DIAGRAMMATIC ICONICITY IN STEM-INFLECTION RELATIONS

JOAN L. BYBEE
SUNY at Buffalo

There is a tendency in languages for utterances to be organized around nouns and verbs, such that the position of the other elements in the utterance is often definable in relation to the position of nouns and verbs. Hence the notion that nouns phrases and verb phrases are major constituents of the clause. There are also a smaller number of elements whose position might be defined in terms of the clause as a whole. It has often been observed that the proximity of elements in a clause follows some natural (iconic) principle whose result is that elements that go together semantically tend to occur close together in the clause.[1] Following this principle, we would expect that elements whose position is defined in terms of the position of the noun would have meanings that modify or relate to the meaning of the noun or noun stem, while elements whose position is defined with respect to the verb would have meanings that modify or relate in some way to the meaning of the verb or verb stem. Similarly, elements whose position is determined with respect to the whole clause would have the entire proposition in their semantic scope.

In this paper obligatory grammatical markers, and more specifically, obligatory grammatical markers that are bound to the verb, are studied in the light of this general principle. Verbal inflections differ with respect to the extent to which they are *relevant* to the verb, that is, the extent to which their meanings *directly affect the lexical content of the verb stem*. The different degrees of relevance of verbal categories that can be inflectional are reflected diagrammatically in three ways: (1) The more relevant a category is to the verb, the more likely it is to occur in a synthetic or bound construction with the verb: (2) The more relevant a morphological category is to the verb, the closer its marker will occur with respect to the verb stem: (3) The more

relevant a morphological category is to the verb, the greater will be the morpho-phonological fusion of that category with the stem.[2]

In a cross-linguistic survey using 50 genetically and areally unrelated languages, I gathered data concerning these three predictions, i.e. data on the frequency of occurrence of inflectional categories for verbs, their order with respect to the verb stem, and their degree of fusion with the stem. These data bear out all of the predictions made by the relevance principle. In the first section of this paper, I will demonstrate how the principle is applied to the various verbal categories. In the second, I will present the data that supports the hypothesis that relevance is reflected iconically in morphological expression. In the third section, I will discuss some seemingly problematic cases.

1. The *relevance* of morphological categories

The particular verbal categories investigated were those that can be *inflectional* in some language, although derivational expression of these categories was noted also, and will be discussed here. An *inflectional category* is one that is *bound* to the stem, and whose expression is *obligatory* in the particular grammatical context.[3] If a category is obligatory the lack of a marker for the category in the context will be taken as signalling one member of the category, i.e., as the zero expression of the category. For example, the Australian language Tiwi (Osborne 1974) has a verbal prefix meaning "at a distance". It is not considered an inflectional prefix, however, because when the prefix is absent, the verb does not mean "close by", but rather says nothing whatever about distance. On the other hand, the Spanish verb *canta* has no marker for person or number, and yet is interpreted as 3rd person singular. Thus person and number are inflectional categories of Spanish.[4]

It was hypothesized in advance that the inflectional categories would be valence, voice, aspect, tense, mood, subject agreement for number, person, and gender, and object agreement for the same.[5] The categories were also defined in advance using definitions from the literature on morphology. It was recognized that these definitions might have to be modified in light of the categories actually found in the fifty languages. However, this was not the case. The definitions proved remarkably suitable for the individual languages investigated, a confirmation of the accuracy of the collective intuitions of linguists, and the inherent comparability of the languages of the world.[6] The few problems that did arise with the definitions will be mentioned below in various places.

The relevance of the category to the verb was also predicted before the survey was begun, by applying the concept of relevance to the definition of the category, in the following way. The inherent lexical content of a verb stem describes an event or state. A category is relevant to a verb to the extent that it directly modifies the event or state described. A category is less relevant if it affects or refers to other elements in the clause instead of or in addition to the verb. Note that in addition to relevance, which refers only to the scope of the modifying category, there is also a difference in the amount of semantic change resulting from the combination of the morphological category with the verb stem. The amount of semantic change ordinarily increases and decreases as relevance does, since the more relevant a category is to the verb, the more profound effect it can have on the meaning of the verb. Thus, for the most part, it will not be necessary to maintain a distinction between them. So for each category we will describe in general terms the extent to which it affects the meaning of the verb stem, as opposed to affecting other elements in the clause.

A confirming diagnostic for relevance was also considered in advance: this is the ability of the semantic notion expressed by the category, or a closely related semantic notion, to be expressed *lexically* as a component of a verb's meaning. Thus changes in valence (the number of arguments a verb can take) are morphological in many languages, but may have lexical expression in English pairs such as *sit* and *set*, *lie* and *lay*, *die* and *kill*. Thus the term *lexical expression* will be used to mean the combination in a single lexical unit of the lexical meaning of the verb with a meaning similar to that expressed in a morphological category.[7] Lexical expression is more likely when a greater meaning change results from the combination of a stem and a modifying semantic notion. Thus lexical expression becomes important in explaining why the most relevant categories are not necessarily the most frequent inflectional categories: a highly relevant category that makes a large meaning change can have lexical or derivational expression, thereby detracting from the number of instances of inflectional expression.

We turn now to the application of the notion of relevance to the verbal categories that can be inflectional in the languages of the world:

Valence refers to changes in the number and the roles of the arguments that the verb stem can take. Valence-changing categories such as transitive, intransitive and causative are relevant to the situation described in the verb stem, in the sense that any changes in the number and role of the participants can have a profound effect on the situation described by the verb stem. While

valence affects the choice of arguments of the verb, it does not refer to the arguments or have the arguments in its scope. Valence only affects the meaning of the verb stem, which then determines the number and role of the arguments selected. (Cf. subject agreement markers, which refer only to the arguments and do not affect the lexical content of the verb stem.) The change in meaning is sometimes dramatic, as in the case of causatives, predicting a tendency toward lexical expression of valence categories, such as the English *die* vs. *kill* and *fall* vs. *drop*. Very often in languages a distinction such as that between transitive and intransitive figures is an important morphological distinction, even if it does not always qualify as an inflectional one. For example, the following intransitive/transitive pairs represent a widespread distinction made in Hebrew (Berman 1978): *avad* 'work' vs. *ibed* 'cultivate'; *yaca* 'go out' vs. *yice* 'export'; and *paxat* 'lessen' vs. *pixet* 'devaluate'.

Voice indicates the perspective from which the situation described by the verb stem is viewed, and in particular, voice distinctions, according to a description by Barber 1975, change the relation that the surface subject has to the verb. In the active, the subject is the doer of the action; in the passive, the subject is affected by the action; in the reflexive, reciprocal and middle, the subject both performs the action and is affected by the action. Voice, then, is relevant both to the verb and to its arguments. In signalling a "deviant function" of the subject, it changes the roles of the NPs in the sentence, as well as the perspective from which the situation described by the verb is viewed. It is not surprising, then, that voice may be morphologically coded on the NPs of the sentence, on the verb, or on both. Distinctions in perspective that resemble voice distinctions also occur lexically, for instance in English verbs such as *buy* and *sell*, *give* and *receive*. Some reflexive verbs in Romance languages, such as Spanish, have taken on unpredictable meanings, and have become lexicalized: *acordar* 'to agree, to decide upon' vs. *acordarse de* 'to remember', *echar* 'to throw' vs. *echarse (a)* 'to begin to', *volver* 'to turn, to return' vs. *volverse* 'to become'. These examples show that the meaning expressed by voice categories is relevant enough to the verb to be combinable in lexical expression, and further that the amount of semantic change is sufficient to lead to lexicalization, at least in some cases.

Distinctions in *aspect* include different ways of viewing "the internal temporal constituency of a situation" (Comrie 1976, taken from Holt 1943). The perfective aspects (inceptive, punctual and completive) view the situation as a bounded entity, and often put an emphasis on its beginning or end. The imperfective aspects in contrast do not view the situation as bounded, but rather as ongoing in either a durative, continuative or habitual sense. Aspect, then, refers exclusively to the action or state described by the verb. It does not affect the participants, nor does it refer to them.[8] Thus, it might be said that aspect is the category that is most directly and exclusively relevant to the verb.

Many languages have aspectual distinctions expressed lexically (*Aktionsart*), such as English *do* vs. *complete*, and *know* vs. *realize*. It is also common to find aspectual distinctions expressed in derivational morphology, as in Latin *facere* "to do" and *cōnficere* "to complete", or inchoative *amō* "I love" and *amascō* "I begin to love", *dormiō* "I sleep" and *obdormiscō* "I fall asleep". These usually express more specific meanings, such as inchoative, as in the Latin example, or completive as in Russian *užinat'* 'have supper', which contrasts with *otužinat'*, which means 'finish supper'.

When *aspect* is an inflectional category, the meaning change effected by it tends, as predicted, to be small. Hopper (1977, 1979) has argued that inflectional aspect serves to indicate how the action or state described by the verb should be viewed in the context of the whole discourse. Background information is expressed by imperfective verb forms, and the foregrounded information of the main narrative line appears in perfective verb form. This discourse use of aspect leaves the basic meaning of the verb unaffected, and only changes its relation to the discourse unit.

Tense is a deictic category that places a situation in time with respect to the moment of speech, or occasionally with respect to some other pre-established point in time. It is a category that has the whole proposition within its scope, and yet it seems to be always marked on the verb, if at all. This is so in part because it is the verb that binds the proposition together, and makes it refer to a situation that can be placed in time. But another reason that tense is marked on the verb rather than on, for example, the nominal arguments, is that, as Givón 1979 has observed, nouns usually refer to time-stable entities, while verbs refer to situations that are not time-stable.[9] Thus it is the verb that needs to be placed in time if the event or situation is to be placed in time, since the entities involved in the situation usually exist both prior to and after the referred to situation. Because tense has the whole proposition in its scope, it is somewhat less relevant to the verb than aspect, but somewhat more relevant than mood and agreement categories.

A *tense* distinction does not affect the meaning of the verb, since the situation referred to by the verb remains the same whether it is said to occur in the present or the past. Consequently, it is rare to find examples of a tense

distinction expressed lexically. To illustrate what a real case would be like, consider English *go* and *went*. They are lexicalized in form, since there is no way to predict the form of one from the other, but they do not constitute a real example of lexical expression, since they must be viewed as a suppletive expression of a general inflectional category of English. A similar pair in a language with no inflectional tense categories would be a genuine example.

Mood distinctions express what the speaker wants to do with the proposition in the particular discourse. This will include expression of assertion (indicative), non-assertion (subjunctive), command (imperative), and warning (admonitive). It also includes other expressions of the speaker's attitude about the truth of the proposition, such as indications about the possibility, probability or certainty of the truth, as well as the source of the information (evidentials). Even when mood is expressed as a verbal inflection, it is clear that it has the whole proposition in its scope, and does not only modify the verb. Furthermore, since it expresses the speaker's attitude, it does not have a direct effect on the situation described by the verb. Both of these properties make mood less relevant to the verb than either aspect or tense. Thus we might expect mood to occur less frequently as an inflectional category of verbs than aspect and tense. Since mood cannot affect the meaning of a verb, examples of lexical expression of mood-like distinctions are rare or non-existent.[10]

Agreement categories in verbal inflection refer not to the situation described by the verb, but rather to the participants in the situation. Thus agreement categories are less relevant that categories that more directly affect the meaning of the verb. Agreement categories commonly include distinct markers for person (usually 1st, 2nd and 3rd), number (singular, dual and plural) and less frequently agreement by gender or classifier. Not all of these agreement categories have the same status with regard to our hypothesis, however. While person and gender categories seem to have little effect on the meaning of a verb, and are, as mentioned above, rarely lexicalized, number is somewhat differnt. The number of participants in a situation, whether agents or recipients of an action, can affect the situation. Thus lexicalized distinctions based on singular vs. plural participants do exist, e.g. English *run* vs. *stampede*, *murder* vs. *massacre*. And, as we shall see below, some examples of systematically lexicalized or derivational expression of number showed up in the cross-linguistic survey.

To summarize this section, a diagram is presented below with the inflectional categories we have discussed arranged in approximate order of degree of relevance to a verb. The categories on the higher end allow lexical as well as inflectional expression, while those on the lower end allow only inflectional.

Expression	*inflectional*	*lexical*
Category		
valence	x	x
voice	x	x
aspect	x	x
tense	x	
mood	x	
number agreement	x	(x)
person agreement	x	
gender agreement	x	

This scale predicts which categories are most likely to be expressed morphologically in conjunction with a verb stem. It alone does not predict which categories are the most likely to be expressed as inflectional categories. To arrive at that prediction, we must take this linear scale and bend it into a bell-shaped curve. The categories in the middle will be the highest points on the curve, that is, the most likely to be inflectional categories for verbs. The likelihood of inflectional expression drops off on either end, but for different reasons. On one end it drops off because the categories become less relevant to the verb. On this end of the scale lie the agreement categories. On the other end the scale drops off because the categories involved make larger and less predictable semantic changes, and are thus more likely to be lexicalized. Such a curve, then, emphasizes nicely the position of inflectional morphology as lying between syntactic expression and lexical expression.

2. Cross-linguistic data

The sample of languages used in this survey is described in detail in Perkins 1980, and summarized in Bybee 1985. One relevant fact will be noted here: because the sample was chosen to be representative of the languages of the world, and free of genetic or areal biases, and *not* chosen for convenience, as most samples are, it happens that in some cases the information about the languages is not complete. It is preferable to tolerate this situation, and take account of poor descriptions where they occur, than to bias the sample by chosing languages on the basis of the availability of information.

Poor documentation is not a serious problem, however, since close to 90% of the descriptions used give a very complete account of the verbal morphology of the language.

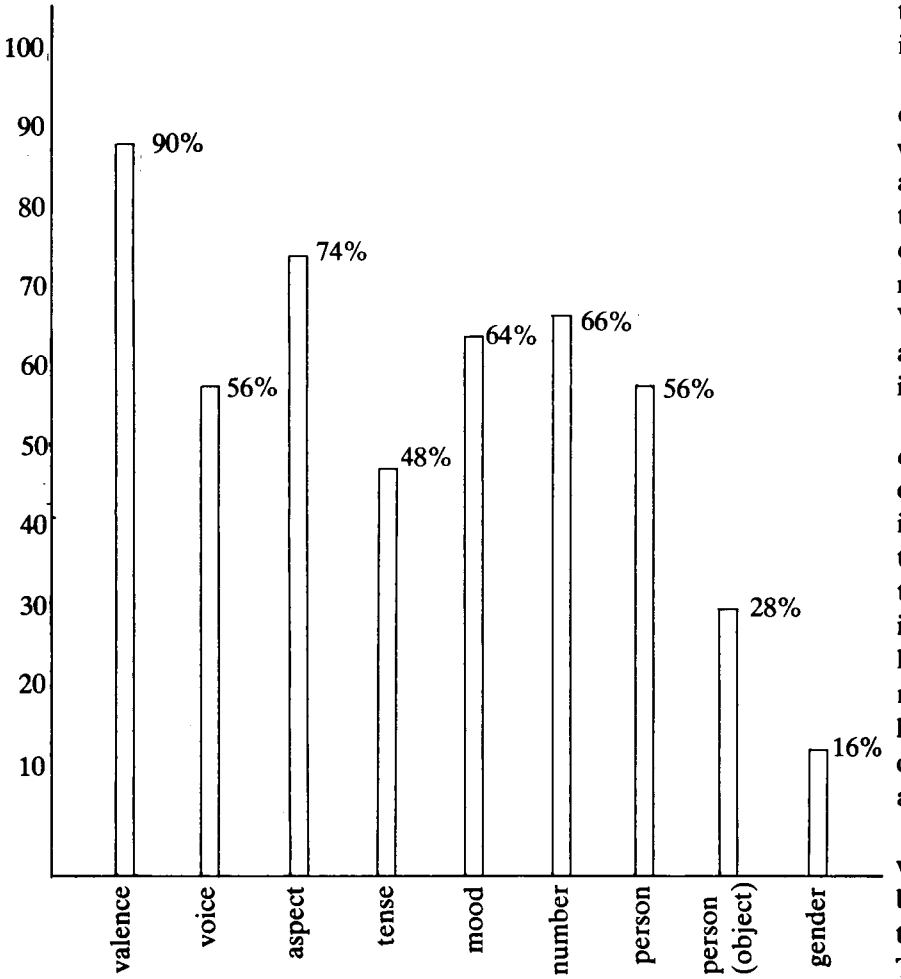

Figure 1. Morphological categories marked on verbs.

The descriptions of the fifty languages were studied and information concerning the verbal morphology was extracted, and coded according to the definitions of categories given above. Figure 1 shows the percentage of languages that have each of the categories as morphological markers on verbs. Note that the markers counted here are not necessarily inflectional, that is, obligatory, but the occurrence of any verbal morphology, whether inflectional or derivational is recorded.

The categories are listed in the order of their relevance to the verb, as established above. The prediction was that the more relevant categories would be more frequent, and there would be a gradual decrease in frequency as relevance decreased. The prediction is upheld in a general way, but important deviations from a simple linear scale point to the need to consider factors other than relevance. Indeed it was not expected that the scale would be any more regular than it is, since the categories differ so much in their functions. What the deviations would be, however, was not always predictable in advance. We will discuss the various surprises as we examine each category individually again.

Perhaps the most striking finding is the near universality of *valence-changing* morphology in the languages of the world. In 90% of the languages of the sample there was evidence for causative, transitivizing or intransitivizing morphology, with causative markers tending to be the most frequent. Of the six languages not included in this 90%, the information about three of them was incomplete, and I suspect that all of these also have valence-changing morphology. There were only two languages which appeared to actually lack valence morphology, and these languages, Haitian Creole and Vietnamese, have very little verbal morphology of any kind. There were six languages which had valence morphology but lacked any other morphological categories for verbs. This means that if a language has any verbal morphology at all, it has valence-changing morphology.

This generalization seems to hold for developing creole languages as well. Mühlhäusler 1980 reports that the first verbal morphology in Tok Pisin, besides the general predicate marker *i-*, which occurs on all verbs, is the transitive verb marker *-im*, which develops into a causative suffix. While Tok Pisin also shows the development of aspect and number morphemes, these morphemes are not bound to the verb. Only the valence morphology is bound to the verb stem.

The centrality of *valence* is evident not only in the frequency with which verbal markers of valence are found, but also in the number of times the

transitive/intransitive distinction is mentioned as a basic organizational dichotomy in the construction of the verb and the clause. It is easy to see why this should be so. One of the most basic manipulations of a situation is a change in the number and role of the participants. Valence changing morphology allows the expression of similarity among situations involving a different set of participant roles, by using the same verb stem, while simultaneously signalling a difference in the situation by adding an affix. When languages do not have valence morphology, they must either use the same unchanged verb stem despite valence changes, as in English *The door opened.* vs. *The clerk opened the door*, or have separate lexical stems for describing similar situations with a different set of participants, as in English *go* vs. *send* or *fall* vs. *drop*. The former solution fails to register the change in the situation in the verb itself, and the latter fails to register the similarity among situations involving different participant roles.

The lower frequency of *voice* categories as verbal markers is probably due to the fact that changes in sentence perspective can be signalled in various ways that do not involve verbal morphology, e.g. by changes in word order, or by markers on nouns. Voice is in the position of having more than the verb stem in its scope. We could say, perhaps that it has the whole proposition in its scope, since it affects the arguments in addition to the verb. However, it differs from mood, which also has the whole proposition in its scope, in that voice can have an effect on the meaning of the main verb of the proposition. That is, an event can be viewed as a different event depending on the perspective, e.g. *buy* vs. *sell*. Perhaps this is a case where relevance and semantic change should be distinguished: voice is less relevant to the verb since it affects the arguments of the verb as well as the verb, but it can produce a meaning change in the verb, which accounts for the possibility of lexicalized voice distinctions.

Aspect morphology is the second most frequent after valence changing morphology. This is to be expected, given the importance of aspect to the verb, and the fact that aspect rarely affects any element in the sentence other than the verb. Note that the languages that were not counted as having morphological aspect might have aspect expressed through auxiliary constructions or other periphrastic means.

Tense, which should be more relevant to the verb, is less frequent as a morphological category in the languages of the world than mood and even the agreement categories of person and number. This is not because tense is expressed in some other way in the sentence. In fact, it is my guess that

tense is not commonly expressed in any way other than by verbal morphology (although this is not something I checked in the survey). Thus it appears that for some reason tense is simply not as common as a grammatical category

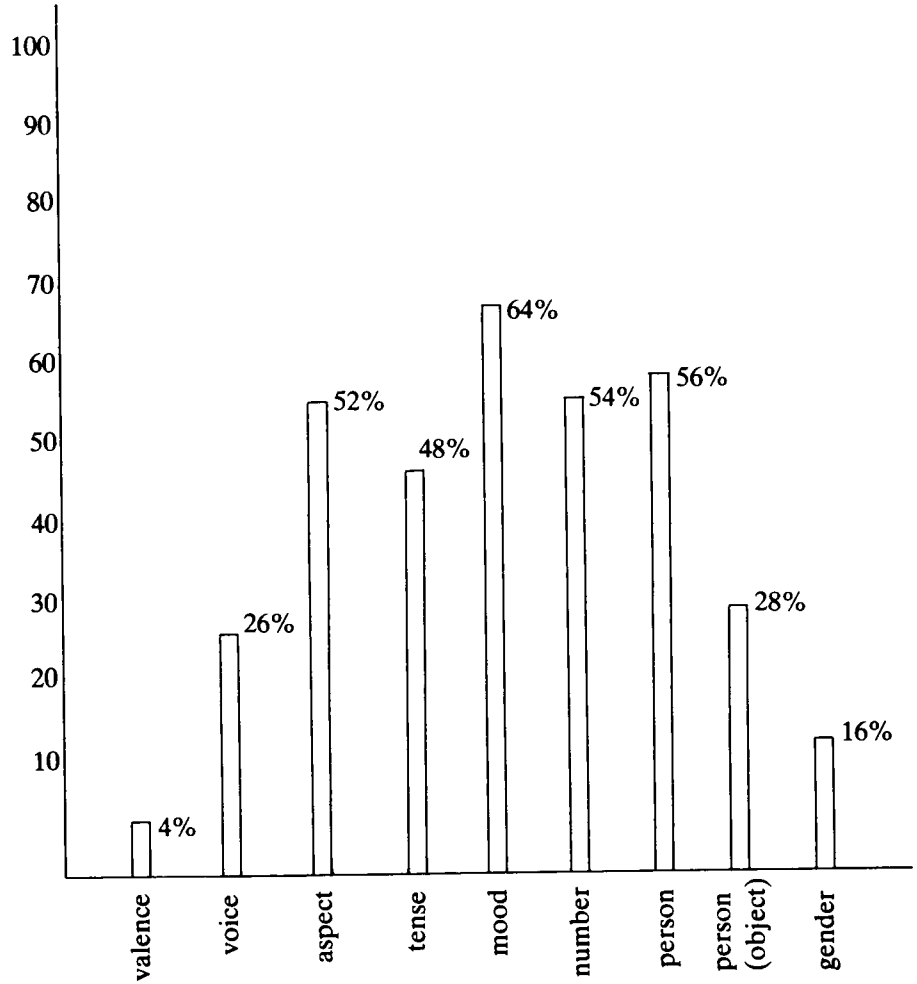

Figure 2. Inflectional categories marked on verbs.

as, for example, mood.

Mood and *number* and *person* agreement with the subject are less frequent than valence and aspect as verbal morphology, as the relevance hypothesis predicts, and agreement with the object, and agreement by gender are even less frequent. Since these categories are almost always inflectional, and rarely derivational or lexical, they will be discussed in conjunction with Figure 2.

Figure 2 shows the percentage of languages that have each category as an *inflectional* category. Recall that in addition to being bound to the verb stem, an inflectional category is defined as one that is *obligatorily* expressed given the grammatical context. While Figure 1 includes both derivational and inflectional expression, Figure 2 includes only inflectional expression. As we said at the end of the last section, we expect the inflectional categories to occur in the middle of the relevance scale, with the likelihood of inflectional expression dropping off at both ends. It drops off at the high relevance pole because of the increased likelihood of derivational or lexical expression, and it drops off at the other end because of the likelihood of periphrastic or syntactic expression. This prediction is nicely supported by the data. Figure 2 shows the predicted bell-shaped curve.

There are differences between Figures 1 and 2 only in valence, voice, aspect and number. These were the only categories that were found to have derivational expression. Valence, voice and aspect were predicted in advance to allow derivational expression, because they are highly relevant to the verb, but number, as an agreement category, would not be expected to occur as a derivational category. This interesting anamoly will be discussed in section 6.

Valence categories hardly ever fit the definition of an inflectional category because there are few cases (if any) of languages in which a specific marker is required on a verb to signal valence, and in which the absence of that marker signals a particular member of the valence category. This rarely occurs because all languages seem to have verbs that are inherently transitive or intransitive. On the other hand, languages that have object agreement marked on the verb have an obligatory expression of valence, but its markers are not uniquely valence markers, since they signal number, person, and gender categories of the object. There are only two languages that were counted as having inflectional valence, Kutenai and Maasai. Both of these languages have benefactive and instrumental markers for verbs. Presumably

these markers occur obligatorily in clauses that have an instrument or a benefactor present, and their absence signals the absence of these arguments. There might still be verbs that do not need these markers, but inherently take instruments or benefactors, in which case the inflectional status of valence would be questionable here, too. In addition, Maasai, a Nilo-Hamitic language, may have other evidence of inflectional valence. Maasai has object agreement. If object agreement is not present on a transitive verb, it is still interpreted as transitive with a 3rd person object. If an intransitive reading is desired, then a suffix must be added to the verb: *arany* "I sing it or them", *adol* "I see it or them", *aranyisho* "I sing", and *adolisho* "I see" (Tucker and Mpaayei 1955). Maasai appears to come very close to having obligatory, bound expression of valence.

Voice is inflectional if there is a general verbal marker for forming non-active voices, and if the absence of this marker necessarily signals active voice. Voice morphemes were counted as derivational in cases such as Diegueño, where two suffixes were described as having a passive-like meaning (be in a state resulting from an action), but were described under the category of "stem-formation" (which usually means derivation). These suffixes appear to be restricted to certain verbs, and further, Langdon points out that in some cases it is not clear whether a form should be analyzed as containing one of these suffixes or not, because the meaning is not transparent enough to be a sufficient clue (Langdon 1970:97).

Where *aspect* is inflectional it usually involves a very general perfective/imperfective distinction, with further distinctions occasionally made in the imperfective. Where it is derivational, it often represents an iterative meaning, with inceptive and durative also occurring. *Tense*, as we said before, is only inflectional and never derivational. The tenses represented in the sample were present, past, future and recent past (or anterior, resultative).

Mood turned out to be the most frequent inflectional category. This is partly due to the high frequency of markers to distinguish imperative from indicative, which occurred in 50% of the languages. It is also related to the large number of contrasts available in the mood category. The following were found to occur in three or more languages of the sample (listed in decreasing order of frequency): imperative, indicative, negative, probable, interrogative, subjunctive, optative, conjunctive, conditional, and dubitative. Further, a single language contrasts up to eight members of the mood category (eight moods were counted in Pawnee and Yukaghir).

The most common type of inflectional system for verbs includes mood and either tense or aspect or both. There were only five languages that have mood but neither tense nor aspect. Among these, however, three have derivational aspect, and in the other two cases the information appears to be incomplete. Languages with tense or aspect and not mood are even rarer. In one of these, the perfective form is used in the imperative and in two cases the information appears to be incomplete. But even when we do not make allowances for derivational aspect and unclear cases, 73% of the languages that have any verbal inflection at all have mood and either tense or aspect as inflectional categories.

Greenberg's 1963 finding that person/number inflection on verbs implies tense, aspect or mood inflection is true of this sample as well. There are no languages that have person or number marking that do not also have either tense, aspect or mood inflections. However, the number of languages that have tense, aspect or mood inflections and do not have person or number agreement is not as great as might be expected. Out of 35 languages that have tense, aspect or mood, only 7 or 20% do not have agreement.[11] It happens, further, that six of these seven languages are SOV languages, and the seventh, Logbara, has SOV word order in imperfective clauses, and SVO in perfective clauses (Crazzolara 1960). Thus it is much more common for a language with tense, aspect or mood to also have person or number agreement than not. In the present sample, all VO languages with tense, aspect or mood inflections also have agreement categories marked on the verb.

3. The order of morphemes

It is often observed that derivational morphemes occur closer to the root to which they attach than inflectional morphemes do. If there is a correspondence between what can be derivational or lexical and its relevance to the root meaning, then we might also expect the degree of relevance in general to predict the order of occurrence of morphemes with respect to a root or stem. More specifically, among the inflectional categories that we have surveyed, we would expect the most relevant to occur closest to the verb stem, and the least relevant to occur at the greatest distance from the verb stem. This type of ordering relation appears to hold for nouns. Greenberg 1963 reports that when both number and case are present on the same side of the noun base, "the expression of number almost always comes between the noun base and the expression of case" (Greenberg 1963:112).

We would interpret this as having a principled basis: namely that the expression of *number* occurs closer to the noun base because it is more relevant to the meaning of the noun. *Number* has a direct effect on the entity or entities referred to by the noun. *Case*, on the other hand, has no effect on what entity is being referred to, but rather, only changes the relation of that same entity to the other entities in the clause.

The prediction concerning the ordering of verbal inflections was tested on the most frequent of the inflectional categories — aspect, tense, mood, and person — in the 50 languages surveyed, and it was found to be a valid prediction with very few exceptions.

Before presenting these results, it is necessary to mention several factors that complicated the test of the ordering hypothesis. First, there are many cases in which it is impossible to discern the relative order of two morphemes because they are fused together in *portmanteau* expression. This was especially true of *aspect* and *tense* morphemes, and of *mood* and *person* morphemes. These cases had no bearing on the test of the hypothesis. Second, in some cases, the two morphemes in question occurred on different sides of the verb stem. These cases were also irrelevant, unless one morpheme occurred adjacent to the stem while the other occurred at least one morpheme removed from it. Then, in these cases, the former was counted as being closer to the stem than the latter. A third situation which rendered a case irrelevant was a situation in which the morphemes in question were mutually exclusive and occurred in the same position. Finally, there were cases in which one morpheme was an affix, but the other was expressed through a modification of the stem, i.e. by reduplication or a vowel change. In these cases, the morpheme expressed by stem modification was counted as occurring closer to the stem than the morpheme expressed by affixation.

The morphemes were examined in pairs to determine their relative order. The results are as follows:

Aspect markers were found to be closer to the stem than *tense* markers in 8 languages, while the opposite order did not occur in the sample. There were a total of 18 languages that have both aspect and tense, but in 10 cases their ordering was not relevant to the hypothesis.

Aspect markers were found to be closer to the stem than *mood* markers in 10 languages, out of a total of 23 that have both aspect and mood. There were no languages in the sample in which the mood marker occurred closer to the stem than the aspect marker.

Aspect markers were found to be closer to the stem than *person* markers in 12 out of 21 languages. In one language, Navaho, the person markers occur closer to the stem than the aspect marker.

Tense markers occur closer to the stem than *mood* markers in 9 languages out of 21 that have both tense and mood. In one languages, Tiwi, the mood markers occur closer to the stem than the tense markers.

Tense markers occur closer to the stem than *person* markers in 8 languages out of the 17 that have both tense and person.

Mood markers occur closer to the stem than *person* markers in 13 languages out of 26. In 5 languages the opposite order occurs.

The position of *number* markers was not tested because in a large majority of languages these markers occur in portmanteau expression with person markers and an ordering of elements is impossible to determine. Thus for the most part, where "person" occurs above, one may read "person and number". This fusion of person and number markers is no doubt due to their diachronic origins as subject (or object) pronouns. We will have more to say below about the diachronic source of the order of morphemes.

The results of this survey give striking confirmation of the hierarchical ordering of aspect, tense, mood and person. The strongest differences are found between aspect and the other categories, and between tense and the other categories, where there are almost no counter-examples to the predicted ordering. The ordering of mood and person is somewhat freer. These results would correspond to the higher relevance of aspect and tense to the verb, and lesser relevance of mood, which has the whole proposition in its scope, and person, which refers to the participants. These results suggest a "diagrammatic" relation between the meanings and their expression, such that the "closer" (more relevant) the meaning of the inflectional morpheme is to the meaning of the verb, the closer its expression unit will occur to the verb stem. This type of diagrammatic relation is also evident in the degree of fusion between the expression of the verb stem and the inflectional morphemes, a topic to which we now turn.

4. Degree of fusion with the stem

If the meaning of an inflectional morpheme is highly relevant to the verb, then it will often be the case that their surface expression units will be

tightly fused, while the less relevant morphemes will have a looser association with the verb stem. This hypothesis can be tested by examining both the effect that the inflectional category has on the surface expression of the stem, and the effect that the stem has on the surface expression of the inflectional category. We are interested here in morpho-phonemic effects that have gone beyond the point of being phonologically conditioned, and are morphologically or lexically conditioned. As examples of cases where the inflectional category has an effect on the verb stem, we will cite languages in which a change in the verb stem is the main signal for an inflectional category or regularly co-occurs with another overt signal of an inflectional category

Aspect conditions changes in the verb stem more frequently than any other inflectional category. In Burushaski and Touareg, vowel and consonant changes in the stem are the primary signals of aspect. In Temiar, reduplication of the stem is the only signal of aspect. In Sierra Miwok and Wappo, stem changes (especially of stress and length in the former language) regularly accompany aspectual suffixes. In Serbo-Croatian, a system of highly fused prefixes and suffixes, accompanied at times by internal stem changes, are the signals of verbal aspect. In Nahuatl, Pawnee, Ojibwa, Zapotec and Navaho there are internal sandhi processes that accompany the affixation of aspectual morphemes. This internal sandhi is often specific to these morphemes, and involves fusion of the affix to the stem by means of consonant and vowel loss or modification.

Stem changes are much less frequent with other categories, but they do occur. Sierra Miwok and Wappo have stem change processes for *tense* that are similar to those for aspect. Nahuatl has stem changes associated with tense in some irregular verbs. As for *mood*, Sierra Miwok has stem changes associated with the Volitional, while Navaho, Pawnee and Ojibwa have internal sandhi associated with the affixation of various mood morphemes. There seem to be no examples in the sample of languages in which the only method of signalling tense or mood is by internal changes in the verb stem.

There are no cases in which simple *number* agreement conditions stem changes as a regular process, but in Acoma and Pawnee there are some verb stems that change in the plural forms. In cases such as Diegueño, where number distinctions by stem change permeate the whole system, number is not so much an agreement category as it is an aspectual one. See the discussion in section 6.

Stem changes with *person* categories are even more rare (Hooper 1979). Acoma has stem changes with non-third person objects in a handful of

verbs, and Navaho and Zapotec have limited internal sandhi with some stems when certain of the person markers are contiguous. Only Maasai has something slightly more spectacular: reduplication of the stem in second person plural of the habitual, and reduplication of the suffix in the same person of the continuous. Further, in second singular and plural, and in first plural, some verbs take an extra nasal after the stem prefix.

There are some languages in the sample that undoubtedly have stem modifications that were not mentioned in the descriptions because the descriptions were brief, e.g. Yukaghir. For that reason, the data presented here are not complete, and are not reliably quantifiable. However, they most likely indicate what would be found in a more complete survey — that stem modifications associated with aspect are about twice as frequent as those associated with other categories.

The effect of the verb stem on the affix, when it is not a purely phonological effect (and perhaps also when it is), may be taken as an additional measure of the degree of fusion of the two elements. Under this heading are cases in which the particular verb stem determines the choice of the allomorph of the inflectional morpheme. For example, in Spanish, the entire verb conjugation system is based on three lexical classes of verb stem — the three conjugation classes. These lexical classes determine the choice of the allomorphs of certain aspects, such as the imperfect, but have no effect on the person or number morphemes. This dependency of the imperfect allomorphy on the verb stem is taken to be an indication of greater fusion.

In the sample, we find lexically-determined allomorphy for *valence* in Ainu, Georgian, Malayalam and Quileute, for *voice* in Nahuatl, Georgian and Quileute, for *aspect* in Serbo-Croatian, Nahuatl and Pawnee, for *tense* only in Malayalam, and for *mood* in Burushaski, Iatmul and Yupik. There are no cases of lexically-determined allomorphy for number or person.

The data, then, seem to support the relevance principle and the hypothesis that the semantic fusion of elements is paralleled in the fusion of expression units. In the case of the effect of the inflectional category on the stem, *aspect* stands out as the category most frequently affecting the stem. In the case of the effect of the stem on the inflectional allomorphy, number and person stand out as the categories most rarely affected by the lexical choice of the verb stem.

5. Explaining the correlations

We have now examined data on the frequency of occurrence of inflectional categories in the languages of the world, the relative order of occurrence of the expression units of these categories within an inflected verb, and the degree of fusion of these expression units with the verb stem. We have found, as predicted earlier, that some categories occur more frequently in the languages of the world, and these same categories tend to occur closer to the verb stem, and exhibit a greater degree of fusion to the stem. These correlations are undeniably strong, but their proposed explanation — that some categories are semantically more relevant to verbs than others — is viable only to the extent that mechanisms can be proposed which suggest how relevance may influence the evolution of inflectional categories. Here we will propose such mechanisms. Since much less is known about the evolution of languages than is known about their synchronic states, this section must of necessity be speculative.

First, it is assumed that inflectional morphemes have their origins in full words that develop a high frequency of use. These frequent items are gradually reduced both phonologically and semantically, and are simultaneously gradually fused, again both phonologically and semantically, with lexical matter contiguous in the syntactic string. The relevance principle predicts that morphemes expressing meanings highly relevant to verbs will be more likely to fuse with verbs than morphemes whose meanings are less relevant. I would claim that there are two reasons for this: first, material that is highly relevant to the verb tends to occur close to the verb in the syntactic string, even before fusion takes place, and second, the psychological restructuring of two words into one depends on the relatedness of the semantic elements being joined, and their ability to form a coherent semantic whole. These two points will be discussed separately.

It seems to be generally true that the order of morphemes within a word reflects an earlier ordering of words within a sentence (Givón 1971, Vennemann 1973). Thus the high frequency of, for example, aspectual inflections, and their proximity to the verb stem, could be traceable solely to the occurrence in earlier times of words expressing aspectual notions in positions contiguous to the verb. This undoubtedly accounts for most morpheme order, but it defers the questions rather than answering it, for we must still explain why words expressing aspectual notions occur close to the main verb. Here we find a wider domain for the relevance principle. As I mentioned at the beginning, it has often been observed that words that function together in the sentence tend to occur together in the sentence. Vennemann cites the "principle of natural constituent structure" proposed by Bartsch, which he

describes as follows:

> This principle says that elements belonging together in the hierarchy of semantic representation tend to be lexicalized and serialized in the surface representation in such a way that hierarchical dependencies are directly reflected in categorial operator-operand relationships

Vennemann illustrates this principle with examples from the ordering of modals and auxiliaries, and the order of elements in a noun phrase. A similar analysis is proposed by Foley and Van Valin 1981 who argue that the ordering of elements in the English auxiliary reflects the increasingly wider scope of the operators. The operator whose scope is primarily the verb (aspect) appears closer to the verb, while the operator whose scope may include the whole proposition (tense) occurs furthest from the verb. If there is a diagrammatic relation between the function of two semantic units and the proximity of their expression units in the clause, then the morphological universals we have discussed here may follow directly from these syntactic principles.

While it is true that a great deal about morphology may be explained by applying the relevance criterion on the level of syntax, we cannot assume that morphology is only fossilized syntax and stop at that. There is a great deal of evidence that speakers actively reanalyze and sometimes restructure their morphological systems, especially during language acquisition. For instance, in Bybee and Brewer 1980 we discuss the restructuring of the preterite in Provençal. In Old Provençal, the segmentation of the preterite forms into clear markers for aspect vs. person and number had become difficult. The only consistent mark of the preterite was the stressed vowel following the verbal root:

canta 'to sing'		*venre* 'to sell'	
cantéi	cantém	vendéi	vendém
cantést	cantétz	vendést	vendétz
cantét	cantéren	vendét	vendéron

Many Provençal dialects restructured these forms by taking a consonant, often the /t/ of the third singular, to be the preterite marker, and adding person/number markers to it (Ronjat 1937:193):

cantéte	cantétem
cantétes	cantétetz
canté	cantéton

In this particular dialect, the third singular form eventually lost its final /t/ due to a regular sound change. However, we can still observe the clear pattern of restructuring, in which /-ét-/ functions as the preterite marker with the person/number markers added after it. It is interesting to note that among all the variations on this restructuring pattern in the many dialects of Provençal, not one added the preterite marker after the person/number markers.

Another interesting example of restructuring that more directly involves the order of morphemes within the verb occurs in Pengo, a Dravidian language (Burrow and Bhattacharya 1970). In Pengo, the past tense has the following conjugation.

Past tense "to see"

	singular		*plural*
1	huṛtaŋ	ex.	huṛtap, incl. huṛtas
2	huṛtay		huṛtader
3m	huṛtan		huṛtar
3f,n	huṛtat	f	huṛtik, n. huṛtiŋ

The perfect was apparently originally formed by the addition of the auxiliary /na/ to the forms of the past tense. In fact, this pattern is still observable occasionally, in forms such as *vātaŋna* 'I have come', *kuccikna* 'they (fem. pl.) have sat down' and *ravtiŋna* '(the rats) have excavated'. However, the more usual conjugation shows forms in the first singular, and in the third feminine and neuter plural in which a person/number marker is added after the perfect marker, with phonological changes in the perfect marker in the third feminine and neuter plural.

Perfect

huṛtaŋnaŋ	huṛtapna, huṛtahna
huṛtayna	huṛtaderna
huṛtanna	huṛtarna
huṛtatna	huṛtiknik, huṛtiŋniŋ

In addition, sometimes the other forms are heard with the person/number suffix added after /na/: *tustannan* '3s has put on', *kuccatanat* 'fem. or neuter sg. has sat down', *temal pantatnat* 'hair has grown long' and *vātapnap* 'we have come'. In a less common paradigm the person/number suffixes occur only once *after* the perfect marker:

huṛtanaŋ	huṛtanap, huṛtanas
huṛtanay	huṛtanader

hurtanan	hurtanar
hurtanat	hurtanik, hurtiniŋ

These examples show that the order of morphemes need not necessarily reflect an earlier order of words, nor the chronological order in which inflectional morphemes develop. (See also Comrie 1980). Cases of reordering of morphemes are not very common, so it will often be the case that morpheme order reflects an earlier order of words, but it is important to recognize that morphology is not immovable fossilized syntax. Speakers will sometimes rework parts of their morphology. Thus the facts that have emerged from the cross-linguistic survey may be interpreted as indicating the existence of universal synchronic principles of linguistic organization. The implementation of these principles, however, must be understood partly in diachronic terms. Thus we have claimed that the order of morphemes is in large part a result of the order of words in the verb phrase, and that the frequency of occurrence of certain categories as verbal inflections is a reflex of their frequent occurrence contiguous to the main verb. We have claimed that the order of words in the verb phrase is at least partly determined by the relevance principle. And this same principle may continue to apply in the active restructuring of morphology that goes on in every generation of language users.

Now we return to the question of whether the frequency of occurrence of categories such as aspect in the languages of the world is merely a reflex of the fact that words expressing aspectual notions often occur contiguous to the main verb. I will claim that the creation of an inflectional category by fusion is not just a mechanical operation that takes place automatically when one word is reduced in the company of another. Rather, the process depends upon the relatedness or relevance of the semantic notions in question, and their ability to form a coherent semantic structure. A reducing morpheme cannot fuse with just any adjacent lexical matter. Its fusion is both phonological and semantic, and the conditions must be right on both levels.

An interesting case that is relevant here is the case of the English auxiliaries, which undergo extreme phonological reduction, attaching themselves to the subject noun or pronoun: *I'll*, *I've*, *I'd*, *I'm*, *he's*, etc. These forms are highly fused phonologically, and yet when children acquire them, they carefully split pronoun from auxiliary, and go through a long stage in which the auxiliaries are produced primarily in their emphatic, whole word forms (Bellugi 1967, Slobin 1973). The fusion of these elements is delayed, or perhaps prevented entirely, by the incompatibility of modifying nominal meanings with tense or aspectual notions. On the other hand, the reduced

form of *have* that follows the modals *should*, *would*, *could*, and *might* has largely lost its identity as the separate aspectual marker *have* for many speakers of English, who, when required to spell this sequence often render *should've* as *should of*, and *would've* as *would of*, etc. Here the *'ve* has come to signal a tense difference, and is well on the way to becoming fused to the modal it follows. The combinability of the tense notion with the modality notions accounts for the possibility of total fusion in this case.[12]

The total fusion of two morphemes into one word, whether it be a lexical and inflectional morpheme or some other combination, depends entirely upon the ability of a generation of language learners to analyze the sequence of morphemes as belonging together in a single word. This means that the sequence must have a meaning that is learnable as a whole. Interestingly enough, the child language literature is full of observations about the very early interpretation of verbs as expressing aspectual notions (Antinucci and Miller 1976, Stephany 1981, Simões and Stoel-Gammon 1979, Bloom et al. 1980), even in languages where aspect is not a part of the inflectional morphology (i.e., in Turkish [Aksu, personal communication], and in Hebrew [Berman, personal communication]). In languages that inflect verbs for aspect as well as person and number, for instance, children mark the aspectual distinctions on verbs long before they mark person/number agreement. It is not that person and number are difficult concepts, because they are mastered in the pronominal system long before they occur on verbs. It is simply the combination of the notions referring to person/number agreement with verbal notions that is more difficult to master. It seems that children exhibit a natural tendency to treat certain notions together. This is a clear manifestation of the relevance principle, and it has an effect on the formation of inflectional morphology.

Consider now the developments in Romance languages, especially Spanish. There is a series of direct and indirect object pronouns which have become clitics and occur in a fixed position right before the finite verb. These pronouns are considered clitics because they are unstressed and do not occur unless the verb is present. They are not considered inflections, however, because they are not obligatory. If full noun phrases for direct or indirect object occur in the sentence, the clitic pronouns need not occur. In other words, the transitive verb is complete without the object pronoun clitics. In another development in Spanish and other Romance languages, the Latin auxiliary verb *habere* in its present and imperfect forms developed into a suffix that marks future tense and conditional mood. These suffixes are bound

to the infinitive, and are an obligatory part of the verb conjugation. If a verb refers to a future activity it must be in the future tense, even if the tense is clear from the context. Incidentally, object pronoun clitics formerly occurred between the infinitive and the form of *habere*. Since the forms of *habere* have become attached the clitic pronouns no longer occur in this position. The clitic pronouns and *habere* are juxtaposed here to suggest that there may be semantic reasons why the formation of inflection has gone to completion where tense and mood concepts are concerned but is delayed where person/number agreement with objects is concerned. Since we have no absolute timetable for the formation of inflection, this case can only be used to illustrate my suggestion, and not as evidence in favor of it.

My conclusion, then, with respect to the frequency of occurrence of inflectional morphemes, as well as their order with respect to the verb stem, is that the relevance principle governs the formation of inflection at every stage. It sets up the syntactic conditions necessary, and in addition governs the likelihood that an actual fusion will eventually take place.

6. Apparent problems — number, negation and object agreement

Number agreement, negation and object agreement would appear to present problems for our hypothesis for the following reason: these categories exhibit properties of both high and low relevance categories. In this section we will find that examining this problem in detail clarifies some of the assumptions behind the previous discussion, and reveals that far from contradicting the relevance principle, the idiosyncracies of these categories lend further support to it. We begin with a discussion of number categories.

Since number marking on verbs is an agreement category referring to the arguments of the verb, and is often fused in expression with person agreement, our hypothesis predicts that number should have little if any effect on the shape of the verb stem and that it should not be found in derivational or lexical expression. However, the survey of 50 languages revealed fourteen languages where number does not behave as predicted. The languages are Acoma, Ainu, Burushaski, Diegueño, Garo, Kiwai, !Kung, Kwakiutl, Maasai, Ojibwa, Pawnee, Sierra Miwok, Tarascan, and Tongan. Six of these languages are North American, but the geographic discontinuity of the others makes it impossible to consider this just an areal phenomenon. In Acoma, Burushaski, Maasai, Ojibwa and Pawnee there are inflectional categories of number agreement for both subject and object, and in addition stem changes accompanying the inflection for a large number

of verbs in Acoma, and for a small number in the other three languages. Diegueno and Kwakiutl do not have obligatory categories for number agreement but signal plural subjects through stem modifications of some irregularity. Ainu, !Kung and Tongan have no real inflectional morphology for verbs at all, and yet have lexical or derivational differentiation of stems for number. Tongan has different stems for a small number of mostly intransitive verbs, but !Kung and Ainu show differentiation according to the number of the object of the transitive verb and the subject of the intransitive for a substantial list of core verbs. In Ainu, some of these are formed with the addition of a suffix *-pa*, as in *ama*, *amapa* "to put or place", *rai*, *raipa* "to die", but there are also a number of suppletive stems, such as *ashte*, *roshki* "to set up", and *raige*, *ronnu* "to kill". In !Kung the singular/plural pairs appear to be morphologically unrelated: *qu* "take (sg. obj.)" and *n/'hwi* "take (pl. obj.)"; *!ei* "die (sg. subj.) and *!ao* "die (pl. subj.)". In !Kung, it appears that the number of the absolutive is a lexical or subcategorization distinction rather than an agreement category.

We have already mentioned that the stem-changing category of plural in Diegueño and Kwakiutl behaves like a derivational category because of the non-obligatoriness, the formal irregularity and the unpredictability of meaning. Pawnee, Sierra Miwok and Tarascan have non-obligatory affixes that signal plurality of the absolutive or object. In Pawnee and Sierra Miwok, the meaning of the derivational affix or process covers plurality of subject or object, but also (and sometimes primarily) iteration of action. In Diegueño distribution of action for objects is included, and in Kwakiutl and Pawnee distribution over space. Some verbs in Diegueño have two plural forms, one distributive and one collective. Consider the following examples (Langdon 1970: 123):

u:cal "he splits it"
u:ca:l "they each split one thing, or he splits it several times" (distributive)
ucəca:l "they (a bunch) each split one, or they (together) split it several times" (collective)

ti:kay "he asks for something"
ti:ka:yp "he is (or they are) a beggar (beggars), i.e., to be in a state resulting form repeatedly asking for things" (possibly a distributive connotation)
təteka:y "they ask for something (or things)" (collective)

In Pawnee the distributive morpheme *wa:* can have the following meanings when prefixed to verbs (Parks 1976:279-280):

/wa: + wiu:s/ *wa:wius* "to defecate here and there"
/wa: + wari/ *wa:wari* "to go about all over"

/ra + wa: + hak/ *rawa:hat* "to pass to (various people)"
/wa: + u/ *wa:?u* "to give (various things)"
/ut...rec + wa:w + is/ *ut — recpa:wis* "to learn things"

About Sierra Miwok, Freeland (1951: 112) says

> Ordinarily, in Miwok there is no expression of plurality in the verb apart from person. The transition between the idea of discontinuous iteration and that of plurality of subject or object, however, is very easy. Many of these iterative verbs that are transitive in meaning convey quite definitely the idea of a plural object.

Some of his examples are the following:

po?a:l- "to slit open"; *po?:al-i:-* "to slit open several"
ma?ta- "to kill"; *ma?:at-i:-* "to kill several"
ha?ta- "to toss"; *ha?:at-i:-* "to toss away repeatedly or several"

These examples demonstrate that when plurality is a derivational (and perhaps also a lexical) category, its meaning extends beyond pure agreement with the arguments of the verb. Plurality of subject or object overlaps with iteration of action. This situation corresponds precisely to the predictions of our hypothesis. When the category has a meaning that more directly affects the lexical content of the verb, that is, is relevant to the event described by the verb, it has an expression that is more highly fused with the verb stem, or it has lexical or derivational expression.

However, a problem still exists because it was claimed earlier that categories on the high end of the relevance scale could have derivational or lexical expression while categories on the low end could not. If number agreement is on the low end of the relevance scale why is it that it is related to meanings that are highly relevant?

This question requires a clarification of what it means to say that the same semantic notion is expressed inflectionally and lexically. I do not mean by this that precisely the same meaning can be expressed in two different ways. There will always be a difference between inflectional and lexical meaning. Inflectional meaning is always very general, it is often redundant in

context, and it is always transparent in the sense that its combination with a stem always produces a predictable meaning. Lexical or derivational meaning sometimes has more content that inflectional, it is often idiosyncratic by lexical item, and it often involves multiple components of meaning. For instance, a few pairs of verbs in English are differentiated according to the plurality of the absolutive argument, e.g. *run* vs. *stampede* and *kill* vs. *massacre*. In both cases, however, there is more to the meaning difference than the fact that the second verb in each case involves plural arguments. Still the number difference is one essential component of the meaning difference in these pairs.

Thus to say that a category can be expressed both inflectionally and lexically is to say that two very closely related, and partially overlapping meanings can be expressed in these two ways. There may be other complicating components of meaning involved as well, as in the cases just cited (i.e. *run* vs. *stampede*), or the semantic correspondence may be very close. A very good example is the inflectional difference in Spanish between Imperfect *sabía* "he knew" on the one hand, and Preterite *supo* which is inceptive and means "he came to know". This distinction is a close as one could possibly get to the distinction in English between the meaning of the lexical items "he knew" and "he found out".

Kurylowicz 1964: 35 argues that there is a correlation of certain inflectional and derivational categories. In particular he says there is a close affinity between aspect and *Aktionsart*; passive voice and derived intransitive verbs; participle and verbal adjective; infinitive and verbal noun; plural and collective (in nouns), etc. I agree that the categories he names cover adjoining and overlapping areas of semantic space. I have claimed further that certain inflectional categories do *not* correlate with or adjoin any derivational categories: in particular, mood, tense, and agreement. These categories, by their meanings, are inherently less relevant to the lexical content of verbs, and thus do not overlap with derivational meaning. Number is different, however, in the following sense. While number agreement itself is not highly relevant to verbs, it overlaps with distributive and iterative meanings in cases where multiple involved participants imply repeated action. This gives number the possibility of exhibiting expression properties of a highly relevant category such as aspect.

The second apparent problem is with *negation*. I have classified negation as a *mood* because it roughly fits the definition of mood as expressing "what the speaker wants to do with the proposition in the particular discourse".

However, it is clear that negation is different from other members of the category *mood* in that it has a quite direct effect on the meaning of the proposition and does not particularly serve to establish the discourse function of the proposition. Negation further differs from other moods in that it may have varying scope: at times the whole proposition is in the scope of negation, and at other times only a portion of the proposition. For these reasons it is not surprising that negation sometimes has lexical or derivational expression in addition to inflectional. For instance, in Korean, the verb *cota* "to like" contrasts lexically with the verb *silta* "to dislike", and *itta* "to exist, be present" contrasts lexically with *optta* "to not exist, be absent". Further, in some languages there are derivational affixes of negation for verbs. These usually occur on stative or adjectival predicates, but they do nonetheless exist.

The third apparent problem is not so easy to dispatch. Object agreement is anamolous with respect to our hypothesis because it is less frequent in the languages of the world than subject agreement, (indicating less relevance), but occurs closer to the verb stem than subject agreement in eight languages, with the opposite order occuring in only three languages, (indicating more relevance). In other cases we found that the frequency of occurrence of a category in verbal morphology correlates with our predictions about the relevance of that category to a verb, and that the order in which the marker of a category occurs with respect to the stem also correlates with predictions about the relevance of the category.

Object agreement for person, number and noun class does not itself directly affect the lexical content of a verb (i.e. is not highly relevant). However, like number agreement, object agreement is related semantically to a category that does have high relevance for the meaning of the verb, i.e., valence. Valence-changing categories deal primarily with the presence vs. absence of an object, and occasionally with properties of that object, such as animacy. Indeed, in a language such as Nahuatl, the structure of the word is rather different for transitive and intransitive verbs owing to the presence vs. absence of object agreement markers. Despite this affinity to valence, evidence that object agreement is highly relevant is lacking. Object agreement never occurs closer to the verb stem than aspectual markers, and occurs closer to the stem than a tense marker in only one language in the sample, and occurs closer than mood in only three languages (cf. subject agreement, which occurs closer than mood in five languages).

Given that object agreement is not totally misplaced on the relevance scale, the question of the relative relevance of subject and object agreement still remains. If subjects are more relevant to the verb, why does the object marker tend to occur closer to the verb stem? If objects are more relevant to the verb, why is subject agreement a more frequent inflectional category? An answer to the latter questions suggests itself more readily than an answer to the former. Besides, the affinity of verbs and their object is well known: in a majority of the languages of the world, the object occurs adjacent to the verb in the basic word order, and forms a constituent with the verb.[12] So it is the higher frequency of subject agreement that has to be explained. There are two factors that might be considered. One is the fact that objects do not occur in as many sentences as subjects do. Thus subject pronouns, which are usually the source of subject agreement markers could more easily become obligatory than object pronouns, which occur only in transitive sentences. The other factor is that subjects are more often topical than objects, and are perhaps more frequently pronominalized. Thus subject pronouns might be more frequent than object pronouns, which would facilitate their reduction to bound morphology.

7. Conclusions

The last example illustrates the multiplicity of factors that will have to be invoked if we attempt a full explanation of why inflectional categories represent the meanings they do, and why they occur where they do. Relevance of the grammatical meaning to the lexical stem has been proposed here as one important factor, and the amount of semantic change brought about by the combination of the two units has also been mentioned. In Bybee 1985, the general applicability of a semantic unit to a wide range of lexical stems is also suggested as a factor that determines the possibility of inflectional expression. This is similar to the explanation just proposed for the higher frequency of subject over object agreement. But if we want to explain why we find valence, voice, aspect, tense, mood and agreement categories marked on verbs rather than the scores of other logical possibilities we will have to investigate a wider range of discourse and cognitive factors for answers.

In 1921, Edward Sapir proposed a classification of concepts which distinguished between "concrete" or lexical concepts and "relational" or grammatical concepts. Among the latter he distinguished three degrees of abstractness that correlated with different types of expression. In his discussion, however, he took care to emphasize that any particular category might appear at any point on the scale in a particular language.

> It is because our conceptual schema is a sliding scale rather than a philosophical analysis of experience that we cannot say in advance just where to put a given concept. We must dispense, in other words, with a well-ordered classification of categories. What boots it to put tense and mode here or number there when the next language one handles puts tense a peg 'lower down' ..., mode and number a peg 'higher up'...? (1921:107)

When the facts discussed in this paper are considered, we must conclude that Sapir was wrong. One and the same concept cannot slide all over the scale. While it is true that, for instance, *aspect* can be expressed in many different ways, i.e., lexically, derivationally, inflectionally and periphrastically, this is certainly not true of *mood* or *tense*, which cannot have lexical or derivational expression. And when *aspect* is expressed inflectionally it exhibits certain uniform properties across languages, especially in its behavior with respect to other categories. I would submit, then, that there is a relation, definable across languages, between the content of categories and their means of expression. Furthermore, when the relation among the expression units of various categories are considered, it is found that these relations are diagrammatic for the relations among the units of content.

NOTES

1) For example, by Vennemann 1973 and elsewhere. See also Behaghel 1923-1932.

2) Other examples of diagrammatic iconicity in morphology and morphophonemics are discussed in Andersen 1980, Hooper 1979, and Bybee and Brewer 1980.

3) A morpheme is considered *bound* if it is inseparable from the stem, and occurs in a fixed order with respect to the stem, with only closed class items intervening between it and the stem. In considering the descriptions of the fifty languages in the sample, if this information was not given explicitly, then the author's decision to write the morpheme as separate or bound was considered adequate indication of its status.

4) A more detailed description of the criteria used in the survey, as well as an extensive discussion of the differences between derivational and inflectional morphology will be found in Bybee 1985.

5) Honorifics were found to occur as verbal inflection in only one language of the sample, and will not be considered here.

6) I approached this task with a great fear of circularity. Since the definitions of the categories I used are rather standard ones, it is possible that the reason I found they worked for the languages I investigated might be because the authors of the descriptions had also started out with these definitions, and described the languages in such a way that the definitions worked. Then when I set out to study these descriptions to see if the standard definitions worked, I would find, sure enough, that they do. This fear was dispelled soon enough in the actual investigation. For a large majority of the descriptions I used, it was evident that the author had a strong sense of the uniqueness of the language he or she was describing, and was at pains to show how different it was from the more familiar languages, often eschewing conventional labels for categories, and inventing new ones. Since I used the descriptions of the *functions* of the categories the authors gave, rather than the labels they put on them, the survey is not biased by this tendency, nor by the tendency of authors to mislabel categories such as aspect and tense.

7) The term *syntactic expression* will be used for periphrastic expressions involving more than one lexical unit, such as *cause to die*, so that the term *lexical expression* can be reserved for concepts expressed in a single lexical unit.

8) An interesting exception is the Finnish genitive vs. partitive distinction. An object marked in the partitive gives an imperfect reading to a sentence, while a genitive gives a perfective reading. Interestingly enough, young children acquiring Finnish do not learn this as an aspectual distinction at first, but rather as a distinction applying to the noun (Melissa Bowerman, personal communication).

9) Languages with tense markers on nouns are rare. Sierra Miwok uses nominalizations extensively and can formulate a past tense proposition, such as "I danced" as a nominal expression translatable as "I was a former dancer". The English prefix *ex-* might be considered the nominal version of a past tense marker.

10) Modalities were originally considered under mood. The epistemic modalities showed up frequently as inflections, but deontic modalities that qualified as inflections were very rare, and perhaps non-existent.

11) Here I leave Gilyak out of consideration because it has agreement markers only on non-finite verb forms. Since Gilyak is SOV, whether we interpret this as agreement or not, it does not affect the generalization.

12) Possible counter-examples to this hypothesis are a few American Indian languages in which agreement markers have different allomorphs according to mood.

13) The ordering of subject and object agreement, however, cannot be attributed only to earlier ordering of subject, object and verb in the clause, because in our sample, seven languages ordered their agreement markers according to the three most common word order, SOV, SVO, and VSO, but five used the rarer patterns, VOS, OVS, and OSV.

REFERENCES

Andersen, Henning. 1980. Morphological change: towards a typology. In Fisiak (ed.) *Historical morphology*. The Hague: Mouton.

Antinucci, Francesco and Ruth Miller. 1976. How children learn to talk about what happened. *Journal of child language* 3.169-189.

Barber, E. J. W. 1975. Voice — beyond the passive. *BLS* 1.16-24.

Behaghel, Otto. 1923-1932. *Deutsche Syntax*. Volume 4: Wortstellung, Periodenbau. Heidelberg: Carl Winter.

Bellugi, Ursula. 1967. *The acquisition of negation*. Harvard Doctoral Dissertation.

Berman, Ruth Aronson. 1978. *Modern Hebrew structure*. Tel-Aviv: University Publishing Projects, Ltd.

Bloom, Lois, Karin Lifter and Jeremy Hafitz. 1980. Semantics of verbs and the development of verb inflection in child language. *Language* 56.386-412.

Burrow, T. and S. Bhattacharya. 1970. *The Pengo language*. Oxford: Clarendon.

Bybee, Joan L. 1985. *Morphology: a study of the relation between meaning and form*. Amsterdam: John Benjamins.

Bybee, Joan L. and Mary Alexandra Brewer. 1980. Explanation in morphophonemics: changes in Provençal and Spanish preterite forms. *Lingua* 52.271-312.

Comrie, Bernard. 1976. *Aspect*. Cambridge: Cambridge University Press.

-----. 1980. Morphology and word order reconstruction: problems and prospects. In Fisiak (ed.) *Historical Morphology*. The Hague: Mouton, pp. 83-96.

Foley, William and Robert Van Valin. 1981. *A semantic explanation for syntactic order in the English verb phrase*. Paper presented at the Annual Meeting of the LSA, New York City.

Givón, Talmy. 1971. Historical syntax and synchronic morphology: an archaeologist's field trip. *CLS* 7.

-----. 1979. *On understanding grammar*. New York: Academic Press.

Greenberg, Joseph. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Greenberg (ed.) *Universals of language*. Cambridge, Mass: The MIT Press.

Hooper, Joan Bybee. 1979. Child morphology and morphophonemic change. *Linguistics* 17.21-50. Also in J. Fisiak (ed.) *Historical Morphology*. The Hague: Mouton, 157-187.

Hopper, Paul J. 1977. Observations on the typology of focus and aspect in narrative language. *NUSA* 4. Jakarta. Reprinted in *Studies in Language* 3.37-64, 1979.

Kurylowicz, Jerzy. 1964. *The inflectional categories of Indo-European*. Heidelberg.

Mülhäusler, Peter. 1980. Structural expansion and the process of creolization. In Valdman and Highfield (eds.) *Theoretical contributions in creole studies*. New York: Academic Press.

Perkins, Revere D. 1980. *The evolution of culture and grammar*. Ph. D. Thesis: SUNY/Buffalo.

Ronjat, J. 1937. *Grammaire historique des parlers provencaux modernes*. Vol. 3. Montpellier: Societe des langues romanes.

Sapir, Edward. 1921. *Language*. New York: Harcourt, Brace and World.

Simões, Maria C. P. and Carol Stoel-Gammon. 1979. The acquisition of inflections in Portuguese: a study of the development of person markers on verbs. *Journal of child language* 6.53-67.

Slobin, Dan I. 1973. Cognitive prerequisites for the development of grammar. In Ferguson and Slobin (eds.) *Studies of child language development*. New York: Holt, Rinehart and Winston, 175-208.

Stephany, Ursula. 1981. Verbal grammar in Modern Greek early child language. In Philip S. Dale and David Ingram (eds.) *Child language: an international perspective*. Baltimore: University Park Press.

Vennemann, Theo. 1973. Explanation in syntax. In John Kimball (ed.) *Syntax and Semantics II*. New York: Academic Press.

LANGUAGE REFERENCES

ACOMA (Keres)

Miller, Wick R. 1965. *Acoma grammar and texts*. Berkeley: University of California Press.

AINU (undetermined)

Batchelor, John O. B. E. 1938. *A grammar of the Ainu language*. Memoir V. Tokyo.

Simeon, George John. 1968. *The phonemics and morphology of Hokkaido Ainu*. University of Southern California Dissertation.

ANDAMANESE (Andamanese)

Basu, D. N. 1952. A general note on the Andamanese languages. *Indian Linguistics* 16.214-225.

-----. 1955. A linguistic introduction to Andamenese. *Bulletin of the Department of Anthropology* 1.55-70. Government of India Press.

APINAYE (Ge-Pano-Carib)

Callow, John. 1962. *The Apinaye language: phonology and grammar*. London University Dissertation.

BASQUE (undetermined)

N'Diaye-Correard, Genevieve. 1970. *Structure de dialecte basque de Maya*. The Hague: Mouton.

BURUSHASKI (undetermined)

Lorimer, David L. R. 1935-1938. *The Burushaski language*. Three volumes. Cambridge, Mass: Harvard University Press.
CAR (undetermined)
Braine, Jean Critchfield. 1970. *Nicobarese grammar — Car dialect*. University of California at Berkeley Dissertation.
DIEGUEÑO (Hokan)
Langdon, Margaret. 1970. *A grammar of Diegueño: the Mesa Grande dialect*. Berkeley: University of California Press.
GARO (Sino-Tibetan)
Burling, Robbins. 1961. A Garo grammar. *Indian Linguistics, Monograph series 21*. Poona, India: Linguistic Society of India.
GEORGIAN (Caucasian)
Vogt, Hans. 1971. *Grammaire de la langue georgienne*. Oslo, Norway: Universitetsforlaget.
GILYAK (undetermined)
Austerlitz, Robert. 1958. Vocatif et imperatif en ghiliak. *Orbis* 7, 477-481.
-----. 1959. Semantic components of pronoun systems: Gilyak. *Word 15*. 102-109.
Jakobson, Roman. 1957. Notes on Gilyak. *In Studies Presented to Yuen Ren Chao*. The Bulletin of the Institute of History and Philology, Academia Sinica.
GOAJIRO (Arawakan)
Holmer, Nils M. 1949. Goajiro. *IJAL* XV. 45-56, 110-120, 145-157, 232-235.
HAITIAN (none)
Hall, Robert A., Jr. et al. 1970. Haitian Creole: grammar, texts and vocabulary. *Memoirs of the American Folklore Society*, v. 43. Mamoroneck, NY: Kraus Reprint Corp.
IATMUL (Central New Guinea)
Staalsen, Philip. 1969. The dialects of Iatmul. *Pacific Linguistics A*. 22.69-84.
-----. 1972. Clause relationships in Iatmul. *Pacific Linguistics A*. 31.-45-69.
KARANKAWA (undetermined)
Gatschet, Albert S. 1891. The Karankawa Indians. *Peabody Museum of American Ethnology: Archaeological and Ethnological Papers*, vol. 1, no. 2. Harvard University.
KHASI (undetermined)
Rabel, Lili. 1961. *Khasi, a language of Assam*. Louisiana State University, Baton Rouge.
KHMER (Mon-Khmer)

Huffman, Franklin E. 1967. *An outline of Cambodian grammar*. Cornell University Dissertation.
KIWAI (Kiwai)
Ray, Sidney H. 1933. *A grammar of the Kiwai language, Fly Delta, with a Kiwai vocabulary by E. B. Riley*. Port Moresby, Australia: Government Printer.
KOREAN (Altaic)
Martin, Samuel E. 1960. Korean reference grammar. *Research and Studies in Uralic and Altaic Languages, Project no. 19*. Bell and Howell, Cleveland.
Ramstedt, G. J. 1939. *A Korean grammar*. Helsinka: Suomalais-Ugrilainen Seura.
!KUNG (Khoisan)
Snyman, J. W. 1970. An introduction to !Xu (!Kung) language. *University of Cape Town, School of African Studies, Communications* no. 34. Balkema, Cape Town, South Africa.
KUTENAI (undetermined)
Garvin, Paul L. 1948, 1951. Kutenai, *IJAL* XIV. 37-47, 87-90, 171-87, XVII. 84-97.
KWAKIUTL (Wakashan)
Boas, Franz. 1947. Kwakiutl grammar, with a glossary of the suffixes. *Transactions of the American Philosophical Society*, vol 37, part 3.
LOGBARA (Nilo-Saharan)
Crazzolara, J. P. 1960. *A Study of the Logbara (Ma'di) language*. London: Oxford University Press.
MAASAI (Nilo-Hamitic)
Tucker, A. N. and J. T. Mpaayei. 1955. *A Maasai grammar*. London: Longmans.
MALAYALAM (Dravidian)
Andrewskutty, A. P. 1971. *Malayalam: an intensive course*. Trivandrum, India: Department of Linguistics, University of Kerala.
George, K. M. 1971. *Malayalam grammar and reader*. Kottayam, India: National Book Stall.
MIWOK (Penutian)
Freeland, L. S. 1951. Language of the Sierra Miwok. Memoirs of *IJAL* or *Indiana University Publications in Anthropology and Linguistics* VI.
NAHUATL (Aztec Tanoan)
Andrew, J. Richard. 1975. *Introduction to Classical Nahuatl*. Austin: Univer-

sity of Texas Press.
NAVAHO (WESTERN APACHE) (Na-Dene)
Edgerton, Faye E. 1963. The tagmemic analysis of sentence structure in Western Apache. *Studies in Athapascan Languages*. 102-48. Berkeley: University of California Press.
Hoijer, Harry. 1945-46. The Apachean verb. *IJAL* XI.193-203. XII.1-13, 51-59.
OJIBWA (Macro-Algonquin)
Bloomfield, Leonard. 1957. *Eastern Ojibwa: Grammatical sketch, texts and word list*. Ann Arbor: University of Michigan Press.
Todd, Evelyn Mary. 1970. *A grammar of the Ojibwa languages: the Severin dialect*. University of North Carolina Dissertation.
PALAUNG (Palaung-Wa)
Milne, Mary Lewis. 1921. *An elementary Palaung grammar*. Oxford: The Clarendon Press.
PAWNEE (Macro-Siouan)
Parks, Douglas R. 1976. A grammar of Pawnee. *Garland Studies in American Indian Linguistics*. New York: Garland Publishing, Inc.
QUILEUTE (Chimikuan)
Andrade, M. J. 1933. Quileute. *In Handbook of American Indian Languages III*. F. Boas, ed. New York: Columbia University Press.
SANTA CRUZ (Papuan)
Wurm, S. A. 1969. The linguistic situation in the Reef and Santa Cruz Islands. *Pacific Linguistics A*. 21.47-105.
SENOI (Semang)
Carey, Iskander. 1961. *Tengleq Kui Serok*. Kuala Lumpur: Dewan Bahasa Dan Pustaka.
SERBIAN (Indo-European)
Partridge, Monica. 1972. *Serbo-Croatian practical grammar and reader*. Belgrade.
SONGHAI (undetermined)
Prost, André. 1956. La langue Sonay et ses dialectes. *Memoires, no.* 47. Dakar, Senegal: Institut Donamental D'Afrique Noire.
SUSU (Niger Congo)
Sangster, Linda and Emmanuel Faber. 1968. *Susu basic course*. Bloomington: Indiana University Intensive Language Training Center.
-----. 1969. *Susu intermediate course*. Bloomington: Indiana University Intensive Language Training Center.

TARASCAN (undetermined)
Foster, Mary Lecron. 1969. *The Tarascan language*. Berkeley: University of California Press.
TASMANIAN (undetermined)
Schmidt, V. W. 1952. *Die Tasmanischen Sprachen: Quellen, Gruppierunger, Grammatik, Worterbuch*. Utrecht-Anvers: Publication 8, Commission d'Enquete Linguistique, CIPL.
TIMUCUA (undetermined)
Granberry, Julian. 1956. Timucua I. *IJAL* XXII. 97-105.
TIWI (Australian)
Osborne, C.R. 1974. *The Tiwi language*. Canberra: Australian Institute of Aboriginal Studies.
TOBELORESE (West Papuan)
Hueting, A. 1936. Iets over de Spraakkunst van de Toberloreesche-taal. *Bijdragen Tot De Taal-Land-en Volkenkunde*. 94.295-407.
TONGAN (Austronesian)
Churchward, Clark Maxwell. 1953. *Tongan grammar*. London: Oxford University Press.
TUAREG (Afro-Asiatic)
Prasse, Karl G. 1972-1975. *Manuel de grammaire touregue* (Tahaggart). Three vols. Editions de l'université de copenhage.
VIETNAMESE (Vietmanese)
Thompson, Laurence C. 1965. *A Vietnamese grammar*. Seattle: University of Washington Press.
WAPPO (Yuki)
Thompson, Sandra A. *Personal communication*.
YANOMANO (Micro-Chibchan)
Migliazza, Ernest Cesar. 1972. *Yanomama grammar and intelligibility*. Indiana University Dissertation.
YUKAGHIR (undetermined)
Jochelson, Vladimir I. 1905. Essay on the grammar of the Yukaghir language. *American Anthropologist 7*. 369-424.
YUPIK (ST. LAWRENCE ISLAND)
Jacobson, Steven A. 1977. *A grammatical sketch of Siberian Yupik Eskimo as spoken on St. Lawrence Island, Alaska*. Fairbanks: Alaska Native Language Center, University of Alaska.
ZAPOTEC (Oto-Manguen)
Pickett, Velma Bernice. 1953. Isthmus Zapotec verb analysis I. *IJAL* 19.292-296. Isthmus Zapotec verb analysis II. *IJAL* 21.217-232.