today.
- (2b) I'll defuse the bomb which exploded yesterday today.

The non-equivalence of (3) guarantees that of (4):

- (3a) I drank the martini and threw it up.
- (3b) I threw up the martini and drank it.
- (4a) I threw up the martini which I drank.
- (4b) I drank the martini which I threw up.

And the commutability of (5) guarantees that of (6):

- (5a) The candidate declared first and he will win.
- (5b) The candidate will win and he declared first.
- (6a) The candidate who declared first will win.
- (6b) The candidate who will win declared first.

The facts of (1)-(6) indicate strongly that the sources of rela-

¹I have profited from comments by Sandra Annear, Emmon Bach, Larry Hutchinson, James McCawley, John Ross and Arnold Zwicky. Space would not allow me to reply to all their questions even if I knew the answers.

tive clauses may coincide with those of certain conjoined sentences. The facts of (1)-(6) are well-known, and the proposal to derive relatives from conjoined sentences is not novel. I wish to breathe further life into it by adducing some structural regularities in addition to the existential ones above, and by showing that independent embedding rules lead to incorrect predictions or require ad hoc restrictions.

To each standard if-then conjunction there corresponds a stylistic variant with the antecedent postposed:

- (7a) If the president recalls the general, the senator will conduct hearings in Washington.
- (7b) The senator will conduct hearings in Washington if the president recalls the general.

The structure of (7b) could be either of (A) or (B).

- | | | | |
|-----|-------------|-----|---------------|
| (A) | Sen If Sen' | (B) | (X If Sen') |
| | | | Sen |

That is, it could be viewed either as a permutation or an insertion. Other conjunctions, but not all of them, behave in similar ways. Such conjunctions may occur in other places besides initially or finally. They may insert (a) after the subject, (b) after the second member of the auxiliary, and (c) within the verbphrase after the last member of the mainverb, thus

- (7c) The senator, if the president recalls the general, will conduct hearings in Washington.
- (7d) The senator will, of the president recalls the general, conduct hearings in Washington.
- (7e) The senator will conduct hearings, if the president recalls the general, in Washington.

but not generally, as far as I can tell, elsewhere in the auxiliary or verbphrase. Since rules for insertion must be constructed in any case, I will assume that the structure (B) above is correct, and one case of the general phenomenon of insertion.

Not all conjunctions may undergo insertion; but is an example. Others, such as and and or, may induce stress phenomena, seem more acceptable if metalinguistic (or so they say, and I think he will), seem to reject positioning in the auxiliary, and generally seem poor in style unless they may undergo reduction. Nonetheless, when acceptable, they appear inserted at just the places where conjunctions of the if-type do.

Certain circumstances allow the insertion to take place elsewhere in the sentence, witness

- (8a) The senator will send the president, if he recalls the general, a telegram of sympathy.

- (8b) *The senator will send a telegram, if he recalls the general, to the president.
- (8c) The senator will send a telegram at noon, if that is a good time, from Washington.
- (8d) *The senator will send a telegram, if that is a good time, to the president.

That is, insertion may occur in places additional to those mentioned above if the matrix and the insert share a nounphrase, after the manner of relatives; and those additional places are just the ones where a relative might occur. (If the relative is present, the insert must follow it.)

Note that, unless insertion is restricted more closely than I have alleged, an and-conjunction (or an or-conjunction) receives a double analysis in some cases. Structure (D) would result from applying the insertion-rule to (C):

(C) Sen And Sen'	(D) (X And Sen) Sen'
-----------------------	--------------------------------

I am inclined to think that the double analysis is correct, and that the insertion-rule is responsible for absence or presence of comma-intonation and the obligatory (as opposed to optional) simplification of the coordinate conjunctions, as in the paradigms

- (9a) If the soldier is missing, (then) he must be wounded.
- (9b) The soldier must be wounded if he is missing.
- (9c) *Then the soldier must be wounded if he is missing.
- (10a) (Both) the soldier is missing and he is missing.
- (10b) The soldier is wounded, and he is missing.
- (10c) *Both the soldier is wounded, and he is missing.

Inserts may themselves contain inserts, but they quickly become unwieldy and must obey heavy and general restrictions. In particular, when the interior insert follows the initial nounphrase, either the conjunctions must differ, or else the nounphrases must differ; as borderline as (11a, b) may appear, (11c) is clearly totally unacceptable.

- (11a) That soldier, if he, and he must obey them, obeys orders, will fire at random.
- (11b) That soldier, if the general, if the sergeant demands it, needs a sacrifice, will volunteer.
- (11c) *That soldier, if he, if he is ordered, fires, will hit the target.

Now, if we consider the relative Wh-formatives to be conjunctions, or reflexes of them, exactly the same condition applies. We have:

- (12a) The soldier who, when he fired, hit the target, won a medal.
- (12b) The soldier who, whoever he is, fired at the target won a medal.
- (12c) *The soldier who who fired hit the target won a medal.

In (12a), the nounphrases differ; in (12b), the relatives differ; but in (12c), neither differ. If the Wh-forms are conjunctive in source, these conditions need not be stated for them; they are met automatically.

Inserted conjunctions, then, behave more like relatives than is at first apparent. At the same time, relatives are not so restricted in occurrence as we usually pretend. They may be postposed, preposed, and inserted in the auxiliary.

- (13a) They will give any man five years in prison who burns his draft card.
- (13b) That man who burned his draft card they gave five years in prison.
- (13c) That man will, whatever he pleads, get five years in prison.

Some details require comment. Pre-position requires the preposing of the matrix nounphrase, whereas postposition refuses it.

- (14a) *Who burned his draft card they gave that man five years.
- (14b) *They gave five years in prison that man who burned his draft card.

Further, preposing blocks in subject position.

- (15) *That man who burned his draft card did get five years?

Only ever-relatives occur in the auxiliary; to (13c) compare

- (16) *That man will, who burned his draft card, get five years.

Recall that, similarly, only if-conjuncts may interrupt the auxiliary. Note also the difference in acceptability between (13a) and

- (17) They will give that man five years in prison who burned his draft card.

For some speakers, (17) is more questionable than (13a), parallel to the reluctance of and to postpose and the freedom of if to do so. It is not so clear that relatives extrapose to other places in the sentence. Some speakers, however, accept (18a) but reject (b), which indicates conformity to the conditions on insertion.

- (18a) Any man will buy the girl a drink who sees her at the bar at any time.
 (18b) *Any man will buy the girl who sees her at the bar a drink at any time.

(The time-phrase is not to be taken as part of the relative.)

Copula sentences do not take postposing without an intonation-break. (The sentence can also be interpreted as containing a non-restrictive relative on the preceding nominal.) Nor do they take preposing.

- (19a) That man is an American(,) who went to Canada.
 (19b) *An American who went to Canada that man is.

In spite of these complications, it seems that the extraposition of relatives is a 'converse' to the insertion of conjuncts. It seems plausible to claim that, essentially, an extraposed relative reflects a non-inserted conjunct, and that a conjunct inserted at the matching nounphrase serves as source for a non-extraposed relative.

I turn now to a detail of postposing which raises questions about embedding rules of the common sort. It is possible to postpone relatives to obtain:

- (20a) Any soldier found a tunnel who tried to.
 (20b) Any tiger will attack a man which is a coward.

We expect, therefore, to find an ambiguity in:

- (20c) Any man will attack a soldier who is a coward.

The expected ambiguity, however, barely arises; the relative is attributive only to the immediately preceding nounphrase unless the intonation is interrupted. In general, a relative may not be postponed if it can be 'accidentally' in agreement with the preceding nounphrase. (This is the condition which creates the difficulty in copular postposing.)

Now consider sentences containing what appear to be sequences of relatives, such as:

- (21) They saluted the soldier who fired who hit the target.

The structure of the last nounphrase is generally assumed to be something like:

- (E) (X (Wh Y) (Wh Z))
 NP Sen Sen'

Unless sentences containing relatives are excluded from further embedding, however, an alternative structure is possible because of postponing:

(F) (X (Wh Y (Wh Z)))
 NP Sen Sen'

Since (21) is not ambiguous, one of (E) or (F) must be somehow prohibited.

Suppose the (F) is excluded. Then, since a sentence such as

(22a) They look for any soldier who will buy
 a girl a drink who sits in a bar.

is ambiguous, we must allow for further embedding in the relative if the matching nounphrase is not the same as that of the dominant relative. Since (22a) is ambiguous, we expect that (22b) will be also:

(22b) They look for any soldier who will buy
 a drink for a girl(,) who sits in a bar.

Our expectation is not rewarded unless the intonation-break is present. I conclude from (22) that postponing has occurred and that (F) is the appropriate structure. The condition for the blocking of coordinate structure would be that there must be no 'accidental' matching of nounphrases. That is, a coordinate structure must not appear to break the condition on postponing. As a consequence of selecting (E), we would be forced to the independent statement of conditions for coordinates and their internal structure, which must elsewhere be given for inserts and subordinates, and would be unable to motivate the intonational facts of sentences like (22b).

Suppose, therefore, that we reject (E) in all cases, adopting (F) and prohibiting coordinate structures for relatives. Then no independent statement is required, provided that relatives arise from conjuncts in the first place.

Under this proposal, the apparent recursive properties of embedding would stem instead from its conjunctive source. The rules for pre-nominal attributives would be cyclic: after conjunction and insertion, a sequence of relativization, reduction and pre-nominalization rules would apply, with further relativization only if reduction had taken place.

Accepting this proposal leads immediately to a further generalization. The fact that only one non-restrictive relative may occur in a nounphrase has been remarked often, but no explanation of this fact has been advanced. Under the present analysis, however, this behavior is not eccentric. Rather, it is the general case for relatives.

It remains to determine more closely the specific conjunctions involved, and the exact mechanism required.

Pre-position of relatives is actually more restricted than I have indicated. In particular, those on nounphrases whose determiner is specific (a, one, this, that, the) are more constrained than those on nounphrases whose determiner is general (any, all, every). Only the first of (23) is acceptable, but all of (24) are.

- (23a) That man whom I saw you spoke to.
- (23b) *That man whom I saw did you speak to?
- (23c) *That man whom I saw please speak to.
- (24a) Any man whom I saw you spoke to.
- (24b) Any man whom I saw did you speak to?
- (24c) Any man whom I saw please speak to.

That is, preposing of specifics is possible only in declaratives, but preposing of generals is possible also in questions and commands. But this is exactly parallel to the behavior of conjuncts in and, which are restricted to clauses of the same type, and conjuncts in if, in which a question or a command as well as a statement can serve as consequent. Each of (24), moreover, has a paraphrase in an ever-relative, and these, you will recall, have a distribution closest to that of if-inserts.

We are thus led to the conclusion that (a) conjuncts in and may relativize if the matching nounphrase is specific, and (b) conjuncts in if may relativize if the matching nounphrase is general.

Before outlining a part of the mechanism of relativization, let me cite one firm piece of counter-evidence. In (25), the conjunct sources would violate the conditions on clause-types for and.

- (25a) Did the soldier who was firing stop?
- (25b) Tell the soldier who is firing to stop.

An explanation is crucial, and I have no serious one to offer.

A consideration of the conjunction but gives insight into the mechanism of relativization. For some people, at least, (26a) is not a paraphrase of (26b),

- (26a) No one but John hit Jack.
- (26b) No one other than John hit Jack.

for in the first, we have it guaranteed that John did, in fact, hit Jack, whereas in the second we know only that, if anybody hit Jack, John did. An adequate paraphrase of (26a) would be

- (27) No one other than John hit Jack,
but John did.

Whatever the ultimate source, (27) yields (26a) by copying in the relevant portions in the but-clause and deleting the original. Now

take note of the rather ridiculous but acceptable

- (28) No one but John hit anyone but Jack at any place but Sue's at any time but midnight.

whose source is surely not more complex than

- (29) No one other than John hit anyone other than Jack at any place other than Sue's at any time other than midnight, but John hit Jack at Sue's at midnight.

What (29) reveals is that the but-conjunct is copied in as many times as the conditions require. Otherwise, we must posit four identical but-clauses in succession.

This behavior is not restricted to but. Under at least one reading, all of the following (and several others) are paraphrases:

- (30a) The soldier bought the girl a drink and she winked at him.
- (30b) The girl for whom he bought a drink winked at the soldier who bought her a drink.
- (30c) The girl for whom the soldier bought a drink winked at him.
- (30d) The girl winked at the soldier who bought her a drink.

Sense can be made of this paraphrase-set if relativization, like the formation of but-phrases, is a process of multiple copying and optional subsequent deletion of all but one instance.

To summarize: the conjunctive source of the surface phenomena known as relative clauses accounts for parallels in the behavior and structure of the two. Only certain conjunctions, if and and, may relativize, and only under certain conditions. The mechanism is a copying process.

I will close with one last audacious suggestion. It would be extremely odd for relativization to be a surface phenomenon of English, and at the same time a part of the deep structure of other languages. I propose very tentatively, therefore, that, for noun-phrases, conjoining is the universal process of recursion in natural languages, and that in some, embedding may be rare or even absent.

LEXICAL INSERTION IN A TRANSFORMATIONAL GRAMMAR
WITHOUT DEEP STRUCTURE

By James D. McCawley
University of Chicago

By 'deep structure', I mean a linguistically significant level 'between' semantic representation and surface syntactic representation. I have argued elsewhere (McCawley 1967a, 1967b, 1968a, 1968b) that there is no such level, i.e. that there is no principled basis for dividing the rules relating semantic representation into two systems such as Chomsky proposed in Aspects of the Theory of Syntax: 'semantic interpretation rules' which relate deep structure to semantic representation, and 'transformations' which relate deep structure to surface structure. In this paper I will assume the following conclusions which I drew in the papers cited above:

1. Syntactic and semantic representations are of the same formal nature, namely labeled trees.
2. There is a single system of rules (henceforth 'transformations') which relates semantic representation to surface structure through intermediate stages.
3. In the 'earlier' stages of the conversion from semantic representation to surface structure, terminal nodes may have for labels 'referential indices' such as were introduced in Chomsky 1965:

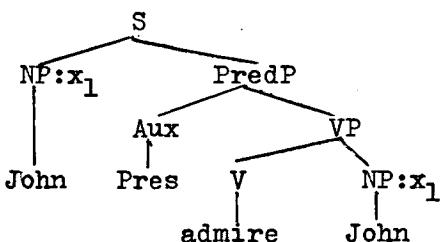


Fig.1. Deep structure of John admires himself, in accordance with unformalized proposals of Chomsky 1965.

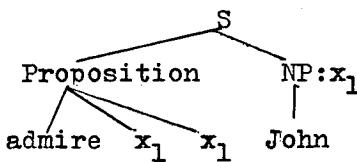


Fig. 2. An early stage in the derivation of John admires himself according to McCawley 1967a.

In semantic representation, only indices and 'predicates' are terminal node labels. The repertory of predicates will be enormous, although not matching lexical items one-to-one, i.e. some lexical items are semantically complex.

The purpose of this paper is to examine one aspect of how a grammar fitting the above conclusions would have to work, namely, the question of how lexical items

get into the representations of sentences. Conclusion 2 implies that the various lexical items get put in by transformations. Indeed, each 'dictionary entry' could be regarded as a transformation, namely a transformation which replaces a portion of a tree that terminates in semantic material by a complex of syntactic and phonological material. At what point in the derivation of a sentence do these transformations apply? It will hardly do to say that these transformations have an extrinsic ordering with respect to each other and to ordinary transformations, since the point at which these insertions take place surely is not a way in which dialects can differ from each other, e.g. surely two dialects of English could not differ as regards whether the morpheme horse is inserted before or after the transformation of extraposition.

I can conceive of four possible ways in which grammars might be constrained so as to prevent the ordering of lexical insertions from being a way in which grammars could differ. The first possibility is that lexical items are not inserted until the very end of a derivation. However, this possibility must be rejected, since the possibility of performing certain transformations depends on the presence of specific morphemes and not just on their meanings, e.g. for is deleted after want but not after desire:

- (1) I want you to win the prize.
- (2) I want very much for you to win the prize.
- (3) *I want for you to win the prize.
- (4) I desire for you to win the prize.
- (5) *I desire you to win the prize.

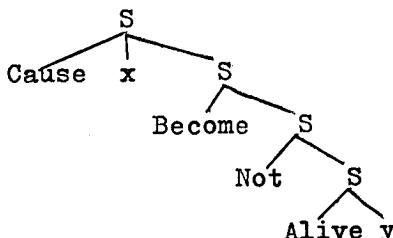
Similarly, particle movement can only affect a verb-particle combination, but a verb-particle combination is often synonymous with a simple verb, e.g.

- (6) He threw Harry out.
- (7) He ejected Harry.

The second possibility is that lexical items are all inserted at the very beginning of a derivation. The bulk of this paper will be devoted to showing that this possibility must be rejected on the grounds that the complex of semantic material which a lexical item corresponds to need not be a constituent of the semantic representation per se but may be a constituent which arises through a transformation, i.e. I will argue that lexical items can be correlated correctly with their meanings only by recognizing prelexical transformations, which apply to trees that terminate in semantic material rather than in lexical material, and stating the conditions for inserting lexical items in terms of the results of these prelexical transformations rather than in terms of the ultimate semantic representation.

What I have just said implies that semantic material is not grouped together in the same way in the semantic representation of a sentence as it is in a lexical item. For an illustration of this, consider the verb to kill. Kill can be resolved into components as cause to die; moreover, at least one of those components, namely die, is itself semantically complex, meaning 'cease to be alive', i.e. 'become not alive'. However, this is not sufficient to describe the semantic representations of sentences involving kill, since a sentence involving it refers to two participants, one of whom causes the event in question and the other of whom dies in that event, i.e. the meaning of x killed y would require a representation along the lines of

Fig. 3



Before kill can be inserted into this structure, it is necessary that Cause, Become, Not, and Alive get regrouped into a unit. An obvious candidate for a transformation that would have that effect is a transformation that adjoins a predicate to the next higher predicate, thus successively converting Fig. 3 into

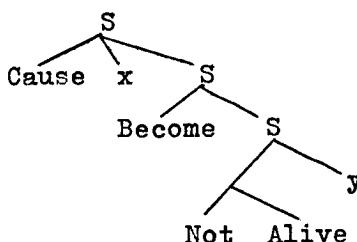


Fig. 4

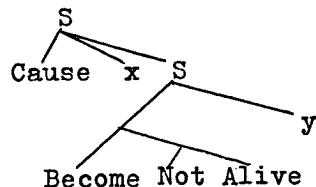


Fig. 5

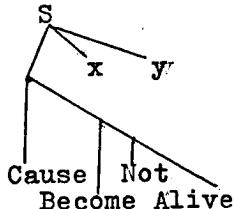
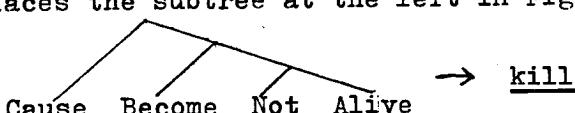


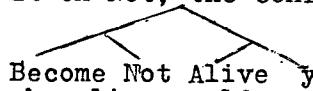
Fig. 6

If lexical insertion did not occur until after the application of that transformation, then the 'dictionary entry' for kill could be expressed as a transformation which replaces the subtree at the left in Fig. 6 by kill:



I would like now to examine some characteristics which the predicate-raising transformation that I have just posited would have to have. First, it would have to

be optional, since there is no need to perform all stages of the last derivation. For example, by failing to perform the last application of predicate-raising, one would obtain sentences such as John caused Harry to die. If one did not perform predicate-raising on Alive but did perform it on Not, the configuration



i.e. y ceased to be alive would result. Secondly, the transformation as I have described it would give rise to many configurations which do not correspond to any lexical item of English. For example, while there is nothing semantically anomalous about the structure in Fig. 7, there is no English lexical item corresponding to the tree of Fig. 8, which would arise if the above

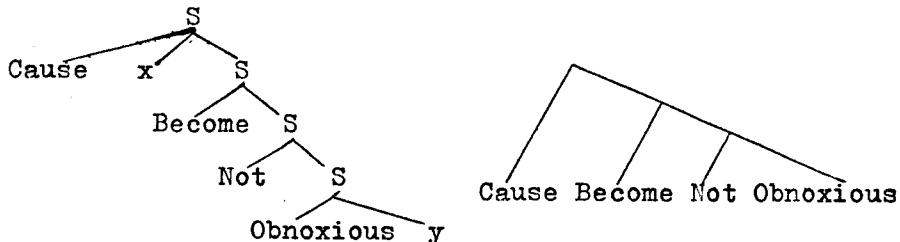
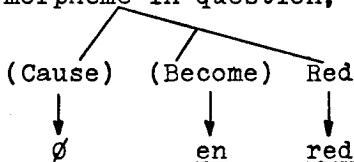


Fig. 7

Fig. 8

steps were carried out. Does this mean that the transformation must be restricted so as to preclude the generation of structures like Fig. 8? It would hardly be possible to formulate such restrictions, since there is very little that can be said in general about the structures to be excluded: indeed, the structures to be excluded are simply those for which English happens to have no corresponding lexical items. Thus, if one were to place such a restriction on the predicate-lifting transformation, he would be adding to the grammar something predictable from the rest of the grammar, namely from the existing gaps in the lexicon. The transformation should thus be unrestricted and the following universal surface structure constraint should be imposed on grammars: a surface structure is well-formed only if all its terminal nodes bear lexical items. Structures in which Fig. 8 had been generated would thus be excluded, since there would be no material in the lexicon to provide a surface realization of them. Note, incidentally, that since there is no regular morphological device in English for marking causatives, the different morphemic realizations of Cause in stative/causative pairs such as red/redden, able/enable, and open/open would have to be part of the dictionary of English anyway. The fact that a given stative adjective, e.g. blue, has no causative does not require that the

morpheme blue be marked as an exception to a transformation but only that there be no dictionary entry for such a causative. The causative of a morpheme can be regarded as the form generated by a composite dictionary entry (along the lines of Gruber 1965) which also generates the morpheme in question, e.g.



(a later suffixation rule would put en after red). Note that the predicate-raising transformation includes as special cases the inchoative transformation and causative transformations of Lakoff 1965.²

Predicate-lifting must thus be a prelexical transformation, i.e. it applies to trees which terminate in semantic matter rather than in lexical matter. Are there other transformations that have to be treated as prelexical? One obvious place to look for such a transformation is nominalizations. Indeed, there is reason to believe that virtually all nominalizations arise through a single prelexical transformation. Specifically, all nominalizations have semantic representations of the form 'the (an) x such that f(x)', e.g.

Agent-nom. The inventor of the wheel.

$(\lambda x)(x \text{ invented the wheel})$

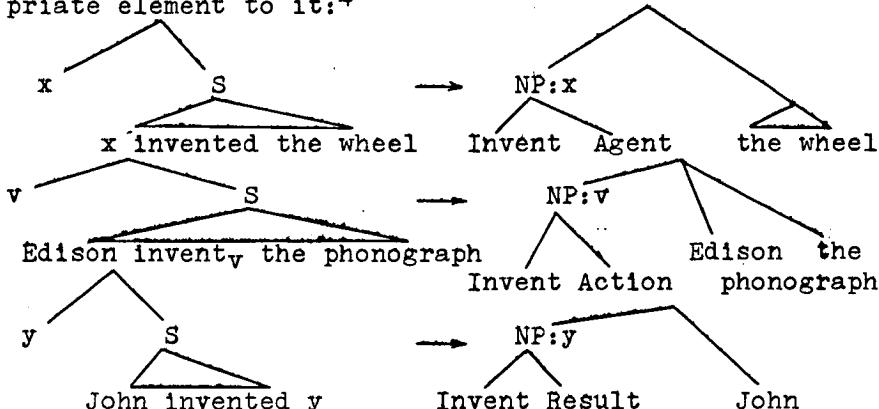
Action-nom. The invention of the phonograph by Edison.

$(\lambda v)(\text{Edison invent}_v \text{ the phonograph})^3$

Result-nom. John's invention.

$(\lambda y)(\text{John invented } y)$

The head noun of the resulting noun phrase consists of the verb of the embedded sentence plus some element that indicates the relation of the variable to the embedded sentence. The nominalization transformation lifts the verb from the sentence and adds the appropriate element to it:⁴



This transformation will also sometimes give combinations of material for which no lexical item exists; for example, there is no agent-nominalization *passer corresponding to pass an examination. In addition, as pointed out by Lakoff (1965), there are lexical gaps of an opposite type, namely items which act like nominalizations but for which no unnominalized verb exists:

- (8) China's agression against India shocked everyone.
- (9) *China agressed against India.

Here there will exist dictionary entries for the combinations produced by the nominalization transformation but not for the 'verb' which the nominalization transformation affected. That these nominalizations must have an embedded sentence is shown by the fact that nominalizations show the same behavior as full clauses when used as the object of a verb such as attempt. The complement of attempt requires an underlying subject identical to that of attempt and loses that subject by the transformation of Equi-NP-deletion:

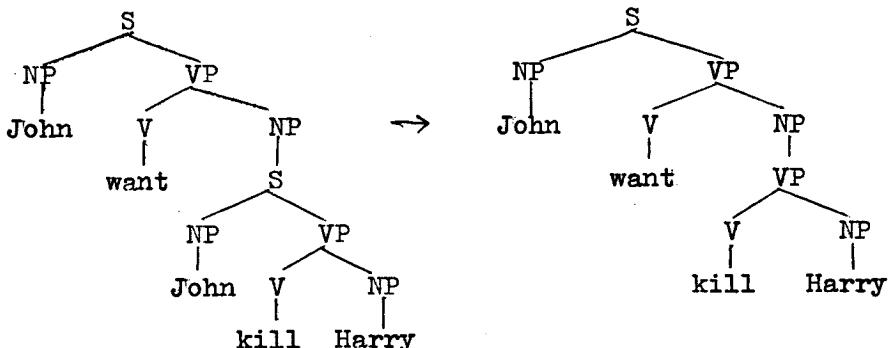
- (10) John attempted to kill Harry.
 - (11) *John attempted for himself to kill Harry.
 - (12) *John attempted for Bill to kill Harry
- The same pattern is exhibited when an action nominalization appears instead of an infinitive:
- (13) China attempted agression against India.
 - (14) *China attempted its agression against India.
 - (15) *China attempted Uruguay's agression against India.

If the nominalization transformation is prelexical, then the following facts (first noted in Warshawsky 1966 and elaborated in Jackendoff 1968) give reason to believe that reflexivization is also prelexical. The reflexive in

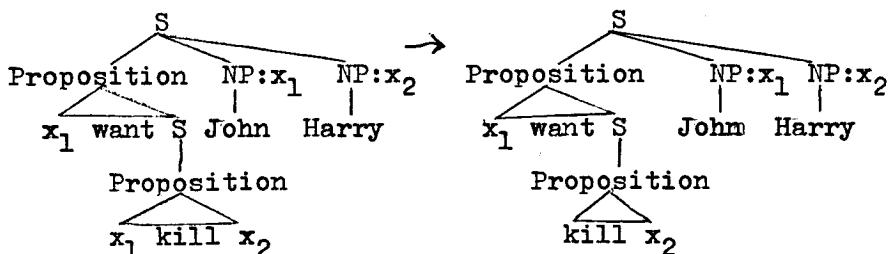
- (16) John showed Harry a picture of himself.
- may refer to either John or Harry, whereas that in
- (17) John showed Harry Picasso's picture of himself.
- may only refer to Picasso, and
- (18)*John showed Harry Margaret's picture of himself.
- is ungrammatical. These facts appear to conflict with the treatment of reflexives in Lees and Klima (1963), according to which reflexives are derived from noun phrases which repeat a noun phrase that occurs earlier in the same simple sentence: that condition would appear to be met not only in (16) but also if there were a copy of John or Harry in place of himself in (17) or (18). Suppose, however, that (following Lakoff 1965) Bill's picture of John, etc. are result-nominalizations of a sentence which exists only in nominalized form and whose verb has picture as a surface reflex. In Bill's picture of John, this verb would have Bill for its underlying subject, but in a picture of John it would have an indefinite pronoun for its underlying

subject. Note that (a) this indefinite pronoun must eventually be deleted, (b) deleting it causes the S-node over the embedded sentence to be lost (by the 'tree-pruning' principles of Ross 1966), (c) the nominalization transformation also causes the loss of the S-node. If reflexivization follows the deletion of the indefinite pronoun but precedes the nominalization transformation, when at the stage of the derivation where reflexivization applies there will still be an S-node over the embedded sentence in (17-18) but not in (16). Thus, this rule ordering plus the Lees-Klima constraint on reflexivization would correctly predict the occurrence of reflexives in (16-18). However, this would entail having reflexivization apply before a prelexical rule and thus that it be prelexical itself.

Another transformation which will have to be prelexical is Equi-NP-deletion. This transformation is usually conceived of as deleting the subject of an embedded sentence if it is identical to a certain noun phrase of the next higher sentence:



Alternatively, however, it could be regarded as deleting a token of a referential index at a stage of the derivation prior to the rule which attaches noun phrases to indices:



The word malingerner has a meaning which appears to result from the latter version of Equi-NP-deletion. Malingerner means 'person who pretends to be sick'. Who

pretends to be sick is what would arise from Equi-NP-deletion from a semantic representation of the form x pretends that x is sick. Thus, treating maligner as deriving from a semantic structure corresponding to 'person who pretends to be sick' entails that Equi-NP-deletion apply prelexically and that maligner be inserted in place of a structure which could only arise through Equi-NP-deletion. The proposal that Equi-NP-deletion apply prelexically is implicit in the proposal of Bach 1968 to derive

- (19) John is looking for a lion.
from the same structure as underlies

- (20) John is trying to find a lion.

Bach points out that having a lion originate in the complement of a verb like try in (19) allows an explanation of the ambiguity of (19) between a referential sense (in which one is talking about some one lion) and a non-referential sense (in which any lion at all would satisfy the quest). The source which Bach proposes for look for thus entails prelexical application of both Equi-NP-deletion and predicate lifting.⁵

The existence of prelexical transformations forces one to reject the second possibility suggested above for how lexical insertions could be transformations but their ordering not be a way in which grammars could differ, namely the proposal that all lexical insertion is done at the very beginning of a derivation. I promised four conceivable solutions to that problem and will quickly describe the remaining two, which I will not choose between, since I have no solid evidence for making such a choice. The third possibility is that lexical insertions are 'anywhere rules', i.e. rules which are not assigned a fixed ordering with respect to other rules but which apply whenever the configuration to which they apply arises. Some evidence for the existence of anywhere rules has been provided in a recent paper by Ross (1967), who argues that the rule which deletes a repeated verb to yield sentences such as

(21) John ordered kreplach and Harry kishke.
must be an 'anywhere rule'. The fourth possibility relates to a controversial question in the organization of grammars, namely whether there is a cycle in syntax. If the transformations of a language divide into two subsystems, a cycle and a system of postcyclic rules, then requiring all lexical insertions to take place after the cycle but before the postcyclic rules would be another way of constraining lexical insertions so that their ordering would not be a way in which languages could differ.⁶ This possibility seems especially worth exploring in view of the fact that the transformations which I have found to be prelexical are also transformations which there is reason to believe are in

the cycle, and since this possibility would constrain the notion of 'possible lexical item' so much more narrowly than would the third possibility. My mind, at least, boggles at the thought of lexical items whose use would be contingent on the prior application of such postcyclic transformations as adverb preposing, question formation, and relative clause extraposition. Proposal four but not proposal three would automatically exclude the possibility of any such lexical items. It should be noted that proposal four would contradict the claim with which I began this paper, that there is no linguistically significant level between semantic representation and surface structure: the stage of representation directly before the postcyclic rules start applying would be such a level, in that that stage would mark a clear break between representations terminating in semantic items and representations terminating in lexical items. However, it would be far from being a breaking point such as Chomsky's 'deep structure' was supposed to be, since (if Lakoff and Ross's current conclusions about a cycle in syntax are correct) a great deal of what have always been considered syntactic rules would apply on the semantic side of that breaking point, e.g. passivization, Equi-NP-deletion, there-insertion, and dative movement.

FOOTNOTES

¹ In the diagrams I capitalize the names of semantic predicates to distinguish them from corresponding lexical items.

² Jerry L. Morgan has called it to my attention that the derivation proposed here for sentences with kill allows an explanation of the ambiguity of John almost killed Harry between 'John almost did something that would have killed Harry', 'John did something that came close to causing Harry to die' (although it didn't affect Harry), and 'John did something that brought Harry close to death': this would be accounted for if the grammar is provided with a prelexical transformation which moves almost into a higher clause.

³ ι is Russell's definite-description operator. A subscript on a verb refers to an event corresponding to that verb.

⁴ Since the structures which I hypothesize to underlie nominalizations are of the form that also underlies relative clauses, it may well be that nominalizations share some of the transformations that are involved in relative clauses: for example, the 'relation marker' may be identifiable with a relative pronoun. However, I

will refrain from taking any position on that question here.

⁵ Jerry L. Morgan has pointed out that prelexical Equi-NP-deletion may explain the difference in meaning between

(22) I asked them when to take off my gloves.

(which has to do with my taking off my gloves) and

(23) I told them when to take off my gloves.

(which has to do with their taking off my gloves).

x asks y Q (with interrogative ask) can be paraphrased as x asks y to tell x Q (with imperative ask). If something of the latter form is in fact the source of interrogative ask, then when to take off my gloves in (22) would be the complement of a verb with me for indirect object, but in (23) it would be the complement of a verb with them for indirect object. Thus in (22) Equi-NP-deletion would apply if the subject of take off is I but would apply in (23) only if the subject of take off is they.

⁶ This idea and those summarized in the remainder of this paragraph are due to George Lakoff.

BIBLIOGRAPHY

- Bach, Emmon. 1968. 'Nouns and noun phrases'. To appear in Bach and Harms 1968
- and Robert T. Harms. 1968. Proceedings of the 1967 Texas Conference on Language Universals. Holt, Rinehart, and Winston, in press.
- Chomsky, Noam. 1965. Aspects of the theory of syntax. M.I.T. Press.
- Gruber, Jeffrey. 1965. Studies in lexical relations. M.I.T. dissertation, unpublished.
- Jackendoff, Ray. 1968. 'An interpretive theory of pronouns and reflexives'. M.I.T., duplicated.
- Lakoff, George. 1965. On the nature of syntactic irregularity. Indiana University dissertation.
- Lees, Robert B. and Edward S. Klima. 'Rules for English pronominalization'. Lg. 39.207-221 (1963).
- McCawley, James D. 1967a. 'Meaning and the description of languages'. Kotoba no uchū, vol. 2, nos. 9(10-18), 10(38-48), and 11(51-57).
- . 1967b. 'The respective downfalls of deep structure and autonomous syntax'. Paper read at LSA meeting.
- . 1968a. 'The role of semantics in grammar'. To appear in Bach and Harms 1968.
- . 1968b. 'Where do noun phrases come from?'. To appear in Jacobs and Rosenbaum, eds., Readings in English Transformational Grammar (Ginn).
- Ross, John Robert. 1966. 'A proposed rule of tree pruning'. Report NSF-17 of Harvard Computation Lab.
- . 1967. 'Gapping and the order of constituents'. To appear in Proceedings of the Xth international congress of linguists.

SOME STRANGE ASPECTS OF IT

Jerry L. Morgan
University of Chicago

It is generally recognized that there exists in English a special use of the pronoun it, usually referred to as "impersonal", as in sentences like (1) through (4).

- (1) It's hard to eat yoghurt in a strong wind without getting your hair in it.¹
- (2) It's raining.
- (3) It's very dark out.
- (4) It's hot in here.

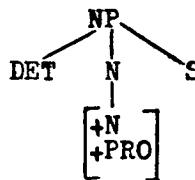
The purpose of this paper is to show first that this lumping together of apparently different uses of it is correct, though to my mind insufficiently recognized and formalized by current transformational accounts, and second that some phenomena directly and indirectly related to it raise some minor and major problems for transformational theory.²

1. On getting it. Otto Jespersen (1965:25) presented an appealing account of the source of this it. According to him, all the above instances are the result of "a grammatical device to make the sentence conform to the type most generally found". That is, in (1) the infinitive is postponed to avoid clumsiness, and (2) through (4) have no real subject. In all four cases, the sentence has no surface subject, so it is inserted to give it one. This analysis has some advantages which have generally been overlooked, which I will return to later.

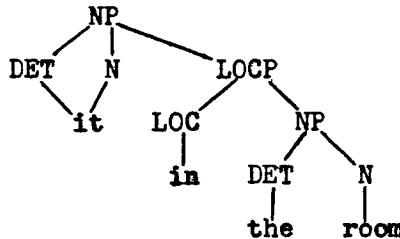
There are two current transformational accounts of the source of it. The first, exemplified jointly in Rosenbaum (1965) and Langendoen (1965), is that the it is the head noun in deep structure, as in (5), and gets deleted if the phrase of which it is the head is not extraposed to the end of the sentence. (6) and (7) are example derivations using this analysis.

(5)

(a)



(b)



(6) base:

[[it] [[John] [came early]]] happened
N NP VP S NP

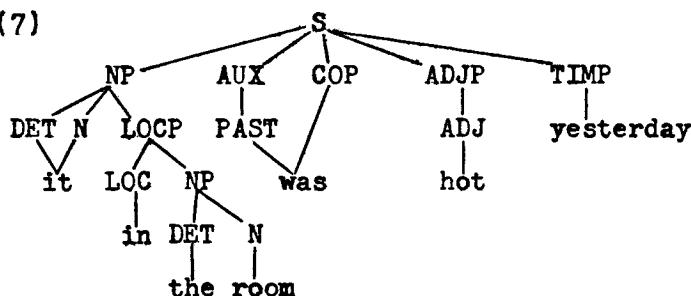
by complementizer placement:

[it] [that [John] [came early]]]happened
NP VP SNP

by extraposition:

[it] happened [that [John] came early]
N NP VP S

(7)



by the rule of locative phrase (LOCP) extraposition:

It was hot in the room yesterday.

Notice that the analyses of Langendoen and Rosenbaum represent the it as head noun of the NP containing the extraposed construction. This seems a rather strained usage of the notion "head noun". As McCawley has pointed out³, the it makes no apparent contribution to the semantic reading of the structure. And if the it were really a head noun, one would expect it to behave like one in regard to relative clauses. But this is not the case, as (8) and (9) show.

- (8) That he is unpopular, which is obvious, doesn't bother John.
- (9) *It, which is obvious, doesn't bother John that he is unpopular.

Also, Rosenbaum proposed that every embedded sentence had an it attached to it in this manner. This seemed to be justifiable, since an it also appears in object position in sentences like (10) and (11).

- (10) John resents it that you have so much money.
- (11) I had expected it that there would be a large turnout.

But Kiparsky and Kiparsky (to appear) have presented strong evidence that the it in these sentences is not the same as the one in subject position, and that they have entirely different sources. Rosenbaum's hypothesis that every embedded S has an attached it therefore seems too strong, since the only place that this it occurs is in subject position.

Postal (1967) has proposed that pronouns are actually surface reflections of what appear at a deeper stage as determiners. Although he does not mention these instances of it, it is possible that his proposal could be interpreted to cover them. Even so, it is doubtful that it would work, since these it's have properties not common to most ordinary pronouns.⁴ These unusual properties become apparent when their behavior in regard to conjunction reduction is examined.

2. On reducing it. The fact that all these apparently different cases of it in subject position are actually the same becomes obvious under conditions of conjunction reduction. It is well known that the rule of conjunction reduction must incorporate, in addition

to criteria of morphological identity, some representation of intended referent, such that reduction is possible only if intended referents are identical. For example, (12) cannot be reduced to (13) unless the two instances of he have the same intended referent.

- (12) He wanted Bill to see the movie but he didn't have any money for tickets.
- (13) He wanted Bill to see the movie but didn't have any money for tickets.

The unusual properties of it in regard to identity appear in such sentences as (14) and (15).

- (14) It is dangerous to believe RFK but obvious that some people do.
- (15) It's easy to respect John but hard to like him.

If this it were an ordinary pronoun, these sentences would be impossible. It appears, then, that Postal's proposal on pronouns cannot be extended to account for this it as a deep determiner without some ad hoc addition. And it is clear here that calling it the "head noun" is unjustifiable, since head nouns, like pronouns, must have some index of intended referent.

But in the case of (14) and (15), the it's are of the same type; that is, they both come as a result of extraposition. Thus it is not surprising that they would be identical in some sense. It might be possible to change Rosenbaum's proposal to say that it is only an empty morpheme, with no index for referent, which is merely left as a marker when extraposition occurs. But this would not account for the fact that the it of Langendoen's rule of locative phrase extraposition is effectively the same as that of Rosenbaum's sentence extraposition, as is shown by (16) and (17).

- (16) The pilot couldn't land his plane at O'Hare because it was dark and snowing and hard to see the runway.
- (17) The pilot couldn't land his plane at O'Hare because it was dark and it was snowing and it was hard to see the runway.

As a matter of fact, Langendoen implicitly recognizes that the two it's are effectively the same, since he posits a single rule to get rid of both. But this does not explain the fact that two rules of English which both move a noun phrase out of subject position both also leave an empty it in place of the noun

phrase, and that the two it's are effectively identical. There is no satisfying way to say that this is anything more than coincidence, since there is no way the two rules can be collapsed. One might try adding a general statement that "rules which permute NP's to the right out of subject position leave it with no referent in place of the NP". But the status of such a statement is not clear, nor is it clear how it would fit into the grammar of English. It is certainly not a universal. Moreover, it is probably not sufficient. The passive transformation, if it exists, moves a NP to the right out of subject position, but leaves no it. Besides, there are other it's effectively identical to these which cannot possibly be the result of extraposition. For example, the first two cases of example (17) would, according to Langendoen, come from locative phrase extraposition, and the third from sentence extraposition. And one could, by stretching things a bit, say that (18) has a locative phrase extraposition whose effects are not immediately apparent from the surface structure. But as far as I can see, no stretch of the imagination would allow the claim that there is any such thing in the derivation of (19).

- (18) When the weather is very hot, it's hard to concentrate and generally unpleasant.
- (19) When you're sick, it's hard to concentrate and generally unpleasant.

Another possibility is that the identity criterion, instead of requiring identical indices, requires only that there be no non-identity of indices. Presumably then, an it with any or no index would match an it with no index to meet the identity criterion. This does not appear to be the case. For example, (20) is bad and (21) is atrocious, at least in the intended sense.

- (20) *When asked about his new-born son, Joe said, "It's a seven-pound boy and hard to believe I'm a father."
- (21) *When asked about his new-born son, Joe said, "It's hard to believe I'm a father and a seven-pound boy."

This reformulation of the identity criterion might be saved by saying that conjunction reduction is blocked because the two it's differ in one or more important features, such as animate or human. Although I have been unable to find conclusive evidence against this, there seems to be a counter-example in (22).

- (22) *The bad thing about smoking cigarettes is that it's a dirty habit and easy to become dependent on it.

from

- (23) The bad thing about smoking cigarettes is that it's a dirty habit and it's easy to become dependent on it.

In this case the two it's would presumably have the same general features. But if the no-referent it had features at all, one would expect it to have quite different ones in such pairs of sentences as those conjoined in (19). And if non-matching features blocked reduction, then (19) would be impossible.

It appears then that the it is totally devoid of any features relevant to identity, and that it has either a unique invariable referent index or none at all, which amounts to the same thing as far as identity criteria are concerned. So we are left with a set of coincidences: that two separate rules, both of which move noun-phrases, leave an empty marker in place of the noun phrase; that constructions such as (24)

- (24) When you have no money it's not very pleasant.

also have an empty marker, even though no movement is involved; that all these markers occur in subject position only; and that the markers in all these cases are morphologically and effectively the same.

But Jespersen's analysis accounts for all these apparent coincidences in a fairly satisfying way. He says in effect that it is a condition of well-formedness for all declarative English sentences that they have an overt subject in normal position, and that where this condition is not met, for whatever reason, an it is inserted. However, a slight modification is necessary. The rule of it-insertion must be formulated so that it applies to all non-imperative sentences, since questions too contain empty it, as in (25).

- (25) Is it really that hard to eat yoghurt in a strong wind without getting your hair in it?

This slightly modified version of Jespersen's analysis has the effect of separating out a single rule of it-insertion, which applies to sentences with no overt

subject. But the rule cannot be interpreted as a surface constraint or output condition. As a matter of fact, it must precede at least the rules which form questions, relatives, and infinitives, and the rule of subject-raising, as is obvious from examples (25) through (28).

- (26) This is the man who it is said invented and marketed the finite transducer.
- (27) John is waiting for it to be announced that RFK has withdrawn his candidacy.
- (28) John wants it to be obvious that he is not Schwartz.

There are some problems here, though, that weaken the argument for a separate single rule of it-insertion. I have found no evidence that any rule must be ordered between sentential or locative phrase extraposition and it-insertion. Such evidence would prove that it-insertion must be a separate rule. A further problem is that if tree-pruning rules are anywhere rules, that is, if they apply immediately whenever their structural descriptions are met, then there must be vestigial NP nodes hanging around to trigger it-insertion, for otherwise the S-node would be pruned and the structural description for it-insertion would never be met. This is no real problem in the two cases of extraposition, since they are usually supposed to have that result anyway, but in the case of sentences which appear to have no real deep subject, this would require an empty NP in deep structure. It is possible, however, that some other analysis of apparently subjectless sentences would produce the proper result.

The modification of Jespersen's analysis has an interesting side effect, in that it explains to some extent the fact that there may be a correlation between those languages which demand an overt subject and those that have a rule analogous to it-insertion. German and English, for example, both require overt subjects, and both have marker insertion rules. Albanian, on the other hand, allows sentences with no overt subject, as in (29).

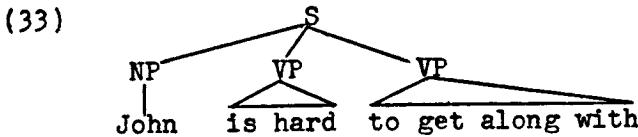
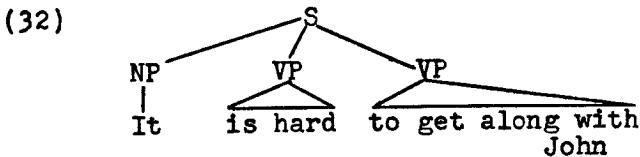
- (29) Kam përfituar nga studimet e dijetëtarëve të ndryshme.
(I) have profited from the studies of several scholars.

And Albanian seems to have a rule similar to sentential extraposition in English, as in (30) and (31).

- (30) Esht e vërtet se jam i lodhur.
Is true that (I) am tired.
(31) Esht e tmerrshme të mos kesh uje për të pire.
Is terrible to not have water to drink.

But where extraposition has applied, no marker is inserted in the vacated position. In fact, any such insertion makes the sentence ungrammatical. I am told that Russian and Romanian, and probably several other languages, resemble Albanian in both not requiring an overt subject and having no rule of marker-insertion. This correlation seems quite natural, and is a candidate for study as a possible universal.

3. On replacing it. Study of this same it leads indirectly to a more serious problem. There is a rule of it-replacement which is proposed to derive sentences such as (33) from (32) by replacing the it by some NP contained in the extraposed clause.



It had been assumed that this rule was a sort of topicalization, with no difference in meaning between, for example, (32) and (33). The rule is appealing in that it accounts for the fact that to every grammatical sentence of type (33) there corresponds a grammatical sentence of type (32), but not vice-versa. The rule of it-replacement also gives the correct derived constituent structure; that is, it explains why (35) and (36) differ in acceptability from (34) in the same way.

- (34) John is eager to go, but Bill is somewhat reluctant to.
(35) ?It's illegal to smoke pot, but it's not dangerous to.

- (36) ?John is easy to please, but Bill is hard to.

But Ross (p. 418) has produced cases which show that the two sentence types have a subtle difference in meaning. His examples are:

- (37) It is hard to play the sonatas on the violin.
- (38) The sonatas are hard to play on the violin.
- (39) The violin is hard to play the sonatas on.

Ross claims that for him (38) and (39) are different in meaning. But a more convincing case for a difference in meaning shows up in (40) and (41), where (41), if it is indeed grammatical, has a meaning quite different from that of (40).

- (40) The crud on my glasses makes it hard to read my notes.
- (41) ?The crud on my glasses makes my notes hard to read.

Also, there is the mysterious fact that (42) is ambiguous for some speakers in the NP all rules; that is whether each and every rule is hard to formulate precisely, or whether the set of all rules is hard to formulate precisely; but (43) does not share this ambiguity.

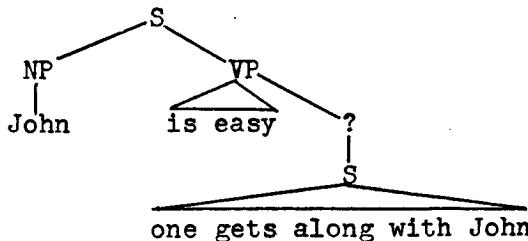
- (42) It is hard to formulate all rules precisely.
- (43) All rules are hard to formulate precisely.

And, since for most dialects of English it is claimed that try requires that its subject and the subject of its complement be identical, (44) appears to be conclusive evidence against it-replacement.

- (44) John tries to be easy to get along with.

Since the idea that transformations can change meaning has been abandoned, at least by anti-lexicalists, Ross is forced to posit a distinctive deep structure, as in (45).

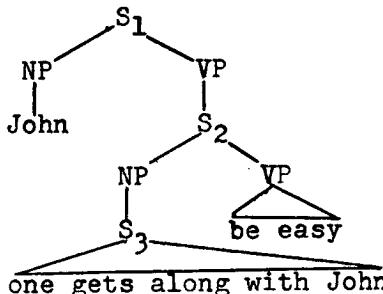
(45)



This deep structure is at first glance reasonable, since it seems that the meaning of such sentences is roughly a predication of a property of the subject; that is, in (45), for example, John has properties such that it is easy to get along with him; or, perhaps, John has the property that it is easy to get along with him. But this is not exactly what Ross' deep structure says. The easy in his structure has an apparently different relation to the embedded sentence from the one in the untopicalized version, and appears to be predicated of John. It is possible that semantic interpretation rules could be patched up to handle this structure, but it seems to me that a very real generality would be missed, in that in both the topicalized and the untopicalized versions of the sentence, what is easy is not John but to get along with him. Also, Ross' structure fails to reflect the relation of inclusion between sentences of type (32) and their topicalized versions. Strong constraints would also have to be placed on the content of the embedded sentence. And his structure would not, as far as I can tell, yield the derived constituent structure necessary to account for the difference in acceptability among (34) through (36).

One might try to rescue the uniqueness of deep structure by modifying the nature of deep structure to allow (46) as the deep structure of "John is easy to get along with".

(46)



In this structure S_2 has a fair amount of independent motivation as the deep structure of "It is easy to get along with John". This analysis more appropriately represents the status of easy, and no additional rules would be necessary to account for the semantic reading of the sentence. If this analysis were adopted, it would seem to move deep structure away from NP & VP to SUBJECT & PREDICATE, since it would be absurd to call a sentence a VP. Deep structure seems to be moving in this direction anyway. But the analysis has some of the same flaws as Ross'. It would require strong constraints on the content of the embedded sentence, and I see no way to get from it the derived constituent structure necessary to account for (34) through (36). Also, both would presumably get rid of the second occurrence of John by the rule of equi-noun phrase deletion. Then I see no way to account for the fact that (47) is grammatical but no topicalized version of it is, as in (48) through (50), even though (51) is acceptable.

- (47) It is hard for John to understand himself.
- (48) *John is hard for himself to understand.
- (49) *John is hard for him to understand.
- (50) *Himself is hard for John to understand.
- (51) Mary is hard for John to understand.

But if topicalized versions were derived through it-replacement, then Postal's cross-over condition (as stated in Ross) would provide a natural explanation for the ungrammaticality of (48) through (50), since it would involve moving a noun phrase over another one co-referential to it.⁵

The only remaining serious objection to it-replacement, other than that it appears to change meaning, is the case of try. But (52) and (53) are grammatical.

- (52) John tried to be arrested.
- (53) Joan tried to get run over.

One can save the hypothesis of identical subjects for try by claiming, probably correctly, that these two sentences contain underlying causatives whose deep subjects satisfy the conditions for try. But it seems to me that (44) has exactly the same status, and that an underlying causative can just as plausibly be posited for it.

Thus it-replacement appears to be the best-moti-

vated rule to account for these phenomena, except for the unpleasant fact that it seems to change meaning. When I took up the subject of it-replacement, it was with the prejudiced intention to find proof that it must be rejected, since it changes meaning. But the opposite conclusion is unavoidable, at least at present. Ross put this rule in hell for that it was a stirrer up of strife. Eccovi! Judge ye! Have I dug it up again?⁶

NOTES

¹I am indebted to Alice Davison for uttering this beautiful sentence in my presence.

²I am grateful to Lloyd Anderson and Georgia Green for enlightening discussions of several points in this paper.

³In course lectures in syntax, 1967.

⁴One possibly related case is the vague general-referent they in sentences such as:

- (i) They grow opium in China, but not in this country.
- (ii) They hate LBJ in North Viet Nam but love him in Texas.
- (iii) They panned the play in Boston but loved it in New York.

⁵I owe this observation to James D. McCawley.

⁶My apologies to Ezra Pound.

REFERENCES

- Jespersen, Otto; 1965. The Philosophy of Grammar,
New York.
- Kiparsky, Paul, and Carol Kiparsky; (to appear). "Fact",
in Bierwisch and Heidolph (eds.),
Recent Advances in Linguistics.
- Langendoen, D. Terence; 1965. "Some Problems Concerning
the English Expletive it",
mimeographed.
- Postal, Paul M.: 1967. On So-called Pronouns in
English, Georgetown University Mono-
graph Series in Languages and Ling-
uistics, number 19.
- Rosenbaum, Peter Steven; 1965. The Grammar of English
Predicate Complement Constructions,
PhD thesis, MIT.
- Ross, John R.; 1967. Constraints on Variables in
Syntax, PhD thesis, MIT

Naturalness Arguments in Syntax

Arnold M. Zwicky, Jr.
(University of Illinois, Urbana)

1. Considerable effort has been devoted, explicitly by generative phonologists and implicitly by many other linguists, to the elucidation of the notions natural class and natural rule in phonological theory. Thus, a set of distinctive features is chosen so that the formal simplicity of the specification for a class of segments corresponds to the naturalness of that class (cf. Halle 1961 and Halle 1962); as an extreme example, the class {p, t, k, b, d, g} is much more natural than the class {p, i, l, ɛ, x, h}—that is, the segments in the former class, but not those in the latter, can be expected to function together in undergoing (or conditioning) morphophonemic processes and to function together in undergoing (or conditioning) phonological change through time—and any system of distinctive features must reflect this fact. Further notational conventions are required to make the formal simplicity of rules correspond to their naturalness. Another extreme example: a rule shifting t to d, p to b, and k to g in position before voiced obstruents is more natural than one shifting t to b, p to g, and k to d in that position, even though both rules refer to the same class of segments. And again: a rule that makes all obstruents voiced before voiced obstruents and voiceless before voiceless obstruents is more natural than one that makes all obstruents voiced before high vowels and voiceless before nasals, even though the classes of segments involved in the two rules are of approximately the same order of naturalness. It is, however, well known that the careful choice of a universal set of distinctive features together with a judicious selection of such notations as parentheses, brackets, subscripts and superscripts, and variables over + and - is far from sufficient for the purposes of distinguishing relatively natural rules from relatively unnatural ones. For example: the triangular vowel system {i, u, a} is quite natural, while the system {ü, i, x̣} is not at all natural, even though the two systems are isomorphic with respect to the featural contrasts and redundancies they exhibit. Further: a rule dropping word-initial h is quite natural, while a rule dropping word-initial th is much less so, even though the standard practice of generative phonology makes little or no distinction between the two rules. Such clear examples are by no means isolated. They indicate the need for major revisions in phonological theory, revisions which have been undertaken by a number of investigators (e.g., Chomsky and Halle 1968 and David Stampe in recent, still unpublished, work) who propose the revival, in some form, of the Prague School notion of markedness (i.e., relative unnaturalness).

In syntax, the notions of natural class and natural rule are frequently employed in arguments for and against particular analyses, but the appeal to naturalness is rarely made explicit.

In what follows we shall examine several types of arguments in which considerations of naturalness are significant, and we shall show that just as the attempt to achieve naturalness in phonological descriptions indicates the need for stronger (i.e., more restrictive) hypotheses about the content of phonological theory, so the attempt to achieve naturalness in syntactic descriptions indicates the need for a strengthening of syntactic theory.

2. Let us first consider arguments referring to natural classes. One criticism that can be levelled against many syntactic descriptions is that although the description indicates (by means of a list or, equivalently, a syntactic feature) that certain items constitute a class, C (either by having similar distributional restrictions, or by being mentioned together in some transformational rule, or both), no explanation is provided as to why these particular items, and not some other (perfectly arbitrary) collection of forms, should class together. The criticism is especially compelling when C must be referred to at two or more places in the description, or when C is a semantic class, just as the claim that a phonological class P is natural is supported by the recurrence of P in the description of a particular language (and in linguistic descriptions in general), or by the character of P as a phonetic class. In any case, the point of the criticism is that linguistic theory should provide a descriptive device for the specification of C, at least to the extent that it should be simpler to refer to C than to refer to a class C' of the same size as C but composed of randomly chosen items.

A criticism of the type just described is the one advanced in Ross 1967a against the analysis of the English verbal auxiliaries first presented in Chomsky 1957. In this analysis the sequence

$$\text{TNU}(\left\{ \begin{array}{l} \text{have} \\ \text{be} \\ \text{M} \end{array} \right\}) ,$$

where TNU is the marker of tense and number and where M represents the class of modals, must be mentioned in a number of transformations--the interrogative inversion rule (Why are you happy? Did he call? How should we go? Has she been hurt?), tag question formation (You are happy, aren't you? He called, didn't he? We shouldn't go, should we? She's been hurt, hasn't she?), one or more rules for placement of adverbs (You are certainly happy. He didn't call. We should soon go. She has never been hurt.), one or more rules of truncation (You are happy, and so is Marvin. He didn't call, and neither did Francesca. We should go, and you should, too. She hasn't been hurt, and I haven't, either.), and a quantifier placement rule (You are all happy. We have both been hurt.). Although it might be possible to reduce the number of references to the

sequence, it seems most unlikely that any amount of ingenuity can reduce this number to less than three (interrogation, adverb placement, truncation). Following Ross, we observe that the parts of the sequence do not necessarily form a complete constituent (in fact, sometimes TNU and be are not even immediately dominated by the same node, except in surface structure; this is the case in John is happy.) and that the formula for the sequence is just as simple as the following absurd formula:

$$\text{PREP} \left(\begin{cases} \text{fish} \\ \text{the} \\ \text{TNU} \end{cases} \right) .$$

That is to say, the usual analysis of the English auxiliary does not treat the sequence as composed of a natural collection of elements, and a new analysis is called for. An interesting point about this conclusion is that it might very well be in conflict with considerations of simplicity, although in the absence of anything but the grossest indication of an evaluation metric for syntax it is difficult to determine whether or not a conflict does in fact exist. At any rate, the analysis which Ross suggests requires at least one new transformation (in exchange for a saving of a few symbols in each of several other rules), and we may conclude that the issue is certainly not clear-cut, especially in view of the fact that the extent of our detailed, systematic knowledge about the syntax of any language is so slight that the consequences of a choice of one partial analysis over another (even when the first appears clearly to be simpler than the second) are incalculable, so that one can never be certain that a simplification at one point in a description is not canceled out by a complication at some other point. In the case at hand, considerations of naturalness presumably outweigh all simplicity arguments except those based on considerable independent evidence.

Another example: the association of English verbs with various types of complements is by no means random, but rather is clearly related to the meaning of the verbs. The assumption of "structuralist" grammar is that this correlation is an accident—an historical accident, perhaps, but nevertheless an accident. To aspire to naturalness in syntactic descriptions is to take such correlations as the paradigm and (in the absence of other evidence) to class deviations from a regular association as exceptions. Thus the English verbs which occur with a marked infinitive in their complements (for example, persuade, want, and plan) are contrasted with the verbs that occur with the present participle (for example, find, imagine, and avoid), despite the considerable formal similarity between the two classes noted in Chomsky 1958 and elaborated in Rosenbaum 1967,

in that the verbs in the former class (We persuaded him to run for the Senate., Eloise wanted to drown the goldfish., Adolph planned to rule the world.) refer to a time preceding the (not necessarily realized) state described by the complement, while the verbs in the latter class (We found Romulus and Remus building a city., Susan imagined Paul wearing a turtleneck sweater., Delilah couldn't avoid cutting Samson's hair.) do not imply such a sequence. Some verbs may have both senses and hence may occur with both types of complements; the result is sometimes a subtle distinction, as in I hate to wash dishes. (my anticipatory loathing is with me always) as opposed to I hate washing dishes. (my loathing accompanies the act, or at least is greatest during the act), sometimes a very clear one, as in She tried to wash her hair in beer. (but didn't succeed) as opposed to She tried washing her hair in beer. (but couldn't stand the resultant stickiness).

To mark (say, by a feature indicating that a transformation may, or must, apply) these verbs occurring with marked-infinitive complements is to provide an account of the facts, but it is not to explain them. An attempt at explanation requires at the very least that the feature uniting the verbs of this class be supplied by linguistic theory. Otherwise, this group of verbs is no more to be expected to function together than is the set { manufacture, infiltrate, crush, amuse, ascertain, vanish, illuminate, pass, furnish, batten, paint, wane } .

The example is not an isolated one. Indeed, virtually every word class that has received attention in the literature on transformational grammar is characterized by a high degree of semantic coherence. A particularly interesting class, that of verbs occurring with sentential that-complements whose truth is presupposed by the speaker (for example, regret, forget, and grasp, as opposed to suggest, think, and suppose) is discussed in some detail in Kiparsky and Kiparsky 1967. In many instances the semantic coherence of word classes can be strikingly demonstrated by the consideration of nonsense words whose meaning is stipulated to be within the range of some known class of English words. It is possible to make extensive inferences about the syntactic behavior of these words solely on the basis of their meaning, and often one's intuitions about the "expected" versus the "unexpected" properties of the words are quite clear. Thus, C. LeRoy Baker has pointed out to me that given the information that the nonsense word griffle means "move smoothly and quickly in an elliptical orbit," we can infer that the verb griffle can occur not only in such sentences as The beachball griffled with unusual speed. and Many planets griffle about Alpha Centauri.., but also in such causative constructions as My paternal uncle can griffle soccer balls with astonishing skill.. It is not a criticism of such examples to suggest that they illustrate a

human ability (or propensity) for analogy; one must have a basis to analogize upon, and such a basis is precisely what is lacking in a treatment of English syntax that supposes that the verbs of pure motion (e.g., move, rotate, sail, roll, shift, and drop) constitute a class merely by virtue of their distribution, just as this basis is precisely what is lacking in a treatment of English phonology that supposes that the English grave stops (p, b, m, k, g, n) constitute a class merely by virtue of their distribution.

3. A second type of argument concentrates not upon the fact that certain distinct forms function together, but rather upon the fact that certain distinct functions are accomplished by the same form. In this way, descriptions may be criticized because they fail to explain why a given word or a given morphological category (such as a case or tense) should represent a particular combination of uses or senses.

For example, McCawley 1967b criticizes the treatment of English "empty it" in Rosenbaum 1967 on the grounds that Rosenbaum's analysis (itself an improvement over the earlier analysis in Lees 1960) of such sentences as It seemed to me that she was stark naked, fails to provide any explanation for the occurrence of the neuter pronoun it, rather than (for example) the pro-verb do or the neuter noun mizzenmast, as the representative of the extraposed clause. McCawley points out, quite correctly, that an explanation for it is available only on the basis of the (or a) rule of English definite pronominalization, but that the required "backwards pronominalization" rule simply cannot apply, given present knowledge about the nature of the conditions on pronominalization (Ross 1967b, chapter 5, and Langacker 1967). McCawley suggests instead a radical revision of syntactic theory along lines suggested in Bach 1967. However, it is argued in Kiparsky and Kiparsky 1967 that McCawley's attempt to unify the "empty" it of extraposition with the "factive" it of I hated it that Kermit kept contradicting Margo, is ill-advised, on the grounds that the latter (but not the former) it blocks the formation of relative clauses and questions (What did it seem to Hera that Zeus had accomplished? as opposed to *What did Hera hate it that Zeus had accomplished?); see the discussion of constraints on reordering transformations in Ross 1967b, chapter 4.

The force of the Kiparskys' criticism is that however desirable it might be to press for one sort of natural analysis of English, in which several closely related uses of the pronoun it are explained as resulting from the application of the same rule, there are nevertheless arguments against this proposal--arguments which, moreover, are weighty principally because they directly involve efforts to construct a consistent description embodying significant generalizations about English.

Criticisms might also be made of standard treatments of the Latin genitive case, or the English "possessive," because the natural unity of the inalienable possessive (my arm), alienable possessive (my chair), subjective genitive (my invention), objective genitive (my trial), and partitive genitive (half of me), among other uses, is not captured. Typical discussions of English transformational grammar call for the generation of a possessive marker at half a dozen or more places (see, for example, Lees 1960); the adjunction of the possessive marker in all of these instances is, however, no simpler than would be the adjunction of [Ep] in one situation, [AwD] in another, [I1] in a third, [S] in a fourth, and so on--or the generation of the noun plural suffix in one case, the past tense suffix in another, the agentive suffix in a third, the repetitive prefix re- in a fourth, and so on. In this case, as in the previous example, there is some evidence (see Chomsky 1967) to indicate that any attempt to unite certain of the senses of the English possessive, or the Latin genitive, is bound to fail, on the grounds that the rules introducing the appropriate case marking must be separated from each other. Similar remarks apply to the problem of providing a uniform explanation for all, or even many, of the uses of such words as as, so, but, any, which, to, or, or even, or even want, keep, tell, or see.

Valid examples of unity of senses or uses can perhaps be appreciated only in the context of an invalid example of unity. Thus, compare, on the one hand, the natural unity of the genitive of alienable possession and the subjective genitive appearing in nominalizations with, on the other hand, the lack of natural cohesiveness in the senses of the verb decline, according to which one can be said to decline either a regular noun or a dry martini.

4. A third type of appeal to naturalness does not necessarily make reference to classes of items which have a single function or to classes of functions which are manifested as a single item, but rather refers to properties which belong to rules themselves (instead of merely to the classes mentioned by rules or created by them). Thus, the formation of relative clauses and the formation of interrogatives have long been considered to be associated with each other in some special fashion because they share certain properties: they both move a specified noun phrase over a potentially infinite amount of intervening material and to the left end of a clause, they both create "wh-words," and they both are subject to a number of special restrictions (although one noun phrase in a series of conjoined noun phrases may be pronominalized--Frederick said Petula had seen Herbert and him.--such a noun phrase cannot be relativized or questioned--*Who did Frederick say Petula had seen Herbert and?, *I slapped the man that Frederick said Petula had seen and him.; for further details see Ross 1967b). That is, it has been assumed that the association of relatives and interrogatives is natural and that the similarities in the form

of the required movement transformation and in certain aspects of the products of the transformation are to be taken as evidence for this natural association.

However, arguments have recently been advanced (in Ross 1967b) that relative clause formation and interrogative formation must be considered as two rules, having different structural descriptions and possibly even ordered at different points in the grammar of English. Ross has argued further that the fact that the two rules are restricted in the same ways can be explained as resulting from the fact that all rules which move specified constituents over variables are so constrained. In other words, the similarity in restrictions is deducible from a partial similarity in structural descriptions and structural changes. The other property uniting relative clauses and questions, their characteristic "wh-words," remains unexplained under the re-analysis suggested by Ross.

Another example: the rule creating do so as a substitute for verb phrases has some properties in common with definite pronominalization (compare You may scream, if you wish to do so., *You may do so, if you wish to scream., If you wish to scream, you may do so., and If you wish to do so, you may scream. with Jerry criticized the woman who interviewed him., *He criticized the woman who interviewed Jerry., The woman who interviewed Jerry was criticized by him., The woman who interviewed him was criticized by Jerry.), but some indication is given in Bouton 1968 that the two rules cannot possibly be considered to be the same rule, but must rather be ordered at different points in the grammar, with at least the passivization rule intervening between the do-so rule and the rule of pronominalization, at least in dialects in which The girl was shot in Boston last night by the same man who had done so earlier in New York. is fully grammatical. As a result, no explanation is provided for the fact that the two rules are subject to the same constraints--having digested Bouton's examples, we are unable to express the natural relation between the rules.

5. For a final set of instances in which rule naturalness is referred to, we return to phonology. A phonological rule in which the conditioning segment is always three syllables away from the affected segment is much less natural than a rule in which the conditioning segment is two syllables away, or one syllable away, or an indefinite number of syllables away (i.e., anywhere within the same word as the affected segment). Also, a rule which affects every other syllable in a word is more natural than one which affects every third, or fourth, syllable.

A syntactic analogue is the (presumable) restriction that the structural description of a transformational rule may mention constituents in two structures one of which is directly embedded within the other or one of which is embedded indefinitely far down within the other, but not in two structures one

of which is always two (or three or four) levels below the other. This distinction in rule naturalness is not captured in current theories of transformational grammar, although many other similar restrictions have been incorporated into theory by means of notational conventions.

On the whole the present theory of syntax (roughly that of Chomsky 1965) is not particularly suited to the natural expression of natural classes and rules. The obvious devices for capturing naturalness lead typically to inconsistent analyses. Thus, one simple way to provide an explanation for the fact that a given item W appears naturally in several different uses is to postulate that W occurs in the underlying structures for the constructions with which these uses are associated. This is the tack taken by McCawley in his discussion of it mentioned above, and also in his more recent treatment of respectively, in which he maintains that he would "regard the addition to linguistic theory of ad-hoc definitions of functions by listing their values a much cheaper price to pay than the alternative of treating respective and respectively as unrelated despite their semantic and morphological identity and their complementary distribution" (McCawley 1967a), but such an approach seems to lead inevitably to contradictions in rule ordering, as we observed above. The type of example mentioned in section 3 is particularly difficult to analyze consistently and appears to require new principles of syntactic theory for its explanation.

An adequate account of natural syntactic classes (section 2) requires more than the assignment of a feature to all members of the class, this assignment being in general the equivalent of a list rather than an explanation for the membership of the class. Two lines of approach suggest themselves. First, it may be possible to press the correlation of semantic and syntactic classes to the conclusion that a syntactic class is no more than a semantic class minus or plus small lists of exceptions. In other words, some natural syntactic classes might be referable to semantic classes in the same way that some phonological classes might be referable to phonetic classes. Second, if a syntactic feature is found (on independent grounds) to play some significant role within linguistic theory, then the assignment by rule of this feature to certain items is not arbitrary or unnatural. Hence, if it is possible to characterize a "rule of pronominalization" as one which assigns the feature PRO, and if there are universal constraints on rules of pronominalization, then the appearance of the same set of pronouns as the products of several different rules (all "rules of pronominalization") would no longer be treated as entirely accidental but would be, to some extent, explained.

BIBLIOGRAPHY

- Bach, Emmon. 1967. Nouns and nounphrases. Unpublished ms.
- Bouton, Lawrence. 1968. Do-so revisited. Unpublished ms.
- Chomsky, Noam. 1957. Syntactic structures. The Hague, Mouton.
- Chomsky, Noam. 1958. A transformational approach to syntax. Proceedings of the third Texas conference on problems of linguistic analysis in English, ed. by Archibald A. Hill, pp. 124-58. Austin, Tex., the Univ. of Texas.
- Chomsky, Noam. 1965. Aspects of the theory of syntax. Cambridge, Mass., MIT Press.
- Chomsky, Noam. 1967. Remarks on nominalizations. Unpublished ms.
- Chomsky, Noam, and Halle, Morris. 1968. The sound pattern of English. New York, Harper and Row.
- Halle, Morris. 1961. On the role of simplicity in linguistic descriptions. Structure of language and its mathematical aspects, ed. by Roman Jakobson, pp. 89-94. (Proceedings of Symposia in Applied Mathematics, 12) Providence, R.I., American Mathematical Society.
- Halle, Morris. 1962. Phonology in generative grammar. Word 18.54-72.
- Kiparsky, Paul, and Kiparsky, Carol. 1967. Fact. Unpublished ms.
- Langacker, Ronald. 1966. On pronominalization and the chain of command. Unpublished ms.
- Lees, Robert B. 1960. The grammar of English nominalizations. (Indiana University, Research Center in Anthropology, Folklore, and Linguistics, Publication 12) Bloomington, Ind.
- McCawley, James. 1967a. The respective downfalls of deep structure and autonomous syntax. Paper delivered before LSA.
- McCawley, James. 1967b. Where do NPs come from? Unpublished ms.
- Rosenbaum, Peter. 1967. The grammar of English predicate complement constructions. Cambridge, Mass., MIT Press.
- Ross, John Robert. 1967a. Auxiliaries as main verbs. Unpublished ms.
- Ross, John Robert. 1967b. Constraints on variables in syntax. Ph.D. dissertation, MIT.

Acculturation as a Late Rule
 Eric P. Hamp
 (University of Chicago)

Albanian typically, and in all known dialects, has no geminate or long consonants.¹ Rearticulation of consonants occurs in a few dialects² where an intervening unstressed schwa between two like consonants is deleted

¹I am ignoring here Italo-Albanian dialects in which the bilingual speakers borrow heavily from Italian and render morphologically adapted loans with complete local Italian phonetics, including the Italian long consonantisms in their correct place in each lexical entry.

Formally, I assume that such bilinguals have essentially a single grammar which differs from that of monolinguals largely in duplicating the content of certain rules, or in containing certain additional rules; and in duplicating or adding to the repertory of lexical entries in like fashion. The so-called 'duplex' lexical entries are marked by two (or more for trilinguals, etc.) sets of phonological features, each indexed by e.g. [+Ital] or [+Alb], but normally by but one set of semantic markings. Thus, one of the sets of phonological markings in an entry may pass through phonetic rules (e.g. Italian style gemination) that fail systematically to apply to the other set. Analogous behavior applies grosso modo to the 'duplex' or bipartite rules.

From our experience with such communities it seems intuitively that duplex lexical entries are much cheaper (i.e. more usual) than duplex syntactic rules. Duplicate semantic markings seem rather rare. This may support the contention that syntax and semantics are fundamentally one.

I consider such duplex indexing in a single grammar to be the normal state of natural bilingualism.

In this framework, a major step of acculturation may be defined formally as the suppression of duplex entries and rules by the deletion of duplex indexing.

It is necessary to clarify this type of situation in order to establish the fact that I am here concerned in the subject of this paper with the situation where no duplex marking is involved. That such a situation of duplex indexing formerly applied to the phonetic phenomena here in question is scarcely to be doubted.

²Vaccarizzo Albanese (Prov. Cosenza) and S. Nicola dell' Alto (Prov. Catanzaro), for instance).

by rule. What I write in this paper with geminate symbols is a phonetically long consonant, not rearticulated.

Words in Albanian may end in either consonants or vowels; in the latter case what is written -ë, and in many dialects may serve largely to preserve voicing or explosion in the preceding consonant or to lengthen the preceding vocalism, in conservative dialects appears in the audible phonetics as a sort of schwa. A few years ago in an oral paper (presented at the spring Foreign Language Conference, University of Kentucky, Lexington, Ky., 1965) I showed that this final schwa (and others) must be assumed for the underlying forms of all dialects of Albanian.³

The village of San Marzano, in Provincia Taranto, is the only Albanophone settlement left in Apulia; two centuries ago there were a dozen or more such villages in the Taranto region. In point of fact, when I first learned of San Marzano I did not yet know whether the language still survived there.

The total publication on the dialect of San Marzano amounts to four items--curiously enough, a broader spread of treatment than for any other variety of Italo-Albanian to date. These four articles were written in the late 19th century by the professional academic philologists Gustav Meyer and J. Hanusz (one each) and by the gentleman-scholar Prince Louis-Lucien Bonaparte (two articles). Bonaparte depended on Italianate transcriptions which he commissioned a local untrained native speaker to write down for him. I have shown in a paper to be published as part of a recent Newberry Library conference on the history of linguistics that Bonaparte's is the most accurate and faithful record.

The dialect of San Marzano, alone among Albanian dialects, shows both phonetic geminates and absence of final consonants adjacent to stressed vowels. Inherited consonant finals now have final phonetic schwa. The effect is obviously Italian in appearance, and must surely be a result of acculturation. (This is not the only, though certainly the most remarkable, phonetic effect of acculturation in the dialect.)

It is our task here to analyze and consider the workings of this effect.

On etymological grounds alone one quickly sees that

³I have personally heard all forms of Albanian dialects which can reasonably exist, except for those of Molise, in Italy, and near Melitopolin the Ukraine.

to words in final schwa (as well as with any following vowel) after stressed vowel in common Albanian the forms of San Marzano respond with geminate (long) consonantisms:⁴

(1)	bukë	:	SM	bukke	'bread'
(2)	hekur	:	SM	hekkur	'hoe' (<'iron')
(3)	bašë	:	SM	balle	'forehead'
(4)	bënë	:	SM	benne	'done, ripe'
(5)	buzë	:	SM	buzzi	'lip'
(6)	khiči	:	SM	khičči	'the key'

Likewise there are many examples (for more see below) showing that to final consonant(ism) after stressed vowel in common Albanian the forms of San Marzano respond with an additional final schwa:

(7)	bark	:	SM	barkë	'belly'
(8)	këpuc	:	SM	kupuce	'shoe'

But these correlations affect not only etymological correspondences; they involve synchronic alternations in important parts of the present-day dialect. Notice the following sets, which involve just the two correlations already observed:

(9)	veš	:	SM	vešë	'ear'
	veši (def.)	:	SM	vešši	'the ear'
(10)	dit (pl.)	:	SM	dite	'days'
	ditë (sg.)	:	SM	ditte	'day'

The following set is pertinent in the same way to our problem even though it is historically not an exact match as the others above have been seen to be:

(11)**kaλ	'kalë'	:	SM	kaλe	'horse'
	kaλi (def.)	:	SM	kaλhi	'the horse'

Here the historically justified form is kalë, which would lead us to expect SM **kaλλe. The fact that we find the result we do is testimony to the regularity of the phenomenon we are inspecting, and a justification of the rules proposed to account for it; and shows that the true regularity is not to be found in the mere matching of surface

⁴For purposes of clarity I am here spelling the common Albanian forms with slight, but easily interpreted, departures from standard orthography.

shapes.

Further grammatically related sets are the following:

- | | | |
|-------|------------------------------------|-----------------------|
| (12) | i-maǵ {m.} : SM i-made | 'big' |
| | e-madde {f.} : SM e-madde | |
| (13) | i-ǵig {m.} : SM i-ǵige | 'ugly' (<'bad,etc.') |
| | e-ǵigë {f.} : SM e-ǵigge | |
| (14a) | pǵak | SM pǵake 'old man' |
| (14b) | pǵek(+e)(pl.): | SM pǵeke |
| (14c) | pǵake {f.} : SM pǵakkë 'old woman' | |
| (14d) | pǵake {f.pl.}: SM pǵakke | |

The -e in (14b) must be analogical after that in (14d), a further testimony to the paradigmatic (though limited) relation of these two nouns. The forms of (13) have the phonetic effect of sounding inverse in SM to the normal values of many varieties of Albanian, so far as the impressionistic length of the stressed vowels is concerned.

In order to relate these obviously related forms two rules are needed; sets (10) and (13) illustrate this strikingly. We need an environment to explain the consonant lengthening; the simplest environment to assume is that of following vowel. We also need a rule to add the final schwa; yet this final schwa cannot be present when the lengthening rule applies, or else we would not observe a single (short) consonant preceding it. The underlying absence of some final schwas is motivated by other considerations. For example, in

- | | | | |
|------|----------------|----------|-----------|
| (15) | vaǵ | SM vaǵə | 'oil' |
| | vaǵ-të (def.): | SM varte | 'the oil' |

the two SM forms can be united only by assuming underlying final -ǵ, which on addition of -të becomes r by rule.⁵

Therefore the two rules proposed must be in the following order:

- consonant lengthening before vowel when consonant is next to stress;
- addition of e to final consonant next to stress.

There are certain apparent exceptions to these two

⁵ Thus, this historical rule is still synchronic for SM. I do not yet see where it is to be ordered with respect to the other rules discussed, and therefore omit it in the ensuing discussion.

rules. First, note that the Albanian 'strong' r [r̩] behaves indifferently long; i.e. under doubling it gives long ('geminate') r, while after schwa-addition the same long r is heard. Thus:

(16) mor <u>̄</u>	:	SM morre	'louse'
(17) bur <u>̄</u> e	:	SM burre	'man' ⁶

It is as if r̄ is already long and cannot be further lengthened. On the other hand, 'weak' r also behaves indifferently, but 'short'; i.e. after schwa-addition we have simply [r̩], while under doubling the same short [r] results:

(18) ar	:	SM are	'gold'
(19) bore	:	SM bore	'snow'
(20) Lëkur <u>̄</u> e	:	SM Lëkkure	'skin'

We see immediately that the situation is not simply the Italian one, whereby some forms simply have as part of their idiosyncratic shape a long consonant, and others a short one. Otherwise, regardless of the behavior of the longs, the short instances should lengthen under (SM Albanian) rule. Clearly, SM still has, like many conservative varieties of Albanian, two kinds of r. Since the relevant context for lengthening, in formal terms, will reasonably contain a feature [-voc] in the segment to be lengthened (especially, e.g. in the obstruents), we can confer a natural distinction to these two varieties of r by characterizing them in their underlying form: /r̩/ [-voc], but /r/ [+voc]. In this fashion, rule (a) will apply vacuously to /r/. A later rule, probably before (b), will then adjust /r/ to [-voc]; before such a rule, and therefore probably a part of (a) all other instances of /r̩/ should be lengthened. Thus a normal phonetic consonantal output for all instances of surface SM r and rr will result without weakening the rules which are clearly justified.

At this point we note one anomalous set in terms of the conclusions just reached:

(21)**zjar	:	SM zjare	'fire'
zjar(ë)(pl.)	:	SM zjarri	(def.)

Here we have a case of r which appears to undergo rule (a). The historical explanation is that the noun, generally anomalous in most dialects, once had a different

⁶ Because bur̄e is masculine, in principle this could be regarded, like (11), as unetymological loss of -ë. But other examples bear this out.

singular (*zjarm probably), and that the definite form must be based on the old plural, which actually had [r̚]. Synchronously, we simply mark this lexeme so that lengthening applies only before vowel. This may mean that phonetic lengthening of /r/ in all positions is not a part of rule (a) suggested above.

A similar situation to that of /r/ and /r̚/ holds also for such forms as

(22) għuha	:	SM għuġa	'the tongue'
------------	---	----------	--------------

Note here that g, when from h, is not lengthened. Thus an underlying h persists in SM. We will mark it the same as all velar segments (which is harmonious with the change h > f before u in SM), except that it is [+voc]. Then a rule after (a) adjusts the features of /h/ to [g].

Note also that in

(23) dru+te	:	SM drute	'the wood'
-------------	---	----------	------------

as opposed to

(24) ðiete	:	SM tiéttə	'ten'
------------	---	-----------	-------

the lengthening does not take place after morpheme boundary.

Clusters introduce further considerations:

(25) huðrə	:	SM fuddrə	'garlic'
------------	---	-----------	----------

with lengthening before r; hence these cases of r are [+voc]. But

(26) i-barðə	:	SM i-barde	'white'
(27) hundə	:	SM fundə	'nose'
(28) mockə	:	SM mockə	'a bit'

where the preceding [-voc] (including these instances of r!) blocks lengthening.

Finally, another special situation obtains for

(29) dēt	:	SM dete	'sea'
dēti (def.)	:	SM deti	'the sea'

Here, in this word (with an unusual vocalism in most dialects of Albanian), the preceding length (which is conveniently assumed on the suggestion from other dialects) blocks (or reverses) consonant lengthening. Reversal seems the easiest assumption to formulate.

Note, additionally, that *v* and *j*, even if derived by rule from underlying /u/ and /i/ (which is likely), become [+voc], since they undergo rule (a):

(30) **k <u>ħ</u> eva	:	SM kħevva	'I was'
ujē	:	SM ujjē	'water'

We now formulate the rules:

1. [+voc] → [-voc] / $\begin{bmatrix} +\text{cons} \\ +\text{diff} \end{bmatrix}$

This expresses the fact that *i,u* → *j,v*. But it may require reformulation if it really forms part of the syllabification rule [-cons] → [+cons], i.e. *i,u* — *j,v*.

2. [-length] → [+length] / [+voc] $\begin{bmatrix} -\text{voc} \\ +\text{voc} \end{bmatrix}$

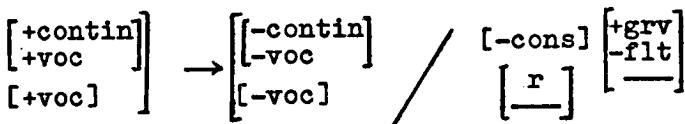
This is rule (a), the consonant lengthening rule in vocalic surroundings. One of the [+voc] segments must also be [+accent].

- 2bis. [-length] → [+length] / $\begin{bmatrix} r \\ -\text{voc} \end{bmatrix}$

This gives length in all positions to /r/.

3. [+length] → [-length] / [+length] —

This length dissimilation rule gives a short consonant in déti. It may require reordering to capture greater generality in conjunction with rule (5) below.



Now that rule (2) has operated, /h/ becomes [g] and

/r/ receives its consonantal non-vocalic characterization. Voicing is not mentioned for /h/; this explains the fact that the medial reflex is heard now as [g], now (more rarely) as [k]. A phonemic view complicates, and fails to explicate, this latter vacillation.

$$4. \quad Q \rightarrow \begin{bmatrix} [-\text{cons}] \\ [+ \text{voc}] \\ [-\text{comp}] \\ [+ \text{grv}] \\ [-\text{flt}] \end{bmatrix} \quad / \quad [+ \text{cons}] \quad - \#$$

This is rule (b), inserting [ə] ~ [ɛ] after underlying final consonants.

$$5. \quad \left. \begin{bmatrix} [-\text{cons}] \\ [+ \text{nas}] \end{bmatrix} \right\} \rightarrow [+ \text{length}] \quad / \quad \left[\begin{array}{c} [+ \text{cons}] \\ [+ \text{accent}] \end{array} \right] \left[\begin{array}{c} [-\text{length}] \\ [+ \text{voc}] \end{array} \right]$$

This inserts the Italian-style 'drawl' on stressed vowels before short consonants. Initial nasals in clusters are similarly affected.

We have shown that the underlying forms for SM are those of general Albanian, in the main, and that the observed striking phonetic result is generated by rule. A small number of rules covers a broad territory of phonetic behaviour; and considerably saves in specifications for lexical entries and grammatical rule content. The synchronic lateness of these rules, which seems to reflect acculturation phenomena, is an interesting and noteworthy characteristic.

Kiparsky has shown (*Languages* 1.73-93) that historically added rules are not always added at the end. It remains to be seen whether the above rules are all equally late en bloc. It seems likely that (2bis) was in the language before (2) was introduced, and before (4).

**A Stratificational Restatement of a Problem
in Manyika Phonology**

Earl M. Herrick
(Michigan State University)

Manyika is a Bantu language, one of the languages of the Shona group, and is spoken in Southern Rhodesia in the area around Umtali. It is number S.13 in Guthrie's classification of the Bantu languages as given in Bryan (1959:144).

Manyika has been studied by several linguists, and the facts about its phonology are relatively well known. An interesting descriptive problem in its phonology is presented by the large number of consonantal syllable onsets that occur.¹ Stevick (1964) has published a description of these onsets in terms of "ultimate phonological constituents", which are the sort of phonological units that Hockett (1947) has called "determining features".² These "ultimate phonological constituents" are very similar in their nature to the phonons of a stratificational description. And although Stevick carefully avoids use of the term "phoneme" in describing these onsets, he uses for them a phonetic broad transcription that can be easily rephonemicized. His description of these syllable onsets therefore gives us the data for a two-level description of a rather complicated and interesting phonological subsystem. This paper will try to show how this part of Manyika phonology can be described as part of a stratificational grammar.

This stratificational description is entirely stated in Figs. 1, 2, and 3. It is composed of two diagrams and one set of lists. The three lists in Fig. 1 state the consonantal phonemes of Manyika, the consonantal phonons of Manyika, and some of the zygmemes which appear in the phonotactics; Fig. 2 states the phonotactics—the allowable combinations of emic units in the phonemic stratum; Fig. 3 gives the phonemic sign pattern—the rules according to which phonemes are realized by phonons or constructions of phonons and by which phonons or constructions of phonons are realizations of phonemes.

¹The 76 consonantal syllable onsets and Stevick's ultimate phonological constituents for them are listed in Fig. 4.

²"The influence of Hockett's thought is evident throughout this paper, and is gratefully acknowledged." (Stevick 1964:58 n.4.)

The diagrams in Figs. 2 and 3 use notation of the geometrical sort that is commonly used for stratificational descriptions. The notation used here does differ from that used in Lamb (1966) in the following details:

1. The brackets for enclosing all emic units are formed by reverse slant lines preceded by subscript Greek letters. The subscript for phonemes is ϕ ; that for phonons is $\phi\nu$.

2. The symbols for emic units are not written alongside the lines of the diagrams, but are written into the lines, so that the lines appear to lead from the emic units of one level to the emic units of the next level.

3. For the sake of convenience in typing and in printing, the lists of phonemes and phonons in the diagrams are arranged vertically rather than horizontally. In Fig. 3, phonemes are on the left and phonons are on the right, so that upward recoding goes from right to left and downward recoding goes from left to right. Because the diagrams have been thus rotated, it has been necessary to write numbers by each ordered node to show the way it is ordered.

4. In the tactic diagram in Fig. 2 and in one of the lists in Fig. 1 there are emic units called zygemes that are written with a subscript ζ . Stratification linguists are quite aware that a tactics for a stratum consists of a hierarchy of ranks, and designations for such ranks have often been written alongside the lines of tactic diagrams (cf. Lamb 1966:24,25,81,84). In the diagrams shown here, these units of rank have been written into the lines of the tactic diagram, in the same way as phonemes and phonons have been written, and they have been given a name and an identifying subscript.

5. In Fig. 2, the phonotactics is said to be in its phonemic form. There is only one tactics for each stratum of a stratification grammar. Within a stratum, however, any occurring group of emes is convertible into a group of ons, and vice versa, by use of the sign pattern. The tactics of a stratum may therefore have two forms, with either emes or ons as its final units. In stratification work published to date, the emic forms of tactics have been used.

6. The phonemic sign pattern, as shown in Fig. 3, contains "upward ordered or" nodes. Nodes of this kind have not usually been used in realization rules, but in this description they play a small but useful role. As can be seen in Fig. 4, many pairs of onsets have the same phonons, except that a certain phonon is present for one onset of the pair but is absent for the other. For example, the onset $\backslash z\backslash$ is realized by $\backslash P1\backslash$, $\backslash Sp\backslash$; its voiceless counterpart

\check{s} is realized by $\wedge P1\backslash$, $\wedge Sp\backslash$, $\wedge Uv\backslash$. In these cases, the upward recoding process presents some difficulties; the rules must not permit the combination $\wedge P1\backslash$, $\wedge Sp\backslash$ to be recoded as both \check{z} and \check{s} . One possible solution is to write separate emic units for all terms of every opposition. This would allow simple realization rules, but would sacrifice the advantages of marked and unmarked terms. Another possible solution is the convention that a recoding must utilize all the emic units present on the known level, and no others. In the example being considered, \check{z} could not recode the combination $\wedge P1\backslash$, $\wedge Sp\backslash$, $\wedge Uv\backslash$ because $\wedge Uv\backslash$ would be left over unre-coded. Unfortunately, use of this convention may result in much searching before the recoding for a group of emic units can be found.

When "upward ordered or" nodes are used in a sign pattern, a rigorous recoding procedure is possible even though some terms are unmarked. All but the last line from the multiple side of the "ordered or" node lead to "and" nodes which are connected to the several marked terms of an opposition. If the recoding cannot proceed through any of these nodes, it then proceeds along the last line. This last line from the "ordered or" node therefore corresponds to the unmarked feature. For downward recoding, these "upward ordered or" nodes have the same function as unordered nodes.

With these six explanations, the lists in Fig. 1 and the diagrams in Figs. 2 and 3, when read according to the ordinary conventions of stratificational geometrical notation, speak for themselves. Except that they are limited to the consonants, they are a complete stratificational description of the phonemic stratum of Manyika.

In addition to the phonemic stratum, the phonology of a language contains both a hypophonemic stratum and a spoken communication channel.³ But this description of Manyika phonology must omit treatment of them, because in the present state of the art a hypophonemic stratum cannot be successfully described for any language. The hypophonemic stratum directly adjoins the spoken communication channel, which is outside language and which embodies the sounds that convey messages from the language. In order that linguistics may successfully describe

³The spoken communication channel is in fact outside language, but it has a linguistic structure provided by the hypophonemic stratum inside language.

this relationship across the phonetic interface of language, the description of the hypophonemic stratum should have a certain simple relationship to the physical measurements used for describing sounds in the spoken communication channel. The correspondence may actually be a one-to-one fit between emic units of the hypophonemic stratum and results found by measuring speech sounds. But whatever the exact relationships, linguists must write their descriptions of hypophonemic strata with acoustics in mind. And, at the present time, acoustics cannot provide linguists with a way to describe speech sounds completely, accurately, and unambiguously. It is the state of the art in acoustics, rather than in linguistics, which now prevents descriptions of hypophonemic strata.

The major difference between this stratificational description and Stevick's description of Manyika phonology⁴ lies in the number of levels of language which they find it necessary to describe. For Stevick, the "ultimate phonological constituents" are the only really valid units of phonology, and everything else is accidental or incidental. He therefore avoids listing or mentioning phonemes. For this stratificational description, both phonemes and phonons are valid units to be described. Stevick states the tactics of the onsets in terms of his constituents. The tactics is given here in terms of phonemes, but if these phonemic tactics were converted to their phononic form they would presumably be equivalent to Stevick's diagrams. The list of phonons given in Fig. 1 is a little longer than that of Stevick's constituents given in Fig. 4. This points out the different measures than economy may have for different theoretical assumptions. If there is only one level in the phonology—whether it is among the phonemes or the phonons or the morphophonemes or elsewhere—that is really important, then it is worthwhile to seek economy at that level, no matter what complications may result at other levels. But in a stratificational description all the levels of language are equally important, and the description must not unduly complicate one level or stratum at the expense of another.

⁴For comparison between the two descriptions, see the parallel listings in Fig. 4. Note that in the columns headed ϕ_v , the phonons for each onset are merely listed; their order does not imply any structure connecting them.

Fig. 1.--Emic units of the phonemic stratum
that occur in syllable onsets

Inventory of phonemes

\ p	t	k	b	d	g	ɓ	ɗ	bw	pf	ڏ	ts	ڏ
ts	ڏ	ts	kx	f	s	s	š	v	z	z	ž	m
n	p	ŋ	w	y	r	l	ڙ	h	h \			

(Note: \ ɓ ɗ \ implosive; \ s z ts ڏ \ whis-
tled; \ bw pf ڏ ts ڏ ts ڏ ts kx \ affricates)

Inventory of phonons

- \ As\ (aspiration)
- \ Cl\ (closure)
- \ Do\ (dorsal articulation)
- \ F1\ (apical flap)
- \ Im\ (implosion)
- \ Lb\ (labial articulation)
- \ Lt\ (lateral release)
- \ Ns\ (nasality)
- \ Pl\ (palatal articulation)
- \ Sp\ (friction)
- \ Uv\ (lack of voicing)
- \ Vz\ (velarization)
- \ Wh\ (whistling)

Inventory of zygmemes for phonemes and phonons
(not necessarily complete)

\syllable\ \onset\ \nucleus\

**Fig. 2.--Phonotactics of syllable onsets
(phonemic form)**

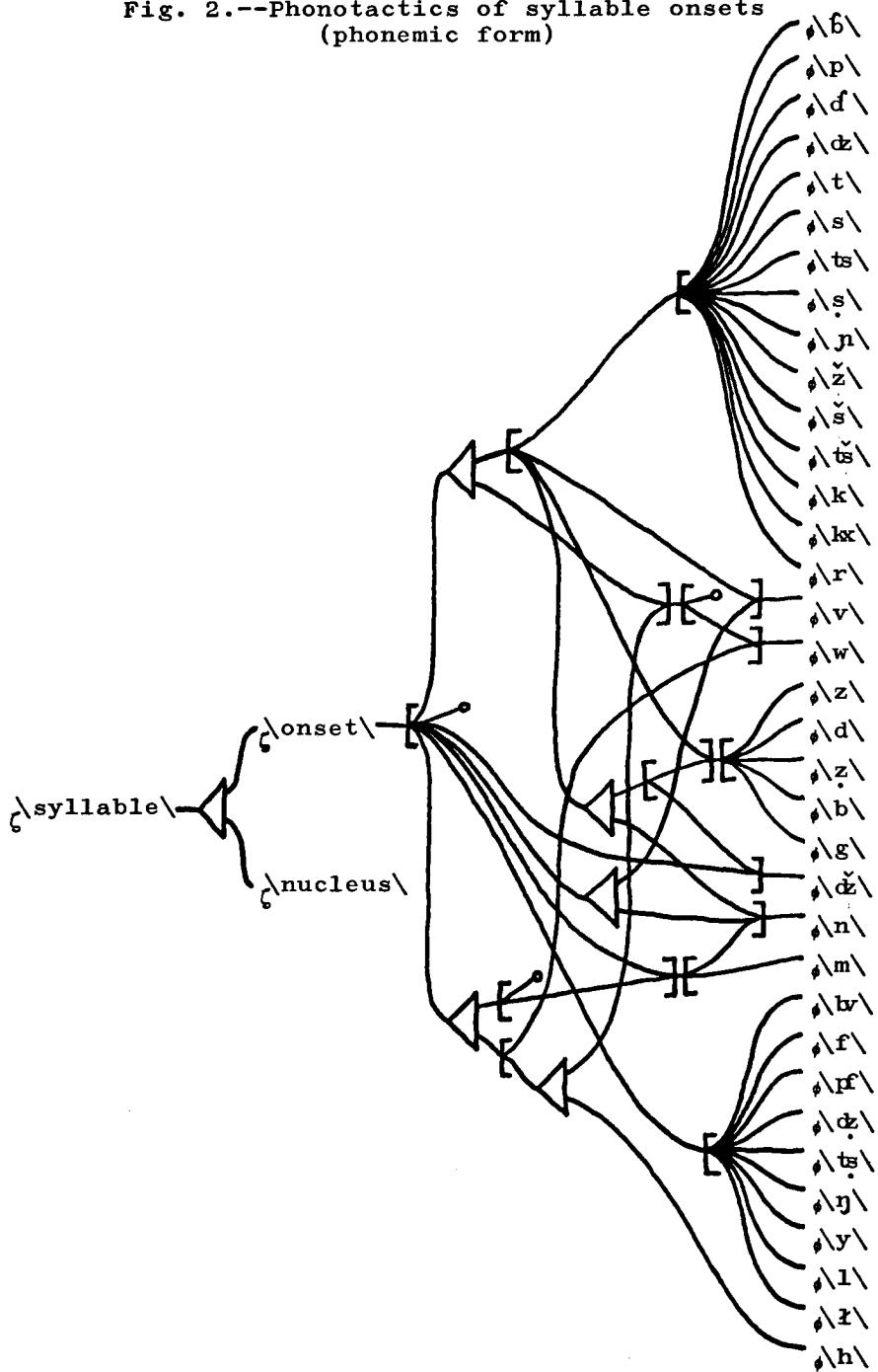


Fig. 3.--Phonemic sign pattern

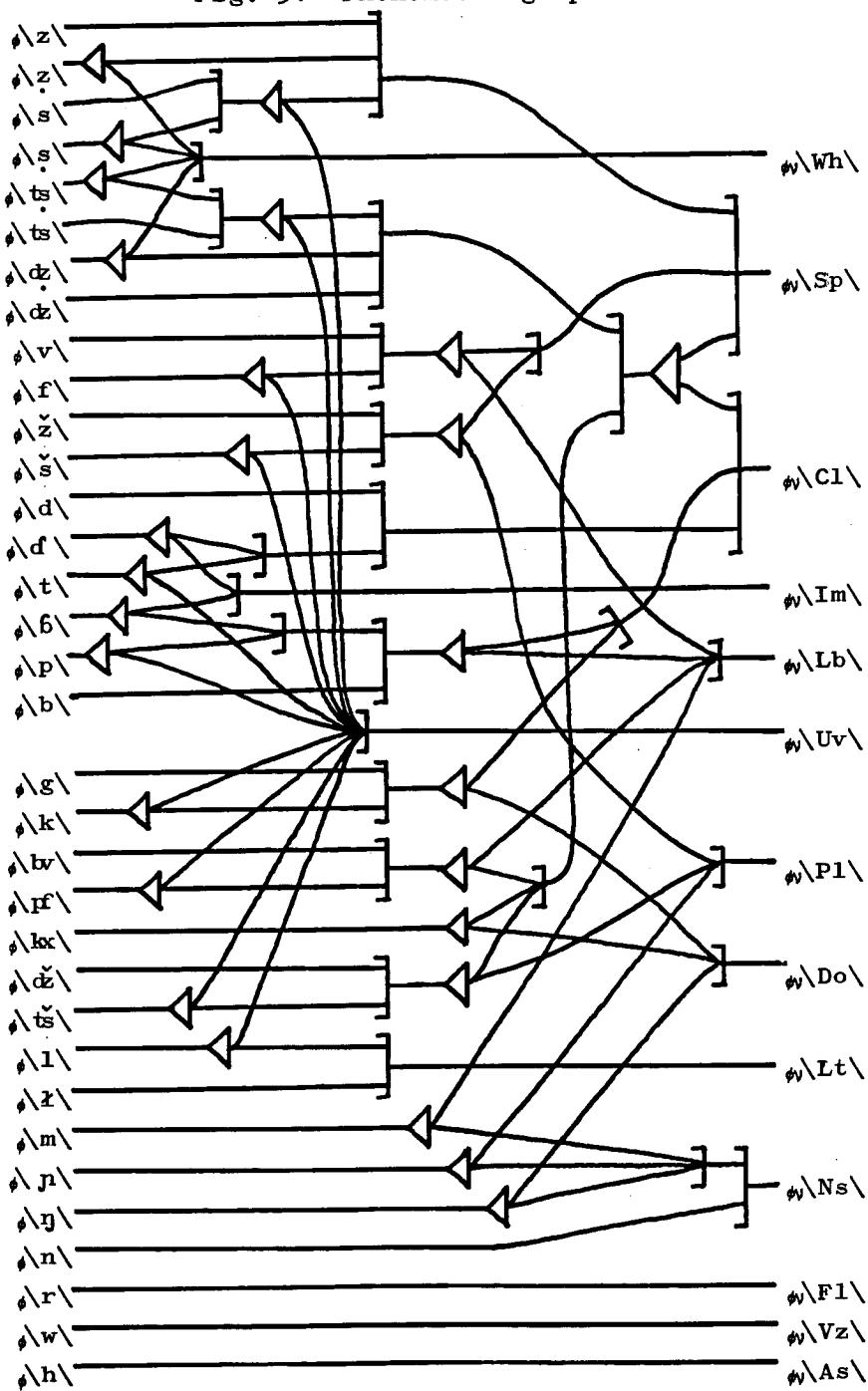


Fig. 4.--Syllable onsets which occur in Manyika

ϕ	ϕ^v	Stevick	ϕ	ϕ^v	Stevick
m	LbNs	LN	tw	C1UvVz	ASQW
b	LbC1Im	LS	sw	SpUvVz	AFQW
v	LbSp	LF	nz	NsSpWh	ANFV
nb	LbNsCl	LNS	dz	C1SpWh	ASFV
nv	LbNsSp	LNF	s	SpUvWh	AFQV
mh	LbNsAs	LNH	zw	SpVzWh	AFWV
bw	LbC1Sp	LSF	tsw	C1SpUvVz	ASFQW
b	LbCl	LSH	ts	C1SpUvWh	ASFQV
p	LbC1Uv	LSQ	nzw	NsSpVzWh	ANFWV
f	LbSpUv	LFQ	sw	SpUvVzWh	AFQWV
mw	LbNsVz	LNW	j	P1Ns	AYN
fw	LbC1VzIm	LSW	z	P1Sp	AYF
vw	LbSpVz	LFW	dz	P1C1Sp	AYSF
pf	LbC1SpUv	LSFQ	s	P1SpUv	AYFQ
pw	LbC1UvVz	LSQW	jw	P1NsVz	AYNW
mhw	LbNsAsVz	LNHW	zw	P1SpVz	AYFW
bw	LbC1Vz	LSHW	ndz	P1NsClSp	AYNSF
nbw	LbNsClVz	LNSW	ts	P1C1SpVz	AYSFQ
n	Ns	AN	sw	P1SpUvVz	AYFQW
d	C1Im	AS	ndz	P1NsClSpVz	AYNSFW
z	Sp	AF	tsw	P1C1SpUvVz	AYSFQW
nd	NsCl	ANS	j	DoNs	DN
nz	NsSp	ANF	g	DoCl	DS
nh	NsAs	ANH	ng	DoNsCl	DNS
dz	C1Sp	ASF	k	DoC1Uv	DSQ
d	C1	ASH	gw	DoCl	DSW
t	C1Uv	ASQ	ngw	DoNsC1Uv	DNSW
s	SpUv	AFQ	kx	DoC1SpUv	DSFQ
nw	NsVz	ANW	kw	DoC1UvVz	DSQW
dw	C1VzIm	ASW	kxw	DoC1SpUvVz	DSFQW
zw	SpVz	AFW	w	Vz	W
z	SpWh	AFV	y	P1	Y
ndw	NsClVz	ANSW	r	F1	R

Fig. 4--Continued

ϕ	ϕ^ν	Stevick	ϕ	ϕ^ν	Stevick
nzw	NsSpVz	ANFW	h	As	H
nhw	NsAsVz	ANHW	rw	F1Vz	RW
dzw	C1SpVz	ASFW	hw	AsVz	HW
dw	C1Vz	ASHW	l	Lt	X
ts	C1SpUv	ASFQ	z	LtUv	XQ

Stevick's Ultimate Phonological Constituents

- A apical or blade articulation
- D dorsovelar articulation
- F friction
- H voiced glottal friction or strong explosion
- L bilabial or labiodental articulation
- N nasality
- Q lack of voicing
- R apical flap
- S stoppage
- V secondary labial friction ("whistling")
- W velarization
- X lateral friction
- Y blade-palatal articulation or vocoid

Note: In the columns headed ϕ^ν , the phonons corresponding to each onset are merely listed; the order among them does not imply anything about the structure which connects them.

REFERENCES

- Bryan, M. A. 1959. The Bantu languages of Africa. (*Handbook of African languages, vol. 4.*) London, Oxford University Press for International African Institute.
- Hockett, Charles F. 1947. Peiping phonology. JAOS 67.253-67.
- Lamb, Sydney M. 1966. Outline of stratificational grammar. Washington, D.C., Georgetown University Press.
- Stevick, Earl W. 1964. Two Bantu consonant systems. Lg. 58-74.

**STEM PHONOLOGY AND AFFIX PHONOLOGY
IN DOGRIB (NORTHERN ATHAPASKAN)**

Robert Howren
(University of Iowa)

Dogrib is a Northern Athapaskan language spoken by some 1200 Indians living in scattered settlements north and west of the Great Slave Lake in the MacKenzie District of the Canadian Northwest Territories. The dialect of Dogrib which I describe here is the Rae dialect in its conservative form. Fort Rae, the largest of five Dogrib settlements, is located on the east shore of Marian Lake, an extension of the North Arm of Great Slave Lake, seventy miles northwest of Yellowknife. The data on which the present discussion is based were gathered by me during the summer of 1965 and, with the assistance of a graduate student, Charles Pyle, during the summer of 1967.(1)

The Phonological Rules

"Classical" phonemic phonology has always emphasized the independence of phonology from grammar, although as everybody knows, the theoretical principle has more often been suspended, for one or another reason, than it has been assiduously observed. In this paper, I present some aspects of the phonological structure of Dogrib which support the proposition that at least for this Athapaskan language, an analysis which is predicated on the independence of phonology from grammar results at best in a superficial description. The surface phonetic texture of Dogrib displays complexities and apparent irregularities which are rooted in quite regular patterns of grammatical structure. Some interesting complexities of the surface phonetics of Dogrib have to do with the divergence and interconnection of stem and prefix phonology, and these are the subject of the present discussion.

It is a commonplace fact in the comparative study of Athapaskan that stems and affixes have a divergent phonological development (see, for example, Li 1933:430). In Dogrib, this divergent development has

¹The research has been supported by The National Museum of Canada, the University of Iowa Old Gold Faculty Research Fellowships, the American Philosophical Society, and the National Science Foundation.

resulted in synchronic structural patterns which are neither immediately apparent from the surface phonetics nor accessible to an analysis which does not make use of grammatical information. The phonological rules will therefore be assumed to operate upon bracketed and labelled strings of morphemes to produce phonetic representations of these strings. Such an approach reveals underlying regularities which are not evident from the phonetic surface, and enables us to operate with minimum inventory of basic phonological entities.

Moreover, the rules which relate the phonetic representation--particularly of verbs--to the constituent structure representation in Dogrib have implications for the comparative analysis of Athapaskan languages, although we shall touch on such matters only in passing. Some rules which we shall formulate seem to be extremely widespread in the family, in essentially the form in which they are stated for Dogrib. Others seem to require modification from one language to another, or to be lacking in some languages. The extent to which phonological rules are shared in Athapaskan as compared with other linguistic families is an interesting subject beyond the scope of this paper.

Representation of Constituent Structure

For the limited purposes of this paper, the phonology of only certain types of verb and noun constructions will be considered. Within these arbitrary limits, however, most of the basic phonological patterns are found.

The verb structures that figure in the discussion may be described by the "rules" (S1-S6), which are not to be taken as actual rules of Dogrib grammar, but only as a convenient means of characterizing the grammatical strings to which the phonological rules apply.

- (S1) Verb → (Modal Prefix) Verb Nucleus
- (S2) Verb Nucleus → Conjunctive Pref + Vstem
- (S3) Conjunctive Pref → Aspect + Pron Subject
- (S4) Vstem → Classifier + Vroot
- (S5) Pron Subject → {Deictic
Personal Prefix}
- (S6) X-Aspect-Deictic-Y → 1-3-2-4

For convenience in the identification of morphemes, I list the morphemes later referred to, according to the categories which appear in (S1-S5):

Modal Prefixes:

nà motion to a specific destination
 k'è nondirectional motion
 šè pertaining to food

Aspect Prefixes:

We} perfective (selected by Vstem)
 e} zero imperfective

Classifiers: d, l, L, zero

Verb Roots:

à	"two move"	k ^w , o	"hear"
Wò	"stay, live"	se	"hunt"
me	"swim"	se	"shout"
t'a	"fly"	ci	"make"
ni	"treat, care for"	e	"paddle"
,	"hide"	ti	"eat"

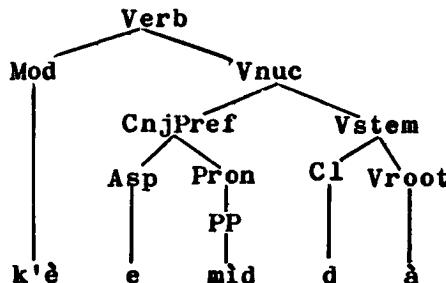
Deictic Pronouns:

c'e indefinite and 1st plural
 ge 3rd plural/dual

Personal Prefixes:

h	1st singular
ne	2nd singular
zero	3rd singular
mid	1st dual (phonetically [wi], with "D-effect" on following con- sonant)
ah	2nd dual/plural

The verb [k'iłt'a] "we two have visited here and there" is typical of the structures characterized by (S1-S6):



The structure of nouns in Dogrib is considerably simpler than that of verbs, and most of our attention will be given to the latter. The nouns mentioned as examples in the discussion will be either free stems (e.g., [We] "belt") or possessed nouns comprising a stem constituent preceded by a possessive prefix (e.g., [se-tà] "my father").

In the remainder of this paper, I hope to demonstrate the relationship of grammatical structure to phonology in Dogrib by (a) showing how a phonemic inventory may be substantially reduced by the use of grammatical information, and (b) presenting an example of the resolution of surface phonetic irregularity in verb stems by positing verb-classifiers in the constituent structure representation of two verbs.

Phonemic and Systematic Phonemic Inventories

An analysis of Dogrib in which the distinctive phonological elements are established strictly by the distribution of segments in phonologically defined units without reference to the grammatical structure of these units gives the following inventory of consonant phonemes (for native forms only):

ⁿ d						
d	t	j	ž	g	g ^w	
t	č	c	č	k	k ^w	
t'	č'	c'	č'	k'	k' ^w	?
L	s	š	š	h	w	
l	z	ž	ž	g	w	
m	n	r	y			

Redundancy rules and phonological rules applied to grammatically labelled strings enable us to reduce this set of thirty-four phonemes to a systematic phonemic set of twenty-seven consonants:

d	t	j	ž	g	g ^w	
t	č	c	č	k	k ^w	
t'	č'	c'	č'	k'	k' ^w	
L	s	š	š	h	w	
l						
m	n		y			

The segments excluded from the systematic phonemic set are accounted for in various ways.

The lenis spirants [z ź g w] are generated from stem-initial /s ſ h W/ by the Spirant Lenition Rule:

$$(1) [+For] \rightarrow [-For] / \left\{ \begin{array}{c} [-For] \\ [C] \\ V \end{array} \right\} + R [\begin{array}{c} [+Cns] \\ [-Int] \end{array}]$$

Where $R [$ is the initial boundary
of a noun or verb root.

The Spirant Lenition Rule yields, for example, [neži] "your song" from $N [ne + N_{stem} [\check{š}i]]$, and [nàwò] "he stays" from $v [nà + v_{nuc} [\emptyset + \emptyset + v_{stem} [\emptyset + v_{root} [Wò]]]]$. It does not apply to the string underlying the first person singular form [nàWò], however, because the verb root is preceded by the first person subject prefix /h/, a fortis consonant (later deleted): $v [nà + v_{nuc} [\emptyset + h + v_{stem} [\emptyset + v_{root} [Wò]]]]$.

Since the Spirant Lenition Rule as stated does not apply to the spirants of prefixes, forms like [wetà] "his father", [Weda] "a living creature lies", [nàwizè] "the two of us hunt", and [nàWizè] "the two of us have hunted", seem to be counterexamples of our formulation of the rule. In prefixes, both fortis and lenis labiovelar spirants occur in analogous or identical environments. Thus if we take the labiovelar spirants of stems and prefixes to be the same phonological entities, we must explain how the fortis and lenis segments come to be in complementary distribution in stems while they are distributed contrastively in prefixes. Since none of the other spirants are distributed in this way, it is in the interest of simplicity and generality to preserve the Spirant Lenition Rule as we have stated it, if possible. The alternative to loosening the rule is to posit a relation between prefixal [w] and some other underlying segment which occurs in stems but not in prefixes. This segment is evidently the nasal continuant /m/, which is realized in stems as a prenasalized bilabial stop [mb] when followed by an oral vowel, as in [k'èmbe] "he swims", and as a continuant [m] when followed by a nasal vowel, as in [dahmò] "roof"--and which does not occur in prefixes.

The evidence of comparative Athapaskan supports the identification of stem [m ^mb] with prefix [w] in Dogrib. In Navaho, for example, both the initial consonant of the third person singular possessive prefix /bi-/ (Dogrib [we-]) and the initial consonant of the noun stem /béš/ "knife" (Dogrib [^mbe]) are identical. Thus we may conclude that Navaho lacks the equivalent

of the Dogrib rule (2), which takes /m/ to [w] in prefixes:

(2) Prefix-Labial Rule

$m \rightarrow w / \text{Pre} [$
 Where Pre [is the initial
 boundary of a nominal or
 verbal prefix.

In Sarsi, another language of the northern group, the cognate of Dogrib [wembe] "his knife" is [bim̥as]. Clearly, Sarsi phonology includes a rule similar to the Prefix-Labial Rule of Dogrib in its differentiation of prefix-initial from stem-initial labials.

Distinguishing between stem-initials and prefix-initials also enables us to eliminate the prenasalized alveolar stop /nd/ from the inventory of basic consonants, and to treat the two nasal continuants /m n/ identically, with a single rule:

(3) Nasal Interruption Rule

$[-\text{Int}] \rightarrow [\alpha \text{Int}] / R [\begin{matrix} +\text{Cns} \\ +\text{Son} \\ +\text{Nas} \end{matrix}] \begin{matrix} -\text{Cns} \\ +\text{Syl} \\ \beta \text{Nas} \end{matrix}$

Rule (3) assigns to a root-initial nasal consonant the interruptedness value which is opposite to the nasality value of the following vowel. Thus we get [k'èm̥be] "he swims" but [emo] "mother"; [ndi] "island" but [s̥n̥i] "I wonder". Systematic phonemic /n/ in prefix-initial position is unaffected by the Nasal Interruption Rule, and is thus automatically associated with the interrupted and continuant nasals of stems: e.g., the prefix-initial [n] of [nit'a] "he flies" with the initials of the stems in [wek'è ahndi] "you (du/pl) take care of him" and [s̥n̥i] "I wonder".

The resonant /r/, which must be included in the inventory of consonants in a phonemic analysis, has a suspiciously limited distribution. It occurs only as the initial of a single prefix morpheme /re-/ , the meaning of which is as yet obscure. On the grounds that [r] sometimes alternates in this morpheme with the lenis alveolar stop [d], I assume that it may be derived from systematic phonemic /d/, which remains a stop as stem initial but becomes [r] as prefix initial. If this is correct, then we must further assume that all prefix-initial lenis alveolar stops are to be associated in the rules with the stem-initial fortis stop [t]. Thus a form like [nàdereh?i] "I am hiding myself" is derived from [nà+fe+de+h+?i] by a series of

rules including (4) and (5):

(4) d-Rhotacization Rule

$d \rightarrow r / \text{Pre}[\underline{\quad}]$

(5) t-Lenition Rule

$t \rightarrow d / \text{Pre}[\underline{\quad}]$

Rules (4-5) are ordered; applied in the order (5-4), they would obliterate the /t d/ distinction in prefixes. The data include no occurrences of the fortis stop [t] as a prefix initial.

Without recourse to grammatical information, the occurrence of the glottal stop cannot be predicted. It contrasts with the glottalized velar stop [k'] in the minimal pair [sek'a] "my fat" and [se'a] "my snowshoe", with the glottalized labiovelar stop [kw'] in [erèhk'w'o] "he is listening" and [goh'w] "he found a compact object", and with zero in [šèahtj] "you two are eating" and [se'aka] "on my snowshoe". But in a generative phonology we exclude it from the set of systematic phonemic segments and introduce it by Rule (6):

(6) Glottal Stop Insertion Rule

$\emptyset \rightarrow ? / R[\underline{\quad}]V$

--i.e., a glottal stop is introduced prevocalically in any noun or verb root which in its base form has a vocalic initial.

Classifiers and Stem Initials

The four Athapaskan verb-classifiers--zero, d, l, and L--have not been preserved as overtly in Dogrib as they have in some of the other Athapaskan languages. But though they have largely disappeared as discrete phonetic segments, they nevertheless underlie a good many of the phonological complexities of the Dogrib verb. The L-classifier is relatively uninteresting, as it is regularly retained as [h] and does nothing more spectacular than to inhibit the operation of the Spirant Lenition Rule. The d-classifier is somewhat more spectacular. It combines in striking ways with a number of root-initial consonants, creating phonologically ambiguous stem initials. The morphophonemics of the d-classifier, the details of which Athapaskanists gather under the heading of "the D-effect", deserves formulation in a rule apart from the common run.

Accordingly, I offer the "D-Effect" Rule:

$$(7) \begin{bmatrix} +\text{Cns} \\ +\text{Int} \\ -\text{Nas} \\ -\text{Syl} \\ \delta F \end{bmatrix} + \begin{bmatrix} \alpha\text{Cns} \\ \beta\text{Int} \\ \gamma\text{Nas} \\ -\text{Syl} \\ \epsilon F \end{bmatrix} \rightarrow \begin{bmatrix} +\text{Cns} \\ +\text{Int} \\ -\text{Nas} \\ -\text{Syl} \\ \epsilon F \end{bmatrix}$$

Where δF = the complementary subset of feature specifications which completes the specification of the segment [d]; ϵF = any complementary subset of feature specifications which completes the specification of a given segment; and $\alpha, \beta, \gamma = +$ or $-$.

In the output segment of the "D-Effect" Rule, the values for Consonantality, Interruptedness, and Nasality are carried over from the segment [d], but all the remaining values (i.e., ϵF) are the same as for the second segment in the original sequence. (2) (Non-syllabicity must be specified to exclude [d+V] sequences from the domain of the rule.) Rule (7) reduces the sequences [d+, d+n, d+w, d+l, d+z, d+ž, d+g] to, respectively, [t', d, gʷ, t, j, ž, g], and reduces any other [d+C] sequence to [C]—i.e., deletes [d]. This rule obviously follows the Spirant Lenition Rule, since sequences of [d]+fortis spirant do not occur.

The persistence of the l-classifier is the least obvious from an examination of the surface phonetic structure of the verb, and is posited solely on the basis of the non-operation in some verbs of certain phonological rules, among them the Spirant Lenition Rule. This "l-effect" can be illustrated with the imperfective paradigms of the verbs hunt and shout.

	<u>hunt</u>	<u>shout</u>
1sg	nàhzè	ehse
2sg	nànezè	neze
3sg	nàzè	eze
1du	nàwizè	wije
1pl	nàc'ezè	c'eze
2du/pl	nàahzè	ahse
3du/pl	nàgezè	geze

²I am indebted to my colleague Larry W. Martin for the initial suggestion that led to the reduction of the rule to its present form.

The initial consonant of the hunt stem is invariant throughout, and is identical with the initial consonant of the shout stem except for the lsg, ldu, and 2du/pl forms. The Spirant Lenition Rule has evidently not operated on the stems of the lsg and the 2du/pl forms of shout, but has applied to the corresponding forms of the hunt paradigm. Moreover, the "D-Effect" Rule has left its mark on ldu [wìje], while ldu [nàwizè] has been unaffected by it. This apparent irregularity can be reduced to order by positing different classifiers for the two verbs--for hunt, a classifier which permits the operation of the Spirant Lenition Rule but blocks the "D-Effect" Rule; and for shout a classifier which, conversely, permits the "D-Effect" Rule to operate but inhibits the Spirant Lenition Rule in forms where the subject prefix ends in [h]. The d- and L-classifiers are rejected out of hand as possibilities (the former would have triggered the "D-effect" throughout both paradigms, and the latter would have remained throughout as [h]). We therefore posit for hunt the l-classifier, which, as a lenis consonant, meets the condition for the application of the Spirant Lenition Rule even in forms which have a pre-stem [h], and at the same time blocks the "D-Effect" Rule in the ldu form (the classifier intervenes between the subject prefix [-wid-] and the root [-zè]). A later rule deletes the l-classifier, thus leaving sequences of [h]+lenis spirant. For the shout stem, we posit the zero-classifier, thereby accounting for (a) the failure of the Spirant Lenition Rule in the lsg and the 2du/pl forms and (b) the application of the "D-Effect" Rule in the ldu form.

In this brief space it has been possible to provide only a glimpse of Dogrib phonology, and of the trying-out of generative phonological theory on an Athapaskan language. Hopefully, the initial proposition--that the phonology of Dogrib is hardly describable apart from the grammatical structure of the language--has been convincingly demonstrated, and the potential fruitfulness of the extension of generative phonology to the comparison of Athapaskan languages clearly suggested.

³ It is not surprising to find that /-sè/"hunt" is indeed historically an l-class stem. In Chipewyan (Li 1946:414) not only has the classifier persisted as a distinct phonetic segment in this verb, but also the Chipewyan verb exhibits the effects of a Spirant Lenition Rule apparently identical in its constraints to the Dogrib rule:

lsg	hesze	3sg	helze
2sg	hilze	2du	hulze etc.

References

Davidson, William, 'A Preliminary Analysis of Active Verbs in Dogrib', in Studies in the Athapaskan Languages, 48-55.

Hoijer, Harry, et al., Studies in the Athapaskan Languages (Berkeley and Los Angeles, 1963).

_____, and Janet Joël, 'Sarsi Nouns', in Studies in the Athapaskan Languages, 62-75.

Li, Fang-Kuei, 'Chipewyan Consonants', Ts'ai Yuan P'ei Anniversary Volume, 429-67 (Peiping, 1933).

_____, 'Chipewyan', in Harry Hoijer, et al., eds, Linguistic Structures of Native America, 398-423 (New York, 1946).

Petitot, LeR. P. E., Dictionnaire de la langue Dène-Dindjié (Paris, 1876).

SUNDAY GREEK

Kostas Kazazis
(University of Chicago)

The purpose of this paper is to discuss some phonological characteristics of Modern Greek, when it is spoken under certain conditions of normative pressure, namely when the speaker is overanxious to impress an interlocutor he considers as a superior. I will try to show that, in the case of at least one Greek speaker, slips of the tongue due to this kind of pressure and involving phonetic segments (rather than suprasegmentals) largely coincide with underlying phonological forms. In other words, the speaker's error consists in his failure to apply one or more obligatory phonological rules.

The speaker in question is an Athenian high-school graduate in his middle twenties, whom we will call 'Socrates'. The material is drawn from about six hours of recorded conversation between the two of us. During the first four hours, we used formal second-plural forms in addressing each other. At the end of the fourth hour, at my suggestion, we shifted to familiar second-singular forms. By that time, however, Socrates had become sufficiently self-confident, so that the use of the familiar form of address does not seem to have affected his verbal behavior too much. In fact all but four of the sixteen slips that I noted occurred during the first hour. Three occurred during the second hour, and only one during the third.

I think there were mainly two reasons why Socrates was somewhat nervous during our conversations. In the first place, he was confronted with someone who taught at a university--such people are considerably bigger shots in Greece than they are in the United States. Note, incidentally, that at the time the conversations were recorded Socrates had been in this country less than three months. The second reason why he was nervous seems to have been the presence of a tape-recorder. Socrates's nervousness decreased as we went along, so that the first hour was also the worst, and especially the first few minutes. For about five minutes, Socrates's intonation was all wrong. His tempo was slower than usual. His choice of words betrayed a strong pre-

occupation with sounding educated and with impressing me: he would select katharevusa words, or at least words generally used in elevated styles of dhimotiki--such words are almost invariably of katharevusa, that is of learned, origin.¹ In other words, Socrates was speaking his Sunday Greek.

After he got a little used to the microphone, his intonation and tempo relaxed considerably, but the lexical, syntactic, and segmental-phonological affectations and blunders persisted for some time, although as I said they decreased as we went along.

We will leave aside the lexical affectations, even though they admittedly constitute a very important component of Sunday Greek. But since the phenomenon is probably universal (at least in literate speech communities), we can simply mention here the fact that it certainly occurs in Modern Greek as well. Suffice it to say that, once during the first hour, Socrates's choice of words was so obviously stilted that he felt he had to go back and use a more everyday synonym. He was talking about a cave by the sea, when he said: "Light penetrates . . . light comes in, because there are certain holes on the roof of the cave." In the original,

(1) Εἰσέρχεται φῶς . . . μπαίνει φῶς, διότι ὑπάρχουν ὥρισμένες τρύπες στὴν ὁροφῇ τῆς σπηλαῖς.

We will also ignore syntactic affectations, such as the use of katharevusa endings, constructions, and the like. Speakers of Modern Greek often sprinkle a more or less generous dose of katharevusa elements in their speech, and few people seem to be bothered by that, other than dhimotiki purists.

As far as segmental-phonological deviations go, we should be careful to distinguish between affectations on the one hand, which are at worst stylistic infelicities but not outright mistakes, and overcorrections,

¹For a brief but accurate discussion of the Modern Greek language question, see Ferguson 1959.

which clearly are mistakes.²

Phonological and other verbal affectations occurring in informal Modern Greek discourse usually involve forms taken from the 'high' or 'learnèd' language, namely from katharevusa. This, of course, is only natural in a speech community with diglossia, like present-day Greece. A typical phonological affectation in everyday spoken Greek is the use of katharevusa consonant clusters, instead of the expected dhimotiki ones. (2), (3), and (4) below are examples of dhimotiki rules (informally formulated here), which yield phonetic representations of some underlying consonant clusters.

(2) spirant → stop / ____ /s/

For example, underlying /taksíðevsa/ 'I traveled' is phonetic [taksiðepsa]. (Underlying phonological forms are given here in a simplified notation, which includes only phonological segments, omitting other information, such as junctures and the like.) Note that katharevusa has [taksiðefsa]. In other words, rule (2) does not apply in katharevusa--only the general rule which devoices obstruents before voiceless obstruents applies, thus here /v/ → /f/ / ____ /s/.

(3) spirant → stop / spirant ____

Thus, for example, underlying /oniréfθika/ 'I dreamt' is phonetic [oniréftika]. Katharevusa has [oniréfθin], that is rule (3) does not apply there--once again only the obstruent-devoicing rule we just mentioned applies, thus /v/ → /f/ / ____ /θ/. We might also note that there are exceptions to rule (3); for example it does not apply to the sequence /sf/: [sfirí] 'hammer' is a good dhimotiki word, which is not of katharevusa origin.

(4) stop → spirant / ____ stop

For example, underlying /plektó/ 'knitted wear' (cf. [pléko] 'I knit') is pronounced [plextó]. Once again, katharevusa has [plektón], which means that its grammar does not contain rule (4).

²For a partly relevant discussion, see Some thoughts on style, by Angus McIntosh, in McIntosh and Halliday 1966.

Note that the underlying phonological representations to which each of the above rules applies coincide with the corresponding phonetic representations in katharevusa. In other words, to get dhimotiki forms in such cases we must apply an additional phonological rule to katharevusa forms. This is by the way a rather common situation in Modern Greek, namely that the underlying phonological forms we set up for dhimotiki will turn out to be very close to katharevusa phonetic forms.

But what about the numerous everyday words which contain sequences of two stops or two spirants, words like [sinélefsi] 'assembly' (not *[sinélepsi]), [apéxθia] 'abhorrence' (not *[apéxtia]), [aptós] 'tangible' (not *[aftós])? And what about the very many doublets, where one member of the pair has a sequence of two spirants or two stops, while the other has spirant + stop? Such doublets are [xθes] and [xtes] 'yesterday', [ásçimos] and [áskimos] 'ugly', [októ] and [oxtó] 'eight'. Should we regard the words with two-spirant or two-stop clusters as katharevusa words? If we did that, we would be broadening the definition of what constitutes a katharevusa word far too much. Moreover, we would be thereby accusing a great many people of being affected, when in fact they are not. Many people namely use the 'katharevusa' member of some doublets natively, that is they have not picked it up later in life for prestige purposes. This is true for example of the numerals [eptá] 'seven', [októ] 'eight', and their derivatives. To be sure, all such words were originally borrowed from katharevusa, but as far as present-day Greek is concerned many of them are perfectly good dhimotiki words. The only vestige of their katharevusa origin is the fact that they are not subject to rules (2), (3), and (4). In other words, somewhere in their respective entries in the lexicon they are marked as being of learned origin. By virtue of this feature they fail to undergo certain major phonological rules, among others rules (2), (3), and (4).

It should be noted, of course, that any two speakers of standard Modern Greek will probably disagree as to whether certain words belong to the lexicon of katharevusa or to that of dhimotiki, but by and large there will be agreement between them. So that the idea that

there are two different codes, one for dhimotiki and one for katharevusa, with among other things important lexical differences, should not be dismissed, even though admittedly there is no clearcut boundary separating them. For not only is there a large middle area containing elements common to both norms of Modern Greek, but there is also heavy mixing of katharevusa elements in dhimotiki and conversely of dhimotiki elements in present-day katharevusa. For the sake of fairness, we ought to mention that the admixture of katharevusa elements in everyday discourse (to say nothing of more formal oral or written discourse) is far from always due to affectation on the part of the speaker. Quite frequently people will use katharevusa endings and constructions along with lexical items originally borrowed from katharevusa, simply because they have been accustomed to seeing or hearing these lexical items in katharevusa settings. This is particularly true of discourse containing technical terms in a variety of fields. Lawyers, engineers, doctors, officers, university students often mix a considerable amount of katharevusa in their speech, when the subject-matter turns to their respective fields.

But to go back to Socrates, it was obviously pure affectation on his part when he said [ísoe] for [íste] /íste/ 'you are' in the sentence [K esís apo tin eláda ísoe] 'you too are from Greece'; or when he said [íce pandrefθí] instead of [íge pandrefθí] /íxe pantrevθí/ 'he had gotten married'; or [taksiðéfsete] instead of [taksiðépsete] 'you travel' (perfective non-past) in the sentence [prókite na taksiðéfsete to kalokéri?] 'will you travel this summer?'

Let us now turn to the real mistakes. First I would like to dispose of a possible, but not certain, case of overcorrection. It concerns such a common word that when I first heard it I could hardly believe my ears. At one point Socrates said [apo fθó], when he meant to say [apo ftó] /apó aftó/ 'from this'. Note that the cognate root in katharevusa is [aft]-, that is it too has [ft] and not *[fθ]--and this is precisely what makes [apo fθó] a mistake, rather than a mere overcorrection. However, that was the only instance of this sort of thing, namely of a false analogy (as all over-corrections are) involving a sequence of spirant + spirant:

(5)	'normal' dhimotiki	katharevusa or stilted dhimotiki
	[ftáno]	[fθáno]
	[aftó]	*[afθó]

'I arrive'
'this'

So perhaps it was a random slip of the tongue, and not an overcorrection. I simply report it here for the record. Note that, even if we had a way of discovering the true nature of this error and had decided that it was indeed an overcorrection, I still would not have gone so far as to claim that it was a deliberate one. I do not think the terms 'slip of the tongue' and 'over-correction' are necessarily contradictory--in the sense that the latter (but not the former) is done deliberately. It may well be the case that an overcorrection is a slip of the tongue in that it 'escaped' from the speaker, who nevertheless was trying to sound more 'refined', however unconsciously he may have done so. And, of course, the general statement that a great many slips of the tongue go unnoticed by the person who made them would still hold true here as well.

The remaining overcorrections are more significant. With one exception, they all involve the failure to voice obstruents in certain environments where voicing is obligatory. Thus there is a rule (6), which voices /s/ to [z] immediately before voiced obstruents and sonorants. Informally, this rule looks something like this:

$$(6) \quad /s/ \rightarrow [z] / _ \left[\begin{array}{l} -\text{syllabic} \\ +\text{consonantal} \\ +\text{voice} \end{array} \right]$$

This rule applies both word-medially, as well as across word-boundaries in close transition. For example, /pinasménos/ 'famished' is pronounced [pinazménos], /merikés vásis/ 'some bases' is pronounced [merikéz vásis], and /merikés méres/ 'some days' is phonetically [merikéz méres].

Socrates failed to voice the /s/ twice before word-boundary in close transition. He said [ekínes tis méres] instead of [ekínes tiz méres] /ekínes tis méres/ 'in those days', and [íster apo líjes méres] instead of [íster apo líjez méres] /ístera apó líjes méres/ 'a few

days later'. Now note that in speaking katharevusa or in reading it aloud, people occasionally fail to voice the /s/ in such environments, even in close transition. There seems to be an unexpressed feeling on the part of Greek speakers that the voiceless spirant [s] is somehow more refined and thus more fit for the 'higher' language, katharevusa, than is its voiced counterpart. As we shall see later, this is not the only instance where voiceless obstruents are felt to be preferable to voiced ones. In informal (dhimotiki) discourse on the other hand, the failure to voice the /s/ in close transition is generally regarded as an aberration.

Even more interesting are the three instances where Socrates did not voice the /s/ word-medially. He said [periorismús] instead of [periorizmús] /periorismús/ 'restrictions' (accus.), and on two occasions [orisméná] instead of the expected [orizméná] /orisméná/ 'certain' (neuter plural). One reason why these examples are particularly interesting is that word-medially Greeks voice the /s/ even in katharevusa.

The next type of overcorrection noted concerns the failure to voice stops after nasals. With the exception of some recent loanwords, Modern Greek voices stops immediately after nasals, both word-medially as well as in close transition across some types of word-boundaries--for instance after negative particles and after the definite article. The relevant rule looks something like this:

$$(7) \begin{bmatrix} + \text{obstruent} \\ - \text{continuant} \end{bmatrix} \rightarrow [+ \text{voice}] / [+ \text{nasal}] _$$

Note that (7) as formulated here makes no provision as to the types of junctures across which the rule applies or fails to apply. This is because I have not yet figured out exactly what those junctures are. Here are some examples where (7) applies across word-boundary: /ðen káθete/ 'he does not sit' is pronounced [ðéy gáθete]; /min pas/ 'don't go!' is phonetically [mim bas]; /tin timí/ 'the price' (accus.) is pronounced [tin dimí]. As we can see, the nasal is assimilated to the following stop as to point of articulation. Moreover, in many varieties of Modern Greek (including the type of Greek that Socrates speaks), the nasal may be subsequently dropped. Thus [ðéy gáθete], [mim bas], and [tin dimí] may also be pronounced respectively as [ðé gáθete], [mi bas], and [ti dimí]. Let us mention

parenthetically that, other things being equal, keeping one's nasals before stops is considered to be 'better' or less colloquial than dropping them. That is why many people who usually drop their nasals will keep them when speaking their Sunday Greek. Incidentally, this is largely true of Socrates as well. The prestige that the nasals enjoy in Modern Greek may also have something to do with the fact that katharevusa has a great many more nasals word-finally than does dhimotiki. Apart from that, in speaking or reading aloud in katharevusa, one is generally expected to retain the nasals before stops, both word-medially and otherwise. Less educated speakers often 'restore' nasals that were never there, saying for instance [tin eláðan] for [tin eláða] '(the) Greece' (accus.)--note that even katharevusa has [tin eláða], without a final [n] in [eláða].

To return to the voicing of stops after nasals, some speakers of Greek seem to consider voiceless stops as somewhat more refined than voiced stops, when there is a word-boundary separating the stop from a preceding nasal. So that in vocalizing katharevusa texts such speakers will occasionally fail to voice stops under these conditions. However, in informal (dhimotiki) discourse, voicing the stop is obligatory, provided the remaining conditions are met, namely that the tempo is relatively rapid, that there are no hesitation phenomena, and of course that the right type of juncture is present. Socrates failed to voice word-initial stops after nasal on about half a dozen occasions. Sometimes he even failed to assimilate the point of articulation of the nasal to that of the following stop, which is another feature of Sunday Greek, but on which we will not insist here. Socrates said for instance [ðén katikúne] instead of the expected [ðéy gatikúne] /ðén katikúne/ 'they don't live (inhabit)'. But what is even worse is that he sometimes also fails to voice stops after nasals within the same word. This is a more serious form of overcorrection, since in that environment the probability of there being open transition is minimal. Socrates does that on three different occasions. He says [ðé borís na kolimpísis] instead of the correct [ðé boríz na kolimbísis] /ðén mporís na kolimbísis/ 'you can't swim'; [kentímata] for [kendímata] /kentímata/ 'embroideries'; [pántio] for [pándio] /pántio/ '[the] Panteios [School]' (accus.). As far as Greek listeners are concerned, the pronunciation of those words with voiceless stops is roughly equivalent to our hearing a native speaker of English pronounce potato with [d:] instead of [æ], as a result of a

false analogy with tomato:

(8) General American	British RP
[təmējDow]	[temā:tow]
[peDējDow]	*[petā:tow]

in an effort to sound more refined, or more British, or what have you.

And now we come to what I regard as the pièce de résistance. At one point Socrates says [ólo vrácia], with a syllabic [i], instead of the correct [ólo vráça] /ólo vráxia/ 'full of rocks', with no [i] whatever, whether syllabic or non-syllabic. Notice that here too what he actually pronounced is almost identical to what I consider to be the underlying phonological shape of the word [vráça]. In (9) we see how the phonetic representation of that word is derived from its phonological representation:

- (9)
- | | |
|--|---|
| a. Palatalize velars before front vowels: | /vráxia/
vrácia
(Socrates's form) |
| b. Unstressed /i/ becomes non-syllabic next to a vowel:
(There are independent grounds why we need (9b) in the grammar. (9b) is obligatory in inherited words, but only optional in words of learned origin.) | vrácia |
| c. Delete non-syllabic [i] after palatal consonant:
(9c) applies only to inherited words.) | [vráça] |

In other words, a pronunciation [vrácia], with a syllabic [i], would have been possible only if [vrácia] had been the plural of a hypothetical neuter noun *[vráçio] of learned origin. No inherited Modern Greek word can possibly be pronounced as [vrácia]. Socrates did use the correct pronunciation [vráça] on several occasions elsewhere on the same tape (first hour). But on the one occasion where he mispronounced that word as [vrácia] he elevated it to the status of a word of learned origin.

The fact that in phonetic representations we do find palatal consonants before non-front vowels, as in [vráça], has led some linguists to posit distinctive palatalization of consonants for Modern Greek.³ For instance, they point out pairs like the ones in (10):

- (10) [xóni] 'he thrusts' [xrónia] 'chronic' (f.sg.)
 [çóni] 'snow' [xrója] 'years'

As I have shown elsewhere, we do not need to mark palatalization of consonants in underlying phonological representations of Modern Greek.⁴ I claim that the four words in (10) have the underlying phonological shapes that we see in (11):

- (11) /xóni/ 'he thrusts' /xrónia/ 'chronic' (f.sg.)
 /xióni/ 'snow' /xrónia/ 'years'

The word for 'chronic' and the word meaning 'years' have the same underlying phonological representation--at least in the simplified notation used here. The differences we find in their phonetic representations are due to the fact that [xrónia] 'chronic' is marked in the lexicon as being of learned origin, and thus fails to undergo rule (9c), while it is optionally subject to rule (9b). On the other hand, [xrója] 'years' is an inherited word and as such is obligatorily subject to both (9b) and (9c), as well as to a further rule (12), which palatalizes /n/ and /l/ before non-syllabic [j]:

- (12) {ⁿ_l} → {^{r̄}_{λ̄}} / — j

To go back to [vráça], Socrates's mispronunciation of that word as [vrácia] seems to confirm the underlying phonological representations I have posited in order to account for the occurrence of palatal consonants before back vowels. Once again, Socrates used a pseudo-refined form which was practically identical to the underlying phonological representation of the word in question.

Thus we have seen that one speaker of Modern Greek, when talking to a relative stranger whom he considered a superior, made a number of overcorrections. The se-

³One work which adopts such a solution is Koutsoudas 1962.

⁴That view was presented in a paper read at the 1966 Annual Meeting of the Linguistic Society of America.

lection of such blunders that I presented here consisted of pseudo-stilted forms which essentially coincide with underlying phonological forms. This seems to indicate that some types of underlying forms are felt to be rather more refined than their phonetic manifestations. Which is not too surprising for Modern Greek, since as I said earlier there is great resemblance between underlying phonological representations of *dimotiki* and the corresponding phonetic manifestations of its learned counterpart, *katharevousa*.

We should point out in closing that similar varieties of Sunday Greek are common in other situations as well. Thus, for example, people often make heavy use of underlying or near-underlying phonological forms when they are under certain conditions of emotional stress, such as fear, and also when they are talking to foreigners or to children--the latter two situations are probably also connected somehow with normative motivations, if not outright with normative pressure. But we cannot possibly enter into all that in a paper of this size.

REFERENCES

- FERGUSON, CHARLES A. 1959. *Diglossia*. Word 15. 325-40.
- KOUTSOUADAS, ANDREAS. 1962. Verb morphology of Modern Greek. Published as Part II of International Journal of American Linguistics 28.4.
- MCINTOSH, ANGUS, and M. A. K. HALLIDAY. 1966. Patterns of language: papers in general, descriptive and applied linguistics. London, Longmans.

SOME OBSERVATIONS ON STRIDENCY IN ENGLISH

Dale E. Woolley
(University of Illinois at Chicago Circle)

In Jakobson, Fant, and Halle's Preliminaries to Speech Analysis and later in Halle's article "On the Bases of Phonology," English [č, ž, f, v, s, z, š, č] are considered strident in contrast to the nonstrident (mellow) [θ] and [ð] and thus form a natural class. But an examination of the distribution of these phones suggests that perhaps they ought not to be classed together. For instance [č] does not occur initially before vowels and does not enter into consonant clusters; [č, ž, z] do not cluster initially; [š] clusters initially only with [r] except in loanwords; and [s] is the only phoneme in English to cluster initially with other obstruents. On the other hand, [f] and [v] have a distribution more similar to that of the other nonstrident obstruents.

An argument can be presented, I think, to show that among the continuant obstruents [f] and [v] are better grouped with [θ] and [ð] as nonstrident, in contrast to the strident [s, z, š, č]. First it is important to notice that the definitions of stridency offered by Jakobson, Halle, and Fant, if applied, actually exclude [f] and [v] from the class of stridents. In the Preliminaries, strident articulations are said to have more irregular waveforms and to display higher intensity noise than nonstrident articulations. This noise is the result of considerable turbulence caused by complex impediments in the vocal tract during the production of stridents. Halle, in the article mentioned above, is more explicit. He says (p. 327) that "strident sounds are produced by directing the air stream at right angles across a sharp edged obstacle or parallel over a rough surface, thereby producing considerable noisiness which is the major acoustical correlate of stridency." If noisiness is definitive for stridency, then [s, z, š, č] are clearly opposed to [f, v, θ, ð], as one can easily detect by introspection. Furthermore, if Halle actually applied his own definition of stridency, which is based on the action of the air stream mechanism in articulation, then he would have to consider [θ] and [ð] as strident also. The air stream in the production of [θ] and [ð], as well as [f] and [v], is directed at right angles across a sharp edged obstacle (the teeth).

Support for the analysis under consideration here can be found in the acoustic data on fricatives presented by Peter Strevens. He discovered that the intensity (noisiness in Halle's terminology) of [f] and [θ] is relatively low, whereas the intensity of [s] and [š] is relatively high. Strevens did not study the corresponding lenis (voiced) fricatives. However, he

argues convincingly that, although each voiced fricative occurs with less intensity than its voiceless counterpart, the intensity of [v] and [ð] would be relatively low as compared to the relatively high intensity of [z] and [ɹ].

Further support for this analysis can be found in the data on perceptual confusions presented by Miller and Nicely. In their study listeners identified utterances of nonsense syllables masked by noise or subjected to frequency filtering. The data clearly show that [f] and [θ], [v] and [ð], [s] and [ʃ], and [z] and [ɹ] form strong perceptual interconfusion pairs. Relatively little confusion in perception occurs between any member belonging to one of these interconfusion pairs and a member of another one of the pairs. That is, for instance, [f] and [θ], and [s] and [ʃ], are frequently confused with each other, respectively, but [f] and [s] are infrequently confused.

Since the argument presented by Jakobson, Halle, and Fant for analyzing [f] and [v] as strident is unconvincing, and since there is a good deal of acoustic and perceptual evidence indicating that [f, v, θ, ð] are sharply distinguished from [s, z, ʃ, ɹ], I will here consider the latter set of phones strident in contrast to the former set, which is nonstrident.¹

Several advantages obtain when this analysis is adopted. First, the distinctive feature configurations reflect the traditional organogenetic separation between slit fricatives (the nonstridents) and grooved fricatives (the stridents). Secondly, the eight fricatives have a successively dichotomous structure, which reduces the amount of information necessary to identify a phone to a minimum. The four stridents are split into two groups, one consisting of two compacts, the other of two noncompacts; one member of each group is tense, the other, nontense. Similarly, the four nonstridents are split into two groups, one consisting of two graves, the other of two nongraves; one member of each group is tense, the other, nontense. But perhaps the most important advantage of the analysis is that it simplifies the statement of a widely applicable morphophonemic rule in English. If [f] and [v] are construed as strident, then another feature has to be specified to subdivide the class of stridents so

¹This analysis was first suggested to me by Lee S. Hultzén, who has been using it for a number of years. Consonant inventories for English displaying this analysis appear in two articles of his listed among the references to this paper. As far as I know, he has not presented a formal argument for the analysis.

that the rule forming the plural and possessive of nouns and the third person singular present indicative of verbs can select just [č, ſ, s, z, š, ž]. This, in fact, is the solution proposed by Halle in his article "On the Bases of Phonology." But employing the analysis proposed here, we can correctly and simply account for the phonetic facts with the following ordered rule:

- (1) {z} → $\begin{cases} [\dot{t}z] & : / \text{strident} \quad \# / \\ [s] & : / \text{tense} \quad \# / \\ [z] & \end{cases}$, where Z = n. pl. or poss., or 3 sing. pres.

Until now our discussion has been mainly phonetic and has been concerned with certain physical and perceptual characteristics of stridency. I would now like to make some tentative remarks on the phonemic status of one of the stridents, namely /ž/ in English. Or, put another way, we are asking to what extent morphemes must be kept separate from other morphemes in the lexicon by contrastive feature representations in a single column of the distinctive feature matrix representations for morphemes, when that column alone will be mapped directly by phonological rules into phonetic [ž]. Let me emphasize that these remarks are tentative and should be considered more as questions than as assertions.

Chomsky (Current Issues in Linguistic Theory, pp. 71-3) has shown that ordered rules such as (2) and (3) are necessary in English.

- (2) (i) [t, d] → [s, z] in the context: ___ + [i, y]
(ii) [s, z] + [i, y] → [š, ž] in the context: ___ Vowel
(3) z → s in the context: ___ + iv²

Rule (2) applied as ordered accounts for permit--permission, exclude--exclusion, etc. Rule (2i) accounts for pirate--piracy, permit--permissive, etc. Rule (2ii) accounts for face--facial, impress--impression, disclose--disclosure, etc.

²Chomsky notes that there are obvious details and certain restrictions which need to be built into these rules, but these do not directly concern us here. For instance, rule (2i) should obviously be generalized to include k. The alphabetic symbolization is only an abbreviation for a set of feature specifications. Chomsky's notation [s₁, s₂] means that only features common to s₁ and s₂ are specified.

Since rule (2ii) is necessary for English and independently motivated, we can consider that such words as azure, leisure, glazier, and treasure contain systematically phonemic z + i or y. Rule (2ii) can then be applied to provide the correct phonetic output. Additional support for this analysis is readily seen in please--pleasure, seize--seizure, use(vb)--usual, and occasion (with underlying d)--casual.

The limited distribution of /χ/ has, of course, frequently been noted. It would appear, however, that that distribution is even more limited than is usually recognized. In fact, /χ/ has only a very marginal phonemic status in English, occurring perhaps only in final position and there in only a handful of morphemes.

One final comment and I shall close. Rule (3) is necessary to account for erosive, implosive, etc. which occur with [s]. Such words have, however, gone through an intermediate stage of derivation by rule (2i), in which they have been assigned a lax (voiced) continuant. Convinced that I frequently hear erosive, implosive, etc. with [z], I have checked a random collection of 22 such words. Jones' pronouncing dictionary of British English and Webster's Third indicate alternative pronunciations with [z] for all but three of these words: incisive, incursive, excursive. It may well be that there is dialect or least idiolect differentiation with respect to rule (3).

REFERENCES

- Chomsky, Noam. Current Issues in Linguistic Theory. The Hague: Mouton, 1964.
- Halle, Morris. "On the Bases of Phonology." The Structure of Language. Edited by Jerry A. Fodor and Jerrold J. Katz. Englewood Cliffs, New Jersey: Prentice-Hall, 1964.
- Hultzeń, Lee S. "Consonant Clusters in English." American Speech, XL (1965), 5-19.
- Hultzeń, Lee S. "Degrees of Difference of English Consonants." Proceedings of the Fifth International Congress of Phonetic Sciences. Basel and New York: S. Karger, 1965.
- Jakobson, Roman; Fant, C. Gunnar M.; and Halle, Morris. Preliminaries to Speech Analysis: The Distinctive Features and their Correlates. Cambridge, Mass.: The M. I. T. Press, 1952.

Miller, George A., and Nicely, Patricia E. "An Analysis of Perceptual Confusions Among Some English Consonants." The Journal of the Acoustical Society of America, XXVII (March, 1955), 338-52.

Strevens, Peter. "Spectra of Fricative Noise in Human Speech." Language and Speech, III (1960), 32-49.

Progress and Problems in Generative Metrics

Joseph C. Beaver

(Northeastern Illinois State College)

The purpose of the present paper is to summarize the status of recent theoretical work in linguistic analysis of meter (to this group, perhaps largely unfamiliar work), and to indicate some of the problems encountered by those doing research in this field.

The theoretical principles underlying the recent metrical analysis were first formally enunciated by Morris Halle and Samuel J. Keyser, in a 1966 study entitled "Chaucer and the Study of Prosody" (College English, December, 1966). As is the case with many other "new" movements, it is possible to relate their work to that of earlier theorists. Seeds of the present theory are to be found as early as George Gascoigne's Certain Notes of Instruction in English Verse, at the very outset of the period of modern English. Skipping many intervening suggestive theoretical statements, we find Otto Jesperson adumbrating one of the key concepts of the new analysis, the concept of the "stress maximum," in his "Notes on Meter", first delivered and published in Danish in 1900, and not translated until the early 1930's. Here Jesperson points out that the first syllable of a line can not be judged to bear ictus until the second syllable occurs--because there is nothing preceding for purposes of stress comparison. For this reason, an initial trochee in iambic verse may be tolerated, for it disappoints --to use Jesperson's word--only in the second syllable, not in the first. If it happens that the third syllable carries even less stress than the second--possible because Jesperson recognizes four degrees of stress--then the line, though it begins with an inverted foot, may have only 10% disappointment, as compared to the 20% disappointment occasioned by an initial iamb in a trochaic line. The recognition of stress relationship to adjacent syllables and the concomitant notion that stress next to nothing can not be optimum stress become cardinal points in the rules evolved by Halle and Keyser.

Halle and Keyser assume that there are rules underlying regular metered verse which will distinguish metrical from unmetered lines. The traditional foot concept does not suffice to make this important distinction: there is available no framework to account for the abundance of exceptional or irregular feet to be found throughout the

history of metered verse. Traditional prosody must be content with a mere listing of these exceptions--it has no principled way to account for them, or to establish a hierarchy of metricality. With certain poets or particular poems, for example, it may be found that the number of inverted initial feet in iambic lines exceeds the number of regular initial feet. Halle and Keyser conceive of the line of poetry as a sequence of positions--in the case of iambic pentameter, a sequence of 10 positions--with certain positions designated as weak, and certain ones as strong--in iambic pentameter, again, the odd positions would be weak, and the even positions strong. Stress maxima may occupy only even (or strong) positions, but not all even positions need be occupied. A stress maximum is defined as linguistically determined stress greater than that found on both adjacent syllables. Since weaker stress is required on both sides for the actualization of a stress maximum, a stress maximum cannot occur in the last position of a line, or in the first--which is why a reversed initial foot in iambic meter is not irregular. Thus the line "Oh that this too too solid flesh would melt" shows both that an inverted initial foot is not irregular, and that not all even positions need be occupied by actualized stress maxima--in fact, only positions 6 and 8 are so occupied.

This is the basic metrical rule, or principle, and it is a prescriptive one. Required also are stress rules. These may change as the language changes, while the rules of meter presumably might remain unchanged as long as the language remains a syllabo-tonic language. In their study of Chaucer's meter, accordingly, Halle and Keyser offer rules of stress only for Chaucerian English.

In an article published early this year (Joseph C. Beaver, "A Grammar of Prosody," College English, January, 1968) I employed the Halle-Keyser principles of meter in conjunction with a set of stress rules of my own devising for purposes of preliminary analysis of some regular metered verse in English over the past three centuries. Syllables are classified in four types: 1) full vowelled monosyllables with non-reducible vowels ("straight," "book," "rush"); 2) stressed syllables of polysyllabics; 3) monosyllabics with reducible vowels (in general, most non-lexical monosyllabics--function words and pronouns); 4) syllables in polysyllabics other than the one with main stress. Stress maxima are then determined by examining any sequence of three syllables occupying three consecutive positions in a line. If the center syllable of the three is of either type 1 or type 2 while the bordering syllables are of type 3 or 4, the center syllable is assumed to carry a stress maximum.

Thus it could be said that the stress rules are lexical, though, since the framework of three slots is a moveable one

applying to any three positions in turn, I prefer to think of the rules as quasi-lexical and quasi-phrasal. In any event, Halle and Keyser propose to supplement lexically based rules such as these with phrasal and clause stress rules. Presumably "The Sound Pattern of English," when it appears, will provide the full set. Such a system will find somewhat more actualized stress maxima, usually if not always in situations where my rules find no stress maxima because of back to back lexical stress. In "he chops wood," for example, my system will not assign prior stress to either "chops" or "wood". A system that assigns clausal stress will presumably assign greater stress to one or the other. It is not at present clear to me that such rules need be a part of the metrical rules determining stress maxima.

Using, then, this system of stress rules and the Halle-Keyser system of metrical rules, I conduct a sampling of English verse and find that actualization of stress maxima for iambic pentameter appears to be about 50 %. This did not vary greatly for any period I examined, though selections from individual poets sometimes ran less or more. Verse lines of lesser length than pentameter (collectively, I call them short-line verse) show a higher density of stress maxima, in fact, in the neighborhood of 65 to 70 % actualization. This phenomenon seems to have something to do with the fit of language units to line length. And incidentally, since trochaic meters show no significantly higher actualization of stress maxima than short-line iambic verse, an answer is suggested to the old, much-debated question of the difference between trochaic and iambic meters--namely that there is no essential difference other than that occasioned by the fact that stress maxima fall in odd positions in the one and in even in the other. Other fancied differences have arisen from an unconscious comparison of short-line verse (in which trochaic meter usually occurs) to long line pentameter verse (the customary habitat for the English iambic).

In stylistic analysis, an examination of ten of John Donne's Holy Sonnets for stress maxima showed a predilection for positions four and eight in Shakespeare, with attendant medial line stability, against two and eight in Donne, with medial instability. More unmetrical lines also are found in Donne, though a further refinement of the rules of synaloepha might possibly reduce the number.

In a forthcoming article, "On the Primes of Metrical Style," (to appear in Language and Style), Donald C. Freeman examines the development of iambic pentameter, with respect to stress maxima and related matters, at the height of the English Renaissance--the last half of the sixteenth century, when modern English blank verse first flowered. He finds it necessary to enlarge the Halle-Keyser rules of meter by

extending the list of consonants over which synaloepha may occur (in Halle-Keyser, only over sonorant consonants) to include the voiced labio-dental, interdental, and alveolar fricatives. He also finds it necessary to add what he calls a principle of "hypersyllabification," which is the occupancy of two positions by one syllable. I suspect that his additions hold good as well for at least certain poets of the first half of the seventeenth century. Working from Gascoigne to Marlowe, Freeman finds that what he calls the neutral (average) realization of stress maxima gradually reduces to about three per line. Specifically, of 400 lines in Gascoigne, 52 % showed four stress maxima per line, and 31 % showed three; 400 lines of Tamurlaine, on the other hand, show this picture reversed, with 34 % having four stress maxima per line, and 43 % having only three.

Freeman suggests that a "roughness" sets in, when the number of stress maxima falls below three per line, and he postulates that from 1600 on, English verse probably tends to swing back and forth between a higher and lower percentage of actualization. This may be right. However, my own perhaps somewhat spotty sampling found that a level of about 50 % actualization was finally reached. Any subsequent period variations I think must be fractional, in terms of one more or one less stress maxima per line. It is quite possible that to the modern ear (meaning the ear of the last three centuries) a variation of 10 % actualization in iambic pentameter is sufficient to produce the effect either of roughness or lock-stepping. In recent scensions of selected poets, I have found from 40 to 45 % in John Donne and in some of Browning's monologues, and up to 57 % in Pope's Essay on Criticism. It is also possible that my somewhat lower figures for actualization as a whole are occasioned by the strictness and comparative lack of flexibility of my quasi-lesically based stress rules. However, I should add that though Freeman does not always indicate the precise nature of the stress rules he uses, I find that my own scensions do coincide pretty well with his.

Freeman's paper continues with some preliminary metrical stylistic analysis. I would like to quote in part his statement of the logical progression for the study of English metrical style indicated by the present theory.

Such a study must begin with a general theory for syllabotonic meters and a description of the typical realization rules for a given period or genre, that is, particular variations of the basic "mapping rules," setting the context for the consideration of metrical style. Studies of metrical style then should assess typical practises of a poet with respect to certain primes of metrical style elicited by this theory.

These primes are a poet's preference for actualization of stress maxima, the disposition of these stress maxima, stress neutralization, catalexis, synaloepha, and ...hypersyllabification. ...Finally, the study of metrical style should proceed from these general backgrounds to consider how the metrical structure created by the disposition of actualized stress maxima and the complications of the neutral pattern works in conjunction with other structures--of rhetoric, of images, or sounds--in an individual poem. ("On the Primes of Metrical Style," mimeographed, 1967.)

Preliminary steps in these directions have already been taken, as I have indicated, by Halle and Keyser for Chaucerian meter, by Freeman for the 16th century, and by Beaver for some poetry since the 16th century. More work is forthcoming. With respect to what density of occupancy is characteristic of different verse types (Freeman refers to this as "most neutral realization"), I have undertaken some additional scansion to confirm earlier generalizations that I was a little afraid of. Working with some 1200 lines of randomly selected poems in iambic tetrameter, trimeter, and dimeter, and with trochaic tetrameter and trimeter, I find that the percentage of actualization in all these short line meters does run substantially higher, on average, than in iambic pentameter. Iambic tetrameter shows 62 % occupancy, while trimeter and dimeter show 69 % and 65 % respectively. (In "A Grammar of Prosody", I found a somewhat higher percentage) These latter figures do not seem to me to differ significantly from the 71 % occupancy I found in the trochaic verse. Since in many stanzaic trochaic poems typically half the lines have masculine endings, with a consequent reduction of the number of positions open for stress maxima, a stanzaic trochaic tetrameter poem seems to me to occupy in this respect a niche somewhere between iambic tetrameter and iambic trimeter. I can only conclude that a trochaic poem is no more lilting, no more jogging, has no more of a "falling rhythm" (Jesperson's phrase) than its iambic counterpart of the same line length. If it is felt to be any more of any of these things, it is because we are unconsciously comparing its density of occupancy by stress maxima to that found in iambic pentameter, with its neutral realization of about 50 % over the past three centuries.

In the balance of this paper I should like to indicate briefly what seem to me the principle current problems facing those who work with a linguistic theory of poetic language which, in Freeman's words, "does more than catalogue the history of tastes--which is, in short, mentalist and

generative rather than empiricist and taxonomic."

One of the issues, as already indicated, is the precise nature of the rules of stress. The rules I have used are of a quasi-lexical variety. Thus in the two trochaic lines from Longfellow,

Till at length a small green feather
From the earth shot slowly upward

my rules find stress maxima only in the third position of the first line and the 7th position of the second line, for only those odd-positioned syllables are linguistically stressed (the first is a full vowelled non-reducible monosyllable, and the second is the stressed syllable of a polysyllabic) and sandwiched by unstressed syllables. Halle and Keyser have contended (personal communication) that phrasal stress rules would find "small" and "slow" to be stress maxima also, and that since these positions are available for stress maxima, being odd positions in trochaic meter, their rules would be stronger than lexically based rules. However, since the rules of meter provide merely that a metrical line is a sequence of n positions in which stress maxima may not occur in even positions (or odd, as the case may be), it is not clear to me how a theory of stress assignment can be said to be stronger merely because it finds more stress maxima; this implies that the more stress maxima the better--but the rule of meter is prescriptive only.

Conceding that phrasal stress rules operate in English, one may still wonder whether the weight provided by phrasal stress on "small" is sufficient to contrast it maximally with "green," or that phrasal stress on the first syllable of "slowly" is sufficient to set it in maximal relief with "shot." It is interesting that Freeman on occasion also adopts, in effect, lexically based stress rules, specifically in situations where he finds that not to do so would result in unmetrical lines. It seems to me that in fact it is phrasal stress rules which undergird traditional prosody, with its sole reliance on the foot concept. Perhaps it is by excessive application of phrasal stress that so many feet may be made to fit the particular foot pattern.

At one further remove stands the issue of whether special contrastive stress is to play a part in determining stress maxima. Halle and Keyser think that it should, though contrastive stress plays only a small role in their study of Chaucer. I have maintained ("A Grammar of Prosody," pp.318, 319) that special meaning stress does not affect disposition of stress maxima. For example, in Hamlet's "Oh that this too too solid flesh would melt," I claim that the very likely special stress on the second "too" does not in fact make it a stress maxima, or even cancel the stress maxima carried by

the first syllable of "solid" for the vowel of "too" has the capacity for reduction. The point is not conclusive: perhaps a study of special meaning stress in metered verse might reveal its actual metrical role, if any. In Browning's lines

...Oh sir, she smiled, no doubt,
Whene'er I passed her; but who passed without
Much the same smile? ("My Last Duchess")

the pronoun "who" probably and the adjective "same" most certainly takes special stress in performance. But would one contend that the latter creates a stress maximum in an odd position, position 3, for an unmeterical line? At present I would not, and in general I would feel that to open the door to taxonomic stress interpretations would be a step backward in linguistic analysis of poetry.

Another problem has to do with the status of secondary accent in polysyllabic words. In the study of Chaucer's meter, no secondary accent is postulated for the Middle English under investigation. In my own study I ignore secondary accent, ostensibly for purposes of simplifying the investigation, but I suspect really because I feel that the stress relief occasioned by secondary stress can not be considered a stress maximum in the more general sense of that word. For example, in Poe's line

To the tintinnabulation that so musically wells

I find only two stress maxima, carried by the stressed syllables of the two polysyllabic words. But if one argued for secondary stress as a potential carrier of stress maxima, possible the first and third syllables of "tintinnabulation" would also be stress maxima, as might the third of "musically." Partial support for the argument against secondary stress is provided, it seems to me, by the fact that secondary stress can nearly always be suppressed (some phrasally assigned stress can evidently be suppressed also). For this reason, among others, I would prefer not to include secondary stress in the theory.

Another problem area is the definition of major syntactic juncture. The theory assumes that stress neutralization occurs at the boundary of a syntactic juncture--that is, a syllable adjacent to a major syntactic break cannot carry stress maximum. Halle and Keyser cite as examples of major breaks those found between interjections and following phrases, or between phrases introducing direct quotes and the direct quotes themselves, or between clauses in a complex sentence, or between items in a series. In most, if not all of these situations, however, it is easy to visualize extensive gray areas. The setting off of phrase units by juncture is in very many cases an optional rule. While it is nice to have an explanation for Wordsworth's line,

Ships, towers, domes, theaters, and temples lie

(one assumes stress neutralization between the items in the series with only the first of "temples" carrying stress maximum, and possibly the middle of "theaters," depending on British pronunciation of that era), how about the adjectives "small" and "green" in the Longfellow line cited earlier? Are they items in a series? If they are, neutralization occurs, which cancels out the stress maximum Halle and Keyser say is assigned to "small" by phrasal stress rules. It is partly because of the strong optional character of many rules of juncture that poets' punctuation is so untrustworthy a guide in ascertaining stress neutralization. What is needed in any event is something more stable than a poet's indication. Emily Dickinson does not punctuate the line "the stiff heart questions was it He that bore," yet surely a juncture is present after questions. For stylistic analysis this a very urgent question, for the density of occupancy of available positions by stress maxima is heavily affected at present by one's interpretation of juncture. It is to be hoped that a solution presents itself more detailed than any thus far offered, and a solution furthermore that is not performance oriented.

Within the stress rules that I have been using, there are problems as to what are the reducible vowels. Others have commented on the same problem. In general, the major categories of noun, verb, adjective, and adverb appear not to permit reduction of the stressed vowel in monosyllables, and in general the function words and the pronouns do. But it is sometimes difficult to withhold stress assignment from pronouns, and it is sometimes difficult to accord it to some adverbs (e.g., now, then, more). It seems counter-intuitive to maintain that "he" in Dickinson's "was it He that bore" does not carry stress maxima. In this particular case contrastive stress would be present on "he", and a possible solution would be that pronouns do not carry stress unless it is assigned by contrast--but not enough data is presently available to support this solution. A similar approach might be used for the monosyllabic adverbs.

The foregoing are in a manner of speaking practical problems arising from as yet imperfectly developed aspects of the theory. In addition to these, there are some more purely formal problems, two or three of which I shall mention only briefly.

It is postulated that the iambic pentameter line consists of 10 positions, to which an extra-metrical syllable may be added (the familiar feminine ending). Ordinarily, then, a stress maximum cannot occur in the tenth position, except

when the optional eleventh is added. However, what about the reality of enjambment, of run-on lines? When one line proceeds directly into the next without syntactic break, are we to say that the last position of the first line does not carry stress maximum, even though it may be bordered by lesser stressed syllables in position 9 of the first line and position 1 of the second line? It seems to me that at least at present we must maintain this. One of the most striking features of the present theory is its spectacular explanation of inverted initial feet in iambic verse. These are entirely regular by the theory. But if we are to say that we may consider bordering syllables across line boundaries in determining stress maxima -join lines together, so to speak- we would lose that explanation, it seems to me. Conversely, the great number of inverted initial feet seem to me to support the abstract concept of the verse line as a sequence of positions only, and a unit unto itself by virtue of that fact.

Secondly, the original statement of the theory allowed for zero occupancy of certain positions under certain conditions, but as it turned out the only position with zero occupancy Halle and Keyser dealt with was the first position (the familiar beheaded line). It seems likely that a fuller formulation of rules may have to reckon with other unoccupied positions, and state conditions for these. This may prove difficult if it is found that poets have significantly used other positions for zero occupancy, as I suspect is frequently the case with the triple meters, anapestic and dactyllic. Freeman, in another context, has already provided an approach to this problem with the concept of hypersyllabification.

Finally, though the theory is not required to find stress maxima in all available positions, still by a reductio ad absurdum it is a little embarrassing to find the possibility of verse with no stress maxima. Halle and Keyser have pointed out (personal communication) that in a three syllable line with strong-weak-strong pattern, for example Tennyson's

Live thy life,
Young and old
Like yon oak,
Bright in spring,
Living gold.

there would be no stress maxima at all, unless some line had a stress maximum in the second position, in which case of course that line would be unmetrical. Among the solutions being considered are to require that in such meters, in place of stress maxima, merely stressed syllables be restricted to occurring in strong positions. This would enable us to establish a hierarchy among realization rules. Another answer that has

occured to me is that real line length does not necessarily correspond to the line length as the poet arranges his verse visually. I think there are many instances of visually long line poems which are in reality faultily divided ("Locksley Hall," and "The Raven," for example), and it is at least possible that there is no such thing as a regularly metered poem with lines of only three positions.

I do not think any of these problems are insoluble. I have indicated possible solutions to some of them, and a number of personas are addressing themselves to these and to other solutions. Meanwhile I believe the theory does provide a logical procedure for metrical analysis, and that it can lead ultimately to meaningful and significant statements, rather than purely impressionistic ones, about metrical stylistics.

PERCEPTION VERSUS COGNITION IN LINGUISTIC THEORY
AND DEVELOPMENTAL PSYCHOLOGY

Wilbur A. Hass
(University of Chicago)

A basic conceptual polarity which linguists have used is that between linguistic material---the "stuff" of language---and its meaning. Such a distinction may be found in the terminology of most linguistic theoreticians: Saussure's (1959) decomposition of the linguistic sign into sound-image (signifiant) and concept (signifie); Jespersen's (1965) correlation of form and notion; Hjelmslev's (1961) polarity between expression and content; Pike's (1954-60) distinction between feature and manifestation modes; and Chomsky's (1966) assertion that linguistics deals with the association between sound and meaning---to give some examples. Once such a distinction is formulated, one tends ---despite emphasizing the interdependence of the dyad ---to suppose that languages must include some rules (or patterns) which concern the organization of the linguistic material, and some which center around the meaning aspect. This distinction is given concrete expression in the actual organization of linguistic descriptions into phonological and semantic sections.

How does a dichotomy of this type relate to psychological theories? That is, if a speaker of a language is presented with a stretch of language, do his operations on it break down into (at least) two sorts: one having to do with the phonological nature of the stimulus, and the other having to do with its semantic interpretation? Or, in other words, does the performance of the speaker-listener rest on or utilize (at least) two kinds of competence, one phonological and the other semantic? Of course, it just might be the case that linguistic descriptions could be efficiently organized in this way, even if it were not a reflection of how language users actually worked. However, the question of "psychological reality" of the distinction---or, perhaps, its "explanatory adequacy"---is the one discussed in the following pages.

Psychological theorizing does include a dichotomy which seems to parallel the one between linguistic material and meaning; it is the distinction between perception (which specifies "what a stimulus is") and cognition (which determines "what it means"). Thus, if someone tells me that he studied tree-pruning at Oklahoma A & M, I both hear what he said (hearing being a perceptual modality) and know what he talked about (knowledge being the end-product of cognition). Both types of activity may continue for a time after the stretch of language has been encountered; I may "re-play" what the person said, or I may think about it.

This dichotomy, although grossly evident in human self-awareness, has by no means been consistently made in psychology; in fact, schools can be sorted out in terms of their treatment of the distinction (Wohlwill 1962). Gestalt psychologists favored treating all dealings with stimuli in terms of concepts and principles derived from work on perception, so that Wertheimer's analysis of syllogistic reasoning talks about the laws of good form. The same tendency is present in the application of "person perception" to areas of social cognition (Tagiuri & Petrullo 1958). Conversely, Bruner has consistently applied terms from the cognitive realm (such as "hypothesis" and "inference") to his analysis of perceptual phenomena; and Neisser (1967) has entitled "Cognitive Psychology" a book dealing mainly with perception. Both trends represent a playing out of divergent epistemological roles, typified respectively by empiricists like Locke and Hume, the latter of whom said that to judge that the sun is shining is to vividly picture the shining sun; and by rationalists like Leibniz and Spinoza, the latter of whom said that seeing something happen is a mental operation analogous to drawing a conclusion.

The mainstream of American behaviorism has not made any major distinctions in the area. Osgood's early discussion (Osgood & Sebeok 1954:127) of the learning of meaning portrays it as "inseparable" from the development of perception, since both are held to involve classical conditioning. Skinner's (1953) position would I believe imply that discrimination and differentiation can apply to stimuli or responses on the basis of any characteristics. Surprisingly enough, the same lack of distinction is implied in Lenneberg's (1967) extended application of the concept of "transformation" to seeing Necker cubes and naming, as well as dealing with ambiguous sentences.

Such confusion might make one suspicious that the material which textbook chapters divide into perception and cognition may be fundamentally the same, so that it might better be lumped together, using any label one happens to prefer. This feeling grows when one realizes that for any basic intellectual content---e.g., space, causality, emotion---the ordinary judgments of an adult may involve weighted combinations of various perceptual and cognitive evidence. If a subject is presented with a coin and asked how big it is, his response will be influenced by what he "knows" about the size of such objects, as well as how big it "looks." Furthermore, the products or results of perception and cognition share a great many features: both involve "coding," produce categories (percepts versus concepts), maintain invariance over change in irrelevant object attributes (constancy versus conservation), are susceptible to "set," can be "satiated," store operations and products, and so forth. Such similarities as these have led Heinz Werner (1948:213-216) to refer to perception and cognition as "analogous processes," which do in fact lead to the same sorts of achievements. In a parallel way, moving about may be accomplished by crawling, walking, driving, riding a bicycle, or many other ways; and in each case one may speak of distance covered, speed, obstacles, etc.

However, also involved in the notion of analogous processes is the stipulation that the means of attaining the results involves different function patterns. This is obviously true in the case of the means of locomotion just mentioned. But is there reason to think that perception and cognition actually differ in the way they operate? Egon Brunswik (1956:89-99) has pointed out certain very evident properties which lead one to reconsider lumping the two together. He gave people the task of determining how big a square patch presented to them was---one group, the perceptual one, by looking at it and saying how big it appeared; another group, the cognitive one, by using a formula to calculate size from data on distance and retinal area covered by the image of the patch. The perceptual group, while they seldom got the size exactly correct, were never far off. On the other hand, the cognitive group were overwhelmingly right to the inch---except for a few who made errors, which led in some cases to fantastically wrong answers. Brunswik concludes that perception is devised, via evolution, to contain safeguards which ensure its usually being "about right," operating by the parallel processing of many interlocking, probabilistically weighted cues; whereas

cognition typically operates along serial "paths" which lead to points which are either quite strikingly right---or wrong.

Also indicative of possible differences is a long line of reaction time studies (*cf.* Woodworth & Schlosberg 1954:ch.2). The task the subject has to perform with respect to the stimulus definitely affects the time of exposure he needs or the time he takes before responding; and the more obviously "cognitive" tasks take longer than ones that seem more "perceptual." A recent example is Oldfield's (1966) study, in which he presented a picture and either asked the subject to name the object portrayed or asked him to tell whether it was the same or different than another picture. Not only does it take longer, in general, to give names than to judge perceptual identity; but latency in the naming task increases with rarity of the name in the language, whereas latency on the perceptual task does not.

The most extensive set of proposals regarding the perception-cognition distinction has been made by Jean Piaget (1950:ch.III; Piaget & Morf 1958). The operations involved in perception, in his formulation, never attain the mobility and flexibility of operational thought. The very patternedness of perceptual Gestalts, which is so prominent, may be seen as a result of this interdependence between components of perceptual acts. Such components of perception, whether simultaneous or serial, have an effect on each other which leads perception to be relative in a unique and distorting sense. For example, Piaget cites lack of reversibility---a percept, once attained cannot be deliberately undone, in any manner similar to the way in which one can trace back a line of thought---as well as lack of associativity---putting one's finger in cold water and then hot does not lead to the same two temperature percepts which are achieved, in reverse order, by putting it first in hot and then in cold.

Some additional bits of evidence bearing on the perception/cognition distinction can be gleaned from the classical experimental areas of learning and motivation, particularly with respect to the operation of "set" and "satiation" mechanisms. However, let us instead return to linguistics and see if what has been said makes any sense. In terms of phonology, it is clear that the sounds of language can themselves become the objects of cognition---or otherwise we would not have an area of phonology to worry about. Usually though, little cognitive

notice is taken of the sounds of any utterance, and they are utilized perceptually, while cognitive operations are being directed toward dealing with the message. Now if Brunswik's characterization of perception holds in general, the extremely messy relation between acoustic cues and phonological distinctive features that has been described by investigators like Delattre (1967) should, unfortunately, be expected---if we have nine perceptual cues available for visual distance, why shouldn't there be several spectrographic characteristics, rather than a single criterial one, involved in perceiving voicing? Likewise, given interdependence as a primary property of perception, one would expect the extreme amount of influence of contiguous sound units that one in fact does find described in increasingly more comprehensive phonologies.

On the other hand, one might expect considerable divergence between phonological and semantic descriptions. As rules governing the cognitive activity of the linguist, one should be able to characterize both areas by Boole's "Laws of Thought"; but as structures describing the phenomena themselves, the "laws of thought" involved in semantics may well turn out quite differently from "laws of perception" embodied in phonology. Of course there is a good deal of wisdom as well as comfort in assimilating the unknown (in this case, semantics) to the better known (in this case, phonology). Certain basic descriptive terms like "feature," "marked/unmarked," and "rule" may well be useful in both areas. But my point is that the linguist may well expect not find strict isomorphism between rules in the two areas.

A crucial question is one which has been carefully avoided until now: where does syntax fit into the picture. At first blush, a syntactic portion of a grammar seems to come into being by virtue of the fact that there is such an obvious gap between phonology (what is heard) and semantics (what is understood). It turns out, however, that most performance models for syntax quite unabashedly use perceptual terms, tasks, or paradigms---or all three. This holds for Lashley's (1951) perceptual-motor schemas, Osgood's (1957) integrative tuning mechanism, Epstein's (1961) chunking device, and Garrett, Bever & Fodor's (1966) active perceiver. Possible exceptions are Yngve's (1960) mechanism, which, although it leans toward what appears to be perceptual features in Miller's "magical number seven," does maintain what I take to be a studied neutrality; and Miller's (1962) schema-plus-correction, although it is derived from Woodworth's discussion of chiefly perceptual data. This predominance of perceptual models seems unfortunate, insofar as the syntactic data themselves appear to

involve far greater independence of components and preservation of nature under re-ordering than one would have expected. More recent attempts to base syntactic analyses on criteria derived from semantic naturalness seems to be a promising corrective. In this connection Piagetian description of cognitive structure in terms of networks is interesting. Lack of time and ignorance restrain me from going further into this most fascinating topic.

Lastly, let us consider the developmental features of the perception/cognition dichotomy. This is of course one of Piaget's main concerns, and divergence of perceptual and cognitive development is probably the strongest evidence for their separation. His own investigations of the development of cognition in the child are well known (cf. Flavell 1963). Less well known, but equally well-founded is his own and others' documentation of perceptual development. These perceptual changes with age do not merely involve increases in veridicality, but alterations in the nature of perceptual activity, as is evident from reactions to presentations of the same perceptual material in company with changing contextual cues. The changes do not seem to involve new forms of perceptual organization (in the sense that concrete and formal operations appear in cognition). The most general perceptual trend is one from a general, unanalyzed reaction to a general, unanalyzed situation (which may be called physiognomic perception) toward greater differentiation of both stimulus and response. The amount of perceptual development actually causes trouble for classical empiricist formulations, which have usually taken perceptual data too much for granted.

The present picture of the infant then is not one of an organism who perceives like an adult, but merely lacks certain additional cognitive operations; rather it is one of a person whose operations on input are best characterized as intermediate between the perceiving and cognizing of adults. Insofar as this is true, the child, in acquiring language is making neither the semantic nor the phonological judgments of an adult. His task in structuring the language he hears and produces is not primarily that of associating adult sound-percepts with adult meaning-concepts, but separating language material (as "stuff" for developing perceptual operations) from language meaning (in the functioning of developing cognitive operations).

This view of early perceptual/cognitive operations in childhood serves as a basis for Werner & Kaplan's (1963) assertion that the young child makes roughly the same sort of physiognomic reactions to language as to the things language comes to represent for him. Language acquisition, then, takes place by utilizing the considerable overlap between "perceptual/cognitive" reactions to what will be the language material on the one hand and to what will be represented by it on the other. Only later in development does the child come to respect the asymmetry which takes sounds as the "neutral" perceptual matter in order to transmit messages about states of affairs (rather than vice versa). This opinion leads to a rather specific and unusual interpretation of the statement that the child is a "natural language scientist"---he is one because he has not learned the importance of the cognitive unimportance of language material in communication.

The question of perceptual/cognitive factors comes up again in connection with syntax acquisition. This matter has recently been discussed from a slightly different point of view (Hass 1968). Here it will suffice to note that many transformations are interpretable as being cued off by semantic markers whose inclusion is determined by the communicative roles involved in the speaker-hearer dyad, so that syntactic development may reflect the development of social-cognitive competence.

In conclusion, I hope to have called your attention to a basic split between perception and cognition that we as psychologists or linguists tend, at least at some times, to make either explicitly or implicitly. There is some psychological evidence, although certainly not a great deal, to substantiate, at least for higher developmental levels, the functional importance of the split. The chief problems for psycholinguistics which arise out of it are: how to know when to resist the pressure toward isomorphic descriptions of aspects of language which lie at different places relative to perception/cognition, and how to deal with the relation between development in both kinds of operations and the nature of language development. The day is not at hand when specific information on the form of linguistic descriptions can be gained from psychological principles, but movement in that direction (from the ad hoc proposals exemplified in the above) can be undertaken.

REFERENCES

- Brunswik, Egon. 1956. Perception and the representative design of experiments. Berkeley, Univer. of California.
- Chomsky, Noam. 1966. The formal nature of language. Biological foundations of language, by Eric Lenneberg, App.A. New York, Wiley.
- Delattre, Pierre. 1967. From acoustic cues to distinctive features. Paper given at International Congress of Phonetics, Prague.
- Epstein, William. 1961. The influence of syntactic structure on learning. Amer. J. Psychol. 74.80-85.
- Flavell, John H. 1963. The developmental psychology of Jean Piaget. Princeton, Van Nostrand.
- Garrett, Merrill, Bever, Thomas, & Godor, Jerry. 1966. The active use of grammar in speech perception. Perception & Psychophysics 1.30-32.
- Hass, Wilbur A. 1968. Syntactic development in children. Paper given at Chicago Linguistic Society.
- Hjelmslev, Louis. 1961. Prolegomena to a theory of language. Madison, Univer. of Wisconsin.
- Jespersen, Otto. 1965. The philosophy of grammar. New York, Norton.
- Lashley, Karl S. 1951. The problem of serial order in behavior. Cerebral mechanisms in behavior, ed. by L. A. Jeffress. New York, Wiley.
- Lenneberg, Eric H. 1967. Biological foundations of language. New York, Wiley.
- Miller, George A. 1962. Some psychological studies of grammar. Amer. Psychologist 17.748-62.
- Neisser, Ulric. 1967. Cognitive psychology. New York, Appleton-Century-Crofts.
- Osgood, Charles E. 1957. A behavioristic analysis of perception and meaning as cognitive phenomena. Contemporary approaches to cognition. Cambridge, Harvard University.
- Osgood, Charles E. & Sebeok, T. A. (eds.). 1954. Psycholinguistics. J. abnorm. soc. Psychol. 49, suppl.
- Piaget, Jean. 1950. The psychology of intelligence. New York, Harcourt, Brace.
- Piaget, Jean & Morf, A. 1958. Les isomorphismes partiels entre les structures logiques et les structures perceptives. Etudes d'epistemologie genetique 6.49-116.
- Pike, Kenneth L. 1954-60. Language in relation to a unified theory of the structure of human behavior. Glendale, Summer Institute of Linguistics.
- Saussure, Ferdinand de. 1959. Course in general linguistics. New York, Philosophical Library.

- Skinner, B. F. 1953. Science and human behavior. New York, Macmillan.
- Tagiuri, R. & Petrullo, L. (eds.). 1958. Person perception and interpersonal behavior. Stanford: Stanford University.
- Werner, Heinz. 1948. Comparative psychology of mental development. New York; International Universities.
- Werner, Heinz & Kaplan, Bernard. 1963. Symbol formation. New York, Wiley.
- Wohlwill, Joachim F. 1962. From perception to inference: a dimension of cognitive development. Monogr. Soc. Res. Child Develpm. 27 (#2).87-112.
- Woodworth, Robert S. & Schlosberg, Harold. 1954. Experimental psychology. New York, Holt.
- Yngve, Victor H. 1960. A model and an hypothesis for language structure. Procs. Amer. Philos. Soc. 104.444-66.

The Notion of Usage in Vaugelas and in the Port Royal Grammar

Keith Percival
(University of Wisconsin-Milwaukee)

In his monograph Cartesian Linguistics, Chomsky among other things attacked the following problem: What are the historical sources of the ideas contained in the Grammaire Générale et Raisonnée, the so-called Port Royal Grammar? His conclusions were as follows: On the positive side the Grammaire Générale was the outgrowth of a new Cartesian approach to language which had two main characteristics: first, an emphasis on the creative aspect of language use, and second the view that language presents two distinguishable and not necessarily congruent aspects, namely deep and surface structure. On the negative side this new approach to language was a reaction against a kind of pure descriptivism which had held sway in the period before the publication of the Grammaire Générale. It is this last of Chomsky's hypotheses that I should like to examine in this paper.

If Chomsky is correct in assuming that the linguistics of Port Royal developed as a reaction to prevailing descriptive approaches we should expect to find that the predecessors of Arnauld and Lancelot shied away completely from explaining the facts of language or discovering any general principles underlying actual linguistic usage, in other words that they confined their attention to presenting actually observed usage. We should also expect to find that the Grammaire Générale would contain criticisms of this fallacious descriptive approach to grammar and an attempt to defend a different approach to the role of usage in grammatical description.

Let us first examine the ideas of Claude Favre de Vaugelas, the French academician and grammarian whose name is mentioned more often than any other in the Grammaire Générale. His book Remarques sur la Langue Françoise (1647) is a long series of discussions about disputed usage. The preface presents a detailed analysis of the notion of usage in all its aspects.

Vaugelas begins by pointing out that usage is the final court of authority in language, 'the master and sovereign of living languages'.¹ I might point out parenthetically that this is a classical point of view familiar from the writings of Quintilian and other Latin authors.² Usage is of two varieties, however, good and bad. Bad usage is the speech of the lower classes, and the speech of the provinces. Good usage is the speech of the Court and the language of the best contemporary writers.³ Literary usage is the seal of approval authorizing the speech of the Court. Thus only those aspects of the

speech of the Court which are in accordance with the usage of the canonical authors are to be considered good and worthy of imitation.⁴

Admittedly, Vaugelas continues, many writers, especially outstanding ones, have a tendency to go their own way and violate good usage. This propensity is to be deplored for if it were indulged in to excess the French language would relapse into its former barbarous confusion.⁵ The only reasonable thing to do is to write like everybody else, and keep one's private inclinations in check.⁶ For Vaugelas, therefore, good usage is something which many individuals, even members of the Court and reputable authors, sometimes violate. While some of these violations are due to ignorance, others are due to consciously indulged whim. Good usage then is a norm of collective linguistic usage which a rational man willingly and consciously conforms to. It is possible therefore for individuals to improve their speech and writing by finding out what constitutes good usage and thereafter conforming to it.

Now if finding out what constitutes good usage were simply a matter of observing words and sentences in actual use this would not be much of a problem.⁷ But too often there are alternative usages which seem equally accepted. In other cases the problematic word or phrase occurs so rarely that no appeal to past observation is possible. What is to be done? In some cases of conflicting usages it is often sufficient to choose the one favored by the majority of speakers. But this becomes an impossible solution in cases where the alternatives are equally common, and where it is a question of adjudicating about a phrase which one has never heard or read before. Here the individual must reach a personal judgment. As for the first situation, choosing among equally acceptable alternatives is a stylistic matter in which considerations of euphony and clarity play a role. But what is to be done in the case of words one has never encountered before? When one is asked to express a judgment about the correctness of an unfamiliar word or phrase one must inevitably base the decision on what is the case in some analogous instance. Analogy is therefore resorted to when observed usage alone is not decisive.⁸ This is perfectly legitimate: analogy is nothing but a generalized form of usage, hence it can be validly applied to individual cases of contested usage.⁹ This is so because there is indeed a sufficient regularity in the things already observed to allow one to construct a model which can then be used to construct new things like the old ones.¹⁰ Analogy is a blueprint (Vaugelas uses the term patron) established not by individual caprice but on the basis of collective usage.

It is therefore not quite true to say that language

is based on usage alone; rather it is based on usage and analogy. But analogy differs from usage no more than a copy differs from its original.¹¹ Therefore one can say loosely that language is based on usage, i.e., already established usage and usage inferrable on the basis of analogy. In practice Vaugelas invokes analogy chiefly in cases of disputed orthography, as for example, inflectional endings such as the -s of the masculine plural which in certain locutions are never reflected in the pronunciation.¹²

But what is the relation between the notion of usage and that of grammatical rule? A rule expresses a regularity. When a rule does not apply where one might expect it to, Vaugelas says that in such an instance usage has triumphed over the rule.¹³ However, he does not leave the matter there. The fact that languages are not completely rule-governed endows them with an esthetic quality, a grace and charm which a completely regular language would not possess.¹⁴ Furthermore it stimulates the curiosity of the investigator to discover the reason why the rule does not apply in some particular instance. Thus in all cases in which language appears to operate irrationally it is pleasant to speculate on the factors, historical or otherwise, which gave rise to the anomaly.¹⁵ If offering a rational explanation for a particular exception to the general rule about the use of the relative pronoun he comments as follows: 'I do not know whether I have made myself understood, or whether if I am understood the reader will be satisfied by this little ratiocination and not find it too abstruse and metaphysical. But the example of the great Scaliger, who produced such beautiful ratiocinations on Latin grammar, emboldened me to do the same for French grammar, which I hope the reader will take in good part'.¹⁶ Scaliger was, of course, the author of a rational grammar of Latin, De causis linguae latinae, which had appeared a century earlier and had by that time become a standard treatment of the subject. (Scaliger is mentioned three times in the Grammaire Générale, for example.) In the preface Vaugelas confesses that if he had written a grammar, rather than a collection of stray observations, he would have organized the material in the order of the parts of speech which is prescribed by nature, 'as Scaliger the Elder admirably demonstrated'.¹⁷ Vaugelas' attitude to rational grammar can hardly be characterized as disparaging.

But were Vaugelas' actual rules mere reports of observed usage? Let us examine one of the rules he sets up, one which decides when articles and prepositions should be repeated before the second of a pair of coordinated substantives or verbs. The rule reads as follows: 'When two substantives joined by the conjunction and are synonymous or nearly so, as for example virtue and generosity, one

should not repeat the article, but when they are contrary or completely different in meaning, as for example strength and dexterity, one must repeat the article and say de la force et de la dexterité.¹⁸ Then in the next section he claims that the same rule serves to explain why two coordinated singular substantives govern a singular predicate if they are synonymous, for example, Sa clemence et sa douceur estoit incomparable, but L'amour et la haine l'ont perdu.¹⁹ Whether Chomsky would regard all this as a display of pure descriptivism I prefer not to surmise. To me at least these rules pose a serious question. Is there really a radical difference in the theoretical orientation between Vaugelas and Port Royal?

To answer this question let us turn to the Grammaire Générale itself and examine the references to Vaugelas contained in it. Do we find the authors explicitly criticizing Vaugelas' basic orientation to grammar? The criticisms they do offer are of two kinds. First, they criticize Vaugelas for failing to mention certain facts. For example, he failed to point out that the appropriate form of the object pronoun after an imperative is moi, toi, etc., and not me, te.²⁰ In other words, they accuse Vaugelas here of observational inadequacy.

The other type of criticism has to do with rules which fail to cover all the facts. Thus Vaugelas formulated a rule to the effect that a relative pronoun cannot refer to a noun which has no article.²¹ One may say, for example, il a été blessé d'une fleche qui estoit empoisonnée, but one may not say il a fait cela par avarice, qui est capable de tout. Vaugelas notes one exception to this rule, namely that if the noun is in the vocative case it may be followed by a relative pronoun, as in Avarice qui causes tant de maux. He emphasizes that this one exception does not invalidate the rule itself. The rule is still a genuine rule although it is subject to this exception. But he then goes on to claim that in this particular instance the exception is only apparent since in the vocative case an article is understood.²² In other words Vaugelas invokes the traditional notion of subaudition, the notion that there are linguistic forms which are not actually pronounced, but which nevertheless exist in the speaker's mind and therefore form part of the sentence.

The authors of the Grammaire Générale examine this particular rule very meticulously,²³ first producing a whole set of counterexamples, and then proposing a reformulation of the rule in more general terms which they believe takes care of all observed cases. The reformulated version reads: 'Qui should not be put after a common noun unless it is determined by an article or by something which determines it just as much as an article would'.²⁴ The concept of determination is then

discussed in general logical terms, and the authors are able to show for example, that the words a kind of serve to determine the words that follow them, and hence that it is correct to say a kind of fruit which is ripe in winter: '*C'est pourquoi c'est bien dit; une sorte de fruit qui est meur en hyver*'.²⁵ Finally they make the following very revealing remarks which bring us back to the topic of usage again: 'There may be other locutions which seem to violate the rule and which cannot be accounted for by any of the preceding observations, but they must be, I believe, remnants of the old style in which articles were almost always omitted. However, a maxim which students of a living language must always bear in mind is that any locution which is authorized by a general uncontested usage must be considered correct even if it conflicts with the rules and the analogy of the language, and that such exceptions should never be cited to cast doubt on the rules and disturb the analogy, nor to sanction other locutions which usage would not have authorized. Otherwise if one pays exclusive attention to the vagaries of usage, and ignores this maxim, the result is that the language will remain for ever uncertain, and lacking general principles it will never attain fixity'.²⁶ That these statements about usage echo Vaugelas hardly needs pointing out.

The reformulation of the rule is of course no mere echo of Vaugelas. What Arnauld and Lancelot succeeded in doing was to explain more facts than he had been able to, and they achieved this result by using a species of semantic or logical analysis which they regarded as their own invention. But Vaugelas was incapable of producing this kind of explanation not because he was totally averse to explanation, nor because he had no linguistic theory whatever to form a principled basis for explanatory attempts, but simply because his linguistic theory, such as it was, could not have accounted for all these facts in one explanatory swoop, as the authors of the *Grammaire Générale* could. Let us not make Vaugelas and the authors of the *Grammaire* seem more dissimilar than they really were. On the primacy of usage and the desirability of revealing explanation they were at least fully agreed.

It is furthermore obvious from some of the discussions in the *Grammaire Générale* that although they were rational grammarians the authors shared Vaugelas' view that not all features of normal linguistic usage can in fact be rationally accounted for. Thus they point out that the masculine and feminine genders were invented to mirror the difference between the sexes. But they hasten to add that inanimate nouns were also assigned genders, but in a capricious fashion (*par un pur caprice, et un usage sans raison*).²⁷ To make matters worse, the Greeks and Romans invented a third gender which they called the neuter, but instead of making all substantives

neuter which did not denote males or females, they again assigned gender capriciously (*par fantaisie*).²⁸

Again while the authors of the Grammaire Générale are able to explain the use of the auxiliary verb être to form the passive, they admit they are at a loss to account for the use of avoir to form the perfect tense, except to suggest that it may have come from German.²⁹ As for the use of être to form the perfect tense of reflexive verbs in French they comment: 'It is difficult to surmise where this usage has come from, for the Germans do not have it, using in this context the verb avoir, as usual'.³⁰ Like Vaugelas, therefore, they regard certain linguistic phenomena as having a raison d'être and as being to that extent capable of explanation, and other phenomena as difficult or even impossible to explain rationally. They differ from Vaugelas in attempting to explain as much as possible, and in the type of explanations they provide. Moreover they had an explicit general theory and they were writing a general grammar. In other words they were not academics laying down the rules of good usage for one language, but universal grammarians attempting to account for as much of usage as they could in all the languages they were familiar with.

In conclusion then the Grammaire Générale was not a reaction against a previously held linguistic theory. Arnauld and Lancelot did not repudiate Vaugelas, they transcended him completely, and in the process incorporated a great deal of his ideas. The term 'pure descriptivism' does not seem particularly apt as a characterization of Vaugelas' attitude to usage, and if Vaugelas is to be dubbed a pure descriptivist, then so must the authors of the Grammaire Générale for they too accept the classical view of the role of usage in grammatical description. And one may legitimately question whether the term 'explanatory', as Chomsky understands it, applies either to Vaugelas or the Port Royal grammarians. Perhaps it would be useful to avoid applying twentieth-century terminology to the intellectual products of the seventeenth century. I should like to recommend therefore that we do something which Chomsky explicitly disclaims any interest in attempting to do, namely try as far as possible to characterize the seventeenth century grammatical theorists as they saw themselves.³¹ I submit that this is not a completely unrewarding enterprise.

FOOTNOTES

1. The phrase Vaugelas uses is "le Maistre & le Souverain des langues vivantes." (Preface, section I). Note that since the preface is unpaginated references to it are made by section number. This characterization of usage amounts almost to a refrain, e.g.: "l'Usage est le Roy des langues, pour ne pas dire le Tyran" (16), "L'Usage est comme l'ame & la vie des mots" (60), "I'ay appris de nos Maistres, & du Maistre des Maistres, qui est l'Usage . . ." (226).

2. Quintilian 1.6.3; 1, 6, 16; Horat. Art. P. 71-2. Aulus Gellius 12, 13.

3. "C'est la facon de parler de la plus saine partie de la Cour, conformément à la facon d'escrire de la plus saine partie des Autheurs du temps" (Preface II, 3).

4. "Mais le consentement des bons Autheurs est comme le sceau, ou une verification, qui autorise le langage de la Cour, & qui marque le bon Usage, & decide celuy qui est douteux. On en voit tous les iours les effets en ceux qui s'estudient à bien parler & à bien escrire, lors que se rendant assidus à la lecture des bons Ouvrages, ils se corrigent de plusieurs fautes familières à la Cour, & acquierent une pureté de langage & de stile, qu'on n'apprend que dans les bons Autheurs". (Preface II, 5). That good usage is a product of artificial cultivation is an idea common to many grammatical theorists of the post-Renaissance period. Thus in his rational grammar of Italian, Delle Cagioni della Lingua Toscana, (1623), Benedetto Buommattei points out: "Il parlar degli uomini adunque è naturale, perche, ha principio dalla natura. Ma il come è artificiale, perchè e' dipende dall'arte, che lo raffina, e preserva dalle corrusioni dell'imperito e sconsiderato vulgo, il quale appoco appoco lo condurrebbe con irreparabil danno a certissimo fino, se la diligenza degli Scrittori non lo sostenesse, e gli fosse riparo continuo . . ." (Trattato settimo, Cap. II).

5. Vaugelas at one point criticizes authors for failing to observe a rule which he considers important and says: ". . . si chacun s'emancipoit de son costé, les uns à n'estre pas si exacts en certaines choses, les autres en d'autres, nous ferions bien tost retomber nostre langue dans son ancienne barbarie . . ." (218).

6. "Mais y a-t-il rien de plus facile que d'accommoder son esprit à la raison en des choses de cette nature, où il ne s'agit pas de combattre des passions, ny de mauvaises habitudes, qu'il est si difficile de vaincre, mais qui veut seulement qu'on suive l'Usage, & qu'on parle & qu'on escrive comme la plus saine partie de la Cour & des Autheurs du temps, en quoy il n'y a nul combat à rendre, ny nul effort à faire à qui n'abonde pas en son sens" (Preface III, 3).

7. Vaugelas distinguishes "usage déclaré" and "usage douteux" (Preface IV, 1).

8. Preface IV, 3-4.

9. "Cette Analogie n'est autre chose en matière de langues, qu'un Usage general & estable que l'on veut appliquer en cas pareil à certains mots, ou à certaines phrases, ou à certaines constructions, qui n'ont point encore leur usage déclaré, & par ce moyen on iuge quel doit estre ou quel est l'usage particulier, par la raison & par l'exemple de l'Usage general . . .".

10. Again speaking of usage, Vaugelas says: "c'est une ressemblance ou une conformité qui se trouve aux choses desia establies, sur laquelle on se fonde comme sur un patron, & sur un modelle pour en faire d'autres toutes semblables" (Preface IV, 4).

11. "De tout ce discours, il s'ensuit que nostre langue n'est fondée que sur l'Usage ou sur l'Analogie, que comme la copie ou l'image l'est de l'original, ou du patron sur lequel elle est formée, tellement qu'on

peut trancher le mot, & dire que nostre langue n'est fondée que sur le seul Usage ou desia reconnu, ou que l'on peut reconnoistre par les choses qui sont connuës, ce qu'on appelle Analogie" (Preface V, 1).

12. See the discussion of prendre à tesmoin, Preface IV, 3, and in the body of the work 563ff. Vaugelas' discussion of analogy is consciously modeled on that of Varro, De Lingua Latina, Books VIII and IX. See Vaugelas' own references to Varro in the Remarques 453, 470.

13. Vaugelas speaks often of "l'Usage qui l'emporte tousjours par dessus la raison" (Remarques 17, 28, 303), and claims that usage can sanction grammatical errors: " . . . que l'Usage favorise souvent des solecismes" (89).

14. "C'est la beauté des langues, que ces façons de parler, qui semblent estre sans raison, pourveu que l'Usage les authorise. La bizarrerie n'est bonne nulle part que là" (Remarques 173). "Sur quoy il est à remarquer, que toutes les façons de parler, que l'Usage a establees contre les reigles de la Grammaire, tant s'en faut qu'elles soient vicieuses, ni qu'il les Faille eviter, qu'au contraire, on en doit estre curieux comme d'un ornement de langage, qui se trouve en toutes les plus belles langues, mortes & vivantes (Remarques 305). Cf. 375, 538. Grammar was thus often regarded as a straight-jacket, a yoke which it is salutary to throw off at times. There is therefore a delicate balance in le bel usage between conformity and bizarrie. There are references in Vaugelas to the pedantry of the school grammarians who invoke "Priscien, & toutes les puissances Grammaticales" (Remarques 311).

15. Discussing the fact that the vowel of the definite article is not elided before heros, contrary to what one would predict, he remarks: "La curiosité ne sera pas peut-être désagréable, de savoir d'où peut procéder cela; car bien qu'il soit vrai qu'il n'y a rien de si bizarre que l'Usage qui est le maître des langues vivantes; si est-ce qu'il ne laisse pas de faire beaucoup de choses avec raison, & où il n'y a point de raison comme ici, il y a quelque plaisir d'en chercher la conjecture" (Remarques 2).

16. "Ie ne sçay si je me seray fait entendre, ou quand on m'entendra, si l'on sera satisfait de ce petit raisonnement, & s'il ne sera point trouvé trop subtil, & trop metaphysique; mais l'exemple du grand Scaliger, qui a fait de si beaux raisonnements sur la Grammaire Latine, m'a donné en la nostre cette hardiesse, que le Lecteur prendra s'il luy plaist en bonne part" (Remarques 388).

17. "Et certainement si j'avois eu à faire une Grammaire, ie confesse que ie ne l'aurois deu ny peu faire autrement, que dans l'ordre des parties de l'Oraison, à cause de la dependance qu'elles ont l'une de l'autre par un certain ordre fonde dans la nature, & non point arrivé par hazard, comme Scaliger le Pere l'a admirablement demontré" (Preface XII).

18. "Quand deux substantifs joints par la conjonction et, sont synonymes, ou approchans, comme vertu & generosité, il ne faut pas repeter l'article, mais quand il sont contraires, ou tout à fait differens, comme force & dexterité, alors il faut repeter, & dire, de la force & de la dexterité (Remarques 215).

19. Remarques 219.

20. Grammaire 62.

21. Remarques 385ff.

22. "Et il est vray que c'est en ce seul cas, où l'on trouvera un nom sans article, avec un pronom qui se rapporte au nom; mais il y a double response, la premiere que cette exception n'empescheroit pas que la Reigle ne fust véritable en tout le reste. La seconde, que mesme la Reigle subsiste encore au vocatif, & n'y souffre point d'exception, parce que l'article du vocatif o, y est sous-entendu, mais l'article n'est point sous-entendu aux autres cas." (Remarques 387).

23. Grammaire 75-83.

24. "Dans l'usage present de nostre langue, on ne doit point mettre de qui apres un nom commun, s'il n'est déterminé par un article, ou par quelque autre chose qui ne le détermine pas moins que feroit un article" (Grammaire 77).

25. Grammaire 81.

26. "S'il y a d'autres façons de parler qui y semblent contraires, & dont on ne puisse pas rendre raison par toutes ces observations, ce ne pourront estre, comme je le croyn, que des restes du vieux stile, où on omettoit presque tousiours les articles. Or c'est une maxime, que ceux qui travaillent sur une langue vivante, doivent tousiours avoir devant les yeux; Que les façons de parler qui sont autorisées par un usage general & non contesté, doivent passer pour bonnes, encore qu'elles soient contraires aux regles & à l'analogie de la Langue: mais qu'on ne doit pas les alleguer pour faire douter des regles & troubler l'analogie, ny pour autoriser par consequence d'autres façons de parler que l'usage n'auroit pas autorisées. Autrement qui ne s'arrestera qu'aux bizarries de l'usage, sans observer cette maxime, fera qu'une Langue demeurera tousiours incertaine, & que n'ayant aucuns principes, elle ne pourra jamais se fixer" (Grammaire 83). Compare the last sentence of this quotation with Vaugelas' statement quoted in footnote 5 above.

27. Grammaire 40.28. Grammaire 42.29. Grammaire 126-8.30. Grammaire 134.

31. "Questions of current interest will, however, determine the general form of this sketch; that is, I will make no attempt to characterize Cartesian linguistics as it saw itself, but rather will concentrate on the development of ideas that have reemerged, quite independently, in current work".
Cartesian Linguistics 2.

BIBLIOGRAPHY

Scaliger, Julius Caesar: *De Causis Linguae Latinae*, 1540.

Buommattei, Benedetto: *Delle Cagioni della Lingua Toscana*, Venezia, 1623.

Vaugelas, Claude Faure de: *Remarques sur la Langue Françoise*, Paris, 1647.

Arnauld, Antoine et Lancelot, Claude: *Grammaire Générale et Raisonnée*, Paris, 1660.

Chomsky, Noam: *Cartesian Linguistics: A Chapter in the History of Rationalist Thought*, New York, 1966.

THE PRONUNCIATION OF ZETA IN ANCIENT GREEK

Charles-James N. Bailey
 (University of Chicago)

The determination of the pronunciation of an orthographic symbol in a dead language needs to be motivated by psychological explanations of the choices of the orthographic symbols and the dialectal and other variations among them. One cannot limit one's attention to phonetic realism and the prior and subsequent phonological developments of the segments represented by such symbols.¹ You cannot determine the pronunciation of ancient zeta without also determining the pronunciation of the double-delta (single delta next to a boundary or consonant) that corresponds to it in Boeotian, Thessalian, Elean, Cretan, Laconian, and possibly in the dialect of Megara. It is also necessary, in this connection, to determine their voiceless congener spelled as double-tau or as double-sigma in various idioms. In Boeotian and Central Cretan (Lejeune 1955:98) double-delta and double-tau are palpable orthographic as well as phonological parallels in the voiced and voiceless orders and need to be explained as such.²

At the outset here, three working principles that will set limits to possible solutions of the present problem need to be enumerated:

i) There are found in Indo-European and other languages the following unidirectional sound changes:

- | | |
|--|--|
| 1) gy --> g'y | ky --> k'y |
| 2) y, dy --> d'y <u>or</u> d ^z : ³ | ty --> t'y <u>or</u> t ^s : ³ |
| 3) g'y, d'y --> d'ž
(= j̄j) <u>or</u> žd'ž ³ | k'y, t'y --> t'š
(= čč) <u>or</u> št's ³ |
| 4) d'ž --> žž (<u>or</u> d ^z :) ⁴ | t'š --> šš (<u>or</u> t ^s :) ⁴ |
| 5) d ^z : --> ðð ⁴ | t ^s : --> þþ ⁴ |
| 6) ž(ž), ð(ð) --> z(z) ⁴ | š(š), þ(þ) --> s(s) ⁴ |

While the changes just tabulated are well-recognized and established, we shall require anyone advocating the following changes first to demonstrate that they are possible:

1) Any unit --> zd or žd⁵

2) zd --> z(z) (Matthews 1954/5:80) or ðð; how could *zd yield both zz and dd (see [2] below)?

3) dz --> dd or nd (Rohlfs 1962:3, 6), ts --> tt (Lejeune 1955:90), t's --> tt (Allen 1957/8:116, 126-27), and so on.⁶

4) ž(z), š(š) --> stop.

On the other hand, we shall not hesitate to allow these changes:

1) zd --> dd; st --> tt.⁷

2) dz --> zz, ðð; ts --> ss, pp.

3) (Probably) dz --> d'ž, žd'ž and ts --> t's, št's, provided these latter already exist and the phonological pattern leaves little choice (Allen 1957/8:127).

What will be regarded as the weakest of all explanations for non-sporadic changes of the sort under consideration in Greek will be metathesis (Witton 1898:429, Allen 1951:945, 1957/8:121 n.40, but 1968:55, Driver 1955: 231; for examples of sporadic metathesis, cf. Rohlfs 1962:5 n. 1, Ladefoged 1967:149; the explanation of metathesis goes back at least to Havet 1878:193); contrast the regular reversal of [α grave] and [- α grave] consonants in clusters which is described by Lejeune 1955:59-60, Buck 1955:74.

Whatever classical Attic zeta may have been, it changed to a fricative in Hellenistic Greek. The change had been made soon after the middle of the fourth century B. C. (Matthews 1954/5:73, Sturtevant 1940:93, etc.). Presumably zeta became žž at this time (see below): at any rate, it later became zz and eventually the modern Athenian z.

ii) Since a fronted velar and a palatalized dental have the same articulatory place somewhere between them--[g'] = [d'] and [k'] = [t'], though the fronted velars lack a concomitant apical articulation--one should not be surprised either at spelling alternations like Mycenaean ke : ze (Mühlestein 1955:130, Lejeune 1958:136, and the citation of Palmer by Lejeune [1958:90 n. 17]; note that voicing was not indicated in most instances in the Mycenaean syllabic script) or "x" : "ss."⁸ And we shall certainly regard a secondary use of letters otherwise symbolizing [d t s] to denote palatal sibilants (including assibilates) in Greek as no less possible than in English verdue, nature, sugar.

The presuppositions will be important when we consider, first of all, the orthographic alternation "tt" : "ss" in ancient Greek. The oldest Attic inscriptions have "tt" from *ky, though the oldest literary use imitated (Lagercrantz 1898:90-91) the Ionic spelling with double-sigma until the orthographic reform in 403 B. C. The logic of the sound development, ky --> k'y --> t's --> šš, demands that we regard the orthographic alternation "tt" : "ss" as representing the phonetic alternation [čč] (= [t's]) : [šš].⁹ It is not necessary to believe that in Attic all examples of sigma from *t (including words like dóxa, dipsa) are to be interpreted as [š]; this sigma may just as well be derived from an earlier *ts.¹⁰ But within the framework of Greek writing, the only way that I see to make sense of the double-sigma spellings exemplified in Attic grápsassthai, eisstén, and Lésson, is to accept the suggestion first put forward by Boeckh (Blass 1888:90) that before a consonant double-sigma represented [š ž]. For examples in other dialects, see Buck (1955:75), and compare "sz" in Attic endészmous (Schwyzer 1953:218). This view would accord quite well with spellings like Zmýrna in Hellenistic Greek and the likelihood that reflexes of *gy and *dy would have passed through the stage [zz] on the way to becoming modern [z].

Before concluding with the voiceless phonological equivalents of zeta, one ought to ascertain whether the assumptions and conclusions that have been stated thus far are compatible with the known history of Greek writing. The results of such a check are very rewarding. In Asia we find the reflexes of *ky and *tw written with Semitic letters. One assumes that such a special writing indicates either the inadequacy of ordinary letters to indicate a given pronunciation; or else that it indicates a different pronunciation from that denoted, ambiguously or unambiguously, by the ordinary orthography. Now Pamphylian writes the segment with the Semitic shin, ש (Thumb-Scherer 1959:187). Here one is to assume that the special letter was used because of the ambiguity of sigma.¹¹ But some Ionic dialects use here the Carian form of the Semitic sade (Lejeune 1955:76, Buck 1955: 18, Allen 1968:57-58; the older views of Witton [1898: 423-24] and Lagercrantz [1898:93-96] require modification), a letter which in Semitic represented the so-called "emphatic" (that is, pharyngealized) sibilant. This presumably happened because the local Ionic pronunciation differed from both the [s] and the [šš] which our previous considerations would lead us to regard as the usual pronunciation in other Ionic locales. The form of the Carian letter was T. The letter for an "emphatic" in Semitic would be as suitable as any, indeed better than any other, to represent a Greek assibilate. In fact, another form of sade ש, called san, represents in Arcadian

what most scholars take to have been [t^s]. The resemblance of the Carian letter to Greek T cannot be overlooked. Indeed, it offers an obvious explanation for the double-tau writing of the same sound, presumably a palatal affricate, in Attic and other dialects. Compare Ephesian téTaras with Attic téttares (Thumb-Scherer 1959:264; for Halicarnassus, cf. Schwyzer 1953:318), and these with Ionian and Arcadian tésseres. The logic of the situation demands a contrast between a phonologically older [čč] and a younger [šš]. Additional confirmation is found in the Thessalian spelling of Thessalós as "Pétthalos," where the affricate pronunciation will seem more or less obvious to different scholars, and in the use of "ss" in Boeotian, where the reflex of *ky was normally spelled "tt," to indicate a foreign, no doubt unaffricated, pronunciation (Lagercrantz 1898:8-9). Finally, spellings like Cretan T(t)ena for Dena (=Ionic Zéna) and tōna (where the initial segment was Proto-Indo-European *y) for what is zōnē in Attic indicate that if either zeta or double-tau was an affricate, the other was also one. Though it is generally agreed that the segments represented by zeta and delta merged into the voiced fricative [ð] in Elean, the spelling "tt" for "z" (Lejeune 1955:97, Buck 1955:71-72) makes one wonder whether this could be true.

Mycenæan may have had čč beside a more common šš as the reflex of *ky. The pronunciational diversity would be expected in a rapidly developing lingua franca, even though Mycenæan curiously enough shows few of these expected diversities. Otherwise, it might be thought odd to find a change (čč > šš) in Mycenæan antedating the same change in Attic by hundreds of years. The view of Mycenæan just stated is predicated on the variation between ke and ze mentioned earlier and the likelihood that "z" represented an affricate (otherwise, why not "s"?), as well as on the use of both "s" and "z" syllabograms for reflexes of *ky, *thy, and possibly *tw (Muhlestein 1955: 122-29; cf. Chadwick 1955:85-87, Lejeune 1958:133, 100 n. 62, and the convenient table provided in Lejeune 1958: 134).¹² The Mycenæan syllabogram no. 85, which Chadwick (1955:90), following Sittig, suggests should be transcribed as si² may really represent či or ši, rather than s(i)ja (where "j," as usually in Mycenæan studies, represents yod). This character always stands word-initial except in a Pylian toponym (a-85-ta²), and may in some cases reflect *tw.

Thus we see that a wide-ranging set of orthographic and phonological facts tie into a neat and consistent bundle when it is assumed that the voiceless equivalent of zeta was a palatal—an affricate when spelled with tau, a non-affricate when spelled with sigma. The phonological expectation in the voiced order would be [ʃʃ], parallel to voiceless [čč]. The orthographic parallel

to double-tau would be double-delta, as some writers (especially Lagercrantz 1898:46) have recognized.¹³ This is in fact what is found in place of zeta in some dialects and in some environments in Attic, as will be shown. The parallelism fails when one looks at zeta itself. The phonological expectation would be [zz], corresponding to "ss" (= [ʃʃ]), where not [jj]. The pronunciation [zz] probably existed in early Hellenistic time; at least Sanskrit transcribed horizon as hariya (Wackernagel 1896: 137, 142). And the Greeks transcribed Iranian šatadrus as Zárdos. But classical zeta as undeniably contained a stop segment as a palatal sibilant segment.¹⁴ This failure of zeta to represent the voiced counterpart of double-sigma or double-tau in orthography or in pronunciation defines our problem in terms of the first two working principles.

iii) The third working principle is one stated some time ago by Sturtevant (1940:61-62) in connection with the orthographic variation between "hr" and "rh" representing Greek voiceless [r]: When psychologically unitary consonantal elements are represented by digraphs, the letters may indifferently be written in either order. (Compare further Ladefoged 1967:149.) In English we find both "wh" and "hw" for voiceless [w]. Unfortunately, Sturtevant neglected his own insight when dealing with zeta; for he assumed that it was [zd] in some instances and [dz] in others. If Sturtevant's principle is correct, and if zeta was a psychological unit in Greek, it follows that we cannot rely on the testimony of the ancients regarding the order of the sibilant and stop elements which classical zeta must have contained.¹⁵ Both Dionysius¹⁶ lend their authority to [zd], though this is denied by Velius Longus (Rohlfs 1962:4,5). Most modern authorities¹⁷⁻⁻ few have added much new to the brilliant exposition of Blass (1888:112-22)--favor the pronunciation [zd] (most recently Allen [1968:53]), or else allow both [zd] and [dz]. But Rohlfs (1962:7) favors only [dz].

It is generally acknowledged that soon after the middle of the fourth century B.C. zeta began to become a non-affricate.¹⁸ The Latins transcribed the Attic sound with "ss" (examples in Sturtevant [1940:93]); so Tarentine Greek (Witton 1898:430). Even in Attic zeta came to be spelled as a geminate, at least in certain cases (Schwyzer 1953:218). Blass (1888:120) spoke of Cretan thálattha as having "spirantic th, a sort of compromise between genuine Cretan thálatta and the usual thálassa," and I suspect that double-theta here shows a development comparable to that indicated by "z" in Castilian esperanza. In late Hellenism zeta became [zz] or the modern [z]. Already in the sixth century B. C.

Argive Greek had hoiz de for hois dè (Schwyzer 1953: 217). This more likely denoted [hoižd'è] than [hoiždè], in view of spellings like Léssbon. Since Doric delta could represent [ȝ] = [d'ž], the pronunciation may have been [hoižd'žè].

If it is clear that zeta was a palatal (further evidence will be found below in the use of zeta to transcribe foreign affricates), that it was voiced, that it contained a stop segment, and that it contained a sibilant segment, it remains only to examine which of the following four possibilities is the right one: (1) The sibilant came only at the end; in this case classical zeta would have to be [dz], for [d'ž] would be identical with the double-delta spelling, and the orthography would no longer make sense. We shall simply examine the general case for a sibilant-final zeta. (2) The sibilant came only at the beginning, as maintained by the most notable grammarians. (3) Zeta had the two pronunciations [d'ž] and [žd]. (4) The sibilant came in both places, the stop forming the middle segment. This would mean that zeta was like the Slavic [žd'ž] (the voiceless equivalent in modern Russian is represented by the letter š'ča, also found for sk' in modern Aegina Greek [Thumb 1910:15]). Developments like *zd or *zdz are unknown, and even Slavic žd in meždu comes from *dy only through *žd'ž. It should be emphasized that the explanation for the development of such a unit in Slavic requires the existence of [žd'] from sources other than simply *d'y, namely from *zd' > *žd' (Diver 1955:231-32). But this condition is met in Greek if Léssbon contained [žb], as argued earlier; for "ssd" would then indicate [žd']. That this is not simply an orthographic possibility, but also a phonological one is confirmed by the development of Sanskrit nīdāh "nest" from *nisdos through *nizdos and *niždos (Meillet 1964:97, Kuryłowicz 1956: 113), though it must be admitted that this development involves specifically Indic retroflexion.

The time has come to present the often-cited evidence for the presence of the sibilant segment of zeta before the stop and after it. First, the evidence for the sibilant after the stop. The evidence is qualitatively and deductively compelling, but empirically weak. (1) First, there is the expectation from our first working principle that *gy and *dy would yield an affricate pronunciation. (2) Then there is the suggestive parallel with the other double letters, ksi and psi. One should also note the pleonastic spelling "xs" for ksi in Chios and Corinth (Blass 1898:115). (3) Thirdly, there are the borrowings in which the Greeks transcribed the Indic place names Ujjayini (> middle Indic Ujjeni) and Pañcāla- as Ozēnē and Pazálai (also as Pássaloi; Wackernagel 1896:137). For Indic Jhoilasa the Greeks wrote Zóilou, and for Indic Bharukacchas they wrote Barýgaza. (Vide further Meillet

[1916:53].) Greek borrowed Latin [d'ž] in coiux and Iulia as zeta in kozous and Zoul(e)ia. However, these last, as well as some of the former, examples must have been borrowed after zeta had ceased to be an affricate. They therefore contribute more to the proof of zeta's palatal character than to its claimed affrication, and not even very much to that. (4) Fourthly, there is the [dz] found in various modern island and Italiot Greek dialects (the evidence is summarized in Rohlfs 1962; cf. Matthews 1954/5:73). Only if something similar existed in ancient times can we make sense of the derivative uses of zeta and "z" in Greek dialects and in Oscan, Umbrian, and the Romance languages-- uses which constitute our next two bits of evidence in favor of a sibilant-final zeta in ancient Attic. (5) In dialects of Greek where zeta was replaced by double-delta or "sd" zeta was free to represent the affricates that developed from *dy in literary Lesbian (for example, kárza = Attic kardíā), Cyprian (kórza), and Pamphylian (korzia; Schwyzer 1953: 330). Most authorities¹⁹ agree that here zeta represented [dz], the only voiced assibilate left if a palatal one already existed (a parallel development is seen in the Slavic so-called second palatalization). The reflex of earlier *ts was represented by zeta in early Central Cretan, while the reflex of *kʷe was spelled tze in Arcadian (Risch 1955:75) and Laconian (Buck 1955: 267). Zeta replaced gamma in azathai̯ (vide Schwyzer 1953:209) with a value that all must agree is "ganz unsicher" (Schwyzer 1953:329). (6) The Oscan and Umbrian native orthographies employed their version of zeta to represent the voiceless affricate [ts]. The Italian use of "z" to denote a voiced or voiceless dental affricate goes back to a vulgar Latin usage no doubt based on Greek dialectal uses. Not only did Latin write zabulus for diábolos and zebus for diēbus; it even transcribed Greek baptízō as baptidio and Óryza as oridia. (7) A final consideration is the following. If zeta represented [zd] or [žd'], rather than an affricate, why was it ever represented by a single letter, rather than the orthography "sd" that was clearly available everywhere?

The evidence for a sibilant segment at the beginning of zeta is quantitatively and empirically much more compelling than the evidence for a sibilant at the end. (1) First is the evidence of the grammarians already referred to. (2) If zeta ended with a sibilant, why are spellings like "sz" found in Attic, but never "zs," like the pleonastic "xs" occasionally found in ancient Greek? (3) Another factor, very important but one whose significance has been misinterpreted by some investigators, is the use of "sd" in literary Lesbian

corresponding to inscriptional zeta. The same "sd" is found in literary Doric²⁰ (whence Oscan Niumsdiēis in the Greek alphabet) instead of the indigenous double-delta. We shall return to the "sd" orthography below. That zeta at least began with some kind of fricative is also attested by the way Aristophanes is said to have represented a Laconian attempt to pronounce Attic zeta at the beginning of a word: Bdeū (=Zeū, probably in Lycistrata, 940 [Allen 1968:54]). Beta here presumably represents a Laconian labial fricative. Since the Laconian reflex of *dy, spelled "d" next to a consonant or boundary, was no doubt [j], the orthography in question must have meant [vjeū]. (4) Fourthly, there is the deletion of the nasal in the preverb syn- "with" before Attic zeta, as in syzō "I live," a phenomenon otherwise found only before sigma.²¹ Compare with this the preposition with sigma in ex Zéas (Blass 1888:114), where one would look for ek before an ordinary consonant. (5) Then there is the use of zeta for earlier *zd in words like ázō "I dry" (compare Czech ozdit), hízō "I sit," and ózos "twig," as well as in Athénaze (from Athenas-de).²² To these examples one may add zeinamen (=sbénnumen) from Hesychius, where zeta represents the older zd of this clearly non-Aeolic dialect (locale not given by Hesychius). (6) Sixth are spelling errors like ézōn for éstōn on a Delphic inscription (Blass 1888:118). Compare Homeric and Ionic mazós for Mast(h)ós in other dialects (Blass 1888:117). The spelling "sz," found in Attica for zeta and in Boeotia for *zd (see under [8] below), attest to a sibilant beginning for zeta. In view of the possible parallelism (Blass 1888:115) with spellings like "sst" noticed earlier, as though "sz" = "ssd," one might see in "sz" further evidence of the palatal pronunciation of zeta. (7) There are borrowings in which Greek writes Semitic Ašdōd as Azōtos, Hellenistic Asdōd, and Iranian Mazdārā and Aura-mazdā as Mazárēs and Oromázes, Hellenistic Oromásdou. The later Hellenistic spellings just cited show a loss of the stop element of zeta. Or do they? (8) The likelihood of another explanation is our last and perhaps main point. We know that in Lesbian the literary "sd" was the same as the inscriptional zeta. Let us see what follows from supposing that here delta could represent [j], as in the closely related Thessalian and Boeotian dialects. This would mean that the literary "sd" represented [žd'ž] on all the grounds that have been cited, and this tallies exactly with the interpretation of zeta which I have been trying to build up! This explanation harmonizes with two Orthographic anomalies, one in Attic, a zeta dialect, and the other in Boeotian, a delta dialect. Blass (1888:116) long ago noted the importance of the poetic word (h)érdo "I work," where a zeta would be expected in place of the delta that is actually found.²³ Now, sibilants were lost between consonants in Greek, so that in *wérgyō it would be quite in order to expect a

reduction of either zd or žd'ž to d or d'ž, respectively.²⁴ Why should one think that the reduction was from žd'ž to d'ž (= [j]), rather than from zd to d? We have seen that sibilants before t, b, m, and so on, were palatals, spelled "ss." Furthermore, we never find zd(z) as the reflex of a phonological unit elsewhere. Another reason will emerge from an examination of the Boeotian writing (Schwyzer 1923:231) of Theószotos (for Theósdotos) and Thiozótā (for Thiosdótā). It has been already seen that Greek clusters of sibilant plus stop developed a palatal sibilant that was sometimes spelled with the double-sigma. A stop d would be attracted to the palatal position in such a cluster, since stops more often than not show sub-distinctive assimilations (Daniel Jone's "similitude") to adjacent fricatives. Boeotian writers evidently felt that "sd" would not symbolize this pronunciation [žd'] (which probably was replaced at once by [žd'ž], but that does not matter) as adequately as zeta, otherwise unused in Boeotian because Boeotian represented the reflex of *gy and *dy, namely [jj] = [d'ž], with the delta spelling.²⁵

If zeta was pronounced [žd'ž] = [jj], it is easy to understand its use in transcribing various sorts of affricates, foreign and domestic, as well as sibilant-stop sequences in Iranian and Semitic. In any case, the evidence leaves no room to doubt that the stop segment in zeta was both preceded and followed by a sibilant element.

Did zeta represent two pronunciations, one with the sibilant segment first and one with it last; or did it represent [žd'ž] = [jj], as I think? Professor E.P. Hamp, to whom I am indebted for comments on this paper, points out that one has to admit both possibilities. If tau or sigma could stand for both non-palatals and palatals, zeta could represent both sibilant-initial and sibilant-final clusters. However, the double pronunciation is from the psychological point of view a bit more difficult. There is the further weight of my interpretation of literary Lesbian zeta as [jj] in accord with what seems to me compelling evidence that delta in the closely related Thessalian and Boeotian could represent [j]. I doubt that zeta could represent both [jj] and anything else, whatever might be the possibilities that it could represent both [žd] and [d'ž]. The existence of sibilant-stop phonological units is in any case very doubtful.

The interpretation of double-delta and zeta preferred here, as [d'ž] = [jj] and as [žd'ž] = [jj], respectively, gives psychological motivation to the orthographies that are found and invokes only phonological developments that are very probable. Our three working principles thus get entirely satisfied. The conflicting evidence and explanations that have previously been written about are now resolved. The explanation of the Boeotian spellings

The ósszotos and Thiozótā given above will also explain the only matter remaining to be accounted for,²⁶ namely the use of zeta in Attic words like ózos, hízō, and Athenaze, and need not be repeated here.

During the Hellenistic period Attic [čč] and [žž] became [šš] and [žž], respectively, and these de-palatalized in late Hellenistic times. When the palatals were de-palatalized, combinations like [št šp žb žm] automatically became [st sp zb zm]. Contemporary dialectal [dz] goes back to [d'ž], an easily intelligible simplification of [žd'ž], though a simplification to [žd], as in Slavic méždu, would also be possible. (If in any modern dialect [dʒ] is descended from [žd'ž], it is explicable because the de-palatalized [dʒ] would be a unit which is unknown in languages, and would therefore be prone to be replaced by [dʒ]. The palatal and dental affricates found in some modern dialects [Thumb 1910:15, 19-20] go back to classical velars and in some cases [Thumb 1910:23-24] dentals.)

My views of zeta avoid the grave weaknesses of the other views: the contradictory proofs concerning the sequence of the sibilant and stop segments, the failure to maintain the palatal status of zeta, the necessity of supposing an unlikely metathesis or the impossible changes *ts, *ts --> tt and *dž, *dz --> dd to get Attic and Boeotian "tt" and Boeotian "dd," and the necessity of an unattested change *zd --> z to get the modern pronunciation of zeta.

FOOTNOTES

- ¹ While Allen (1957/8) strives for phonetic realism, he fails adequately to motivate the psychological choices of the orthographic symbols. A decade ago he was of course unaware (Allen 1957/8:123) of the possibility of allowing for synchronic rule-ordering in the sense of generative phonology, let alone of the possibilities offered by rule-re-ordering for lexical exceptions (Bailey 1968). His manner of dealing with junctures (Allen 1957/8:124) needs greater clarity.
- ² The complex dialectal developments are tabulated for the voiceless order by Lejeune (1955:90). In Attic some instances of original *ty underwent the minor rule, ty --> t'y = k'y; otherwise, *ty --> *[t^s:].
- ³ For example, Slavic. The replacement in (2) by a dental assibilate (exemplified in the Slavic so-called second palatalization) seems to depend on the current existence of palatal affricates. On the other hand, we see in the history of French change (4) being caused by a simultaneous change (3). For tabulations of expected changes which differ from mine, cf. Allen (1957/8:115, 126), Matthews (1954/5:77), Witton (1898:430). Various writers have expressed the view that zeta (other than late literary zeta in Lesbian and zeta in Cyprian, Arcadian, and Cretan) at least passed through a palatal affricate stage (Risch 1959:219, Roussel's view cited in Matthews [1954/5:64], and Meyer's view cited in Blass [1888:120 n. 450 b]). For its voiceless equivalent spelled double-tau and double-sigma the palatal affricate stage is more widely recognized (Walters and Conway 1928:76, Hamp 1959:190, Allen 1968:57; cf. Pedersen 1923:113, 115).
- ⁴ For example, Romance. The change d'ž --> dz is exemplifiable in Venetian mazore (= standard maggiore) and za (= standard già). The parallel change in the voiceless order is more difficult to exemplify. For ts --> b, cf. Castilian esperanza.
- ⁵ Matthews (1954/5:68, 78); Slavic žd in méždu apparently did not come directly from *dy, but through the mediary of *žd'ž. Anyhow, the Slavic change does not prove that *dy can become a non-palatal zd, despite attempts to justify the latter with the former. Note that žd can yield z, at least before m, as in English asthma(tic).
- ⁶ Proto-Greek *ts easily gave s(s); for the Boeotian reflex (double-tau), vide infra.

7

Matthews 1954/5:80, Rohlfs 1962:4, Thumb-Kieckers 1932:85. These changes are uncommon, but nevertheless are found in Greek dialects. Compare ette for éste at Orchomenos and late Laconian ettan for estàn (Schwyzer 1953:216, Buck 1955:73-74). Prof. Hamp points out the change from st to t through the intermediate ht found in Spanish dialects. Note also the change z --> d in English isn't, doesn't, wasn't.

8

Vide the words cited by Lejeune (1955:85 n. 8; cf. possibly p. 63 n.1); and cf. Sanskrit "ks" (= [k's]) with the fact that in English "ak Chicago" sounds just like "at Chicago." For verbal forms, cf. Buck (1955: 72,116), Ruipérez (1953:257-58); note Cretan "tt" for "x" in logíttomai (Matthews 1954/5:71). For "x" instead of expected "s" (= "ss") in Mycenæan, cf. Lejeune (1958:94; but vide p. 96 n. 45). Schwyzer (1953:318) notes the use of ksi for expected "tt/ss" at Iasos and Teos. Lack of space forbids an analysis of the sound development in dóxa and dípsa. Xeus in Tharan is a purely orthographic change (Buck 1955:18) and is therefore irrelevant here. Finally, it is worth noting that Latin transcribed Greek malássō as malaxo (but vide Postgate 1906). I shall not attempt to comment on the spelling krábaktos, cited from the Codex Sinaiticus by Schulze (1895:375 n.4).

9

Geminate spellings were simplified next to a consonant or word boundary. So "tt" became "t," "ss" became "s," and "dd" became "d," as in Boeotian Deus. Word-initial alternations between "t" and "s" indicate palatal segments: t/sémeron, t/sétes (see list in Witton 1898:428-29; cf. Pedersen 1923:113). Compare Gortynian kartónans (Lagercrantz 1898:45), with "t" from *ty. Compare also sáō with diattáō. Even word-initially, the palatals became geminates after preceding vowels, like the palatals in modern Italian. Besides the last examples, one should notice Homeric ássa, made by false division from neuters ending in -a plus čča (from *tya). Seúō often made position (Witton 1898:428), though sometimes not only zeta but even "sk" did not (Blass 1888:116). The double-zeta found in later Greek is due (except perhaps in foreign names; Schwyzer 1953:218) to the change of classical zeta to a plain sibilant fricative.

10

Lack of space prohibits exhausting the orthographical alternations of "t" : "s" even in Attic, where both represent the development of *t before yod or front vowels. It is doubtful that in ánti and mántis tau represents an affricate ([č]); compare pístis, esti. (All dialects have éti, árti Holt 1937:178-79). Both in básis and in pósis (where *t belonged to the root) the sigma in the

middle probably goes back to a proto-Greek alveolar affricate. In these words the rule merging *ty with *ky was inoperative. Following n it is phonetically difficult to keep ts and s or t's and š apart. Cf. the comments by Witton (1898:421-22) and his citation of English conscientious; cf. further Schwyzer (1953: 322). I take sigma to stand for [s] in Thessalian and Cretan pánsa, Attic pása, and Lesbian páisa, as well as in the feminine present active participles, no less than in mésos and tósos. The same is true of sigma in dóxa and dípsa.

- 11 Sigma itself comes from shin also. One is tempted to speculate that the ancient pronunciation of sigma had the retroflexion often found in modern Greek. For s → š in Modern Greek, vide Thumb (1910:para.28).
- 12 It is doubtful whether za = dià in Mycenæan; cf. Lejeune 1958:129.
- 13 Blass (1888:120) and Lagercrantz (1898:46) both recognize the importance of (h)érdō (vide infra) in this connection. The parallelism is reinforced by such spellings as Cretan T(t)ena for expected Dena (= Ionic Zēna), as well as fluctuations in Cretan and Elean between "tt" and "z" (Buck 1955:71-72).
- 14 Matthews (1954/5:72-73) is incorrect in asserting that Archinus "defined δ and ζ as homorganic," that is, as dentals. What Archinus actually says is that both are apical, a statement quite in harmony with the fact that palatals in modern Romance languages, for example, are often no less apicalized (pronounced with the tongue tip on the teeth) than the plain dentals.
- 15 The only dissenters are Witton (1898:430, 436), who favors [ž] for zeta, and Lagercrantz (1898:93-96), who suggests [ð] and other possibilities.
- 16 Their statements are Englished in Sturtevant (1940:90-91).
- 17 Their views are conveniently summarized by Matthews (1954/5:63-64) and Rohlfs (1962:1-2); vide also the references in Schwyzer (1953:329 n.1).
- 18 Aristotle (Metaphysics 993a 5-7) says, "For some say that za is (made up) of s, d, and a; others say that it is another sound, and not one of the well-known ones." (Sturtevant [1940:93 n. 106] says that this has been emended.) Whether an affricate or a non-affricate [žž] is being described here, the statement may imply a transitional period. But I think that little certain can be gleaned from this statement.

- 19 Otherwise, vide Witton (1898:433), referring to the views of Meyer.
- 20 Doric hísdō (= Attic hízō, both from *sízdō) "I sit" could be authentic.
- 21 I have been able to find no relevant examples of the treatment of the nasal in question before delta corresponding to zeta in appropriate dialects like Boeotian and Thessalian.
- 22 Blass (1888:116) and subsequent scholars take éraze and khamáze to have been created by false analogy. See Schwyzer (1953:330) for býzēn, etc.
- 23 Attic prose has rhézō with a different gradation of the root.
- 24 Nevertheless, Mycenaean has zeta in wo-ze = wórzei. It is not wholly impossible that the word contained a syllabic [r] in Mycenaean.
- 25 Authors who have maintained that "dd" represents an earlier pronunciation than zeta are no doubt wrong. There is no phonological reason for supposing [j̪j] to be an earlier reflex of *gy and *dy than [žj]; on the contrary, [j̪j] is a later simplification of [žj].
- 26 Except the Boeotian reflex of *ts spelled "tt." Allen (1957/8:127) suggests that *ts may have gone to tš. One must not close the door to this possibility. In any event, both *ts and *ty in all words yielded Boeotian [č(č)], probably because of some pattern merging. The merger of *ts with [čč]=[t's] makes sense, however, only if "tt" represented [čč]!

- ALLEN, W. Sidney. 1951. "Some prosodic aspects of retroflexion and aspiration in Sanskrit." Bulletin of the School of Oriental and African Studies, University of London, 13:937-46.
- . 1957/8. "Some problems of palatalization in Greek." Lingua 7:113-33.
- . 1968. Vox graeca: the pronunciation of classical Greek. Cambridge: At the University Press.
- BAILEY, Charles-James N. 1968. "An untested idea on lexical exceptions to the regular ordering of the phonological rules of a language." Chicago Journal of Linguistics, 2:59-64.
- BLASS, Friedrich. 1888. Über die aussprache des griechischen. 3d rev. ed. Berlin: Weidmannsche Buchhandlung. Englished by W. J. Purton (trans.), Pronunciation of ancient Greek. Cambridge: At the University Press, 1890.
- BUCK, Carl Darling. 1955. The Greek dialects: grammar, selected inscriptions, glossary. Chicago: The University of Chicago Press.
- CHADWICK, John. 1955. "La représentation des sifflantes en grec mycénien," in Lejeune (1956).
- DIVER, William. 1955. "The problem of old Bulgarian št." Word, 11:228-36.
- HAMP, Eric P. 1959. "Notes on early Greek phonology: 1. Assilated clusters in Proto-Greek." Glotta, 38:187-90.
- HAVET, L. 1878. "Notes de grammaire III: Sur les divers sons du grec: renseignements tirés des langues italiotes." Mémoires de la société de linguistique de Paris, 3:192-96.
- HOLT, Jens. 1937. "Remarques sur l'assibilation grecque." Aarsskrift for Aarhus Universitet IX, Acta Jutlandica IX. København: Levin & Munksgaard.
- KURYŁOWICZ, Jerzy. 1956. L'apophonie en indo-européen. Prace Jezykoznawcze 9. Wrocław: Zakład Imienia Ossolińskich Wydawnictwo Polskiej Akademii Nauk.
- LADEFOGED, Peter. 1967. Three areas of experimental phonetics: Stress and respiratory activity, the nature of vowel quality, units in the perception and production of speech. London: Oxford University Press.

LAGERCRANTZ, Otto. 1898. Zur griechischen lautgeschichte.
Upsala Universitets Arsskrift. Upsala: Akademische Buchdruckerei Edv. Berling.

LEJEUNE, Michel. 1955. Traité de phonétique grecque. 2d rev. and cor. ed. Paris: Librairie C. Klincksieck.

(editor). 1956. Études mycéniennes: Actes du Colloque International sur les Textes Mycénien (Gif-sur-Yvette, 3-7 avril 1956). Paris:

1958. "Les sifflantes fortes du mycénien." Minos, 6:87-137.

MATTHEWS, W. K. 1954/5. "The pronunciation of Attic Greek ξ in the sixth and fifth centuries B. C." Lingua, 4:63-80.

MEILLET, A. 1916. Review of M. Fasmer, Izslēdovanie v oblasti drevnegrečeskoi fonetiki. Bulletin de la Société de linguistique de Paris, 20:52-55.

. 1964, a reprint. Introduction à l'étude comparative de langues indo-européennes. Preface by George C. Buck. Alabama Linguistic and Philosophical Series # 3. University, Ala.: University of Alabama Press.

MÜHLESTEIN, Hugo. 1955. "Zur mykenischen schrift: die zeichen za, ze, zo." Museum Helveticum, 12:119-31.

. 1958. "Einige mykenische wörter." Museum Helveticum, 15:222-26.

PEDERSEN, Holger. 1923. "Das auf einen t-laut zurückgehende σ und ττ im griechischen," in Ἀντίδωρος: Festschrift Jacob Wackernagel zur vollendung des 70. lebensjares am 11. Dezember 1925, gewidmet von schülern, freunden und kollegen, mit einem bildnis. Göttingen: Vandenhoeck und Ruprecht.

POSTGATE, J. P. 1906. "On malaxo and μαλάσσω." Classical Review, 20:443.

RISCH, Ernst. 1955. "Die gliederung der griechischen dialekte in neuer sicht." Museum Helveticum, 12:61-76.

. 1959. "Frühgeschichte der griechischen sprache." Museum Helveticum, 16:215-27.

ROHLFS, Gerhard. 1962. "Die aussprache des z (ζ) im alt-griechischen." Das Altertum, 8:3-8.

RUIPÉREZ, M.C. 1953. "Sobre la prehistoria de los dialectos griegos (A propósito de un libro de F. R. Adrados)." Emerita, 21:253-66.

SCHWYZER, Eduard. 1923. Dialectorum graecarum exempla epigraphica potiora. Leipzig: In aedibus Salomonis Hirzelii.

_____. 1953. Griechische grammatic: auf der grundlage von Karl Brugmanns Griechischer grammatic. Handbuch der altertumswissenschaft, Div. II, Part I, Vol. I. 3d ed. München: C. H. Beck'sche Verlagsbuchhandlung.

SCHULZE, Wilhelm. 1895. "Samstag." Zeitschrift für Vergleichende Sprachforschung auf dem Gebiete der Indogermanischen Sprachen, 33:366-86.

STURTEVANT, Edgar H. 1940. The pronunciation of Greek and Latin, 2d ed. William Dwight Whitney Linguistic Series. Philadelphia: Linguistic Society of America, University of Pennsylvania.

THUMB, Albert, 1910. Handbuch der neugriechischen Volks-sprache: Grammatik, texte, glossar, 2d cor. and enlarged ed. Strassburg: verlag von Karl J. Trübner. Englished by S. Angus (trans.) A handbook of the modern Greek language: Grammar, texts, glossary. Chicago: Argonaut, Inc., Publishers, 1964.

_____, and KIECKERS, E. 1932. Handbuch der griechischen dialekte. 2d augm. ed. Indogermanische bibliotek, Div. I, Sect. I, Vol. VIII, Part I. Heidelberg: Carl Winters Universitätsbuchhandlung.

_____, and SCHERER, A. 1959. Handbuch der griechischen dialekte. 2d augm. ed. Indogermanische bibliotek, Div. I, Sect. I, Vol. VIII, Part II. Heidelberg: Carl Winter--Universitätsverlag.

WACKERNAGEL, Jacob. 1896. Altindische grammatic: I. Laut-lehre. Volume I. Göttingen: Vandenhoeck und Ruprecht.

WALTERS, C. Flamstead, and CONWAY, R.S. 1928. Deigma: A first Greek book. 2d ed. London: John Murray.

WITTON, W. F. 1898. "On -ΣΣ - and -Ζ-." American Journal of Philology, 19:420-36.

GRAMMATOLOGY AND GRAPHEMICS

I. J. Gelb
 (University of Chicago)

0. Introductory Remarks.

The object of this paper is to discuss the concepts and to define the terms used in connection with writing as one of the systems of signs, and with the study of writing within the scope of grammatology, linguistics, and philology. For a fuller discussion of the topic see my "Methodology of the Study of Writing" soon to appear in one of the fascicles of the Enzyklopädie der Geisteswissenschaftlichen Arbeitsmethoden, ed. by Manfred Thiel.

1. Writing as a System of Signs.

Man, like animal, interacts through communicative behavior by means of signs or symbols used conventionally. The term "conventional" implies that the signs or symbols when used by some individuals, have the potential of being reacted upon, or understood, by other individuals, namely, those on the receiving end of the communication.

A system of signs is an assemblage of organically related signs conventionally used for the purpose of communication, such as language, writing, or gesture language. A sign is a component of a system of signs, such as a word in a system of signs called "language," a written mark in a system called "writing," or a gesture in a system of signs called "gesture language." A symbol is similar in meaning to a sign, but it does not form part of a system; e.g., the symbols "cross" for Christianity, and "anchor" for the concept of hope. For differing opinions on the meaning of "sign" and "symbol" see Sebeok in Garvin 1963, 49 f. and 53 f. (with references to A.L. Kroeber and J.H. Greenberg).

There is no good expression to cover all of the conventional means of communication through signs. The French linguists use le langage in this sense, while they call the auditory language la langue. In Anglo-American usage "speech" often stands for auditory language, and "language" for all the means of communication through signs.

In a study of the systems of signs a basic distinction must be made between the raw materials (subject matter) and the systems deduced from them. Spoken (and/or written) utterances represent the raw materials of a language; a phonemic or morphemic analysis, a paradigm,

a grammar, or a lexicon presents a linguistic system reconstructed from these materials. Similarly, written texts represent the raw materials of writing; a graphemic analysis of a logo-syllabic writing, a reconstructed alphabet, a chart, or a list of signs, gives a graphemic system of the underlying written materials. Thus the relationship of language systems to language materials or of a graphemic analysis to the underlying written texts is that of *langue* to *parole* of de Saussure's terminology, or of "code" to "message" (or "speech-act") sometimes found in American usage (see Stankiewicz in Sebeok et al. 1964, 241 and 266).

2. Subdivisions of Writing.

Writing is a system of human inter-communication by means of visual marks used conventionally. Viewing writing in its broadest aspects, we can distinguish three main classes or types:

Subgraphy, sub-writing, semasiography, or ideography. These are primitive attempts at communication by means of visual marks which have no set correspondences in language.

Phonography or full writing. These are systems of writing in which written signs generally have set correspondences in language.

Paragraphy or para-writing. These are various devices which are used beside or in addition to the generally formulated systems of writing.

In this discussion of the different types and subtypes of writing the time element is disregarded. Sub-writing is characteristic of the earliest attempts at visual communication, but examples of sub-writing can be attested both in some aspects of full writings and also in modern times in the writings of the so-called "primitive" societies. Full writing is attested from the earliest historical times in the logo-syllabic writings of ancient Mesopotamia and Egypt to modern times in our alphabetic writing. Para-writing is a para-feature of both sub-writing and full writing in all their aspects.

3. Semiotics.

The general science dealing with the various systems of signs is called "semiotics" or "semiotic." The term "semantics," which deals with the meaning of linguistic elements, should be carefully distinguished from the much broader term "semiotics." For these and many other terms see Read 1948, 78-97, and Sebeok et al. 1964, 5 f. and 275 f.

A tremendous amount of information on various semiotic devices used in human inter-communication has been presented by various scholars in a book edited by Sebeok et al. in 1964, as by Ostwald (mainly on psychiatry), Mahl and Schulze (psychology), Hayes (language teaching), La Barre (cultural anthropology), Stankiewicz (emotive language), and Mead (generally on the communicative process).

4. Fields of Study of Writing.

Of the various fields of study concerned with writing, the most important are: philology, which deals mainly with written and oral source materials; and linguistics and grammatology, which deal mainly with systems of signs, which in turn are based on written and oral source materials. The use of the word "mainly" means that, while certain disciplines deal predominantly with certain aspects of study, there are no sharp divisions between them, and some areas of study can be treated by different fields.

5. Grammatology.

The field of study which deals with writing in the broadest sense was called "grammatology" by Gelb 1952, 23, following partially the term "grammatography," which was used in 1861 in a title of a book on writing by Ballhorn published in England. The original German edition, from which the English translation was made, does not use the term.

Equally appropriate as "grammatology" are the terms "grammatonomy" used by Boedberg 1957, 113, and "graphonomy," used by Hockett 1958, 539. Not acceptable are "graphology" of Halliday 1961, 244, because this term already has another, well-established meaning; and "graphemics" (see below), because this term is too narrow.

6. Approaches to Grammatology.

We can distinguish two main approaches to grammatology: the descriptive-historical and typological-structural.

The traditional descriptive-historical approach to the study of writing is by far the most common. Here belong such standard old manuals on writing as Lenormant, Faulmann, Taylor, and Berger, as well as the more recently written Jensen, Diringer, and Février. What they all have in common is a simple narrative approach to the description of writing in its historical evolution. Their apparent shortcoming is the general lack of systematic typology. Good studies on individual writings, such as hieroglyphic

Egyptian or the Greek alphabet, are not wanting. What is entirely missing is theoretical and comparative evaluation of the various types of writing, such as discussions of various types of syllabaries, alphabets, word signs, and logo-syllabic writings. The present confusion in the typological classification of writing can be best illustrated by the term "transitional" given to such important writings as Mesopotamian cuneiform and Egyptian hieroglyphic, which lasted for about 3,000 years and whose exact place in the classification of writing can be established without any difficulty.

The historical approach is further vitiated through confusion with considerations of a geographic nature, as evidenced by such chapter titles as "Asiatic writings" and "American writings," or "writings of Asia" and "writings of America," which are frequently found in the standard manuals on writing.

The typological-structural approach, initiated by Gelb, 1952, is based on the realization of the importance of structure and typology in the study of writing. In contrast to the traditional approach, according to which the writings of the world are described in their evolutionary progress, the new approach requires first a thorough analysis of the structure of individual writing systems, and then their classification within the typological framework of writing in general.

Based mainly on the descriptive-historical approach but utilizing gains attained by the typological-structural approach are Cohen 1958, Istrin 1961 (especially in its 1965 edition), and Friedrich 1966.

7. Subdivisions of Grammatology.

The three main subdivisions of grammatology are: subgraphemics, graphemics, and paragraphemics.

The field of subgraphemics deals with sub-writing, subgraphy, semasiography, or ideography, as represented by primitive forerunners of writing characterized by the use of visual marks having no set correspondences in language.

The field of graphemics deals with full writing or phonography, that is, systems of writing in which the written signs generally have set correspondences in elements of language. Graphemics is concerned with the graphemes of a system, that is such signs as phonograms, syllabograms, and morphograms, which have set correspondences in language in the form of phonemes, syllables, and morphemes.

Instead of "graphemics," other scholars use the terms "graphics" (Francis 1958) or "graphic linguistics" (Crossland 1956). All three terms are generally misused by those linguists who limit the terms to the study of alphabetic writings with their correspondences to phonemes of language, overlooking or paying scant attention to all the other types of writing, such as the logo-syllabic and syllabic systems.

Trager 1958, especially page 8 assumed that certain communication devices which set a background for language proper (such as voice set) or serve as accompaniments of language proper (such as voice qualities and vocalizations) fall in the class of para-language. Hamp 1959 pointed out the existence of parallel characteristics in writing, and called the study of these elements "paragraphemics." This is still a very obscure field, and its relationships to both sub-writing (semasiography) and nonsegmental elements in full writing need a full elaboration.

8. Philology and Linguistics.

There is much confusion regarding the aims and methods of the fields of philology and linguistics.

Philology, a study of a language or a linguistic group and its linguistic sources, forms one of the basic means for comprehension of the respective cultures. It is largely involved in the study of literature (whatever its exact meaning), mainly written, less oral. Philology deals with the formal aspect of writing under the heading of epigraphy and paleography.

Linguistics represents an analytical discipline dedicated to the study of the underlying language systems, e.g., phonology or morphology. It is involved more in the study of oral than written sources. Linguistics deals with the study of the structural aspect of writing under the heading of graphemics.

The field of graphemics is concerned with writing after it became a secondary transfer of the spoken language, a vehicle by which elements of the spoken language were expressed in a more or less permanent form by means of visible signs used conventionally. This took place for the first time about five thousand years ago in the Sumerian and Egyptian writings.

9. Writing in Relation to Language.

Very little work has been done in the field of the study of writing in its relation to language. The philologists have been concerned mainly with the historical

evolution of writing and have paid little attention to the interrelations between writing and language. The linguists have been more concerned with the spoken language than with the written language. When interested in written languages, they have often limited their study to living written languages, neglecting the rich sources of information which can be culled from ancient written languages, and with it of pre-alphabetic systems.. The few scholars who have not limited their study to modern alphabetic systems often show a lack of factual knowledge in relying for basic facts on manuals with an antiquated, often Kircherian, approach to writing. As examples I should like to single out: lack of a clear understanding of the function of logography in contrast to ideography and the misguided notion of the allegedly alphabetic, rather than syllabic, character of the West Semitic writing.

The question of the relationship of writing to language has been pursued in recent years mainly by scholars with a background in linguistics. Because of their interest in modern languages and writings, this implies generally relations between the alphabet and language. A general treatment of the subject can be found in the respective chapters of the introductory manuals to linguistics by Bloomfield, Gleason, Francis, Hill, Hall, Hoenigswald, and Lehmann. A more detailed treatment is offered in the articles by Vachek 1939, Vachek 1945-49, Vachek 1959, Uldall 1944, Pulgram 1951, Pulgram 1965, and McIntosh 1956. Due to the preference for synchronic-descriptive approaches, rather than for diachronic-historical, linguists generally have stressed the independent character of writing and have preferred to study it as an independent system, rather than as a system ultimately derived from and related to the underlying language. Both approaches seem justified to me. Scholars have as much right to point out the close interrelations between writing and language as they have to study writing as a relatively closed system, without being involved in matters of relationships between writing and language and the degree of their interdependence.

BIBLIOGRAPHY

Boedberg, Peter, "The Chinese Script: An Essay on Nomenclature (the First Hecaton)," Studies Presented to Yuen Ren Chao (Taipai, Taiwan, 1957).

Cohen, Marcel, La grande invention de l'écriture et son évolution. Three volumes: Texte, Documentation et index, Planches (Paris, 1958).

Crossland, R.A., "Graphic Linguistics and Its Terminology," Mechanical Translation, III/1 (1956); republished in Proceedings of the University of Durham Philosophical Society I B (1957).

Francis, W. Nelson, The Structure of American English (New York, 1958).

Friedrich, J., Entzifferung verschollener Schriften und Sprachen (Berlin, 1954; 2nd edition, 1966). Also American and Russian editions.

Garvin, Paul L., editor, Natural Language and the Computer (New York, 1963).

Gelb, I.J., A Study of Writing: The Foundations of Grammatology (London and Chicago, 1952; revised edition, 1963).

Halliday, M.A.K., "Categories of the Theory of Grammar," Word XVII (1961), 241-292.

Hamp, Eric P., "Graphemics and Paragraphemics," Studies in Linguistics XIV (1959) 1-5.

Hockett, Charles F., A Course in Modern Linguistics (New York, 1958).

Istrin, V.A., Razvitiye pis'ma (Moskva, 1961), 2nd edition Vozniknovenie i razvitiye pis'ma (Moskva, 1965).

McIntosh, Angus, "The Analysis of Written Middle English," Transactions of the Philological Society 1956, 26-55.

Pulgram, Ernst, "Phoneme and Grapheme: A Parallel," Word VII (1951), 15-20.

Pulgram, Ernst, "Graphic and Phonic Systems: Figureae and Signs," Word XXI (1965), 208-224.

Read, Allen W., "An Account of the Word 'Semantic'," Word IV (1948), 78-97.

Sebeok, T.A., A.S. Hayes, M.C. Bateson, editors, Approaches to Semiotics. Cultural Anthropology, Education, Linguistics, Psychiatry, Psychology (The Hague, 1964).

Trager, George L., "Paralanguage: A First Approximation," Studies in Linguistics XIII (1958).

Uldall, H.J., "Speech and Writing," Acta linguistica IV (1944), 11-16.

Vachek, Josef, "Zum Problem der geschriebenen Sprache,"
Travaux du Cercle Linguistique de Prague VIII
(1939), 94-104.

Vachek, Josef, "Some Remarks on Writing and Phonetic
Transcription," Acta linguistica V (1945-49), 86-93.

Vachek, Josef, "Two Chapters on Written English," Brno
Studies in English I (Praha, 1959), 7-34.

ON THE EVOLUTION OF THE RUNIC ALPHABET

Peter H. Salus

(University of Massachusetts, Amherst)

The Germanic runic script is alphabetic: in principle each character represents a single feature matrix; in actuality some symbols represent several different matrices.

The origin of this system of writing still remains something of a mystery. There have been a number of theories advanced, some probable, some truly absurd. Behrens (1941:41) typifies the most absurd in attempting to derive 'all writing . . . from the rune-hoard of the Stone Age'.

The two related questions I want to consider here concern the origin of this Germanic runic alphabet and its development.

There are three basic theories as to the origin of the Germanic runes: derivation from the Greek alphabet (von Friesen 1904), from the Latin (Wimmer 1874; Agrell 1938), and from the Etruscan-North Italic (=Alpine) scripts (Marstrander 1928; Hammarström 1930; Arntz 1944; Elliott 1959). It is this last theory which has the widest acceptance today, and I would like to supply additional evidence to support it.

The theory of the derivation of the runic alphabet from the North Italic scripts is based upon the fact that about two-thirds of the 'Common Germanic' runes have shapes which can be traced to characters of the Alpine alphabets, while fewer than half of the characters correspond with either the Greek or Latin alphabets. My argument will not really be based upon graphemic, but rather upon phonological evidence, and it will involve the 16-character Scandinavian runic alphabet rather than the more frequently cited 24-character 'Common Germanic' one.

Before going into the evidence, it will be necessary to glance at the significant allophones of North Germanic. (Table 1)

There are four major series of Scandinavian runes: Danish, Swedish-Norwegian, Orkney, and Hälsinge. Each of these sets contains sixteen symbols: f u þ a r k h n i a R s t þ m l (see Table 2). Elliott (1959:23) points out that most of these -- as must be apparent from inspection of the table -- 'carried multiple phonetic values'. Thus, þ denotes both +voi and -voi, as do k and t; þ denotes +cont, +voi as well as -cont, +voi and -cont, -voi. The vowels show considerable

TABLE 1
The Allophones of Germanic

Consonants

	f	p	x	h	t	k	b	v	d	g	g	s	z	l	m	n	n	r	R
consonantal	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
sonorant	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	*	+	+
grave	+	-	+	-	+	+	+	-	-	+	-	+	-	+	-	+	-	-	-
diffuse	+	+	-	+	+	-	+	+	+	-	-	+	+	+	+	+	-	+	+
strident	+	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	+
nasal	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-
continuant	+	+	+	+	-	-	+	-	+	-	+	-	+	+	+	+	-	-	-
voiced	-	-	-	-	-	-	+	+	+	+	+	-	+	+	+	+	+	+	+
tense	+	+	+	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-

Vowels

	i			u	
+diff					
-diff		e		o	-comp
					+comp
		a			
-grav	+grav				
		-flat	+flat		

TABLE 2

	Etruscan	Alpine	Scand. (Orkney)	Common Germanic	English (Thames)
f	Φ	F	Φ	Φ	Φ
v	YY	Λ	C	C A N R V	C A N R C
b		BB	C A T R		
a	A	Λ	R	(a)	
r	Ω	D D D	R		
k	>K	<K	R		
g					
w					
h	B				
n	N				
i	I				
j			*	S	S
e			T	L	L
p	q	M L			
z	I #	↑			
s	Z Z	↑			
t	T +	↑			
b		B			
e	Y	Φ			
m	M M	W			
l	✓	↑			
n	Ω				
a			(s)		
o		Φ			

(+ a, o, y, and ēa in English)

(æ)

(x)

(b)

(d)

(æ)

variation. u may be used for any +flat vowel, i for any -grav, -comp vowel. à is both +comp and +nas.

If we scan the Latin and Greek alphabets, we cannot help but note that there are separate symbols for +voi and -voi (e.g. C and G in Latin, or T and Δ in Greek). Looking at Etruscan, however, we note that whereas there are symbols for allophones which are +cons, -cont, -voi, there are none for allophones which are +cons, -cont, +voi (Pallotino 1955:259; Jensen 1958:480). In the Alpine scripts (Rhaetic, Leopontic, Venetic) we also find no +cons, -cont, +voi allophones represented in the script (Jensen 1958:486).

The numbers game finds support in the Alpine scripts, but, more important, the orthographic convention of representing +voi and -voi with one symbol can only have been derived from North Italic, it cannot have been derived from either Greek or Latin. We can find further phonological support in the fact that Etruscan conflates the two +flat vowels, just as the Scandinavian runic alphabets do. (Pallotino 1955:261)

Having found phonological support for the graphemic evidence, we can now concern ourselves with the development of the runic alphabet. And here we must deviate from the customary approach.

Most scholars have attempted to derive the 16-character Scandinavian scripts, the 25-character set of the Saleby stone (ca. 1228 A.D., Sweden), the 28-character English series, and the 33-rune Northumbrian set from the 24-character 'Common Germanic' series. It is my contention that, despite the dating of the various inscriptions, the 16-rune Scandinavian alphabet represents the oldest stage of runic writing, and all the other sets are amplifications and expansions of this alphabet.

To demonstrate I will attempt to 'generate' the 16-rune set from North Italic and the 24-rune set from these Scandinavian runes.

First of all, f u þ à k h n i s t þ m l are almost the same in Etruscan-North Italic and the Scandinavian runic scripts. The r, a and R are not so simple, though R and Alpine z are fairly close (R and Z). As the feature matrix of R is fairly close to that of z (R is +son, -cont where z is -son, +cont; they are identical in all other features), the use of the shape may be an adaptation. The Scandinavian a-rune is the same as the 'Common Germanic' j-rune, which finds no parallel in either Etruscan or North Italic. It is closest to the Latin symbol G, but also identical to one of the pre-runic Germanic symbols which has no known phonetic value. r resembles the r in one Alpine alphabet, but

not any other. It is quite similar to Latin R. It is possible that both a and r came into North Italic from Latin and thence into runic, or that they were borrowed directly from Latin into runic. Unless some inscription illustrating some transition phase is unearthed, there is no way of solving this problem. At least 12 of the 16 Scandinavian runes are directly traceable to Etruscan-North Italic and two others are most likely derived from an Alpine script. The remaining two are more like Latin than Greek, but little more can be established with certainty.

The 'Common Germanic' 24-rune alphabet adds g w ê p e n, o ð to the Scandinavian series. Of these, w and o have analogues in North Italic (but not in Etruscan). g resembles Greek χ , but is not at all like the symbol for [kh] in Etruscan. e, ê and p have no known analogues in any script. ð is very like the North Italic character for s, and as these sounds differ only in voice and stridency we may consider them fairly close. n, has an analogue in Etruscan (but not in North Italic) in the symbol which represents q (the k-allophone before +grave vowels). Here, obviously, there is little phonetic correspondence. However, it must be noted that there are pre-runic Germanic symbols which are very much like the runes for g ê n, ð and o, and it is possible that they were borrowed from the pre-runic, non-phonetic signs.

The Old English runic alphabet (as found on the Thames scramasax and in the Vienna codex) adds four more symbols to this system. It shifts the rune for a to æ (a -diff, -grav, -flat, +comp vowel) and invents new symbols for a o y and êa (f , m , n , and w). The new a and o are obvious modifications of the old a-rune, as though the two +grav, -diff vowels were seen as having something in common. The new runes for y and êa are not related to any other symbols that survive.

The answers to the two rhetorical questions posed earlier must now be obvious: the Germanic runic alphabets find their origin in the Etruscan-North Italic scripts, and this origin can be documented not only by using the graphemic shape-similarities usually adduced, but also by means of the allophonic patterning of the feature matrices represented by the symbols; secondly, in attempting to trace the development of the runic writing system, the 24-rune 'Common Germanic' system ought not be used as the base, but rather the 16-rune set found in Scandinavia, from which, in turn, both the 'Common Germanic' and Old English runic alphabets can be derived.

¹ I am indebted to James E. Cathey for his comments on this paper.

REFERENCES

- Agrell, S. 1938. Die Herkunft der Runenschrift. Lund.
- Arntz, Helmut. 1944. Handbuch der Runenkunde. 2nd ed.
Halle/Saale.
- Behrens, E. 1941. Zur Herkunft der Runen und zu ihrer
Verwandtschaft mit vorgeschichtlichen und ge-
schichtlichen Schriften. Leipzig and Strasbourg.
- Elliott, Ralph W. V. 1959. Runes: an introduction.
Manchester.
- von Friesen, Otto. 1904. Om runenskriftens härkomst.
Uppsala.
- Gelb, I. J. 1963. A study of writing. Rev. ed. Chicago.
- Hammarström, M. 1930. Om runenskriftens härkomst. Stud.
i nord. filol. 20.1ff.
- Jansson, Sven B. F. 1962. The Runes of Sweden. Trans.
by P. G. Foote. London.
- Jensen, Hans. 1958. Die Schrift in Vergangenheit und
Gegenwart. 2nd ed. Berlin.
- Marstrander, C. J. S. 1928. Om runene og runenavnenes
oprindelse. NTS 1.85ff.
- Pallotino, Massimo. 1955. The Etruscans. Harmondsworth,
Middlesex and Baltimore.
- Wimmer, L. F. A. 1874. Runeskriftens oprindelse og ut-
vikling i norden. Copenhagen.
- 1887. Die Runenschrift. Berlin.

ON THE ORIGINAL DIRECTION OF THE BRAHMI SCRIPT¹

Michael C. Shapiro (University of Chicago)

The theories concerning the origin of the Brāhmī script may be divided into two classes, namely those which consider it to be of indigenous Indian origin and those which do not. First I will consider those theories which maintain that the script is of native Indian origin.

Dowson² and Cunningham³ are the proponents of what is probably the most widely held theory of the Indian origin of the Brāhmī alphabet. Briefly, it reads as follows:

1. The phonetic arrangement of the characters (which he attributes to Pāṇini) is an indication of the prior development of the alphabet.
2. The claims made as to the similarity of the Brāhmī letters to the characters in certain Semitic alphabets are invalid and far-fetched.
3. The Indian system of vowel notation is indigenous.
4. The old Indian alphabet did not have a retroflex series and that series was a later addition.

Furthermore, in Vol. I of the Corpus Inscriptionum Indicarum, Cunningham traces the Brāhmī from a hypothetical set of hieroglyphs.⁴

Pandey,⁵ on the other hand, maintains that the Brāhmī could only have been derived from the Indus script. Shamasastri,⁶ on the other hand claims that the Brāhmī is an outgrowth of Tantric signs. The above-mentioned hypotheses, as well as the others to be mentioned in this paper, will be discussed in a later section.

The following theories consider Brāhmī to be of non-Indian origin. Halévy⁷ presents a theory in which he claims Brāhmī to have evolved out of a mixture of Kharosthi, Greek, and Arabic characters. This theory would have it that eight consonants were taken directly from the Aramaic alphabet of the fourth century B.C., six consonants, two initial vowels, the medial vowels as well as

the anusvara from the Kharoṣṭhi, and five consonants from Greek. Burnell⁸ considers Brāhmī to be ultimately of Babylonian or Bactrian origin.

The remaining theories consider Brāhmī to be of Semitic origin. These theories may be divided into two classes, namely those maintaining that it is of North Semitic origin, and those saying it is of South Semitic origin.

Deecke⁹ feels Brāhmī to be ultimately derived from Assyrian cuneiform characters through an ancient South Semitic alphabet, which was also the parent of Himyaritic. Taylor¹⁰ proposes that Brāhmī was taken from a South Arabic alphabet, also the parent of Himyaritic. As Taylor's views on this subject are widely held, I shall summarize them here. Basically, they read as follows:

1. The Indo-Bactrian and the Aśoka must have been developed out of two earlier alphabets, both of which possessed an insufficient number of characters for the necessities of Indian speech.¹¹
2. Two distinct and ancient types of the primitive Semitic alphabet were independently introduced at distant points of Indian territory and both underwent a gradual evolution, and finally a systematic redaction and arrangement was effected by scientific grammarians who were acquainted with both alphabets.¹²
3. The "absolute uniformity" of the Aśokan characters, from the most distant provinces, is difficult to explain except on the hypothesis of a comparatively recent importation of a foreign alphabet.
4. Brāhmī was introduced through commercial intercourse.
5. The form of South Semitic script which reached India was Sabæan and must have branched off from Phoenician some time not later than the sixth century B. C.¹³

Taylor's chart of correspondences between Brāhmī and South Semitic is reproduced on figure 1. Weber¹⁵ and Bühler¹⁶, on the other hand, maintain that Brāhmī is derived from North Semitic. As Bühler's views on this matter are widely held, particularly by such modern scholars as Jensen and Dani¹⁸, I will treat them

Name.	Value.	Safa.	Saba.	Primitives.	Derivatives.	
Alf	'a	X X	^K	K a	K a	1
Bet	b	ג ב	נ נ	ו b	ו b	2
Gēmel	g	T	ת	א g	א g	3
Dent	d	ד ב	כ כ	ב d	ב d ב dh	4
Hoi	h	ו	ו	ו h	ו h	5
Wawe	w	ו ו	ו ו	ו u	ו u ו o	6
Zai	z	ח	ח	ח jh (dzh)	ח jh (dzh)	7
Kharm	kh	ש	ש	ו gh	ו gh	8
Tait	t	ח	ח	ו th	ו th ו dh	9
Yaman	y	ו י	ו י	ו v	ו v ו y	10
Kaf	k	כ כ	כ כ	כ kh	כ kh	11
Lawe	l	ל	ל	ל r	ל r	12
Mai	m	ב	ב	ב m	ב m	13
Nahas	n	ל	ל	ל n	ל n ת n צ n ח n	14
Sat	s	ח ח	ח ח	ו s	ו s	15
Ain	'a	א	א	א e	א e א i א i	16
Ef	f	ו	ו	ו p	ו p ו ph	17
Tsadai	ts	ת	ת	ת z	ת z מ z	18
Qaf	q	פ +	פ	ב chh	ב ch + k	19
Rees	r	ו ש	ו ש	ו l	ו l	20
Saut	sh	ס א	ס א	ס j	ס j ס sh	21
Tawe	t	خ خ	خ خ	خ t	خ t	22

Figure 1

further here. They may be summarized as follows:

1. A certain methodology must be followed in the comparison of forms of different scripts. Briefly,
 - a. In the comparison of characters thought to be related to one another, the oldest or fullest forms must be used and the originals from which they are derived must belong to the

types of one and the same period.

- b. The comparison may indicate only such irregular equations as can be supported by analogies from other cases where nations have borrowed foreign alphabets.
- c. In cases where the characters show considerable differences from the supposed prototypes, it is necessary to show that there are fixed principles, according to which the changes have been made.¹⁹
- 2. The legend of the Eran coin, which runs from right to left, and the letters which appear turned around in the Asoka edicts and more frequently in the Bhāttiprolu inscriptions, point to the correctness of the view that Semitic signs are the prototypes of the Brāhmī characters.²⁰
- 3. The derivation of Brāhmī from North Semitic meets the conditions of 1. c above. Bühler states the following:
 - a. The letters are set up as straight as possible, and with occasional exceptions in the case of ta, tha, and ba, they are made equal in height.
 - b. The majority consist of vertical lines with appendage attached mostly at the foot, occasionally at the foot and at the top, or rarely in the middle; but there is no case in which an appendage has been added to the top alone.
 - c. At the top of the letters appear mostly the ends of vertical, less frequently short horizontal strokes, still more rarely curves on the tops of angles opening downwards, and quite exceptionally in ma and one form of tha, two lines rising upwards. In no case does the top show several angles, placed side by side, with a vertical or slanting line hanging down, or a triangle or a circle with a pendant-line.²¹
- 4. Owing to what Bühler calls "the inclinations and aversions of the Hindus," the heavy tops of many

Semitic letters had to be gotten rid of, by turning signs topsy-turvy or laying them on their sides, by opening the angles, and so forth.

5. Finally, the change in direction necessitated a further change in as much as the signs had to be turned from right to left, that is to say, flipped over an imaginary vertical axis, as in the history of Greek script.

Using some earlier material from Bühler, Dani elaborates on the above and shows clearly which symbols are taken from the Semitic and which are in turn derived from other symbols, i.e., the clear interrelation between the symbols for e and i and between those for th and th̄. This is summarized in Figure 2.

The above-mentioned theories represent only a small part of the literature dealing with the origin of the Brāhmī. Both Prinsep²² and Senart²³ maintain the Brāhmī to be of Greek origin. The indigeneous origin of Brāhmī has also been suggested by Lassen,²⁴ Thomas,²⁵ Jayaswal,²⁶ Ojha,²⁷ Hunter,²⁸ and Sircar.²⁹ The Semitic origin of Brāhmī was suggested as far back as Sir William Jones,³⁰ a view which was narrowed to pinpoint Phoenician by R. N. Cust³¹ and Benfey,³² Aramaic by Diringer,³³ and South Semitic in general by Sethe.³⁴

II

All of the theories listed above have come under attack at one point or another. I shall here attempt to list some of the more cogent arguments against each of them.

The best that can be said for General Cunningham's theory is that there is no reason for believing it. Although it is not a priori impossible that derivations such as the one which relates the Brāhmī  to a supposed earlier pictogram representing the word for yoni are valid, there is no evidence to demonstrate that this is indeed the case. Jensen has rightly pointed out that we have no trace of such an earlier alphabet either in inscriptions or in tradition,³⁶ and in the absence of such, any derivation becomes as valid as any other. Any of the infinite set of internally consistent reconstructions would be as correct as any other and nevertheless be on about the same level of validity as any one of the so-called decipherments of the Indus script.

Altsemitisch		Älteste indische Schrift		
Lautwert	Zeichen	Zeichen	Lautwert	Neugebildete Zeichen
a'	א	א	a	ା
t	ת	ତ	ta	ତ୍ତା
g	ג	ଗ	ga	ଗ୍ରା
d	ד	ଦ	da	ଦ୍ରା
h	ח	ହ	ha	
v	ו	ବ	va	ବୁତ୍ତା
z	ז	ଶ	ga	ଫ୍ରା
b	ବ	ବ	gha	
t̄	ଠ	ଠ	tha	ଥ୍ରା
j	ଜ	ଜ	ga	
k	କ	କ	ka	
l	ଲ	ଲ	la	ଲ୍ଲା
m	ମ	ମ	ma	ମ୍ମା
n	ନ	ନ	na	ନ୍ନା
s	ଶ	ଶ	sa	ଶବ୍ଦା
c	ଚ	ଚ	ca	ଚାଇଃି
p	ପ	ପ	pa	ପର
f	ଫ	ଫ	ca	ଫା
q	କ	କ	ka	
r	ର	ର	ra	
ସ	ଶ	ଶ	ଶା	
t̄	ଠ	ଠ	ta	

Figure 2

Shamasasty's theory has less working inherently against it in as much as he has the Tantric formulas and signs with which to work. Nevertheless, Shamasasty's theory is invalidated by chronological considerations. As Upasak has stated, "The evidence produced by him comes from the Tantric texts alone, which are of very late origin. Moreover, the symbols and terms used in such texts are very ambiguous and obscure in nature. A theory built upon such a material leads us to completely erroneous con-

clusions; and hence not much reliance can be placed upon it."³⁷

Some of the other arguments in favor of an indigenous origin for Brāhmī are based on linguistic factors. The arrangement of the alphabet based on articulatory criteria as well as the derivation of certain of the characters from certain other ones on phonological grounds are said to presuppose the existence of a school of phonetic thought; moreover, since this patch-work was done with the intent of making some script suitable for indigenous Indian languages, it has been supposed that this alphabet was clearly created by Indians. Such views are held by Dowson,³⁸ Jayaswal,³⁹ and Ojha.⁴⁰

These linguistic arguments for the indigenous origin of the Brāhmī are misleading on several counts. There can be no doubt that the arrangement of the Indian alphabet was carried out on phonological principles. This arrangement is unquestionably the work of men highly sophisticated in phonetics. The concept of varṇa, according to which the consonants are arranged in classes consisting of voiceless unaspirate, voiceless aspirate, voiced unaspirate, voiced aspirate and nasal at a given place of articulation seems clearly to indicate the presence of a pandit's mentality. But the fact that some scholarly minds were at work in the arrangement of characters in no way indicates that they had anything to do with the creation of these characters. To draw a simple analogy, one could arrange the Roman characters in an order which would correspond to the articulation of the sounds corresponding to them, i. e., the vowels first, followed by the velar consonants, followed by the alveo-palatals, followed by the velars etc. In fact, when we describe the Sanskrit phonological system in various scholarly works today, we often describe the sounds with transliterations into Roman characters, which are then arranged in the articulatory pattern used by the Sanskrit grammarians to categorize the sounds of Sanskrit. In short, the arrangement of symbols into a given pattern by pandits in no way indicates that these symbols were created by them.

The argument maintaining that certain Brāhmī signs were derived from others and that these derivations were carried out on phonetic grounds, i. e., the interrelation of i, e, and ai, e and ai being the guṇa and vṛddhi of i, and that this shows that the creation of Brāhmī was carried out by pandits, clearly is misleading. As Dani has shown,⁴¹ what this demonstrates is that certain of the signs were derived by the pandits. Yet, if we subtract the number of derived signs from the total number of characters, we are left

with 22, the number of consonants in North Semitic. And it is these 22 graphemes which show the strongest resemblance to Semitic forms.

The arguments proposing the Indus script to be the origin of Brāhmī are equally unsatisfactory. Since the Indus script is untranslatable at our present state of knowledge, and furthermore, since we do not even know if the script was alphabetic, syllabic or ideographic, the attempt to derive Brāhmī from it seems unverifiable, particularly in light of the long chronological period between the extant examples of the two scripts.

Halévy's views have often come under heavy fire. Jensen has noted that his theory is both geographically and chronologically impossible.⁴² Bühler dismisses Halevy on literary and paleographic evidence "which makes it more than probable that Brāhmī was used several centuries before the beginning of the Maurya period, and had had a long history at the time to which the earliest Indian inscriptions belong."⁴³

There has also been extensive criticism of the view taken by Deecke and Taylor. Jensen mentions that their comparisons are not convincing and that it is unmethological not to always treat the earliest forms, especially when they conflict.⁴⁴ Bühler also notes that Deecke and Taylor's theories do not meet the prerequisites which he set up for the explanation of correspondences between two different scripts. He states,

Deecke and Taylor do not fulfil the absolutely necessary conditions and it probably would not be possible to obtain satisfactory results, even if all the impossible equations are given up, and the oldest Indian signs in every case are chosen for comparison. It would be necessary to assume that several Sabaean letters, such as aleph, gimel, zain, teth, goph, and resh which show strong modification of the North-Semitic forms, had been again made similar to their prototypes on being converted by the Hindus into a, ga, ja, tha, pa, kha, and ra. In other cases, it would be impossible to show any connection between the Sabaean and the Indian signs. These difficulties disappear with the direct derivations of the Brāhmī from the oldest North Semitic alphabet, which shows the same type from Phoenicia to Mesopotamia.⁴⁵

Furthermore, we have no instance of the early form of Aśokan which would serve as a link between South Semitic and extant Brāhmī. Taylor writes "The actual ancestral type of the Aśoka alphabet is unknown, but there is no reason why it should not be ultimately discovered in the unexplored regions of Oman or Hadramaut, or among the ruins of Ormus, Bahrein, Gerrha, or some other center of primitive commerce on the shores of the Persian gulf."⁴⁶ Bühler questions the likelihood of such a discovery.⁴⁷

Likewise, Taylor raises charges against the adherents of a North Semitic origin for Brāhmī. He states "A very superficial examination will suffice to show that the Aśoka alphabet, though it offers hardly any appreciable resemblance to any of the North Semitic alphabets, agrees in a very remarkable way with the general type of the alphabets of the South Semitic family."⁴⁸

The question of the origin of the Brāhmī script will be taken up again in the next section in conjunction with material concerning the original direction of this script.

III

One of the central issues involved in the determination of the origin of the Brāhmī script is the matter of the original direction of this script. It would, for instance, be much simpler to maintain the Semitic origin of Brāhmī if we could show that this script was originally written from right to left. The argument in favor of saying that Brāhmī was originally written from right to left are based upon a limited amount of material. I shall attempt to present some of it here.

The first relevant source of information is an inscribed coin from Eran.⁴⁹ The six characters on this coin seem to be the reverse of the standard form of Brāhmī; that is, if each character is revolved 180° over a theoretical vertical axis, the result will be the standard Brāhmī form. The forms on the coin read

K D J Y U J instead of K D L Y J L . Diringer seems to feel that this indicates an earlier stage of Brāhmī than we find in the other inscriptions.⁵⁰

Also relevant are the Bhāṭṭiprolu casket inscriptions. Bühler states in his article on these caskets that "the dental media d exactly resembles the Maurya letter but is turned round like the Andhra da and that of the modern Devanagari. The same remarks apply to the fourth labial bh."⁵¹ Any analysis of the casket itself will place this statement in a new light. Thus, although it is true that on the lid of the third casket we have the reversed symbol for d, on the edge of the box of the same casket we have the normal form for the same sign. (See Figures 3 and 4.) We may then say that in this inscription we get an alternation between the two forms.

Sircar,⁵³ in Select Inscriptions also notes some relevant information in the Duwe Gala Cave inscription, no. 7. In this inscription each grapheme faces in the standard direction, but the order of the graphemes reads from right to left. Thus the actual text reads मवनः गणेशित्रयं लप्यते, while the Sanskritized reads वाम गणेशित्रयं लप्यते. Sircar attributes this inscription "to the error, ignorance or ideo-syncrecy of the engraver as is the case of early medieval inscriptions engraved negatively or to be read from bottom to top to bottom."⁵⁴

The last important source of material is the Erragudi minor rock edict.⁵⁵ This inscription has been called a "confused mass of writing."⁵⁶ As Upasak has noted,⁵⁷ of the 26 lines of text, as many as eight are written in reverse. Furthermore, there are many startling changes of direction appearing in mid-line. Figure 5 is a line representation of this inscription. Many of the theories called upon to account for this inscription are absurd. R. B. Daya Ram Sahani's theory that the inscription reflects an earlier boustrophedon style of writing is unfounded.⁵⁸ First of all the inscriber is totally inconsistent in his own usage. Boustrophedon is as regular a style of writing as is any other and one cannot simply describe an inconsistently inscribed text as being in boustrophedon because of this fact. Furthermore, there is no other substantiating evidence to support this claim. B. M. Barua's suggestion that the scribe executing the inscription was used to inscribing Kharoṣṭhī and thus kept falling back to Kharoṣṭhī while inscribing Brāhmī is totally unfounded also.⁵⁹ The refutation of this theory is obvious. Were the scribe to suddenly forget himself and fall into writing Kharoṣṭhī he would surely be brought back to his senses by the obvious fact that he would have to inscribe over rock which had already been inscribed on, that is, to change direction would necessitate going back over ground already covered. In order to have a clean space on which to inscribe, the scribe



Figure 3

NORMAL	BHATTIPROLU
þ	q + þ
I	

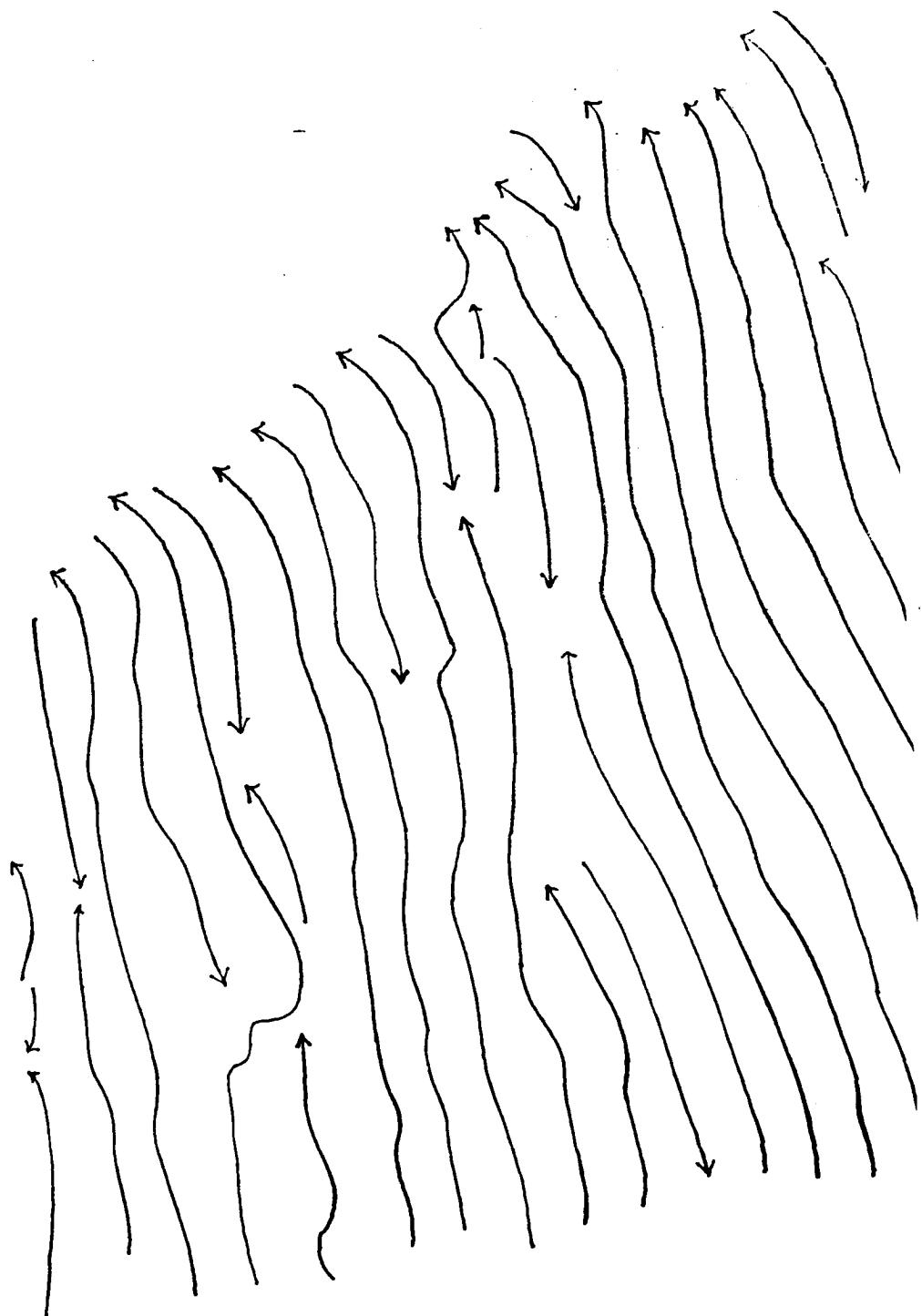
Figure 4 52

would have to make a sizeable jump, by which time he surely would have remembered in which script he was writing. It is most difficult to disagree with Upasak when he says concerning the Erragudi Minor Rock Edict "Thus it is very difficult to pronounce anything about the style of writing with certainty."⁶⁰

It is thus difficult to draw valid conclusions concerning the original direct of the Brāhmī script. Nevertheless, some of the evidence does allow certain tentative statements to be made. In the case of the Eran coin, we have each grapheme reversed in direction. The Bhattprolu inscription has certain individual characters standing both in proper and reversed direction. The Duwe Cave Inscription has characters which individually stand in the right direction but whose order is reversed. The Erragudi inscription has letters in the normal form but as was noted above the linear arrangement of them is terribly confused.

It seems to me that of all of the theories presented above, the one having the most validity is that proposed by Weber and Bühler. With the exception of Halévy's, which has been shown to be impossible, none of the other theories is a priori out of the question, but lack evidence to substantiate them. Bühler is the only investigator to set up rules which when applied will actually produce the Brāhmī from an earlier script, and which has, furthermore, the actual scripts in question extant. Moreover, his comparisons are convincing. Most of the indigenous origin theories are mere speculations, interesting hypotheses but, for the most part, methodologically insecure. Taylor's views are more difficult to refute, and in fact, all one can say is that Bühler's derivations look more convincing, and furthermore are not based on scripts yet to be found.

In the light of the above, the Eran coin and the Duwe Cave Inscription might serve as important substantiating evidence for the claim that the Brāhmī is of Semitic origin. The coin might demonstrate that at an early stage the Brāhmī script was still written from right to left and this direction was later changed in accord with the principles laid down by Bühler. Similarly, the Duwe Cave Inscription might conceivably reflect the form in which the Brāhmī script was introduced into Ceylon.⁶¹ Nevertheless, all of the material concerning these sources is uncertain and, as must be the case with such a small amount of material, must be used with great caution.

Figure 5⁶²

NOTES

1. An earlier version of this paper was written at the University of Pennsylvania in November, 1967 in a course taught by Professor A. L. Basham. I would like to thank Professor Basham for his kind help and encouragement. I would also like to thank Arlene R. K. Zide and Ram Dayal Munda of the University of Chicago and Professor Peter H. Salus for their helpful suggestions.
2. "The Invention of the Indian Alphabet," JRAS, n. s. XIII (1881), p. 119ff.
3. Coins of Ancient India, Varanasi, 1961, pp. 38-42.
4. 1877.
5. Indian Palaeography, part 1. Varanasi, 1952, p. 31.
6. "A Theory of the Origin of the Devanagari Alphabet," Indian Antiquary, XXV (1906), pp. 253-67, 270-90, 311-24.
7. Revue Semitique, 1895, pp. 225, 372.
8. Elements of South Indian Palaeography, London, 1878.
9. "Über das indische Alphabet im Zusammenhang mit dem übrigen Südsemit. Alphabeten." Ztsch. d. dtsh. morgenl. Ges., XXI (1877), 598-612.
10. The Alphabet, Vol. 2, London, 1883.
11. Ibid., p. 201.
12. Ibid., pp. 303-4.
13. Ibid., p. 316.
14. Ibid., p. 320.
15. Ztsch. d. dtsh. morgenl. Ges., X 389ff.
16. Indian Palaeography, Calcutta, 1959, 24-7 and in On the Origin, pp. 53-91.
17. Geschichte der Schrift, Hanover, 1925, pp. 144-48 and in Die Schrift in Vergangenheit und Gegenwart, Berlin, 1958, p. 346.
18. Indian Palaeography, Oxford, 1963, p. 28.
19. Ind. Pal. p. 24.
20. Ibid., p. 25.
21. Ibid., pp. 25-6.
22. In Essays on Indian Antiquaries, ed. E. Thomas, London, 1858.

23. Les inscriptions de Piyadasi, vol. 2, Paris, 1881
24. Cited in Taylor, Vol. 2, p. 305.
25. Ibid.
26. Cited in Upasak, The History and Palaeography of the Mauryan Brāhmī Script, Patna, 1960, p.13.
27. Ibid.
28. The Script of Mohenjo-Daro and its connection with other Scripts, London, 1931.
29. Indian Epigraphy, Delhi, 1965.
30. Cited in Taylor, Vol. 2, p. 305.
31. Cited in Upasak, p.10.
32. Cited in Taylor, Vol. 2, p. 305.
33. The Alphabet, New York, 1962, pp.144-5.
34. Cited in Upasak, p.10.
35. Ind. Pal., German ed., p.12.
36. Geschicht der Schrift, pp.144-8.
37. Upasak, p.13.
38. Cited in Taylor, Vol. 2, p. 305.
39. Cited in Upasak, p.13.
40. Ibid.
41. pp. 23-30.
42. Geschicht der Schrift, p.145.
43. Ind. Pal., pp. 24-5.
44. Geschicht der Schrift, p.145.
45. Ind. Pal., p. 25.
46. pp. 321-2.
47. Origin, pp. 35-43.
48. p. 318.
49. See Cunningham, Coins of Ancient India, plate XI:16; also reproduced in Diringer, L'Alfabeto nella Storia della Civilità, Firenze, 1937, p. 622.
50. L'Alfabeto, p. 627.

51. Epigraphica Indica, II, part xiv (1898), p. 323-9.
52. Ibid., plates.
53. Calcutta, 1965, p. 242.
54. Ibid.
55. Reproduced in Arch. Survey of India, Annual Report, 1928-9, p. 164.
56. Upasak, p.141.
57. Ibid.
58. Cited in Upasak, p.141.
59. Indian Historical Quarterly, IX, p.113ff.
60. Upasak, p.141.
61. Sircar, p. 242.
62. Upasak, p.320.

BIBLIOGRAPHY

- Barua, B. M. Indian Historical Quarterly, IX, pp. 113ff.
- Bühler, G. Epigraphica Indica, part xiv (1898), pp. 373-9.
- _____, Indian Palaeography, Calcutta, 1959.
- _____, On the Origin of the Indian Brāhma Alphabet, Strassburg, 1898.
- Burnell, A. C. Elements of South Indian Palaeography, 2nd ed., London, 1878.
- Cunningham, A. Coins of Ancient India, Varanasi, 1963.
- _____, Corpus Inscriptionum Indicarum, 1877.
- Dani, A. H. Indian Paleography. Oxford, 1963.
- Das Gupta, C. C. The Development of Kharosthi Script, Calcutta, 1958.
- Deecke, W. "Über das indische Alphabet im Zusammenhang mit das übrige Südsemitische Alphabet," Zeit. d. dtsc. morgenl. Ges., 31 (1877), pp. 598-612.
- Diringer, D. The Alphabet, New York, 1962.
- _____, L'Alfabeto nella Storia della Civiltá, Firenze, 1937.
- Dowson, —. "The Invention of the Indian Alphabet," JRAS, n. s. XIII (1881), pp. 119ff.
- Halevy, M. J. Revue Semitique, 1895, pp. 223ff, 372ff.
- Hunter, G. R. The Script of Harappa & Mohenjo-Saro and its Connection with other Scripts, London, 1934.
- Jensen, H. Geschichte der Schrift, Hanover, 1925.
- _____, Die Schrift in Vergangenheit und Gegenwart, Berlin, 1958.
- Pandey, R. N. Indian Palaeography, Part 1, Varanasi, 1952.
- Prinsep, J. Essays on Indian Antiquities, ed. E. Thomas, London, 1858.
- Sahni, D. R. Arch. Survey of India, Annual Report, 1928-9, pp. 166-7.
- Senart, E. Les inscriptions de Piyadasi, 2 vol., Paris, 1881.
- Shamasastri, R. "A Theory on the Origin of the Devanagari Alphabet," Indian Antiquary, XXXV (1906), pp. 253-67, 270-90, 311-24.
- Sircar, D. C. Epigraphica Indica, XXXI, pp. lff.
- _____, Indian Epigraphy, Delhi, 1965.
- _____, Select Inscriptions, Calcutta, 1965.
- Taylor, I. The Alphabet, London, 1883.
- Upasak, C. S. The History and Palaeography of Mauryan Brāhmī Script, Patna, 1960.

A BRIEF SURVEY OF WORK TO DATE ON THE INDUS VALLEY SCRIPT

Arlene R. K. Zide (The University of Chicago)

1.0 Introduction.

1.1. The discovery of the cities of the Indus Valley Civilization in 1921 and 1922 by Banerji and Sahni at Harappa and Mohenjo-daro opened up a new chapter in the prehistory of man in India. The extensive excavations by Sir John Marshall and Ernest Mackay unearthed among other things, about 800 seals of steatite, and many copper plates incised with an unknown script of a pictorial, but, clearly, fully developed nature.

1.11 The script consists of approximately 300 signs, which can be divided into basic signs, about 250 in number, and various additional auxiliary marks which do not stand alone but which are used in combinations with the basic signs. These combinatory marks may be determinatives, vowel marks, punctuation, or other indications of gradation but it is not possible to determine at this point which analysis is correct. In addition to these signs, there occur many compound signs which are made up of the basic signs combined with each other, or basic and auxiliary signs combined. The number of signs has been variously presented as ca. 250 to ca. 375 in number; the discrepancies occur as a result of the manner in which one regards the auxiliary and compounded signs, the correct assignation often being a moot point. In any case, the number of signs, regardless of these various interpretations, would indicate that the script cannot be alphabetic; nor is it likely to be merely syllabic, but it is most probably assignable to a category which is known as logo-syllabic, i. e., composed of logograms ("ideograms") which are used syllabically as well as logographically. Thus, if we employed our English logogram 2 pronounced tu with a logogram for "day" ☼ to produce 2-☼ "today" i. e., tu-dey in rebus fashion, this would be a syllabic use of logographic symbols.

1.2 In evaluating the attempts at decipherment of the Indus Valley script (or of any script, for that matter), one must bring into operation a series of criteria by which one may fairly and carefully judge the validity of the conclusions. These criteria involve basically two aspects of the decipherment; a. the procedural and methodological evaluation of the script itself, and b. in the event of comparison with other extant scripts, the methodology of the comparison.

1. 21 Thus, a would-be decipherer must have a thorough knowledge of a) the historico-geographical area and its cultures, b) the field of grammatology, and c) the structure and typology of writing systems. His procedures must include assessment of the statistics of the script. A count of the signs yields some of the most valuable information about the structural nature of the script itself. It is known, for instance, that an alphabetic system would have 50-130 signs, and a logographic, or more properly, a logosyllabic system, since there are no "pure" systems, would generally be in the neighborhood of 400 or more signs, usually around 600 signs. It has been shown that even the highly logographic system usually attributed to Chinese is based on a system containing 540 "keys," or basic signs from which the vast number of Chinese signs are composed.

Further, the decipherer must make sign-lists, in order to establish the frequency and order of signs which yield important information about the graphotactics—the relevant position and sequence of the signs. In addition, word-division should be established, and numbers, being easy to recognize, must be assessed.

1. 22 In the comparison of scripts, the oldest or fullest forms should be used, and the originals from which they are derived must belong to the types of the same period; irregularities must be supported by analogies from other cases where foreign alphabets have been borrowed, and where the characters show considerable differences from their supposed prototypes, changes must be shown to have been made according to fixed principles.

1. 23 Comparisons of pictorially represented writing, especially, must conform to these criteria, since the comparison of outer form is for the most part useless; pictorial representations would necessarily have to have many points of resemblance, since the objects which are represented are basically the same for all people. "Thus, men, and parts of the body, animals and plants, tools and weapons, buildings and structures, sky, earth, water and fire are everywhere represented by pictures characterized by great simplicity in form. . . . There is no need to claim for these signs one single origin."¹

1. 3 The civilization of the Indus Valley has been dated as spanning approximately 500 years, though there have been some who would assert that a shorter chronology of about 3-400 years would be more

¹I. J. Gelb, A Study of Writing, p. 218.

in keeping with the archeological evidence.¹

1.31 S. N. Kramer would place the upper limit at 2800 B.C. on the basis of evidence from Susa and Kish,² since it would have taken several centuries for the Indus civilization to grow to the size it had become by 2500 B.C.

1.32 The chronology of the Indo-Aryan civilization in India on the other hand, has as its upper limit the date 1200 B.C., but recent reexamination of the literary evidence of the Rg Veda and Atharva Veda would place the limit as more probably around 1100 B.C., or even 1000 B.C. The archeological evidence as well points to such a conclusion, and to a definite break between the Indus Valley civilization and the advent of the Aryan invaders.³

2.0 The attempts at decipherment: most are invalid.

2.1 Most of the attempts to decipher the script must be considered to be invalid because of their lack of attention to the criteria as outlined above. Many of the attempts lacked completely in sound, rigorous methodological approach; others were unsuccessful because of their complete inadmissability on the grounds of chronological incongruity. Thus, any attempt to interpret or read the seals of ca. 2300 B.C., or possibly earlier as a form of Sanskrit, (even "archaic" Vedic), as proposed by such would-be decipherers as Pran Nath, Ray, Gadd, Karmarkar and so on, is immediately discredited on chronological grounds.

2.11 Similarly, attempts to connect the script of the Indus Valley with the symbols of the Tantric Code, (as for instance, do Barua, and Sankarananda), fail to take into account, the great discrepancy in time lapse between the end of the Indus Valley culture, and the emergence of the Tantric Code many centuries later. In addition, one should note that the symbols and terms used in the Tantric texts are extremely ambiguous and obscure in nature; thus reliance on a theory of connection based on the "Tantric" "code" would be, at best, tenuous.

¹V. N. Misra, "Prehistory and Protohistory," Review of Indological Research in Last 75 Years, ed. Chinmulgund and Mirashi, Poona, 1967, pp. 375- , and A. Ghosh, "Archaeology in India," Expedition, Vol. VI, no. 3, 1964, pp. 12-17.

²"The Indus Civilization and Dilmun, the Sumerian Paradise Land," Expedition, Vol. VI, no. 3, 1964, pp. 44-52.

³Ronald B. Inden, "Early India: Literary Evidence for a Revised Chronology," University of Chicago, unpublished lecture ms.) 1968, and A. Ghosh, op. cit.

2.12 Hrozny's "decipherment" of the Indus Valley script as Hittite, is invalidated for chronological reasons, aside from his obvious lack of sound procedure. As Diringer has pointed out, Hrozny's attempt is largely based on untenable strings of hypotheses, and unsupported revisions of chronological data, proposing a derivation of the Indus Valley script from the Hittite hieroglyphics of at least 1000 years later.

2.13 The contention by Waddell that the early Sumerians are Aryans and that the Indus Valley people were Early Sumerians (i.e., "Vedic Aryans" sic) is another excellent example of the disregard of chronological data. Further, his methodology is highly questionable, to say the least. His derivations of English, Greek, Hindi and Sanskrit words from Sumerian betrays a lack of linguistic sophistication as to the nature of reconstructions and the need for rigorous, scientific method in accomplishing a decipherment.

2.2 Heras, similarly, betrays a certain amount of linguistic naivete in his reconstructions of "proto-Dravidian" as the language of the Indus seals. Although his notion of chronology is far more sophisticated than that of the preceding authors, his reconstructions of "proto-Dravidian" are not scientifically credible, since they are not based on sound historical and comparative techniques. His "proto-Dravidian" bears far too close a resemblance to Old Tamil than would be likely for a language of the (projected) age of "proto-Dravidian". Proto-Dravidian is probably one of the more likely languages to be considered as a candidate for the language of the seals of the Indus Valley, since the chronological and archeological data would not discredit such an hypothesis, at least. But Fr. Heras' "decipherment" leaves much to be desired in the way of scientific accuracy and methodological procedures. Lacking a solid reconstruction of "proto-Dravidian" forms, a scientifically and linguistically accurate notion of the structure of the writing system, and rigorous systematic procedures, Heras' "decipherment" cannot be taken seriously.

2.3 The comparison of scripts involves many pitfalls as well. Thus, the comparison of the Indus Valley script with the so-called "script" of Easter Island by de Hevesy, taken up later by others such as Heine-Geldern and Bilimoria is a striking example of fuzzy thinking. The Easter Island script is not a script at all. Unlike a fully developed script, whether alphabetic, syllabic or logographic, the Easter Island "writing" cannot be statistically assessed. Thus counts ranging from 120 to several thousands have been made for it. The range of variation of a sign is enormous, pointing not to a modicum of simplicity usually associated with a writing system which after all is presumably used for the purpose

of communication, but to at most, semasiographic, pictorial representations which cannot be called a "system" at all. The distribution of the signs and the lack of groups exclude the possibility of phoneticism and the limited number of different symbols eliminates the possibility of the Easter Island writing being a pure pictography.¹

2.31 The 200 or so signs which de Hevesy supposedly connected with the Indus signs were shown by Alfred Metraux, to be mostly inaccurate representations, with those resemblances remaining, being only of the oddest sort, not of the most characteristic signs. "Hevesy fails to explain how 2 scripts separated in time by 4000 years at least, can present minute and complicated resemblances in trifling details and at the same time be so completely different in all the essential elements." In his comparison of signs he chooses from the thousands of Easter Island signs small variations which appear once or twice, thus violating the principles of valid comparison of scripts as outlined above in 1.22. Further, here again chronology would indicate the impossibility of connection of two "scripts" over a period of at least 4000 years, preserved throughout the vicissitudes of wars and migrations, only to re-appear on a Spanish oar of the 19th century.

2.32 Another example of faulty comparison is that of Piccoli,² who compared 40 isolated signs on Etruscan utensils with simple geometrical signs found in the Indus script. Like all comparisons of this sort, resemblances are bound to appear, but the signs of the Indus Valley compared are just the widespread variety basic to any pictographic or even syllabic script, e.g., circles, plusses, a zigzag line, a square with a plus in it, etc. Nothing conclusive could conceivably be inferred from such scanty evidence of identity. Further, of the various unknowns, Etruscan, is an unknown language which is in a known writing, thus these marks are presumably potters' or owners' marks and not script at all. The validity or even usefulness of a proposed connection between such diverse civilizations, based on the doubtful "evidence" of identity of a few minor, widespread, uncharacteristic signs is to say the least, questionable.

¹That is, in Gelb's terminology a forerunner of writing, using pictures as signs, e.g., as among the American Indians or Eskimos, pp. 250-1.

²Giuseppe Piccoli, "A Comparison Between Signs of the Indus Script and Signs in the Corpus Inscriptionum Etruscarum," Indian Antiquary, Vol. LXII, p. 782.

2.4 There have been two attempts at what might be termed "interpretation" of the seals, rather than actual decipherment. Readings are made without the actual reduction of the signs to phonological elements, or words of a particular language.

2.41 The treatment by Meriggi of the Indus Valley script is one of these interpretative attempts. Meriggi classifies the script as ideo-phonographic, presumably by this term he means, logo-syllabic, and attempts to classify the signs of the inscriptions into determinatives, ideograms, sic etc. He asserts that the script most closely resembles the Hittite hieroglyphic but does not go so far as Hrozny in identifying the language of the inscriptions as Hittite, or even Indo-European. Thus he does not postulate any phonetic values for the signs, and his work must come under the category of preparatory work towards a decipherment, rather than any actual attempt at decipherment.

2.42 Like Meriggi, Petrie interprets rather than deciphers the seals of the Indus. His interpretation is based on proposed similarity of the writing of titles in Ancient Egypt to what Petrie contends are the administrative seals of the Indus Valley. A major objection to such an interpretation of the function of the seals, as merely seals of the officials of the government, is that there would have been as many officials as there were people in the Indus; often several officials per family, since several seals were often found in the same rooms and houses. Though it is certainly plausible that many of such seals were those of contemporary officials of the government, not all could have been, unless they were hereditary, possibly titular seals of authority, which passed from generation to generation.

The interpretation of these seals as administrative seals as do Meriggi and Petrie results in far too many administrative seals; if this defective theory is correct, then everyone in Mohenjo-daro for instance, was an official of the government.

2.5 Both Hunter and Langdon refrain from claiming that they have deciphered the script, and Hunter in particular, should be commended for his methodical, scientifically researched lists and tables of signs. They both, however, would assert the connection between the script of the Indus Valley and the much later Brāhmī script. At the point Hunter argued for this connection, part of his assumption was based on the lack of knowledge of the lower limits of the Proto-Indian civilization. These limits are far clearer now and would thus militate against such a proposed connection because of the great gap in time. In addition, as Mr. Shapiro has shown, Brāhmī is derivable (as a consonantal skeleton with vowel

markers) from North Semitic and is in structural character far removed from the Indus system, which is clearly logo-syllabic.

3.0 There have been, several serious and successful attempts at treatment of the Indus script. Those treatments which could be considered valid, however, are not full decipherments, or even interpretations of the scripts, but sound conclusions about various aspects of the system based on clear, demonstratable principles, often with analogies from the known writing systems of the world. Thus, the work by B. B. Lal of recent years on the direction of the script, as based on discernible overlaps of incised signs on sherds, and the analysis of the possible numeral signs by A. Ross should be noted for their adherence to sound procedure.

3.1 Ross, is the only author who makes inferences from the number of different signs and one of the handful who employs scientific, evaluative procedures rather than rhetoric in his examination and conclusions about the nature of the script, or as in his particular treatment the nature of the numeral system. His conclusion is that the chief numeration-system is decimal, i. e., that the base of the script itself is greater than 8 and not 11, and that 12 is not a change-point but is specifically indicated. Further, that the chief function of the numeral-signs appears to be a syllabic one, based on position of sign-groups and the proximity of certain numeral signs.

3.11 Lal's treatment indicates without doubt that the direction of writing of the Indus Valley inscriptions is from right to left, as has been heretofore suspected but never conclusively demonstrated. Overlap of incised signs, and the careful evaluation of a seal, inscribed along three edges, furnish the concrete proof for this assertion.

3.12 Hunter also, as indicated above, aside from his untenable derivation of Brahmi from the signs of the Indus Valley, has presented an excellent and full set of tables and lists, and evaluation of sign-groups, based on sound principles.

3.13 Smith, in his chapter on the "Mechanical Nature of the Early Indus Writing, in Mohenjo-daro and the Indus Civilization makes a careful and useful evaluation of the structure of the writing system and the stratification of the seals, though correctly, he draws no conclusions from this early, brief study.

4.0 In summary, we may conclude that for the most part, the attempts at the decipherment of the inscriptions on the seals of the Indus Valley civilization have been unsuccessful; except for a

handful of methodic evaluations as outlined in section 3, . . . the "decipherments" have been based on untenable hypotheses, lack of chronological accuracy, unsound reconstructions because of linguistic naïvete, and in general, preconceptions of the language of the inscriptions which were not based on any systematic or rigorous evaluation but on a priori and untenable assertions. The linguistic principles of decipherment as presented in the early part of this paper were ignored or violated.

4.1 The script of the Indus Valley seals as we have it, one must conclude, is virtually undecipherable. The task of deciphering an unknown language in an unknown script is, to say the least, formidable. The inscriptions which are left to us, are far too short, and too limited in nature to infer any readings; further, the names of the civilization are lost to us, and as names usually provide the keystone of a decipherment, the possibility of breaking the script is extremely unlikely. What can be inferred about the script, is limited to certain structural conclusions such as its nature as a type of logosyllabic writing, the direction of the writing, and possibly the numeral-signs as outlined by Ross. A thorough evaluation of sign-groups incorporating the newer material from Lothal and Kalibangan would be a useful study preparatory to a possible later decipherment. Its connection with another known script is possible, but, as it is a pictographic script, conclusive, rigorous demonstration would be necessary to prove that the resemblances were more than the predictable fortuitous ones, expected between any two pictographic scripts representing similar objects.

4.2 What is needed for a successful decipherment of the script are either bilingual texts, or longer inscriptions containing names, or place-names which could be identified and thus provide a key to syllabic values of certain of the signs.

4.21 For such aids, I believe, we must look to further excavations not in the Indus sites themselves, but in those of the earliest Sumerian, Ubaidian, and Elamite remains in Iran and places such as Susa, Jemdet-Nasr, and Kish where it is clear there is contact with the Indus Civilization at the early stages. Further, if, as Dr. Dales has suggested,¹ the lack of longer texts should be attributed to the use of perishable writing materials such as wax, we cannot hope to find longer inscriptions in India or Pakistan. Until such aids to decipherment are found,

¹George Dales, "South Asian's Earliest Writing," Expedition, Vol. I, No. 4, 1967, pp. 34-39.

the most useful research on the Indus seals would be a careful preparatory compilation, perhaps by computer analysis, of all the sign-groups of seals including those from the recent excavations at Kalibangan, Lothal and other sites in India.

Select Bibliography

- Barua, B. M. "Indus Script and Tantric Code," Indo-Iranica, Vol. I, 1946, pp. 15-21.
- Bühler, Georg. Indian Paleography. Calcutta, 1959.
On the Origin of the Indian Brahma Alphabet. Strassburg, 1898.
- Dales, George. "The Mythical Massacre at Mohenjo-daro," Expedition, Vol. VI, No. 3, 1964, pp. 37-43.
- Diringer, David. Writing. London, 1962.
- Gelb, I. J. A Study of Writing. Rev. ed. Chicago: University of Chicago Press, 1963.
- Ghosh, A. "Archeology in India," Expedition, Vol. VI, No. 3, 1964, pp. 12-17.
- Heras, Henry, S. J. Studies in Proto-Indo-Mediterranean Culture. Vol. I. Bombay, 1953.
- de Hevesy, M. G. "Sur Une Ecriture Océanienne Paraissant d'origine Neolithique," Bulletin de la Société Préhistorique Française, Vol. XXX, 1933, pp. 434-449.
- Hrožny, Bedřich. Ancient History of Western Asia, India, and Crete, n. d. Prague, in English.
"Inschriften und Kulture der Proto-Inder von Mohenjo-daro und Harappa, I and II, Archiv Orientalni, Vol. XII, XIII, 1941-42.
- Hunter, G. R. The Script of Harappa and Mohenjo-Daro and Its Connection with Other Scripts. London: Kegan Paul et al. and company, 1934.
- Jensen, Hans. Die Schrift in Vergangenheit und Gegenwart. 2nd ed. Berlin, 1958.
- Lal, B. B. "The Direction of Writing in the Harappa Script," Antiquity, Vol. XL, 1966, pp. 52-55.
- Marshall, Sir John. Mohenjo-Daro and the Indus Civilization. London: Arthur Probsthain, 1931, 3 volumes.
- Meriggi, Piero. "Zur Indus Schrift," Zeitschrift der Deutschen Morgenlandischen Gesellschaft. Leipzig, Vol. XII, 1934.
- Metraux, Alfred. "The Proto-Indian Script and the Easter Island Tablets (A Critical Study)." Anthropos, Vol. XXXIII, 1938, pp. 218-39.
- Petrie, Sir Flinders. "Mohenjo-daro," Ancient Egypt, 1932, pp. 33-40.

- Ross, Allan S. C. "The Numeral-Signs of the Mohenjo-daro Script," Memoirs of the Archeological Society of India, No. 52, Delhi, 1938.
- Vats, Madho Sarup. Excavations at Harappa. Vols. I, II. Calcutta: Government of India Press, 1940.
- Waddell, L. A. The Indo-Sumerian Seals Deciphered. London: Luzac and Company, 1925.
- Zide, Arlene R. K. The Indus Valley Script: A Selectively Annotated Bibliography. Chicago: University of Chicago unpublished ms., 1968.
- Zvelebil, Kamil. "Harappa and the Dravidians: An Old Mystery in a New Light," New Orient, Vol. IV, No. 3, 1965. Prague.

INDUS SCRIPT.

HEVEZY'S VERSION

EASTER ISLAND TABLETS.

1	U U	U U	T T Q Q
2	>	O O	O
3	W	J J	J
4	X X	K K	
5	TTT T	L L	L L L L
6		C C	Q Q
7	Y	J J	U U
8	O Y	O Y	O J J
9	X X	X X	Q Q Q Q
10	M M	M M	
11	K K	I I	Q Q
12	W W	W W	Q Q Q Q Q Q

Figure 2. Comparison Between the Symbols of the Easter Island Tablets and the Writing of the Indus Civilization, Man, January, 1938.

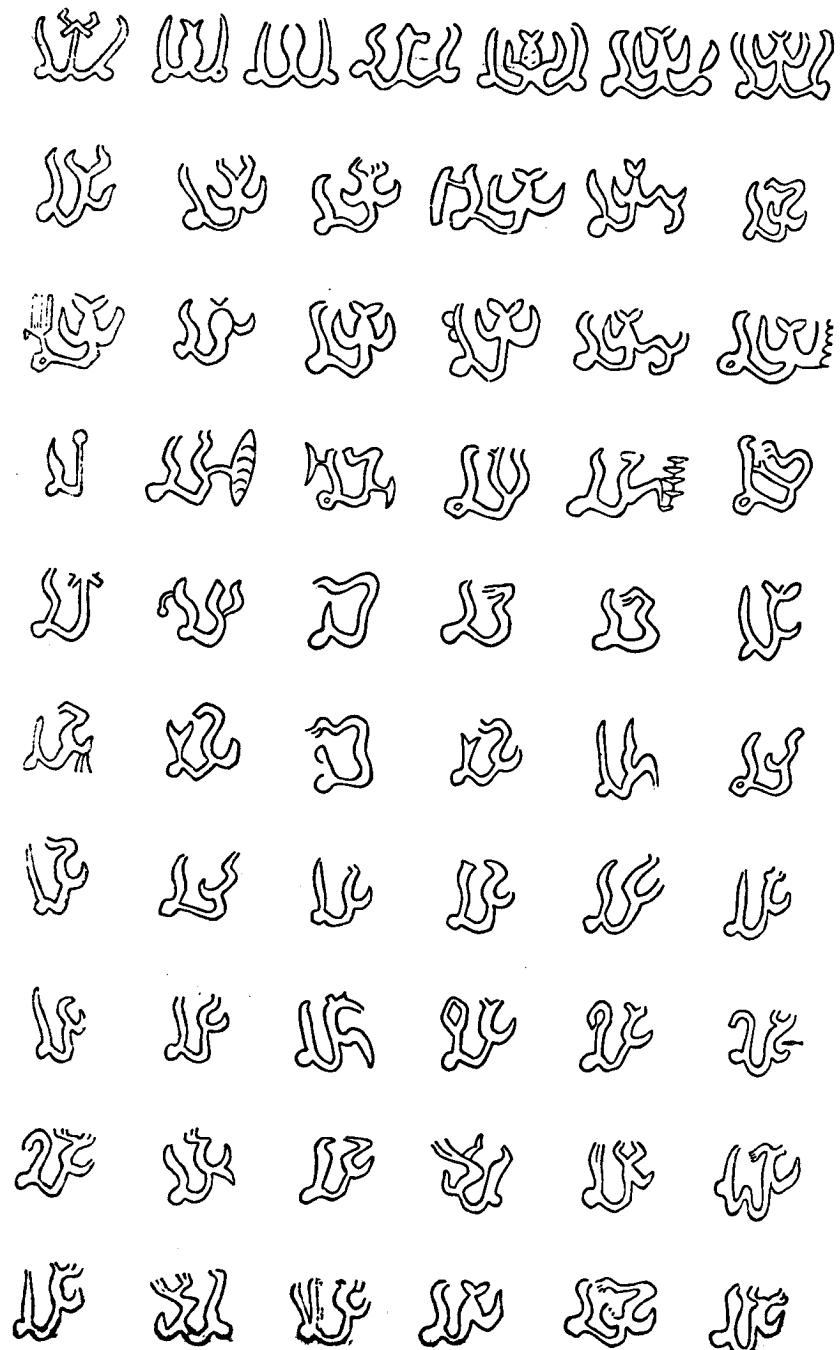


Figure 3. Range of Variation of a Symbol, Metraux,
in Anthropos, 33, 1938.

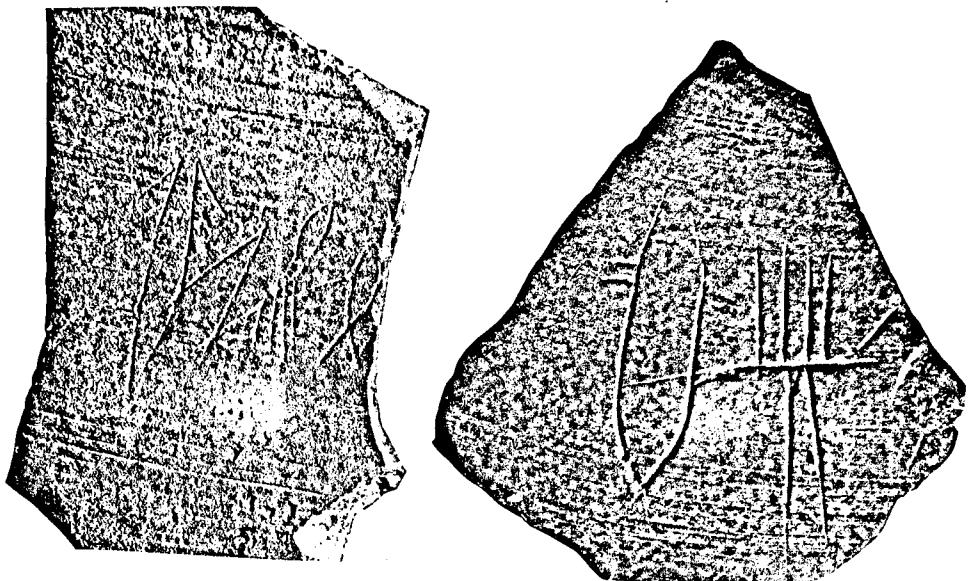


Figure 1. Inscribed Potsherds from Kalibangan (Rajasthan) in "The Direction of Writing in the Harappan Script" by B. B. Lal, Antiquity, Vol. XL, 1966.

GRAPHEMIC SYSTEM IN THE OL CEMED SCRIPT

Norman H. Zide
 (University of Chicago)

Ol Cemed (or Cemet) is the name of a script used to write the Santali language, a Munda language spoken chiefly in the states of West Bengal, Bihar and Orissa in India. This script was invented about fifty years ago by a Santal, Pandit Raghunath Murmu.

It is an alphabetic system with nothing of special interest in its alphabeticity principles; it is much like English (which undoubtedly served as a model for some of its features) and unlike the other local scripts which the deviser of OC must have been familiar with (Hindi [Devanagari], Bengali and Oriya) in having no syllabary properties and in writing its vowels linearly, i.e., after the appropriate consonants, not--as in Hindi, Bengali, and Oriya--either before, after, over, under or before-and-after for some one or more vowels.

Another feature worth noting is the morphophonemic--or systematic phonemic--representation reflected in the Ol Cemed graphemes for the 'checked' (preglottalised) consonants which the OC transcriptions of Santali but not the other transcriptions (i.e., Hindi, Bengali, Oriya, Roman) has.

Another notion worth mention exhibited only for a few of the letters is that of trying to make the letter shape represented by a kind of visible speech the position of the vocal organs in making the sound the letter represents. This is done unsystematically and only for three of the thirty-odd letters (g, ŋ and o).

Of more interest and indicative of more linguistic ingenuity on the part of the inventor is something I described elsewhere: the schema for arranging the letters in an array of columns and rows both of which can be characterised phonologically. The characterisations owe something to the analogous arrangements in Sanskrit, but there are interesting modifications in the principles governing these arrangements. Also discussed in this earlier paper are the more or less hypothetical or fictitious derivations of the present letter shapes as conventionalisation of pictures of objects also serving as the letter names.

What I will do here is to try and describe the system characterising the shapes of the graphemes or

letters of OC — of the majority if not of all of them.

Before doing this, let me say something about the cultural context of the invention of the script, and the multiscriptal character of the Indian written language scene. India has the greatest script diversity of any country in the world today. There are eleven different scripts recognised at some regional level as official in India: the Devanagari script used for Hindi, and also with additional diacritics for Sindhi (and its close relative the Balbodh script for Marathi), the Gujarati script for Gujarati, the Bengali script for Bengali (and its close relative the Assamese script for Assamese), the Oriya script for Oriya, the Gurmukhi script for Punjabi, the Telugu script for Telugu, the Tamil script for Tamil, the Malayalam script for Malayalam, and the Kannada script for Kannada. In addition to these--which are all derived from Brahmi and work on roughly similar principles--there are the Perso-Arabic script for Urdu, and with additional diacritics for Kashmiri, and the Roman script for English. An analogous situation in Europe would exist if English, Welsh, Dutch, Danish, Finnish, French, Polish, Italian, Spanish, Portuguese and Hungarian all had--and insisted on maintaining--separate scripts. In India this maintenance has been in the face of repeated efforts to make one script official for all the languages, the one script being either Roman (which was pushed more before independence) or Devanagari (which is more often proposed now). The reformers not only tried to persuade the regional script advocates of the general advantages to be gained by adopting **one script** for all the languages but presented transcriptions of all the regional languages in Roman (or Devanagari) to show how nicely and easily this could be done. The script diversity remains what it was--and, at a sub-official level, may even be increasing. The diversity stems from old regional traditions where each region usually had its own language and script, where these scripts were reasonably efficient writing systems for their languages. The language and script traditions were never overridden by any central authority, and have become even stronger recently with the introduction of the local regional languages--not English or Hindi unless Hindi is the local regional language--as the medium of public education at all lower levels, and most higher. Thus, in India any newly literate group of some size (at least 100,000) aspiring to 'high culture,' if it has a language of its own feels the need for a script of its own as well.

This means that India is a good place to look for new writing systems, for scripts in statu nascendi, and in fact at least four new scripts have been invented in this century. One areal exception to what I have said

about felt needs for new scripts relates to the NEFA region and adjacent parts of Assam where Roman is the accepted regional script of these--mostly Christian--'tribals.' Roman and English were introduced by missionaries, and the groups who have become literate in the last century or in this are mostly Christian, and have English as their 'culture language.' For them, an orthography for their own language (previously unwritten, usually) in Roman is all that is wanted. Which is to say that the local Hindu regional subcultures, and the Islamic groups that compete with these roughly in the same terms require special scripts that go along with their special languages, whereas the Roman (script)-based Christian 'tribals' do not.

One requirement for the scripts of the groups wanting and devising them is that these scripts look obviously different from any of the other local scripts. If at some less superficial level there are fundamental similarities, this is all right so long as there can be no confusion of their own script with anyone else's nor any mutual intelligibility (without proper training) among them. This dissimilarity requirement is, I think, characteristic only of neography in certain complex cultural contexts like the one I have just described. I assume this sort of consideration did not apply in the formation of writing systems.

This conspicuous distinctive orthography and conspicuous literacy for newly literate groups in India obviously introduces its own problems. Santali is written in five scripts, Hindi, Bengali, Oriya, Roman and OC. Any high school student in Orissa or Bengal who has to learn, say three of these is spending more of his time on scripts than he can well afford. One result of this is that OC is falling out of use where it was formerly more or less accepted. Other newly derived scripts have lapsed entirely, or, in some cases, have retained a ceremonial or other not primarily orthographic value. That is, the having a special script is itself enough; it serves an important prestige function. It can be shown off, and used in one way or another (if not actually written and read) on certain occasions (as is e.g., Saurashtri), but for everyday writing only the local script (or scripts) is current. An extreme aberration of prestige-motivated script production and dissemination can be seen in NEFA where several of the languages have recently been provided with (Roman-based) orthographies, and for each of the languages a number of books has been published, this in spite of the fact that for some of these languages there have been more books printed than there are literate readers for them.

What I claim, and will try to demonstrate in the

remainder of this paper has to do with the printed forms of the letters. This is not because these necessarily have special priority or fundamentality but because the only data available in valid quantity are printed materials. I think that what can be said about the printed graph shapes is important, whatever the historical priorities of cursive forms may be. However, I would not be surprised if the printed forms of OC were devised first, or concomitantly with the cursive forms of the letters.

I talk about graph features (or components). I had said earlier of these features that it was convenient for these purposes to speak of them as simultaneously generated in one typographic act. It is not clear that this is the case. If we want--in talking of the simplicity of certain OC letters--to say that they (also) can be written in two strokes or less (i.e., by lifting the pen from the paper once at most), this sort of claim is not allowable if the whole letter is produced instantaneously, with no starting at one place and moving the pen to another, etc. This means, of course, that I am not confining myself to printing procedure and printed forms, but engage in what I think to be necessary reductionism to handwritten forms and how these are made. The results of our feature analysis give us a partly fictitious stroke count and order of writing the strokes, partly fictitious because alternatives rejected on the grounds of accounting for the system simply are used (and can be justified later, but we don't want to enter into such justification here) in writing them, and also on the ground that the assumption that all letters are written in these imitation printed forms 'lifting the pen' for new strokes, rather than going over lines already written and then continuing them in new directions (which is, of course, how most cursive scripts including cursive OC are written) itself is fictitious. It does describe the 'ideal forms' which I claim to be the fundamental ones, the ones from which the rest can be most simply derived. These ideal forms can be considered to be carefully written or drawn version of the printed letters.

I claim that most of the OC letters are made up of four (or six) graph components, that the letters are more simply described by having these in fixed position, and having one or both of the transformations apply to them. The fundamental space occupied by an OC letter is oval-shaped: more less of the circumferent line is used as part of the letter, and lines in or across the oval serve to distinguish the letters from each other. Additional restrictions on how these lines can come together are given later. Further, the letters that are not describable in terms of four (or six) fixed position components and 180° rotation and 'mirror image' reversal trans-

formations tend to be those which resemble their 'objects' (e.g., the mushroom 'ut', the sickle 'ir').

I will classify the letters as a) simple b) half-simple and c) complex. The simple letters are those that can be derived from the set of components, mapping these onto the standard oval outline with or without application of one or more of the transformations specified. b) The half-simple are those that (i) take the transformations on a non-simple element or (ii) that use the standard components and transformations but use more than four components, or involve new modifications of these. I use as standard forms of the letters the forms (without the flourishes and serifs) of the booklet ol urum. w and h--the last two letters of the last row--are properly diacritics, and graphically distinguished as such. I comment on these later. Of the (remaining) twenty-eight letters, three u't, uc and (l)o are clearly complex; of the semi-simple (this being very much of a leftover category), of type (i) there are two letters derived from the non-simple 'sausage' element o'g, uy (the diacritic ow is also derived from this); of type (ii) (l)u, ep (which could also be considered complex), on, ir, (l)c and (l)a. Of these last two, the latter is clearly derived from the former in the one obvious borrowing in the letter shapes. In Hindi, Bengali and Oriya the 'short a' (roughly e in Hindi, o in Bengali and Oriya) plus a vertical stroke (in the full forms of these letters) gives the (full form of) 'long a'. In OC the parallel line in (l)a is not vertical, but is a curve parallel to the long vertical curve of (l)o. The remaining sixteen letters are simple.

The letters, the features and my derivations are given at the end of this paper.

All these 'simple' letters are 'vertically balanced,' that is they have either one long vertical line (Sets A-ii, iii, D, E and F), or, if seen as physical objects, 'balance,' or 'stand upright' (Sets A-i, C). The one exception is the oval (Set C), which like other exceptions noted elsewhere resembles its object (the round earth made elliptical by graphic convention) more closely than do most of the other letters. None of these letters is 'horizontally balanced.' No OC letters consist of lines not joined (e.g., simple parallel lines, or more complex figures) nor of lines that only intersect, i.e. touch somewhere other than at their endpoints (e.g., like English X). All the graph features (straight and curved lines) can be considered to join at their ends except those where the first stroke of their baseform is a vertical (i.e., Sets E and F). It is true that the letter p of Set F has a slight extension beyond the point of junction with the vertical and for the four members of Set B--in some printing--the lower trough extends slightly beyond its junction with the vertical. In the

former this is accounted for by the required 'initial onset' in writing a letter (see the cursive forms for this onset exaggerated). For Set B (where this extension occurs) one must state that the full A curve is written at the bottom, although by doing so one violates the general rule, and over-shoots a little the point of junction with the vertical.

I have stated that all these letters have a basic outline or plan (I use the word in the sense of floor plan not plan of behavior) and that each letter has as one of its features a line which coincides with part of this outline; by this the orientation of the letter to the outline can be set.

It can be seen that if the more or less oval plan or envelope were extrapolated or projected for all these letters it would not be an identical figure in all cases. Particularly for Set F (and its half-simple relatives, e.g. (1)u) would it be distorted. Nevertheless, for all of these simple letters, and even for complex ones such as uc the oval envelope is partly outlined by the letter, and its orientation suggests how the rest of it would fit inside the extrapolated envelope.

There are Indian scripts whose (consonants) characters do suggest that the letter shapes have a common outline, partly filled in in all the (well-formed) letters, but for e.g., Oriya (which has a semicircular arch as the top of its consonant letters), the outline filler (i.e., the arch) is invariant, and non-distinctive. What distinguishes say, b from c in Oriya is something occupying a small portion of the letter space in the lower half of the letter--which makes Oriya relatively hard to read. Similarly for Hindi, the wordlong top bar--interrupted for a few letters, e.g., dh and bh--is non-distinctive, and thus unnecessary. It is dispensed with in many styles of handwriting, and in the fairly similar Gujarati script. The sort of fixed plan I speak of is different from a fixed amount of space that any normal form of a letter must fit into and fill to some degree -- which is common. Certain scripts, Indian scripts such as Tamil among others, do still maintain differences in size (width) of letters. Gil'arevskii and Grivnin show a number of scripts (Burmese and its derivatives Khamti and Karen look likeliest) that may resemble OC in using a standard size outline figure (in the case of these scripts a circle) for almost all of their letters, differentiating one letter from others by differential filling in of the standard outline and by lines made inside the circle (mostly the latter, apparently) The Burmese script looks as if its circle filler plan (i.e., the circles or circumference pieces that fill or cover this plan) is more invariant--and thus less informative--than the OC oval fillers. (I say 'looks as' etc. because I have not access to expositions of how Burmese, Karen and Khamti work, and am speculating from the appearance of the letters as printed in Gil'arevskii and Grivnin.)

What can we presume about the history of the script, and these characteristics of it, and what can we say of its advantages (in and out of the light of this speculative history)? The one obvious borrowing from Hindi-Bengali-Oriya was noted before. Set F would seem to owe something to the English b-d p-q set, but also given the above desiderata (and the rejection of enclosed or partly enclosed 'horizontal space') there are not many possibilities of the system not used. If the non-vertical component D has a short form, d, one can look for a short c from C, although its form, not surprisingly does not keep to the oval outline but is, instead, symmetrical. Given the c, and already required d, such letters as those in Set F are not outside the basic possibilities of the system. Such hypothetical letters as I , J , Q , and A which do not introduce new components can be considered unsatisfactory because: (for the first) lines join at points other than those at their ends when the attachment of lines not at the endpoints is not made onto a vertical, for the second that it has horizontal enclosed space, for the third it has two verticals and also looks too much like the complex letter ep. Only for the last do none of these apply. I will not offer here various alternative ad hoc constraints to justify--or try to justify--a claim that it too is not well-formed, a claim we might like to make since this letter does not occur, and we have as yet no reasons to expect this non-occurrence. In fact, given a criterion such as maximising the regularity of the (projected) oval, would be preferable to the members of Set F. One could, of course, claim that no further distinctions--graph component combinations--are needed, but why is A not accepted before Set F. One reason for excluding A --more about this can be gotten from the notes to the chart at the end of this paper--would be that it uses a vertical which is not a complete stroke, but (in the expected standard from of writing) CV (one stroke), D (second stroke) a continuation of the previous component. It would also require a single stroke diagonal, not required elsewhere on OC. Alternatively, if it was written C (one stroke) VD (second stroke), the continued vertical condition would still hold, and we would require a single stroke C, and single stroke C is not require elsewhere (nor is any other single stroke curve (i.e., A) or short component (c,d) required elsewhere).

The advantages that can be claimed for the OC system of simple letters are: the letters have a standard simple outline, which is filled in distinctively in different letters and added to (within the enclosed space) by a small number of features in a small number of fixed positions. The number of these positions can be further reduced--probably to zero--by treating certain letters as transforms of others. This makes for easy reading

and given standard ways of making the letters quickly writable (and there seems to be nothing difficult about providing these) easy writing as well. All of these serve to distinguish one letter from the others by using the whole letter (i.e., making maximum use for differentiation of the whole letter space) not as, e.g. in Oriya, by at most half of it.

The features have little or no relation of phonology. It is interesting that the notion of representing mouth position for a few letters did occur to Raghunath Murmu (who was, as far as OC shows, not familiar with Sanskrit articulatory phonetics beyond the general classifications derived from it and exemplified in the Sanskrit letter arrangement schema) but these were not exploited in his system. Of letters that I would guess Santals regard as looking like their objects (these as drawn in some identifiable and acceptable fashion), there are not many; most of these, not surprisingly, are non-simple: the mushroom u'd, the thing pointed upward (mountain?) uc, the earth st and probably the traditional sign of meeting ep; and among the simple is the plough beam (which, in later printed forms, has been regularised over earlier more realistic representations of the lower A). A relationship between the drawn object and the letter form is maintained by the letter's name being also the name of the object, but I would assume that a letter name --say delta or daleth--once learned would no longer 'cause the letter to be seen as' say, a door even if it looked like one (and most of the OC letters resemble their 'objects' only via fanciful etymologies which finished serving their purpose once the script was completed and justified in the eyes of its inventor).

Appendix.

In this appendix I give a short bibliography of the less obvious works found to be useful in the writing of this paper, and some remarks more or less specifically footnoting statements in the paper.

As to bibliography, in addition to the standard books of Gelb, Diringer and Jensen, I have found interesting and useful Smalley's, Orthography Studies (Amsterdam, 1964)--including its bibliography, Ettemble's, The Orion Book of the Written Word (English translation, 1961), Gil'arevskii and Grivnin's, Opredelitel' Jazikov Mira (Moscow, 1961), and also M. Eden, "On the Formalising of Handwriting," in R. Jakobson (ed.), Structure of Language and its Mathematical Aspects (Providence, 1961).

On the modern Indian scripts, useful examples (in addition to those in Gil'arevskii and Grivnin) of most of these can be found in V. Kannaiyan, Scripts in and Around India (Madras, 1960), Parivardhit Devanagari, (in Hindi), New Delhi, 1967; and Report of the Official Language Commission, 1956 (New Delhi, 1957). How the scripts work can be learned from grammars of the written languages in question. H.M. Lambert has an introduction to some of the north Indian scripts which is very useful. I have seen at one time or another unpublished papers by A.K. Ramanujan (on Kannada), and H.A. Gleason (on several of the modern Indian scripts) which provided more and better analysis than any of the above. On the 'tribal scripts' Zide has a paper on Ol Cemed, and Pinnow one on the Ho script (see my article on Munda in Current Trends in Linguistics, V [The Hague, 1968] for bibliography). To my knowledge the Kurukh and Sora scripts are nowhere described in print. On the background for the invention of the OC script, see M. Orans, The Santal, A Tribe in Search of a Great Tradition (Detroit, 1965). The earlier paper on OC is Norman H. Zide, "The Santali Ol Cemet' Script," in Languages and Areas: Studies Presented to George V. Bobrinskoy (Chicago, 1967).

What seems to be an opposite desideratum (to script dissimilation) has been reported by R.A. Miller in his report on recent work on Tibeto-Burman in Current Trends in Linguistics V (The Hague, 1968). Miller mentions the work of Nishida on the Hsi-Hsia script (for the language of the same name) which deliberately assimilated its (different) script to the Chinese, because it wanted its own writing to look like as much like Chinese as possible, which assimilation made its present interpretation all the harder.

Raghunath Murmu was probably familiar with a well-known Hindi book on paleography, G.H. Ojha's, Bharatiya

Appendix - continued.

Prācīn Lipimālā (1st ed. 1894, 2nd ed. 1918, reprinted in 1959), and one might account for the developments from the pictured objects to the final-letter shapes in terms of 'evolution' (or 'decay'), Ojha's book suggesting the kinds of changes in letter shapes that might be considered likely to have occurred in such a script if it had had a long history. 'Conceivably,' intermediate forms were derived by Murmu in tracing or manufacturing this synthetic telescoped history, but if so we can only guess at what these may have been. I am convinced that there has been some 'built-in evolution' on the Brahmi-to-the-modern-Indian-scripts model in the making of this script, and that this justified to Murmu in part the final, conventionalised (and unrecognisable as their 'objects') shapes of the letters.

On the passing comment in the paper that ow and oh are properly diacritics and not letters, and this is evident in how they are written: both of these consist of letters, aw and ih respectively, and an additional (new) feature, a raised dot for ow and a final hook or loop for oh. The raised dot occurs elsewhere as a nasalisation diacritic. The other new component -- the hook 'aspiration diacritic' -- does not. It looks as if the whole letter (and not just the final component) functions as a diacritic, but Murmu did not treat consonantal aspiration as such perhaps because English writes aspirated consonants as two letters. Hindi, Bengali and Oriya have special characters for individual aspirated consonants, and do not represent aspiration of consonants as such by some particular letter or diacritic. A relationship somehow similar to that between aw and ow (i.e., /w/ and /v/ seems to be implied for ih (=i? ?) and oh (where ih indicates postpausal or postvocalic h (and postvocalic l ?), and oh indicates postconsonantal h, i.e. consonant aspiration).

The OC numbers -- and, of course OC like all the other Indian scripts has its own set of numbers -- cannot be handled (except for zero) simply with these features. Possibly, a nest system of separate number features could be set up. The numbers look like (Roman, Hindi, Bengali

Appendix - continued.

and Oriya) numbers, and one could not mistake a number for a letter easily. It also looks as if the numbers in general have borrowed much more (specifically for particular number representations, and generally) than the letters.

Chart 3.

ଡ	ଓ	ଶ	ଢ	ପ
ଅ	(ଓ'ଟ)	(ଓ'ମ')	(ଓ'ନ')	(ଓ'ଲ')
ଅ	(ଆୟ) e	(ଆଯ') ଏ'	(ଆୟ) e	(ଆଯ) ଏ
ଅ	(ଆର) ର	(ଆର') ର'	(ଆର) ର	(ଆର) ର
ଓ	(ଓ'ତ) t	(ଓ'କ') k'	(ଓ'ନ') n'	(ଓ'ଲ') l'
ଘ	ବ	ଙ	ଙ	ଘ
ଆ	(ଆକ) k	(ଆଜ') ଜ'	(ଆମ) m	(ଆବ) b
ଆ	(ଆକ) k	(ଆକ') କ'	(ଆମ) m	(ଆବ) b
ଆ	(ଆର) ର	(ଆର') ର'	(ଆମ) m	(ଆବ) b
ଆ	(Ak) k	(Ae') e'	(Am) m	(Ab) b
ଇ	ଈ	ଉ	ଉ	ଇ
ଇ	(ଇମ) m	(ଇହ') h'	(ଇନ) n	(ଇର) r
ଇ	(ଇମ) m	(ଇହ') h'	(ଇନ) n	(ଇର) r
ଇ	(ଇମ) m	(ଇହ') h'	(ଇନ) n	(ଇର) r
ଇ	(ଇମ) m	(ଇହ') h'	(ଇନ) n	(ଇର) r

The OC letters are on top and beneath them Hindi, then Bengali, then Oriya and lastly English, i.e. Roman.

Chart 3
(cont.)

	ବ	ଧ	ଫ୍ର	ମ	ର
ଶ	(ଶ୍ର.)ଚ	(ଶ୍ର.)ଦ୍	(ଶ୍ର.)ର୍	(ଶ୍ର.)ର୍	(ଶ୍ର.)ର୍
ଷେ	(ଷେହ.)ଛ	(ଷେହ.)ଦ୍	(ଷେହ.)ନ୍	(ଷେହ.)ର୍	
ରୁ	(ରୁର.)ର୍	(ରୁର.)ର୍	(ରୁର.)ର୍	(ରୁର.)ର୍	
୪	(୪୦)୯	(୪୦୫)୯୍	(୪୦୩)୩୍	(୪୦୩)୩୍	(୪୦୩)୩୍
୪୦	(୪୦)୯	(୪୦୫)୯୍	(୪୦୩)୩୍	(୪୦୩)୩୍	(୪୦୩)୩୍

	ଈ	ଇ	ଏ	ରୀ	ରୀ
ଏ	(ଏପ.)ପ୍	(ଏଡ.)ଡ୍	(ଏନ.)ନ୍	(ଏର.)ର୍	(ଏର.)ର୍
ୱେ	(ଏଗ.)ଗ୍	(ଏଡ.)ଡ୍	(ଏନ.)ନ୍	(ଏର.)ର୍	(ଏର.)ର୍
ୱେଇ	(ଏଗିଏପ.)ପ୍	(ଏଡିଏଡ.)ଡ୍	(ଏନିଏନ.)ନ୍	(ଏରିଏର.)ର୍	(ଏରିଏର.)ର୍
ୱେଏଇ	(ଏଗିଏପ.)ପ୍	(ଏଡିଏଡ.)ଡ୍	(ଏନିଏନ.)ନ୍	(ଏରିଏର.)ର୍	(ଏରିଏର.)ର୍

	ଠ	ଠୀ	ଙ୍ଗ	ଙ୍ଗୀ	ଙ୍ଗୀ
ପ୍ରୋ	(ପ୍ରୋଟ.)ଟ୍	(ପ୍ରୋକ୍ରିଟ.)କ୍ରୀ	(ପ୍ରୋଟିକ୍ଲାର୍)ଟ୍	(ପ୍ରୋଟିକ୍ଲାର୍)ର୍	(ପ୍ରୋଟିକ୍ଲାର୍)ର୍
୳	(ପ୍ରେଟ.)ଟ୍	(ପ୍ରେଟିକ୍ଲାର୍)କ୍ରୀ	(ପ୍ରେଟିକ୍ଲାର୍)ଟ୍	(ପ୍ରେଟିକ୍ଲାର୍)ର୍	(ପ୍ରେଟିକ୍ଲାର୍)ର୍
୳େ	(ପ୍ରେଟିକ୍ଲାର୍)ଟ୍	(ପ୍ରେଟିକ୍ଲାର୍)କ୍ରୀ	(ପ୍ରେଟିକ୍ଲାର୍)ଟ୍	(ପ୍ରେଟିକ୍ଲାର୍)ର୍	(ପ୍ରେଟିକ୍ଲାର୍)ର୍
୳େଇ	(ପ୍ରେଟିକ୍ଲାର୍)ଟ୍	(ପ୍ରେଟିକ୍ଲାର୍)କ୍ରୀ	(ପ୍ରେଟିକ୍ଲାର୍)ଟ୍	(ପ୍ରେଟିକ୍ଲାର୍)ର୍	(ପ୍ରେଟିକ୍ଲାର୍)ର୍
୳େଏଇ	(ପ୍ରେଟିକ୍ଲାର୍)ଟ୍	(ପ୍ରେଟିକ୍ଲାର୍)କ୍ରୀ	(ପ୍ରେଟିକ୍ଲାର୍)ଟ୍	(ପ୍ରେଟିକ୍ଲାର୍)ର୍	(ପ୍ରେଟିକ୍ଲାର୍)ର୍

I am grateful to R.D. Munda for tracing the Chart 3 forms. He points out that Hindi i is sometimes used wrongly for Bengali ି in the ol urum booklet.

CHART 1

O (ɔ)	T (t̪)	K (k̪)	N (ŋ)	L (l̪)
a (a)	A (ə)	K (k̪)	C (s̪)	M (m̪)
I (i)	S (s̪)	H (h̪)	N (ŋ̪)	R (r̪)

CHART 1 (continued)

ꝝ    

ꝑ    

				
<i>u [ju]</i>	<i>u c</i>	<i>u d</i>	<i>u n</i>	<i>u y</i>
<i>U (w)</i>	<i>C (c)</i>	<i>T (t)</i>	<i>R (n)</i>	<i>Y (y)</i>

				
<i>e [le]</i>	<i>e p</i>	<i>e d</i>	<i>e n</i>	<i>e r (rel)</i>
<i>E (e)</i>	<i>P (p)</i>	<i>D (d)</i>	<i>N (n)</i>	<i>R (r)</i>

				
<i>o (o)</i>	<i>o t</i>	<i>o b</i>	<i>o w</i>	<i>o h</i>
	<i>T (t)</i>	<i>B (b)</i>	<i>W (w)</i>	<i>H (h)</i>

SIMPLE LETTER DERIVATION CHART

		<u>Components</u>	<u>Transfor-</u> <u>mation(s)</u>	<u>Strokes</u>	<u>Let-</u> <u>ters</u>	<u>Letter</u> <u>Name</u>
Set A	(i)	A D A		(1)	Z	(l)e
	a)	A D V		(2: AD;V)	↗	(l)i
	(ii)	b) A D V	180°, REV	(2: V;DA)	↙	is
	a)	A D A V		(2: ADA;V)	▷	er
	(iii)	b) A D A V	180°, REV	(2: V;ADA)	▷	en
Set B		A C A C		(1)	O	ot
Set C	a)	C d d C		(1)	Ⓜ	un
	b)	C d d C	180°	(1)	Ⓝ	in
Set D	a)	C V C		(1)	⓪	o'b
	b)	C V C	180°	(1)	⓫	ed
Set E	a)	V d A		(2: V;dA)	↑	ot
	b)	V d A	180°	(2: V;dA)	↓	a'i
	c)	V d A	180°, REV	(2: Ad;V)	↶	am
Set F	a)	V d c d		(2: V;dcd)	p	ol
	b)	V d c d	180°	(2: V;dcd)	b	ak
	c)	V d c d	180°, REV	(2: dcd;V)	q	in

Components:

A: ↗ ↘ D: / \ C: () V: |||

(Short) d X c { }

Note that the forms in Chart 1 are photographic copies of handwritten originals. The letters given in the above chart are based on the printed forms given in the booklet ol urum (with serifs omitted.) The ol urum forms of the letters with their Hindi, Bengali, Oriya and Roman equivalents (as given in ol urum) are given in Chart 3. The Chart 1 versions are much less regular than the others.

Given the following constraints on baseform letters and on the strokes used in making them (as specified in such formulae as A D A in the chart) one does not have to specify a particular order or sequence of components, nor to specify which D or V or d (of two, three or four) is the one called for in the particular formula. One can give the constraints applying to transforms as well as to baseforms (baseforms all exist as actual (terminal output) letters) but this will not be done here. A few of these will be noted below in some comments on why the non-occurred transforms do not occur. There are only two transformations: 180° rotation, and mirror image reversal.

1) (LIMITS) Every simple letter has one component which coincides with the oval boundary outline, which outline limits the direction of a component so that it does not extend beyond that outline (except for certain minor deformations, notably of c in Set E).

2) (DIRECTION) The first non-vertical component is finally right-going and initially rising. By 'finally', I mean in its second (or final) half (i.e. as in (in C-(a)) if not for the whole component. Verticals are made with a downward stroke.

3) (CONTINUITY) A vertical never continues or is continued by a preceding or following component in the same stroke. A non-vertical always continues or is continued (except where the adjacent component is a vertical).

4) (JUNCTION) A new component continuing a line (i.e. in the same stroke) joins that line at its end. Where a line is continued in a new stroke (i.e. where the previous stroke was a vertical -- as in the baseforms of Sets E and F) it is (as required by the direction constraint on initial non-verticals) right-going. This would mean it must be D or d. D does not occur as a stroke-initial component, so that d is the only component that does. d to keep within limits joins the preceding vertical at about its midpoint.

Since the transformation possibilities in the simple letters of OC have been exploited fairly thoroughly, it would seem useful to find out which possible transforms do not occur, and, if possible, why. (I exclude from consideration here transforms which are identical with letters already in the system. Note that for all the transforms in the chart the reversal transformation

operates only on the output of a 180° rotation transformation. Set D-(b) is not interpreted as D-(a)/Rev (which would look the same) to maintain this regularity.)

The non-occurred transforms can be grouped into two groups: those that do not take terminal reversal transformations as applied directly to the baseforms, and those that do. The former group has three members: ADA/180 (\equiv ADA/Rev), *ADV/180, *ADAV/180, the latter four: *ADV/Rev, *ADAV/Rev, *VdA/Rev, and *Vdc/Rev. One can capture the regularity of the first set simply enough in a rule (e.g. 'letters with A D ... do not take 180° rotation transformations') but more explanation emerges if one actually writes these non-letters. They do not occur 'because' they violate our set of constraints; to preserve continuity within the set limits, etc. of *S one would have to violate the direction requirement. One does not start a letter writing from right to left in OC. Note that if *S occurred, specifying D with no further qualification would not be enough to distinguish it from Z which would also be ADA (unless we further specified A), etc. (This constraint on direction is violated in one letter; F-(c), which is otherwise also less well-formed than the rest of the simple letters. One could claim for the initial stroke of F-(c) that the stroke is finally right-going, which is not the case for the three non-occurred reversal transforms and for ADV/180 $^\circ$ and ADA/180 $^\circ$. The remaining two non-occurring transforms are in a separate class. Since their rotated transforms (which do occur) are by our criteria less well-formed than the rest, one would like to see how they are more well-formed than their non-occurred relatives. I can see no way at all in which their surface representations are better formed. All that can be brought in to account for them is the generalization -- thus additionally important -- that (terminal) reversal transformations made directly onto baseforms (NOT onto rotated baseforms) do not occur. If it is said that the non-occurrence of VdC/Rev and Vdc/Rev results from the accidents of 'object' occurrence -- i.e. that, accidentally, there were no 'objects' (things or actions portrayed in the drawings from which the letters presumably derive) that would have been conventionalised to *d and *n, while there were for q and Q, one must point out that the above-stated generalisation fits the results of these accidents (and others) very nicely.

