PROBLEMS OF MORPHEMIC ANALYSIS

CHARLES F. HOCKETT
CORNELL UNIVERSITY

I. INTRODUCTORY

1. This paper develops further the theory of morphemic analysis presented by Zellig S. Harris in 1942.[1] Morphemic analysis is the operation by which the analyst isolates minimum meaningful elements in the utterances of a language, and decides which occurrences of such elements shall be regarded as occurrences of 'the same' element.

This operation does not constitute all of grammatical analysis: when it is completed, there remains the task of describing the arrangements in which the minimum meaningful elements occur, and—where the same elements are observed to occur in more than one arrangement with a difference in meaning—the features other than morphemes ('tagmemes') that are involved. For this latter phase, I have proposed the term 'tactics'.[2]

The fact that *John hit Bill* and *Bill hit John* have different meanings,[3] or that *old men and women* is sometimes approximately the same as 'old men and old women', sometimes rather comparable to 'women and old men',[4] proves that features other than morphemes have to be recognized—unless, of course, we wish to redefine 'morpheme' to cover features of this type too. There is other evidence: a graduate student in a hurry to prepare himself for a French reading exam, or a scholar deciphering a dead language written in a non-phonetic or semi-phonetic orthography, may achieve good control of the tactics and semantics of the language, but remain in almost total ignorance of anything submorphemic. To do this he needs some mnemonically satisfactory device for keeping morphemes apart. The device probably consists of speech sounds;[4a] but these may be purely private. Thus a western sinologist may know Confucius backwards and yet stumble in passing the time of day with any speaker of a modern Chinese dialect.

Although, then, morphemics and tactics are both necessarily involved in grammar, we nevertheless have considerable range of choice in drawing the line between them.[5] Faced with a language of a certain degree of complexity, we may

[1] Zellig S. Harris, Morpheme Alternants in Linguistic Analysis, LANG. 18.169–80 (1942).

[2] Review of Eugene A. Nida, Morphology: the Descriptive Analysis of Words, LANG. 23.273–85 (1947).

[3] Leonard Bloomfield, Language §10.4 (New York, 1933).

[4] Rulon S. Wells, Immediate Constituents, LANG. 23.81–117, esp. 93 ff. (§§30 ff.) (1947).

It would be possible to say that this ambiguity of meaning of *old men and women* was grammatically irrelevant; but features of order of the type involved in *John hit Bill* versus *Bill hit John* cannot be ignored. This being so, Bloomfield's term 'tagmeme' for a feature of meaningful arrangement is useful.

[4a] This is a reasonable assumption because of man's million years or so of natural selection, in which ability in aural memory and oral mimicry has been a factor making for survival.

[5] Cf. Zellig S. Harris, From Morpheme to Utterance, LANG. 22.161–83, esp. 162–3 (§2) (1946).

prefer to describe it with simple morphemics and complicated tactics, or conversely, or somewhere in between. The language is not disturbed by our choice; its complexities remain whether itemized in one part or another of our description. But the resulting descriptions may vary a great deal in the clarity with which they depict the situation. Presumably we should try to obtain that distribution of data between morphemics and tactics which produces the greatest clarity. In this paper we assume, without steadfast conviction, that this end is achieved by the simplest possible tactics, whatever submorphemic complications may be necessitated.

2. The same assumption was apparently involved in Harris's formulation of 1942. Yet Harris realized that this cannot stand as the only assumption. We must have, also, a set of principles on the basis of which we identify, or refuse to identify, different stretches of speech as morphemically the same. The great value of Harris's paper lies in this: that although he does not add any individual method of morphemic identification to those currently used, he demonstrates how all the superficially diverse methods can be regarded as cases of one general procedure. This general procedure we outline herewith, with such minor modifications of terminology as will be useful to us:

STEP 1. The utterances of a language are examined.[6] Recurrent partials with constant meaning (*ran away* in *John ran away* and *Bill ran away*) are discovered; recurrent partials not composed of smaller ones (*-way*) are ALTERNANTS or MORPHS.[7] So are any partials not recurrent but left over when all recurrent ones are accounted for. The citable case most nearly approaching this is the *cran-* of *cranberry*, which does indeed recur, but always with *berry* following. By definition, a morph has the same phonemic shape in all its occurrences. Because we operate with whole utterances, morphs are not always composed of continuous uninterrupted stretches of phonemes,[8] but they are always composed of phonemes. Every utterance is composed entirely of morphs. The division of a stretch of speech between one morph and another, even if the two are simultaneous, overlapping, or staggered, we shall call a CUT.

STEP 2. Two or more morphs are grouped into a single MORPHEME if they '(a) have the same meaning,' (b) never occur in identical environments, and (c) have combined environments no greater than the environments of some single alternant in the language',[9] e.g. *-en* in *oxen*, /z/ in *cows*, and various others, all meaning 'noun plural', with combined environments, or RANGE, paralleling the range of zero with meaning 'noun singular'.[10]

[6] Obviously not all of them, but a sampling which we hope will be statistically valid. By working with successively larger samplings, and by predicting on the basis of each what else will occur, we approach, at least asymptotically, a complete description.

[7] A convenient term, because it (1) eliminates the lengthy expressions 'morpheme alternant' and 'morpheme unit,' and (2) suggests a valid analogy *(allo)phone : phoneme = morph : morpheme*.

[8] The possibilities are investigated by Harris, Discontinuous Morphemes, LANG. 21.121–7 (1945)—but the added complication of this is avoided in the examples of the present paper.

[9] Harris, Morpheme Alternants §7.1.

[10] The zero element with meaning 'noun singular' is one of Harris's parallels (Morpheme Alternants §2.2). Such a morpheme has very dubious status, having no alternant of other

STEP 3. The differences in the phonemic shape of alternants of morphemes are organized and stated; this constitutes MORPHOPHONEMICS. Morphophonemic statements may involve morphophonemes—that is, the symbols used for phonemes, plus supplementary ones, with special definitions as to phonemic value under varying circumstances—or they may not; often lists are more convenient, and sometimes they are unavoidable. But regardless of the methods used in describing them, such alternations are morphophonemic.

3. In several ways a rigorous adherence to Harris's system as here stated is troublesome.

(1) Sometimes we are confronted with a set of alternants with apparently identical meaning which are almost, but not quite, in complementary distribution. So with the two alternants meaning 'noun plural' in *hoofs* and *hooves*, or *laths* with /θ/ and with /ð/. These would forbid the tactically desirable conclusion that there is but one noun-plural morpheme in English.

(2) Sometimes a set of alternants with identical meaning and completely in complementary distribution have to be kept apart because we can find no single alternant whose range parallels that of the given group. In Latin, for instance, there is no case-number combination represented after all noun stems by the same suffix; therefore we may not legitimately (by Harris's criteria) speak of a single 'nominative-singular' morpheme, or a single morpheme for any other case-number meaning.

(3) Sometimes a stretch of speech may be cut at either of two places, so as to produce equally satisfactory—and equally unsatisfactory—morphs. In Menomini, when an element ending (otherwise) in a consonant precedes, in the same word, an element beginning (otherwise) in a consonant, an /e/ appears between them. Do we cut before or after this /e/? Either cut will do; either choice is arbitrary. Harris proposes that we cut in both places, and regard /e/ as an alternant of /-/ 'morpheme juncture'.[11] In this proposal he does not adhere to his own rules, for morpheme juncture has no meaning, and is not a morpheme; yet any Algonquianist will say that his solution is correct, and the problem is to readjust the rules so that the interpretation does not violate them.

(4) Since there is no way in which French /o/ 'to the (masc.)' can be cut, we must take it as a single morph. But the tactical survey suggests rather that it be taken as two successive morphemes, *à* 'to' plus *le* 'the (masc.)'. There is at present no way in which the latter conclusion can be reached without doing violence to our criteria.

(5) As we perform step 1 of Harris's procedure, only morphs of overt phonemic content turn up. It is suggested that the definition of morph be extended to cover also the following: minus-features, such as that which added to French *bonne* 'good (fem.)' produces *bon* 'good (masc.)'; replacement features, as in *man* : *men*; zero features, as in *sheep* (sg.) : *sheep* (pl.); and combinations of these, such as the difference between *child* and *children*. This is a difficult maneuver,

than zero shape (see fn. 37 and reference cited there). Harris lists also the parallel *-ful*; given the modification of criteria proposed in §13 of this paper, one could add also *'s* 'genitive'.

[11] Harris, Morpheme Alternants §4.2.

however desirable; Harris (within the scope of his paper) tells us neither under what conditions it is called for nor how to perform it.

The items just listed are not criticisms, but points on which improvement is clearly possible within Harris's general framework. The first two difficulties are easily handled; the remaining three are more serious, but respond to a single modification in plan of attack.

Many of the problems of morphemic identification met with in dealing with any language are trivial. Before turning to the full-scale discussion of the five difficulties listed above (in Parts III and IV below), we attempt to show how the more trivial problems can be solved quickly and easily, in a fashion that sheds light on the more intricate questions to which one must eventually turn.[12]

II. Preliminary Normalization

4. Let us assume that we have before us a display of a large number of utterances of a language, in a phonemic notation. As we begin the search for recurrent partials, we may discover that a phonemic notation other than the one we have used—for there are always several mutually convertible possibilities—would simplify the task.

In Yawelmani,[13] for example, the point of syllable division is phonemic. One way to write Yawelmani is to use a hyphen for syllable juncture; then the phonemic content of syllables can be indicated with a relatively small number of vowel and consonant letters. If our display of utterances is in this notation, we find such obviously related forms as /g?ads?/ 'obsidian' : /g?ads?-ni/ 'obsidian (dative)' : /g?a-ds?a/ 'obsidian (accusative)'. The second form contains a stretch identical with the first, plus /-ni/ 'dative'; the third form contains, before /a/ 'accusative', a stretch identical with the first part of the other two save for an inserted hyphen. If we are to identify the non-identical stretches /g?ads?/ and /g?a-ds?/ as being morphemically the same (whether one morpheme or more is another matter), and make similar identifications in other cases where the presence or absence of a hyphen is the phonemically differentiating factor, then we must handle this evanescent hyphen in our morphophonemic statements.

Yawelmani can also be written phonemically without the hyphen. If we want to write it so, we must use unit symbols for certain consonants which otherwise might be interpreted either as belonging wholly to a single syllable or as being divided between two: for example, /Vg?V/ would be an ambiguous notation for both /V-g?V/ and /Vg-?V/; but if we replace /g?/ in a single syllable

[12] We propose to say both 'the morph x occurs in such-and-such an utterance', and 'the morpheme x occurs in such-and-such an utterance'. By our definition, a morpheme is a class of morphs, so that the latter type of expression, without further qualification, is logically invalid. We render it valid by stating that an expression of the form 'the morpheme x' shall be taken in some cases as a class-name, in other cases as a variable indicating the appropriate though unspecified member of the class, depending on what the context requires. No ambiguity results; this is customary usage in linguistics; but it is a point on which more care is needed than is usual.

[13] Stanley S. Newman, The Yokuts Language of California (New York, 1944). The phonemicity of the point of syllable division is my conclusion from the evidence he gives.

by /k'/, the ambiguity is removed.[13a] With all such changes as are necessary, we reach a notation which does not write the point of syllable division with a separate symbol, but which nevertheless indicates it unambiguously: when a single consonant symbol stands between two vowel symbols, a point of syllable division falls before the consonant symbol; when two consonant symbols stand between two vowel symbols, a point of syllable division falls between them.

In this notation, the forms given above appear as /k'ac'/ : /k'ac'ni/ : /k'ac'a/. The partial /k'ac'/ appears to be identical in all three forms. Phonemically, of course, it is not; but the only phonemic difference has been relegated to a status of notational predictability, and can be ignored in our further manipulations. There are so many intricate problems in Yawelmani morphemics that any advantage of this kind that we can obtain is greatly to be desired.[14]

5. A notation is phonemic if it indicates, in every position, only those phonemic contrasts which occur in that position, but indicates all of them. Once one has found the morphemically most desirable phonemic notation, one can often handle certain additional simple morphemic problems by modifying it in such a way that, in addition to indicating unambiguously all the phonemic contrasts occurring in a position, it also indicates in certain positions contrasts which are NOT there phonemic.

If our display is of Navaho utterances,[15] we notice, sooner or later, that vowel symbols do not occur before pause (P). The sequence /VhP/ (with V for any vowel symbol) does occur. When we examine the display for recurrent partials, we find certain stretches with constant meaning that end in /Vh/ both before P and elsewhere, others that end in /Vh/ before P but without the /h/ elsewhere. Thus /bìtàhP/ 'among them' : /bìtàh níhP/ 'among them, he says', but /dò·dàhP/ 'not' : /dò·dà níhP/ 'he says no'. The morphemic identification of /dò·dah/ (P) and /dò·dà/ (no P) is elementary, as are other such cases. So we modify our notation throughout the display, by erasing certain pre-pause *h*'s—namely, those at the end of stretches that occur medially with the same meaning but without the /h/. In our notes we enter the memorandum: both *V* and *Vh* before pause represent phonemic /Vh/. Thereafter, save when reading off our transcription with Navaho speech sounds, we ignore the memorandum; the new /dò·dà/ now has the same shape, to the eye, before P and elsewhere.[16]

[13a] In this notation the letter k and the apostrophe ' are meant to constitute one symbol together. Similarly c and ' below.

[14] The second notation is that used by Newman. It may be wondered why anyone would be led to investigate the potentialities of our first notation, the one that we decided to reject. But in Southern Athabascan (see citations in fn. 15) an entirely similar problem arises, and Hoijer chooses an orthography comparable to our first Yokuts orthography, not to our second. The complexity of morphophonemic statement which results is considerable, and could be rendered measurably easier if a phonemic notation were used in which syllable division is marked indirectly instead of overtly.

[15] Harry Hoijer, Navaho Phonology, University of New Mexico Publications in Anthropology I (Albuquerque, 1945); a similar phenomenon in Chiricahua Apache: Harry Hoijer, Chiricahua Apache, Linguistic Structures of Native America 55–84 (New York, 1946).

[16] Hoijer's working notation incorporates this normalization, though he calls the contrast between 'constant' pre-pause *h* and the evanescent type phonemic instead of morpho-

Or suppose that we are dealing with Latin. We find pairs like *ars* : *artis*, *noks* : *noktis*, *urps* : *urbis*, *re·ks* : *re·gis*, *niks* : *niwis*. The semantic and morphemic difference between *ars* and *artis* recurs with the other pairs, but the difference in phonemic shape between the members of a pair is not so constant from one pair to another. Whatever may be our ultimate morphemic conclusions (e.g. that *artis* is *ars* plus something, or that *ars* and *artis* are both *art* plus something), they will be more easily reached if we can make the difference in shape to the eye parallel the morphemic difference.

This can be done. From phonemics we know that the sequences *rts*, *kts*, *rbs*, *gs*, *gws* do not occur before word juncture,[17] and that *gw* does not occur intervocalically. We may therefore rewrite the forms with precisely these non-occurrent sequences: *arts* : *artis*, *nokts* : *noktis*, *urbs* : *urbis*, *re·gs* : *re·gis*, *nigws* : *nigwis*. We note that in the modified orthography, *rts* and *rs* (before word juncture) are both representations of phonemic /rs/, and so on.[18] In the new notation, the second form of each pair differs from the first only in the presence of an *i* before the final consonant.

6. Sometimes it helps to perform this type of normalizing operation more than once.

In Potawatomi, a first normalization introduces, at certain points within utterances, a mark (say a space) indicating POTENTIAL PAUSE. Then we examine the stretches between successive points of potential pause to see which ones recur in various positions (relative to actual pause, or to adjacent stretches of varying structure) with the same meaning and the same or almost the same phonemic shape. Neither preceded nor followed by (actual) pause, we find *kak* 'porcupine', *k·we* 'woman', *muk·* 'beaver', *k·uk·* 'bucket'. Preceded but not followed by pause, we find rather *kak*, *kwe*, *muk·*, *kuk·*. Followed but not preceded by pause, the forms with the same meanings are *kak*, *k·we*, *muk*, *k·uk*. Both preceded and followed by pauses, the forms are *kak*, *kwe*, *muk*, *kuk*.

Now observation shows that the phonemes /p·/, /t·/, /č·/, /k·/ occur neither directly after nor directly before pause. Therefore we normalize all such forms as those itemized above, in all positions in which they occur, to *kak*, *k·we*, *muk·*, *k·uk·*, with the necessary memorandum that when either preceded or followed by pause, both *k* and *k·*, both *p* and *p·*, etc., represent phonemic /k/, /p/, etc.

This same normalization also accounts for all other alternations turning on the presence or absence of pause. When we have retranscribed our entire display of forms, we look further, and discover such pairs as *nkutšuwe* 'he wins a race' : *nnuktušwe* 'I win a race', *kwtumočke* 'he's fishing' : *nkwutmočke* 'I'm fishing',

phonemic. A similar normalization leads to the writing, within word borders, of phonemes both of the *s*-series and of the *š*-series, although, at least in rapid speech, only those of a single series occur within the stretch bounded by word junctures; see Zellig S. Harris, Navaho Phonology and Hoijer's Analysis, IJAL 11.239–48 (1945).

[17] Though it is not clear what word juncture is in Latin: it may be a non-phonemic matter introduced by a previous notational normalization.

[18] Full phonemic information is still given, since such a graph as *rts* before word juncture stands ALWAYS for phonemic /rs/, never anything else. In the new notation we have multiple writings for certain phonemic sequences, but only one phonemic sequence for each writing.

pmos·e 'he's walking' : *npums·e* 'I'm walking', *msun?ukun* 'paper' : *nmusnu?kun* 'my paper'. If we want to, we can begin at this level to cut our forms up into smaller recurrent partials. Clearly *n* means 'I, my'; the remainder of the first form, *nuktušwe*, presumably means 'win a race', but it is not the same morph as *nkutšuwe* '(he) win(s) a race', because of the difference in phonemic shape. For that matter, each of these may be more than one morph. Whatever comes of this, our morphophonemic statements are going to have to be complicated at every stage by an alternation of the positions in which vowels appear.

So instead of continuing our comparison and cutting, we can try first to make a further notational normalization that will take care of the alternating vowels, or many of them. We do this by rewriting the original forms, writing a vowel in BOTH forms of a pair wherever it appears in EITHER form, and in a few other places for good measure: *nUkUtUšUwe* : *nUnUkUtUšUwe*, *kwUtUmočUke* : *nUkwUtUmočUke*, *pUmOs·e* : *nUpUmOs·e*, *mUsUnU?UkUn* : *nUmUsUnU?UkUn*. Our memorandum this time states how these graphs are to be interpreted, not directly into a phonemic notation, but into the supraphonemic notation achieved by the previously applied normalizations: the first and each alternate one of a series of capitalized vowel symbols counts as zero, unless it precedes a final consonant.[19]

The second form of each pair now consists, to the eye, of *nU* (presumably 'I, my') followed by something that is identical with the first form of the pair. All the remaining problems of morphemic isolation and identification are rendered simpler.

7. There is no real drawback to counter the advantages of this kind of preliminary notational normalization, but there is a caution which must be observed. Our notational changes make morphs of differing phonemic shape look alike—indeed, that is why we make them. But the ultimate problem of the grouping of such morphs into morphemes is one which must be solved in a manner consistent with our handling of less patent cases—that is, on the basis of Harris's criteria (Step 2, §2) or of some other set. Performed as we have here suggested, notational regularizing is not apt to obscure more desirable morphemic identifications; but in extensive work with any specific language, one needs to check back over such preliminary operations from time to time to make sure.

III. Revision of the Grouping-Requirements

8. In step 2 (§2) are stated Harris's three GROUPING-REQUIREMENTS—the three conditions which must be met by two or more morphs if they are to be regarded as belonging to the same morpheme.. Some of the troubles itemized in §3 result from the particular way in which these grouping-requirements are formulated. The first of them, involving meaning,[20] is obviously the most difficult

[19] We are forced to use capitals or some other device for evanescent vowels, because other vowels, phonemically the same, are not evanescent. This fact marks these alternations as NON-AUTOMATIC. Where no extra symbols are needed—where the symbols already used phonemically are merely extended to positions in which they do not phonemically occur—the alternations are AUTOMATIC.

[20] In a manuscript not yet published, Harris demonstrates how, at least in theory, this criterion can be eliminated, thus appealing to semantic considerations at only one step of

to handle. We attempt no revision of it here, and any choices dependent on it under the Harris procedure will remain in the present scheme. The second and third requirements are purely distributional, and more easily subject to analysis and modification.

Preliminary to our proposal for a modification of the second grouping-requirement we define NON-CONTRASTIVE DISTRIBUTION. Two elements of the same kind (i.e. both allophones, both morphs, or the like) are in non-contrastive distribution if either (1) they are in complementary distribution, or (2) they are in partial complementation, and in those environments in which both occur, they are in free alternation. By free alternation is meant (a) that one cannot predict, save perhaps statistically, which form will occur in a particular instance, and (b) that the occurrence of one, rather than of the other, does not produce an utterance different in meaning.

In phonemic analysis, non-contrastive distribution is often used as a criterion permitting the grouping of two or more allophones into a single phoneme. Thus the unaspirated [t] of *stick* and the aspirated [t'] of *tick* are both found at utterance-final and in certain other positions: *He's in the skit* may end with [t] or with [t']. But this utterance is 'the same' utterance whether the aspiration is present or not; similarly with any other pair differentiated only in this respect. The occurrence of both allophones in certain environments does not deter us from classing both in the same phoneme, /t/.

We propose, then, to revise the second grouping requirement from 'never occur in identical environments'—which is another way of saying 'are in complementary distribution'—to read 'are in non-contrastive distribution'.[21]

The examples which follow (§§9–11) will demonstrate both the way in which this change increases the efficiency of our analysis, and also a danger inherent in it.

9. In Modern West Armenian[22] a number of morphs occur with meaning 'genitive singular': /o/ and /0/ with one stem of preceding noun, /van/, /u/, /an/, and /i/ with other stems. The environments in which these occur can be differentiated in terms of the nouns which immediately precede, and such nouns fall into a series of classes (of purely morphophonemic importance) by virtue of the morph or morphs of this meaning which follow them. One occurs only with /o/ and the oblique stem: /asdəvaʒ/ 'God' : /asduʒ.o/ 'of God'. Some occur only with /0/ and the oblique stem: /kuyr/ 'sister' : /kəroč/ 'of sister,

---

the whole process of descriptive analysis: the step at which one must decide whether two UTTERANCES, as historic events, are 'the same' or not (Bloomfield's fundamental assumption of linguistics, Language §5.3, §9.5). The first grouping-criterion (same meaning) thus becomes a practical shortcut; as such it is used here.

[21] In Yokuts Structure and Newman's Grammar, IJAL 10.196–211 (1944), Harris makes explicit use (§6) of the second grouping-requirement as modified, but without theoretical discussion.

[22] Information and forms were kindly supplied by Gordon H. Fairbanks. It does not matter for the present discussion whether the stem differences are submorphemic within the stem or are a part of the suffix. Some other complications, which do not alter the picture materially, are omitted here. In the cited genitive forms, a dot separates stem from ending.

sister's'. Some occur only with /i/ and the singular stem; this is the most common pattern: /atoř/ 'chair' : /atoř.i/. Most of those which occur with /van/ and the singular stem, for example /irigun/ 'evening' : /irig.van/, are also observed to occur with /i/ : /irigun.i/. The same is true for most of those which occur with /u/ or with /an/: /meg/ 'one' : /meg.u/ or /meg.i/; /axčig/ 'girl' : /axčəg.an/ or /axčig.i/.

The morphs meaning 'genitive singular' are thus not in complementary distribution, and by the original form of the second grouping-requirement could not be combined into a single morpheme. However, there is no observable difference in meaning between, say, /axčəgan/ and /axčigi/; nor is a speaker's choice of one or the other of these on any particular occasion predictable, save perhaps statistically. Therefore, within the limits of semantic judgment available to us at present, the various morphs in question are indeed in non-contrastive distribution, and by the modified second grouping-requirement—providing the third grouping-requirement is also met—are classifiable as a single morpheme.

10. In Peiping Chinese there are, as the elementary texts usually put it, two words for 'two': *èr* and *lyǎng*. The latter graph, by a preliminary normalization that need not concern us here, subsumes two distinct forms, *lyǎng* and *lyáng*. Unit numerals (those for 'one' through 'nine') occur in Chinese in the following positions: before a measure; after a group numeral ('ten' etc.); before a group numeral; after an ordinal demonstrative (*dì* 'the ...th'); preceded and followed by unit numerals (in counting 'one, two, three, ...'). *Sān* 'three' appears in all these positions with a single phonemic shape. *Èr* occurs in the second, third, fourth, and fifth positions, *lyǎng* in the first and third. In the third position, before a group numeral, only *èr* occurs before *shí* 'ten', but either may occur before *bǎi* 'hundred', *chyān* 'thousand', and *wàn* 'ten thousand' and its multiples. The choice here is free. *Èr* also occurs in position one, to the exclusion of *lyǎng*, before the measure *lyǎng* 'tael, ounce'; and before a few other measures either *èr* or *lyǎng* may be found, with no difference in meaning.

This statement of distribution is not quite exhaustive, but a completely exhaustive one reveals the same facts. Clearly, *èr* and *lyǎng* cannot be regarded as a single morpheme under the old form of the second grouping-requirement, but can be under the new form.

The case of *èr* and *lyǎng* is unique in Chinese; most other sets of morphs differing in phonemic shape but classed nevertheless as belonging to single morphemes have at least some phonemic features in common. If this factor is to deter us in grouping *èr* and *lyǎng* together, then some additional grouping-requirement, not mentioned by Harris, and apparently quite difficult to formulate in a strict fashion, must be involved. We mention this possibility because we feel the Chinese example 'instinctively' to be somewhat different in nature from the Armenian.

11. As indicated in §3 (1), Harris's example of English noun-plural morphs will not hold by a strict application of his criteria, because of such pairs as *hoofs* : *hooves*. One of the more interesting of such pairs is *brothers* : *brethren*. Cases like *hoofs* : *hooves* present no difficulty under the modified second grouping-requirement because the morphs involved (*hoof* and *hoove-*, /s/ and /z/) are in

non-contrastive distribution: *hoofs* and *hooves* do not differ in meaning. This is not true of *brothers* and *brethren*.

There are several possible ways of handling the problem. One way, which fits both versions of the second grouping-requirement but seems not too pleasing, is to group the /z/ of *brothers*, along with most other morphs meaning 'noun plural', into a single morpheme, but to exclude from this morpheme the morph found in *brethren*, both because it is not in free alternation with /z/ in this environment and because it has a different meaning: 'plural, with semantic specialization, producing a form of address for fellow lodge- or church-members of the male sex'.

Another solution is to postulate two distinct, though homophonous, morphemes *brother*: *brother*$_1$ 'male child of same parents', and *brother*$_2$ 'fellow lodge- or church-member of the male sex'. The plural of *brother*$_1$ is *brothers*; that of *brother*$_2$ is either *brothers* or *brethren*, in free alternation. The morphs meaning 'plural of noun' in these two cases, together with others of the same meaning, are now in non-contrastive distribution, and can be grouped into a single morpheme.

This breakdown of *brother* into two homophonous morphs, in order to achieve a greater differentiation of the environments in which various morphs meaning 'noun plural' occur, may seem artificial; but if one starts with the plural forms and then works to the singular, it seems less so. For *brethren* occurs in larger environments in which *brothers* also occurs, whereas *brothers* occurs in some larger environments in which *brethren* does not occur, e.g. *I want you to meet my* ( ), *John and Bill.* If one groups the cases in which both may occur and contrasts the non-linguistic environment of these cases with that of the cases in which only *brothers* occurs, the semantic difference is fairly clear. Extension by analogy to the singular forms then seems justified. The source of difficulty here, as often, lies of course in the complexity of manipulating any type of semantic criterion.

12. The proposed revision of the second grouping-requirement leads rather clearly to simpler tactics; but it raises a problem for which I have no answer.

There is a generally accepted working assumption in descriptive analysis to the effect that while there may be homophonous morphemes, there are no exactly synonymous ones.[23] No matter how subtle the difference in meaning may be between, say, *twenty* and *score*,[24] the difference in phonemic shape implies non-identity morphemically. Now the revised version of the second grouping-principle implies that we WILL violate this working assumption when the evidence leads us to believe that the violation is desirable. But I can make no statement as to the formal conditions under which the principle should be suspended. In every case of not quite complete complementation, we have to examine the positions in which more than one morph of the set appears, and decide whether in these positions they are in free alternation or not. In every case, this decision seems to turn on semantic considerations. If this is true, then for a long time to come such decisions are going to be partly a matter of individual taste. This

[23] Bloomfield, Language §9.5.

[24] Harris, Morpheme Alternants §2.2.

need not deter us; for in any such case we need only suspend judgment, state both or all the alternative analyses, and indicate that our choice of one for further analytical purposes is only tentative.

13. The modification of Harris's third grouping-requirement that we propose is somewhat simpler; cf. §3 (2). Instead of requiring that the morphs to be grouped 'have combined environments no greater than the environment of some single alternant in the language', we require that they have a total range which is not unique. The range of a morpheme is the class of all environments in which the member morphs of that morpheme occur. Our revised requirement still stipulates that a morpheme obtained by grouping several morphs together shall have a range identical with (or paralleling) that of some other morpheme, but no longer requires that the second morpheme (the test morpheme) shall consist of a single morph.

The tactical advantage to be gained by either form of the requirement is that we thereby avoid the need to list separately the range of individual morphemes; we prefer to handle them in terms of classes having identical or closely parallel ranges. With either form of the requirement, there may remain morphemes containing only a single morph, which have unique ranges; but this we cannot handle in the present connection.[25] For those morphs which perhaps can be grouped into complex morphemes, the tactical advantage is worthwhile.

The revised form of the requirement enables us to gain this tactical advantage in cases where it is impossible under the older form. Latin case-endings are a clear example. Since no single case-number category is represented after all noun stems by the same morph, it is impossible under the old form of the requirement to group all the morphs for any single case-number combination into a single morpheme. Under the new requirement, we may do so for one case-number combination, providing we also do so for at least one other case-number combination; the natural conclusion is to do so for every such combination. The set of eight or ten case-number morphemes can now be handled tactically as a class: they occur after noun stems; and a noun stem occurs before a case-number morpheme.

IV. Morph and Morpheme

14. We now attempt to remove the source of the remaining difficulties mentioned in §3 (3–5). This we do by a single rather fundamental alteration of the relationship between morph and morpheme.

Both before and after this alteration, an utterance consists wholly of morphs: every bit of phonemic material in an utterance is part of one morph or another. Before the alteration, every morph belongs to one and only one morpheme, so that there are as many morphemes in an utterance as there are morphs. After the alteration, the number of morphs in an utterance and the number of morphemes therein may not be identical: some of the morphs, and hence some bits of phonemic material, of some utterances, are morphemically irrelevant.

[25] In his unpublished material Harris shows how this can be handled. His example is English /tuw/ (*to, two, too*), which in the absence of semantic criteria first appears as a single morph.

How this change is brought about, and with what utility, will be demonstrated presently. In making it, we must conform to a principle which Harris does not state but which he adheres to rigorously: the principle of TOTAL ACCOUNTABILITY. Every morph, and every bit of phonemic material, must be determined by (i.e. predictable from) the morphemes and the tagmemes (if any) of which the utterance is composed.[26]

Two morphemic analyses of an utterance are TACTICALLY EQUIVALENT if they give the structure of the utterance in terms of the same morphemes and tagmemes—whatever the differences in the handling of submorphemic matters. For example, according to one analysis, Fox[27] *poonimeewa* 'he stops talking to him' consists of the morphs *pooni* 'cease', *m* 'act by speech on an animate object', and certain succeeding elements which do not concern us. A different analysis breaks the form into *poon* and *im*, with the same meanings. These two analyses are tactically equivalent. By the first one, the morpheme 'cease' has form *poon* before morphs beginning with a vowel, form *pooni* before those beginning with a consonant, and the morpheme 'act by speech on an animate object' has everywhere form *m*. By the second analysis, 'cease' has everywhere form *poon*, and the second morpheme has the two forms *m* and *im*, depending on what precedes. In either case, the sequence of morphemes involved can be indicated as {poon} + {m}; it is only below the tactical level that there is any difference.

If, on the other hand, we divide the given form into *poon* 'cease', *i*, and *m* 'act by speech etc.', and consider each of these a morpheme as well as a morph, the analysis will not be tactically equivalent to the first two. For in this case the sequence of morphemes involved must be indicated as {poon} + {i} + {m} —there are more morphemes in the word than by the first two analyses. It is easy to see also why this analysis is tactically inferior to the first two: the statement as to the occurrence of the morpheme *i*—to which no meaning can be assigned—will have to operate in terms of SUBMORPHEMIC (phonemic) properties of environments, whereas on the tactical level we should like to be able to state environments of occurrence and non-occurrence of classes of morphemes in terms of other classes of morphemes, without regard to submorphemic matters.

But can we find any valid basis for preferring the first of the above alternative treatments to the second, or vice versa? Clearly, there can be no TACTICAL reason for choosing any one of two or more tactically equivalent analyses. If any reason at all is discoverable, it will be within the submorphemic realm: a matter of patterning, or perhaps simply of greater convenience. And although

[26] No defect of many older grammars of less-well-known languages is more marked than the confusion, or at best fuzziness, which results from a neglect of the principle of total accountability. Of course we do not condemn their writers for being 'men of their times rather than of ours'; for one thing, this doctrine could hardly be stated explicitly until the phonemic principle had been discovered.

[27] I choose Fox rather than Menomini because the examples are a bit easier to cite; the same principles apply. The Fox forms are from Leonard Bloomfield, Notes on the Fox Language, IJAL 3.219–32, 4.181–226 (1924–7), and from the same writer's Algonquian, Linguistic Structures of Native America 85–129 (New York, 1946).

convenience is a legitimate basis for a choice, we must recognize such a criterion as different in kind from others, and as more open to disagreement. A more convenient analysis tells us nothing more about a language than a less convenient one that is otherwise equivalent; but what it does tell it tells more clearly.

15. The alteration by which the number of morphemes in an utterance fails in some cases to coincide with the number of morphs consists of recognizing two special kinds of morphs: EMPTY MORPHS, which have no meaning and belong to no morpheme; and PORTMANTEAU MORPHS, which belong simultaneously to two (or, theoretically, more) morphemes, and have simultaneously the meanings of both.

If for some submorphemic reason (patterning or convenience), the breakup of Fox *poonimeewa* into *poon* + *i* + *m* (+) is to be preferred to either of the two alternative procedures outlined in §14, this breakup can be made tactically equivalent to the latter two, rather than to the analysis which requires the occurrence of *i* to be taken care of on a tactical level, by calling *i* an EMPTY MORPH. Total accountability is maintained because we say, on the submorphemic level, that when a morph ending in a consonant is followed in the same word by one beginning with a consonant, the empty morph *i* appears between them.

The simplest example of a PORTMANTEAU MORPH is French /o/ 'to the (masc.)' (§3). If this be taken as a single morpheme, tactical difficulties ensue. What other morpheme has a range of positions of occurrence parallel to the range of this one? On the other hand, since /o/ is a single phoneme, it is hardly possible to make a cut and produce two morphs. But if we interpret it as a portmanteau morph, the representative of the morpheme sequence {à le}, we not only eliminate a forlorn morpheme, but round out the distribution of {à} and of {le}, both otherwise somewhat defective. For *à* 'to' parallels to a great extent the distribution of *sur* 'on', *après* 'after', and other morphs, but—unless the proposed interpretation is accepted—does not occur in one important position where the others occur: before *le* 'the (masc.)' when the following noun begins with a consonant. Similarly, the suggested treatment of /o/ makes the parallelism between *le* and *la* 'the (fem.)' much neater. The case is so clear-cut that there is nothing remarkable in the fact that *au* has been traditionally so interpreted.

It is to be noted that our morphemic expansion of /o/ to {à le} involves not only the morphemes {à} and {le}, but also a specific ORDER thereof: /o/ is not morphemically {le à}. This specific order, like the morphemes themselves, is given not by the portmanteau as such, but by its distribution and that of the morphs to which we propose to relate it. Choice of the order {à le} leads to the parallelism indicated above; choice of the reverse order leads to nothing at all.

16. The simple examples just given speak, it is believed, for the naturalness of this approach; but as yet we have given no formal statement of the conditions under which an empty morph or a portmanteau morph is to be set up.

Because of the possible importance of submorphemic patterning, it will be necessary to consider the typical phonemic shapes, or CANONICAL FORMS, of morphs. It is a well-recognized fact that in any particular language, if we examine and classify those cases of morphs which do not patently involve the

questions here being raised, we find that many different morphs have much the same general phonemic shape.[28] Fijian affords an elementary example.[29] A large number of morphs have the shape $\#C_1V_1C_2V_2(C_3)$, where # is word juncture, the C's indicate any consonant (or none), the V's any vowel, and C_3 is lacking when word juncture or a consonant follows: *koro* 'village' (C_3 = zero) *sala* 'path', *dina* 'true, truth', *selev* 'cut, knife', *ðabet* 'go up', *kaðiv* 'call, announce'. A second, much smaller, group have the shape $\#C_1V_1C_2V_2C_3V_3(C_4)$: *taɣane* 'man, male', *yalewa* 'woman, female'. A third group have the shape V or V_1CV_2: *a* 'transitive with common object', as in *ðabeta* 'go up (a hill)', *raiða* 'to see (a child, etc)'; *i* 'transitive with proper object', as in *raiði* 'to see (John, me, etc.)'; *aka* 'transitive indirective', as in *ðabetaka* 'carry (someone) upwards'. Lastly, there is a group of structure #CV#, occasionally $\#C_1V_1C_2V_2\#$: *na* 'the (common)', as in *na koro* 'the village'; *ni* 'of the (common)', as in *na yaða ni koro* 'the name of the village'; *i* 'that connected with the act of', as in *na i sele* 'the knife'; *ko* 'the (proper)', as in *ko viti* 'Fiji' or *ko ðei* 'who?'.

In some languages the variety of canonical forms is far greater than in Fijian, but in every language the total number—however assessed, for there is some choice in the process of abstraction from specific phonemes to symbols like C and V—is relatively small. In English many morphs have the shape of a single syllable with # preceding[30] (*girl, act*); others consist of a single consonant, or of a single syllable with initial vowel, with no preceding # (*-s, -ing, -ed, -or*). In both Fijian and English, and probably generally, some canonical forms can be expressed as the 'sum' of certain smaller ones: Fijian #CVCVCV(C) as #CVCV-(C) plus V; the English type of *author* as the type of *act* (or *watch*) plus the type of *-or* (or *-ing*). Moreover, in these cases the 'sums' occur as sequences of several morphs (*actor*) as well as in single morphs (*author*). Those canonical forms which cannot be so expressed may conveniently be called MINIMUM.

17. If in analyzing the morphemics of a language we make a preliminary classification of canonical forms, based only on those morphs whose status is perfectly clear, this classification can serve as a guide in handling the less obvious cases.

Multiplicity of analytical choice turns on two things: the location and number of cuts to be made in certain utterances; and the classification of the resultant morphs as ordinary, empty, or portmanteau. When faced with alternatives, we base our decision, wherever possible, on the relative desirability of the resulting tactics. It is on this account that the treatment of French /o/ 'to the (masc.)' as a single morpheme, or of Fox connective *i* as a morpheme, is rejected. When this factor cannot play a part, we turn next to morphophonemic simplicity. Morphemes of constant phonemic shape are simplest; when we cannot find these, we look next for sets of morphemes showing similar alternations in phonemic

[28] A point discussed in detail by Benjamin L. Whorf in various unpublished material, and orally.

[29] C. Maxwell Churchward, A New Fijian Grammar (1941).

[30] This avoids the risky complications which result from calling word juncture a morpheme, as Rulon S. Wells does in his Immediate Constituents §64 (see fn. 4). The semantic contrast between *Thank you* with word juncture and the same without it means that word juncture is MORPHEMIC, but in such cases it might just as well be concluded—I think, a little better so—that ABSENCE of word juncture is the morpheme.

shape, since then we can describe the alternations of many different morphemes at once. When this also is not decisive, we turn to canonical forms, and prefer that analysis which produces morphs most closely conforming to the canonical forms already established—if possible, to minimum canonical forms. It may be that the second and third of these considerations should be assigned the other order of priority; apparently they are not often in conflict.

When we are confronted with three tactically equivalent alternatives for Fox *poonimeewa* (*pooni*|*m*, *poon*|*im*, and *poon*|*i*|*m* with *i* as an empty morph), we need only proceed to the second consideration to reach an answer. If we make either the first or the second choice, one of the morphemes involved will have two alternants (*poon* and *pooni*, or else *im* and *m*). If we make the third, both *poon* and *m* become morphemes of constant phonemic shape. If this were not enough, the third criterion would show us that failure to set up the *i* as an empty morph would force us to recognize some morphs, beginning or ending in *i*, of canonical forms not otherwise required (though not in the case of the elements in the particular word *poonimeewa*), whereas the decision to set up the *i* by itself produces only morphs of shapes necessary anyway.

Likewise in the French case: tactical considerations rule out a monomorphemic interpretation of /o/, but do not decide whether we must take it as a single (portmanteau) morph or may cut it further. Now by a criterion mentioned in §2 under Step 1, a morph must have overt phonemic content. In order to cut /o/ into two morphs, we must break up the phoneme /o/ into two components, say mid-back tongue position and lip-rounding. Neither of these components fits into any otherwise necessary canonical form of morphs in French (though in some other languages some morphs do have a shape definable in components rather than in phonemes). On the other hand, /o/ taken as a single morph fits into a canonical form, represented by such clear cases as /o/ 'water', /e/ 'and', /u/ 'or', /a/ 'to'. The vote is clearly for the interpretation as a portmanteau morph.

In the succeeding sections of Part IV we give further examples in which these same principles call for the recognition of empty or portmanteau morphs.

18. In Nootka[31] a word consists—submorphemically, as we shall show—of a stem plus one or more suffixes. (This statement is circular: a stem is a morpheme which begins with word-juncture, a suffix one which does not; but it will suffice for orientation.) Certain suffixes, which we may call LOCATION SUFFIXES, occur both after ordinary stems and after a stem *hina-*, *hin-*, *hita-*,[32] which Sapir and Swadesh label an EMPTY STEM.

Thus *ƛ̓iḥ-* 'red' + *-(q)oˇ(ɬ)* = *ƛ̓iḥo·ɬ* 'red on the face'; *sixʷ-* 'sores, pox' + the same suffix = *sixo·ɬ* 'having sores on the face'; but *hina-* + the same suffix = *hino·ɬ* 'on the face, being on the face'. Similarly, *ƛ̓iḥ-* 'red' + *-(ʔ)akso(ɬ)* =

[31] Edward Sapir and Morris Swadesh, Nootka Texts (Philadelphia, 1939), esp. Part III, The Primary Structural Elements of Nootka; Morris Swadesh, Nootka Internal Syntax [sic], IJAL 9.77–102 (1936–8). The specific examples cited were generously supplied by Swadesh.

[32] The alternation among these three shapes of the element will not concern us; it is covered by statements on a lower level of morphophonemic treatment.

ʔxiḥaksoł 'red at the lips'; *hap-* 'hair, fur' + the same suffix = *hapaksoł* 'having a moustache'; but *hina-* + the same suffix = *hinaksoł* 'at the lips or mouth, being at the lips or mouth'. Finally, *maχ-* 'tied' + *-aḥo·p* 'cause momentaneously to be in front' = *maχaḥo·p* 'ties in front'; but *hina-* + the same suffix = *hinaḥo·p* 'places in front'.

The empty stem has no meaning. Our tactics are just as well suited, and our morphophonemics are not complicated, by interpreting each form of the empty stem as an empty morph. The remaining stems constitute a class of morphs which begin with #. Suffixes other than the location suffixes constitute a class of morphs which do not begin with #. When a location suffix is preceded by a stem, neither # nor any other phonemic material intervenes. When a location suffix is not preceded by a stem, it is preceded instead by the appropriate (predictable) form *# hina-*, *# hin-*, *# hita*, which in any case is meaningless and tactically irrelevant. The principle of total accountability is not violated; and the empty morphs conform to canonical forms. The alternative of taking *# hina-* etc. as a stem (a morpheme) is undesirable because of the meaninglessness of this element. The alternative of taking *hinaḥo·p* as an alternant going with *-aḥo·p*, and similarly for every other such combination, produces a greater complication of canonical forms.

19. In most of the central Algonquian languages occurs a phenomenon which we shall here illustrate with Potawatomi examples. Nouns appear in both unallocated and allocated forms: *wUkUma* 'chief' : *nUtOkUmam* 'my chief' or *kUtOkUmamwa* 'your (pl.) chief'. Some nouns, however, appear only in allocated forms; so *nos·* 'my father', *kos·* 'your (sg.) father'. A noun in an allocated form contains a personal prefix before the noun stem, and after the noun stem one or more of several suffixes (including, if the allocation is plural, a personal suffix) and various inflectional endings. Some of those nouns which occur in both unallocated and allocated forms contain, after the noun stem in an allocated form, a morph *m*, *Um*, *im*, *om*; so, for example, the forms for 'my chief' and 'your chief' above. Other such nouns appear sometimes with a morph of this shape, sometimes without it, but with no semantic contrast—the presence and absence of the morph are in free alternation. Still others, including all nouns which appear only in allocated forms, never occur with the *m* element.

The *m* elements are more satisfactorily regarded as morphs than as parts of the preceding morphs, particularly since the choice among the various forms of the *m* element depends on the environment in much the same way as does the choice among the alternants of the morpheme {k} 'locative' and (though with less similarity) among the various alternants of a number of other suffix morphemes. But the *m* elements are meaningless, even where forms appear both with and without one of them, and it is tactically convenient to eliminate them from the picture before tactical discussion begins. So we take them to be empty morphs.

20. The English interjections written conventionally as *hm!*, *eh?*, and the like, consist phonemically of an intonation-sequence, a stress, and a segmental 'carrier' for these features. In my dialect, this segmental component may have any vocalic quality (whether this occurs elsewhere or not), or any oral closure or

closures, but it must be nasalized. Such a segmental structure is atypical in the wide range of nondistinctive variation, but the articulatory feature involved distinctively—nasalization—is one which does recur in more typical segmental structure, for example as that which distinguishes /m/ from /b/, /n/ from /d/.

If we compare the utterance *hm* (intonation *32*)[33] with *yes*(*32*) and with *hm*(*24*), we see that the meaning of *hm*(*32*) is that of *yes*(*32*) minus the meaning of the intonationless abstraction *yes*; between *hm*(*32*) and *hm*(*24*) there is no semantic similarity. *Hm* itself, then, apart from the intonation which it serves to carry, has no meaning at all.

We may conclude that the *hm* part of such interjections is an empty morph. The intonational morphs which accompany it are also found spread through such morphs or morph-sequences as *yes*, *I know*, *no*, *maybe*, *he didn't come*, and so on. Accountability is maintained: if an utterance consists morphemically of an intonational morpheme alone, the empty morph *hm* will be present; otherwise *hm* is absent.

The tactical implications are interesting: the only free morphemes of English, in Bloomfield's sense of 'free', are intonational morphemes, and the only monomorphemic utterances of the language are those consisting of such a morpheme.

21. In certain Spanish verb forms there appears, between the stem and the endings, an element often called a CONJUGATION VOWEL: the *á*, *é*, and *í* of *amar* 'to love', *beber* 'to drink', *vivir* 'to live'; the *áb*, *í* and *í* of *amábamos* 'we loved', *bebíamos* 'we drank', *vivíamos* 'we lived', etc. Which vowel (or in one case vowel plus consonant) appears, depends on the stem and on the ending: the infinitive ending *r*, for example, requires *á* after a stem of the first conjugation, *é* after one of the second, and *í* after one of the third. The three conjugations are classes of stems, in fact, based precisely on this feature of behavior.

The conjugation vowels have no meaning. The meaning of *amar* is that of two component morphemes, stem *am* 'love' and the infinitive ending, whether we treat the latter as *ár* in this case, alternating with *ér* and *ír* elsewhere, or simply as *r*. The latter alternative relegates the *á*, and all other such conjugation vowels, to the status of empty morphs.

Not all the post-stem vowels which occur in Spanish verbs have this status. The *a* of *amas* 'thou lovest', for example, is the only thing which distinguishes this form from *ames* 'that thou lovest' (subjunctive). Here the *a* is no empty morph, but an ordinary morph with meaning 'present indicative'. In one possible analysis of Spanish verbs, which would perhaps be the simplest from the morphophonemic standpoint, distinctions such as that between *a* 'present indicative' and the meaningless *á* of *amar* are not made. But this somewhat greater morphophonemic ease is outweighed by more complicated tactics.

22. Our first additional example of a portmanteau morph comes from Yokuts.[34]

[33] Following Pike (The Intonation of American English; Ann Arbor, 1946) and Wells (Immediate Constituents §79) in the assignment of figures, and numbering the four levels from top down.

[34] See fn. 11. The capital letters at the beginning of cited suffixes are components of the vowels in the part of the word which precedes the specific phonemes of the suffix. Thus the stem *me·k'i* 'swallow' contains two consonants and parts of two vowels: after the first consonant, the vowel components high-front and long, and after the second consonant, the

In the Yawelmani dialect there are about a dozen morphemes which occur after a verb stem and before a finite or gerundial suffix. One of these is WZ*a·la·* (with alternants WZ*la·*, FR*la·*, WW'*e·*, WL*e·*, FZWZ*la·*, *ila·*, *la·*, WS*la·*, variously apportioned among different types of preceding element, but in non-contrastive distribution) 'cause someone to x': *tisa·la·hin* '(he) caused (it) to come out, (he) took (it) out', with stem *tisi* 'come out' and finite suffix *hin* 'aorist'. Another is WA*da·* (with variants *da·*, R = reduplication) 'x often or repeatedly': *sodoxdo?* 'will throw him repeatedly', with stem *sodox* 'throw' and final suffix *?* 'future'.

In some cases two of these elements occur in succession, within the position of occurrence stated above. Indeed, the alternant R of the morpheme 'x repeatedly' occurs followed by the alternant FZWZ*la·* of the morpheme 'cause to x': *muhmuhlat* 'was made to dive repeatedly', with stem *muhu* 'dive' and final suffix *t* 'passive aorist.' But other alternants of the morpheme 'x repeatedly' do not occur before any alternant of the morpheme 'cause to x'. No semantic gap results, however, for there is an element WE·*lsa·* ~ WE·*sa·* of the same positional class, meaning 'cause to x repeatedly': *nine·lsa·hin* 'got him to keep still several times', with stem *nine·* 'keep still, become quiet', and final suffix *hin* 'aorist.'

It is far from convenient, within the submorphemic economy of Yokuts, to cut WE·*lsa·* ~ WE·*sa·* into smaller morphs; each alternant subsumed by this notation is best taken as a single morph. The distribution and meaning lead one to interpret each of these morphs as a portmanteau representative of the sequence of two morphemes 'x repeatedly' + 'cause to x'.

23. In finite forms of the Spanish verb the tense-mode is usually indicated by one morph, and the person-number by another, in that order: *amáb|a|mos* 'we loved', *ama|re|is* 'you (pl.) will love'. In a few cases, it is difficult or impossible to separate the element meaning a tense-mode from that meaning a person-number; in these cases, we may regard the undivided endings (after any conjugation vowel that may occur) as portmanteau morphs: *o* 'present indicative + first person singular', as in *amo* 'I love', *é* (with verbs of the first conjugation), *í* (with those of the second and third conjugations) 'preterit indicative + first person singular,' and *ó* (first conj.) *ió* (second and third conj.) 'preterit indicative + third person singular.' This treatment, combined with the empty-morph interpretation of conjugation vowels, reduces all finite Spanish verb forms to a uniform structure: stem + tense-mode morpheme + person-number morpheme.[35]

24. In Fijian there is a construction consisting of any of certain particles followed by a noun or pronoun: *na sala* 'the path', *na ðava* '(the) what?', *ni koro*

vowel components high-front and short. When this stem occurs with the suffix FR*it* 'passive aorist', the component F merges with the first group of vowel components in the stem to give *e·*, and the component R merges with the second group of vowel components in the stem to give zero; the resulting form is *me·k'it* 'was swallowed'. With a different set of components contributed by the suffix WA*?an* 'durative present', the resulting form is *mik'a·?an* 'is swallowing'. For the details of this, see Zellig S. Harris, Yokuts Structure and Newman's Grammar, IJAL 10.196–211 (1944).

[35] The 'irregular' verbs present more complex cases of both portmanteau and empty morphs, but are tactically quite the same, save where one or another form is missing.

'of-the village', *ko viti* '(the) Fiji', *vei au* 'to me'. One of the particles is *vei* 'to, with, of', as in the last example above and in *vei keda* 'with, to, of us', *vei Joni* 'with, to John'. One of the pronouns is *koya* 'he, she', as in *ko koya* '(the) he, she', *nei koya* 'of him, of her'. But the specific combination of *vei* and *koya* seems not to occur.[36] Where semantically it would be expected, one finds, instead, the portmanteau morph *vuaa* 'with, to, of him or her', as in *au na vosa vuaa na ŋone* 'I future speak to-him the child' = 'I shall speak to the child.'

25. We may best approach a consideration of the fifth difficulty of §3 by examining some English cases. On the tactical level, it is certainly desirable to consider *men* as consisting of the morpheme *man* plus the morpheme *s* 'plural'. When cutting utterances containing *men* into morphs, we will not be led to cut *men* into smaller pieces; it fits a canonical form and if broken further the smaller fragments do not. So one solution, and certainly the most obvious one, is to regard *men* as a single portmanteau morph, representing the morpheme sequence {man} + {s}.

It is true that there is a phonemic similarity between *man* and *men*—the identity of initial and final consonants—which we do not want to lose sight of. This places *men* in a different category from French /o/, Yokuts WE·(*l*)*sa*·, Spanish *o*, or Fijian *vuaa*, for in the latter cases the resemblance of the portmanteau to other alternants of either of the constituent morphemes is negligible. Even if *men* were an isolated case in English, this resemblance would be worthy of mention. But it is, of course, far from isolated; we have also *mouse* : *mice*, *foot* : *feet*, *woman* : *women* (if *woman* is a single morpheme), *slide* : *slid*, *sing* : *sang*, and many others.

The portmanteau interpretation of such bimorphemic forms need not obscure the phonemic resemblance of which we are speaking. In our morphophonemics we have to mention all portmanteaus. By assembling, in one section of our description, all portmanteaus which have this feature of partial phonemic identity with one of the constituent morphemes, and by organizing them into groups on the basis of the specific phonemic difference, we give ample attention to the matter.

Some may nevertheless prefer to reinterpret portmanteaus as bimorphic as well as bimorphemic, even though to do so one must extend the definition of 'morph' to cover elements of other than overt phonemic content. If this is considered desirable, then in the notion of portmanteau we have at least achieved a more rigorous way of extending the coverage of the term 'morph' in such a manner, as follows:

In our initial cutting of utterances, we obtain only morphs of overt phonemic content. Further examination, along the lines detailed in this paper so far, reveals the possibility that certain of our morphs are portmanteaus; but for our present purpose we may call them rather TENTATIVE PORTMANTEAUS. We then examine each tentative portmanteau and compare its phonemic shape with that of the other alternants of the constituent morphemes. If we find that the tenta-

[36] Churchward is not entirely clear on the matter: he says (op.cit. I.24.3) that *vei koya* is 'unusual'. If it does indeed occur, then the interpretation proposed is wrong; rather *vei koya* is like English *with it* and *vuaa* like *therewith*.

tive portmanteau has some phonemes (or components) in common with one of the non-portmanteau alternants of one of the constituent morphemes, we may set up the entire non-portmanteau alternant as one constituent MORPH of the form which has tentatively been regarded as a portmanteau, and the alternation from this shape to that of the tentative portmanteau as the other constituent MORPH. Alternatively, we may set up the tentative portmanteau, as a whole, as an alternant of that constituent morpheme which it resembles phonemically, and set up a ZERO MORPH as an alternant of the other constituent morpheme.[37] For example, our initial cutting produces *men*, which does not look like more than one morph. The sequence *man* plus *s* does not occur. *Men* fills the tactical place which one might expect to be filled by the sequence *man* plus *s*. *Men* is therefore morphemically {man} + {s}. But—so runs the argument that would set up alternation morphs—*men* and *man* resemble each other in phonemic shape, both containing *m-n*. So *men* is not a portmanteau. One morph in *men* is *man*. The other is the alternation $a \sim e$. Or—arguing now for a zero morph—*men* is not a portmanteau, but consists of an alternant *men* of {man} plus an alternant /0/ of {s}.

If a language contains only a few isolated instances of this kind, probably everyone would agree to reject the last steps of the above argument and return to the portmanteau interpretation, relying on the organization of one's morphophonemic statements to put the matter of partial phonemic resemblance into clear relief. But if the language contains a sufficient number of such cases that one is warranted in setting up a canonical form for morphs like $a \sim e$, or like /0/, then some may prefer the extension.

Somewhat similar considerations apply to French *bon* 'good (masc.)' and to English *sheep* (pl.). In the course of examination, the portmanteau interpretation is that which first presents itself; from it we may proceed to the recognition of morphs of other than overt phonemic content if we find factors comparable to those in the case of *men*. It is to be emphasized that when portmanteaus are eliminated in this way, the new definition of 'morph' is no longer that with which we began; perhaps, therefore, it would be advisable to distinguish terminologically between, say, 'primary morphs' (those of overt phonemic content) and 'extended morphs' (including primary ones and morphs of the zero, replacement, or subtraction type).

English *children*, however, remains recalcitrant. Obviously it is morphemically {child} + {s}; so that whatever submorphemic interpretation we chose, the tactical picture is clear. The first part of *children* resembles *child*, and the last part is identical with one of the alternants of {s}, namely the *-en* of *oxen*. The alternative analyses are (1) *child|ren*, (2) *childr|en*, (3) *child|r|en*, (4) *child* + vowel change and *-ren*, and (5) no cut, i.e. portmanteau. The first gives a morph

[37] This second alternative is that proposed by Bernard Bloch, English Verb Inflection, LANG. 23.399–418 (1947). Bloch rejects all alternation or subtraction morphs, and interprets all tentative portmanteaus as an alternant of one of the constituent morphemes plus a zero alternant of the other. One special criterion is introduced for dealing with zero alternants: no morpheme is postulated which has ONLY a zero alternant.

/čild/, the difference between which and *child* recurs in other contexts, e.g. *slide* : *slid*, *bite* : *bit*; but then the morph *-ren* is unique. The second gives a morph *-en* which recurs; but then the difference between /čildr/ and *child* is unique. The third has the merits of each of the first two, without the defects, but involves an empty morph *r*, which is not observed to recur and therefore requires a special statement for this occurrence. The fourth produces a morph (vowel change and *-ren*) which fits no canonical form, unless the vowel-change-plus-*en* of *bitten*, *hidden*, and others is grouped with it from the point of view of shape. Apparently this is one of the cases in which all our preferential criteria (§14) fail, and nothing remains but a resort to convenience.[38]

V. Conclusions

26. We now summarize the procedure of morphemic analysis worked out in the course of our discussion, and end with an example from English which illustrates several of the points that have been made. Our summary of the procedure is given in steps, as in §2; but in actually working with a particular language one has to skip back and forth, operating by trial and error.

Step 1. We assemble the utterances of the language before us, recorded in some phonemic notation. If preliminary examination reveals that a different (also phonemic) notation would make the task simpler, we retranscribe them all. If further preliminary examination shows that some normalization of notation, maintaining all phonemic distinctions but adding thereto, would further simplify the task, we retranscribe again, and perhaps again. As we proceed to other steps, we check back from time to time to be sure we have not involved ourselves in contradictions.

Step 2. The utterances are now examined in the notation finally chosen. Recurrent partials with constant meaning are discovered; those not composed of smaller ones are morphs. So are any partials not recurrent but left over when all recurrent ones are accounted for; therefore every bit of phonemic material belongs to one morph or another.[39] By definition, a morph has the same phone-

[38] The unsolved case of *children* is discussed in detail for a reason. There is no merit in an analytical procedure which 'eliminates' all but one of a set of alternative analyses simply by fiat—by saying that when such-and-such types of alternatives present themselves we shall accept the one which has certain characteristics and reject the others. Our aim is to achieve the most accurate and clearest picture possible of the workings of a language, on all levels—phonemic, morphemic, and tactical; in some cases this is attained not by giving a single treatment, but precisely by indicating the alternatives. For in some cases a range of choice is determined not by our approach, but by the nature of the language; and when this is so, the existence of a range of choice in a particular portion of the language is one of the facts about the language that ought to be portrayed in our description. In one sense, any method of description which conforms to the principle of total accountability is correct; if we nevertheless discuss the relative merits of one procedure or another within this fundamental framework, the purpose is to attain greater mutual intelligibility among the writers of grammars and, in terms thereof, more accurate pictures of the languages we describe.

[39] We say 'phonemic' for simplicity's sake; if our notation has been normalized, then more accurately this should read 'every bit of orthographic material'.

mic shape in all its occurrences; and (at this stage) every morph has an overt phonemic shape, but a morph is not necessarily composed of a continuous uninterrupted stretch of phonemes. The line between two contiguous morphs is a cut.

STEP 3. Omitting doubtful cases, morphs are classed on the basis of shape and the canonical forms are tentatively determined.

STEP 4. Two or more morphs are grouped into a single morpheme if they fit the following grouping-requirements: (a) they have the same meaning; (b) they are in non-contrastive distribution; (c) the range of the resultant morpheme is not unique. Some morphs, however, may be assigned to no morpheme at all, and some may be assigned simultaneously to two (or more) morphemes. An empty morph, assigned to no morpheme,[40] must have no meaning, and must be predictable in terms of non-empty morphs. A portmanteau morph must have the meanings of two or more morphemes simultaneously, and must be in non-contrastive distribution with the combination of any alternant of one of the member morphemes and any alternant of the other (usually because no such combination occurs).

STEP 5. Where there are alternative possibilities, choice is based on (a) tactical simplicity, (b) morphophonemic simplicity, and (c) conformity to canonical forms, in this order of priority.

STEP 6. The differences in the phonemic shape of morphs as alternants of morphemes are organized and stated; this (in some cases already partly accomplished in Step 1) constitutes morphophonemics. In particular, portmanteaus are compared with the other alternants of the morphemes involved, and if resemblances in phonemic shape and the number of cases warrant it, morphs of other than overt phonemic content are recognized, some of the portmanteaus being thus eliminated.

27. Our final example is the system of personal pronouns in English (including *who, whom, whose*).

At least in certain dialects, the morphs *I* and *me* (and similarly *we* and *us*, *he* and *him*, etc.) are in non-contrastive distribution; in some dialects, indeed, the complementation is probably complete. We may suspect that if it were not for the Latinizing school tradition, the complementation would be complete for most speakers: *I* initially except in isolation, *me* directly after a verb or a preposition and in isolation. Actual exceptions to this are either on the Latin pattern (*It's I*, or *Who's there?—I*, instead of *Me*), or are overcorrections (*between you and I*). For many speakers whose usage of *I* and *me* does not put them in complete complementation, there is no contrast between, for example, *It's I* and *It's me*. In other dialects and styles, on the other hand, the forms are in contrast: literary English, schoolteachers' on-duty English, and certain

[40] All the empty morphs in a language are in complementary distribution and have the same meaning (none). They could, if there were any advantage in it, be grouped into a single empty morpheme—but one which had the unique characteristic of being tactically irrelevant.

whimsical styles.[41] The remainder of this discussion applies only to a dialect in which the distribution is non-contrastive.

My and *mine* (and similarly *our* and *ours*, *your* and *yours*, etc.) are in complete complementation: *my* occurs when a noun follows without pause, *mine* otherwise.

If the above statements are to hold, we must split the occurrences of *her* into those which parallel those of *his* and those which parallel those of *him*; the former, her_1, is morphemically identical with *hers*, while her_2 is morphemically identical with *she*.

Paralleling *John* in *John came*, *Bill saw John*, *John's book*, *the book is John's*, and virtually every other utterances containing the morpheme *John*, we have *I came*, *Bill saw me*, *my book*, *The book is mine*, etc. *John's* is two morphs and two morphemes; we conclude that *my* and *mine* are two morphemes each, though each is only a single morph.

We conclude, therefore, that the English personal pronouns have the following morphemic structure:

{I} *I, me*	{she} *she,* her_2
{I} + {s} *my, mine*	{she} + {s} her_1, *hers*
{we} *we, us*	{it} *it*
{we} + {s} *our, ours*	{it} + {s} *its*
{you} *you*	{they} *they, them*
{you} + {s} *your, yours*	{they} + {s} *their, theirs*
{he} *he, him*	{who} *who, whom*
{he} + {s} *his*	{who} + {s} *whose*

The forms *it*, *its*, and *whose* are the same morphically and morphemically; the others illustrate one or more of the grouping-requirements that we have discussed. Together, the twenty-six forms are analyzed into only nine different morphemes.[42]

The tactical implications are considerable. Except for the category of number, the pronouns are now exactly like any proper noun in their tactics, and can be classed as a subdivision of proper nouns. There is no longer any justification for speaking of case in English; for the distinction between subjective and objective 'cases' (under whatever name) disappears as soon as *I* and *me*, etc., are shown to belong to the same morpheme. A form with added *-'s* is not a case-form either, but simply a form with added *-'s*: the *-'s* is simply another morpheme, with a statable range of positions in which it occurs.

[41] For example, that style in which one says *me, myself, and I* as if the reference were to three people. This is not unrelated to a style which obviously has to be excluded, both here in the discussion of English pronouns and in any other discussion of morpheme alternants: the style of the discussion itself, in which such forms as *me* and *I* contrast because they are used as names of particular morphs.

[42] We might go further, interpreting *we, us* as {I} + pluralizing {s}, with a similar treatment for the other plural pronouns. We are deterred from this step not because plural *you* is identical with singular *you* (since after all *sheep* and other nouns manifest this property), but because {he} + {s}, {she} + {s}, and {it} + {s} would all add up to *they, them*.