CHAPTER 2:
SEMANTIC DETERMINANTS OF INFLECTIONAL EXPRESSION

1. *Lexical, inflectional and syntactic expression.*

This chapter proposes a set of principles that contribute to an eventual answer to the question of what are the possible inflectional categories in the languages of the world. The hypothesis is set in the larger framework that includes a consideration of what semantic notions can be expressed lexically and syntactically in the languages of the world, although no complete answer to any of these questions is proposed. The proposed principles do, however, make predictions about the behavior of inflectional categories: their frequency of occurrence in the languages of the world, their order of occurrence with respect to a stem, and the morpho-phonemic effect they have on the stem.

In order to state the hypothesis, it is necessary first to define three major ways in which semantic elements may be combined into expression units.

(a) First, two or more semantic elements may be expressed in a single monomorphemic lexical item. This is *lexical* expression. For instance, the lexical item *kill* at some level of analysis combines the semantic elements of 'die' and 'cause'; *drop* combines 'fall' and 'cause'.

(b) In *inflectional* expression, each semantic element is expressed in an individual unit, but these units are bound into a single word. Inflectional expression may be in the form of affixes added to a stem, or in the form of a change in the stem itself, e.g. English regular Past Tense, *walked*, vs. irregular *brought*. Inflectional expression is by definition very general. A morphological category is inflectional if some member of the category obligatorily accompanies the radical element when it occurs in a finite clause. Thus, an inflectional category must be combinable with any stem with the proper syntactic and semantic features, yielding a predictable meaning.

(c) In *syntactic* expression the different semantic elements are expressed by totally separable and independent units, that is, in separate words. Thus *come to know* is the syntactic expression of 'inchoative' and 'know', while

realize is the lexical expression of the same notions. This type of expression is also often termed "periphrastic" expression.

These three expression types do not constitute discrete categories, but rather mark off areas on a continuum. Intermediate expression types also exist, and are important to the general hypothesis. Between lexical and inflectional expression lies *derivational* morphology. Derivational expression resembles lexical expression in that derivational morphemes are often restricted in applicability and idiosyncratic in formation or meaning. It resembles inflectional expression in that two distinct morphemes are combined in a single word. The hypothesis to be outlined here will contribute to a better understanding of the inflectional/derivational distinction, which is discussed in detail in Chapter 4.

Between inflectional expression and syntactic expression are various types of units that have properties of grammatical morphemes, that is, they belong to a closed class and occur in a fixed position, but which are not *bound* to any lexical item, and thus are not inflections. These units are variously named *clitics*, *particles* or *auxiliaries*. Examples are the cliticized object pronouns of Spanish or French, or the modal auxiliaries of English (*may, can, will*, etc.). These *free grammatical morphemes* resemble inflections in that they make up contrast sets that are obligatory in certain environments, and they have positional restrictions. They resemble periphrastic expressions in that they are not *bound* to lexical stems. Free grammatical morphemes are not studied in the present work, although their existence is recognized by the theory to be proposed.

These expression types form a continuum that ranges from the most highly fused means of expression, lexical expression, to the most loosely joined means of expression, syntactic or periphrastic expression:

lexical - - - derivational - - - inflectional - - - free grammatical - - - syntactic

< ----------

greater degree of fusion

Given these basic types of expression, the question this chapter considers is whether there are any generalizations to be made concerning the *content* of the semantic elements that may be expressed in each of these ways. Since there seem to be few constraints on what may be expressed syntactically, it is more interesting to focus attention on what may be expressed lexically and inflectionally. In fact, the hypothesis of this chapter is concerned primarily with the most constrained expression type — inflectional expression.

When one looks around casually at the languages of the world, one is struck by the fact that the same or very similar inflectional categories appear in one language after another. Verbal systems quite commonly inflect for aspect and/or tense, mood, and person and number agreement with the subject. Somewhat less frequently, it seems, one finds inflectional causatives, negation, voice, and object agreement. For nouns, number, gender or other types of classifiers, case and sometimes deixis are expressed inflectionally. These lists are not exhaustive, of course, but the fact that it is possible to come up with a relatively short list of semantic elements often expressed inflectionally indicates that there must be some general principles governing inflectional expression. The remainder of this chapter is devoted to explicating two of these principles, *relevance* and *generality*.

2. *Determinants of inflectional expression*

2.1. *Relevance*

A meaning element is *relevant* to another meaning element *if the semantic content of the first directly affects or modifies the semantic content of the second*. If two meaning elements are, by their content, highly relevant to one another, then it is predicted that they may have lexical or inflectional expression, but if they are irrevelant to one another, then their combination will be restricted to syntactic expression. Notice that the hypothesis does not predict that a semantic element *must* have lexical or inflectional expression in a particular language, it only delineates the set of elements that *may* have lexical or inflectional expression. Let us consider some examples. In English we have a lexical item *walk* which means "to go on foot by taking steps". If we add the meaning "through water", we can express both meanings together in one verb *wade*. These two semantic notions may be expressed together in a single lexical item because whether one has one's feet on dry land or in water is quite relevant to the act of walking. In contrast, whether the sky is sunny or cloudy is not usually relevant to the act of walking, so the hypothesis predicts that languages will not have separate lexical items for "walk on a sunny day" vs. "walk on a cloudy day". The latter two combinations would be more likely to have syntactic expression.

Relevance depends on cognitive and cultural salience: no matter to what extent an entity, event or quality is decomposable into semantic features, if it is perceived as discrete from surrounding entities, events or qualities, it can have a lexical item applied to it. So two semantic elements are highly relevant to one another if the result of their combination names something

that has high cultural or cognitive salience. If cognition shapes both language and culture, and cultural and linguistic variation is patterned and constrained by cognition, then, we would expect lexicalization patterns to differ across languages, but in a principled and patterned way. This point is illustrated for motion verbs by Leonard Talmy (1980), who identifies three major lexicalization patterns for motion verbs in the languages of the world. In the type familiar from English, but which also occurs in Chinese and Caddo (an Amerindian language), an indication of the *manner* in which the motion takes place is characteristically included in the meaning of a verb. Thus we have motion verbs such as *walk, swim, fly, slide, roll, swirl, bounce, jump* and many more. Romance languages exemplify the type in which the *path* of the motion is included in the meaning of the verb. This pattern is also found in Semitic, Polynesian and Nez Perce. For example, Spanish has motion verbs such as *entrar* 'to go in', *salir* 'to go out', *bajar* 'to go down', *subir* 'to go up', *pasar* 'to go by', *volver* 'to go back', etc. The third, perhaps least common type is found in northern Hokan languages, in Navajo, and in American Sign Language (McDonald 1982). In this pattern, verbs of motion are not distinguished from verbs of location, but rather both are expressed by verb stems whose meaning specifies the kind of object or material being moved or located. Partially comparable English verbs would be *to rain*, *to spit*, or *to drip*, which give an indication of the type of material that is in motion. These three patterns illustrate a range of variation, yet all within the boundaries laid out by the requirement that semantic elements lexicalized together be relevant to one another. Manner and direction are inherently related to motion, and types of motions can be distinguished on these parameters. Shape or kind of object or material is not as exclusively relevant to motion, and consequently not restricted to motion verbs in the third language type, but applicable also to verbs of location.

In order to apply the relevance criterion to inflectional categories, it is necessary to refer to a distinction made by Sapir 1921 between *material* and *relational* content. Sapir was also interested in the relation between content and expression, and he distinguished two types of concepts: basic, material or concrete concepts "such as objects, actions (and) qualities" which are "normally expressed by independent words and radical elements (and) involve no relation as such", and less concrete or relational concepts which may be expressed by affixes or changes internal to a root (Sapir 1921: 101). In discussing the relevance of inflectional categories, we will be referring to the relevance of the relational concept (expressed by an affix or internal change) to a material concept (expressed by the radical element).

Among inflectional categories, we can distinguish degrees of relevance of the concept expressed inflectionally to the concept expressed by a radical element, in this case a verb stem. A category is *relevant* to the verb to the extent that *the meaning of the category directly affects the lexical content of the verb stem*. For instance, let us compare aspect to person agreement with the subject. Aspect represents different ways of viewing the internal temporal constituency of an action or state (Comrie 1976:3). Since a verb stem describes an action or state, aspect is highly relevant for verbs. Subject agreement is somewhat less relevant to the verb, since it refers to an argument of the verb, and not to the action or state described by the verb itself. Two general predictions emerge from this difference between aspect and subject marking.

The first concerns lexical expression: we expect to find lexical expression of aspect as a common phenomenon, but lexical expression of subject agreement by person much more rarely. This prediction seems to be correct. English has lexical pairs that differ only by inchoative aspect, such as *know* vs. *realize*, and also pairs that differ by completive aspect, such as *do* vs *complete*. Latin had aspectual distinctions expressed by derivational morphology, such as the inchoative in *amō* 'I love' vs. *amascō* 'I begin to love' *caleō* 'I am warm' vs. *calescō* 'I get warm' and *dormiō* 'I sleep' vs. *obdormiscō* 'I fall asleep'. Examples of lexical aspect are so common that Lyons (1977:705) asserts that there are few languages that do not have aspect. Lexical (or derivational) expression of person agreement, however, is extremely rare. English has a single example: in the verb *to be*, 1st and 3rd singular are distinct in *am* and *is*. In this case, however, the lexical distinction is the residue of an older inflectional distinction. Examples of the lexical or derivational expression of person agreement on the verb independent of a diachronic source in inflection do not seem to occur.

The second prediction concerns inflectional expression: we expect to find aspect expressed as an inflectional category for verbs more often than subject agreement for person. Greenberg (1963) has tested a similar hypothesis, and found that the presence of person/number inflection on verbs in a language implies the presence of tense, aspect or mode inflection. This supports our general hypothesis, since tense and mode are also more relevant to the verb stem than person agreement is.[1]

The prediction our hypothesis makes, then, is that semantic elements that are highly relevant to one another are likely to be packaged together

and expressed lexically, or will be the most common inflectional or derivational categories. Less mutually relevant semantic elements will be expressed less frequently in morphology, and mutually irrelevant semantic elements will come together only syntactically. Relevance, which is a semantic criterion, makes predictions concerning the *degree of fusion* of formal elements. Particular correlates of the degree of fusion of stem and affix will be discussed in sections 6 and 7 of this chapter.

Another consequence of relevance, which we discussed above in connection with lexical expression, is the fact that concepts formed by the combination of other mutually relevant concepts are highly distinguishable from one another. This applies to the relevance of inflectional categories as well. The application of a highly relevant inflection to a stem produces a greater semantic change than the application of an inflection with lesser relevance. For instance, the following Spanish sentences differ only in the aspect of the verb:

Anoche Juan leyó una novela. (perfective) "Last night John read a novel."

Anoche Juan leía una novela... (imperfective) "Last night John was reading a novel..."

In these two sentences, however, the verbs describe different events, in that in the first, the event described is the reading of a *complete* novel, while in the second, the event is the process of reading some part of a novel. Consider in contrast two sentences that differ only in the person of the verb:

Anoche leí una novela. "Last night I read a novel."

Anoche leyó una novela. "Last night s/he read a novel".

In this case, the events described are at least potentially the same, only the participants have changed.

The degree of semantic differentiation, then, is another consequence of relevance, and it has as a formal correlate *degree of relatedness* in a morphophonological sense. We will not discuss the degree of relatedness in any detail in this chapter, but will take it up again in Chapter 3, where we will see the consequences it has for the structure of verbal paradigms, and again in Chapter 4, where we will see how the degree of relatedness conditions the likelihood of inflectional and lexical splits.

2.2. *Generality*

The second factor that needs to be taken into consideration in determining what can be an inflectional category is *lexical generality*. By definition,

an inflectional category must be applicable to all stems of the appropriate semantic and syntactic category and must obligatorily occur in the appropriate syntactic context. In order for a morphological process to be so general, it must have only minimal semantic content. If a semantic element has high content, i.e. is very specific, it simply will not be applicable to a large number of stems. Consider the notions of *path* or *direction* that are sometimes expressed lexically in motion verbs. Many languages also have morphological means of indicated direction with motion verbs. For instance, Latin prefixes indicated direction: *eō* 'I go', *exeō* 'I go out', *transeō* 'I go across'; *ducō* 'I lead', *producō* 'I lead forth', *traducō* 'I lead across', *reducō* 'I lead back'. But each of these prefixes has a limited lexical applicability, for they are only appropriately added to verbs indicating motion of some sort. Their semantic content prevents them from meeting one of the criteria for inflectional status.

The diachronic counterpart to this is the semantic generalization that takes place in the evolution of inflection. In the passage of a word to a clitic and eventually to an inflection, both its phonological shape and its semantic content must be reduced. In order for the form to occur commonly enough to be reduced and become bound, it must have a meaning that is widely applicable — that is, general enough to be appropriately combinable with any stem of the syntactic category. In addition, it must have a meaning that is communicatively useful enough to ensure a high frequency of occurrence.

Most potential categories that are highly relevant to verbs are not general enough to attain inflectional status. The reason for this is that high relevance tends to *detract* from generality. Because relevant categories produce derived words that are more distinct in meaning from their bases than the ones produced by less relevant categories, the combinations of relevant notions tend to be lexicalized. There are two ways that lexicalization detracts from generality. On the one hand, if the combination of meanings that would be expressed morphologically is already expressed lexically, the morphological process does not usually apply. For instance, since English has a lexical reversative for *freeze*, which is *thaw*, the formation **unfreeze* is unnecessary and indeed, not possible. On the other hand, a morphological category may have a generally applicable semantic content, but still make a rather substantial semantic change in the stem when the two are combined. If this occurs then the resulting words will tend to become lexicalized. These lexicalizations detract from the generality of the category.

Causatives may serve as an example here. The causative meaning is highly relevant to verbs, since it affects quite directly the event or state being

described by the verb stem. However, a causative meaning combined with a verb stem describes quite a different action than the verb stem alone does.[2] For example, dying and causing to die (killing) are two quite different activities. Further, the causative notion itself describes quite different activities according to the verb stem it occurs with. This can easily lead to a situation in which the products of a morphological causative process could become unpredictable semantically and therefore lexicalized. When many of the words resulting from a morphological process become lexicalized, it becomes more and more difficult for speakers to learn to apply the process productively, and the process might eventually lose its productivity.

Consider as an example Causatives in Luganda, a Bantu language. One of the causative formations of Luganda is quite general, and widely applicable, due to the fact that it is used to express the occurrence of either an agent, instrument, reason or purpose in the sentence (Ashton et al. 1954). For many verbs the meaning of a verb stem plus causative is predictable, although some causative verbs are ambiguous between an agent or instrument interpretation. For example, *kùsalà* means 'to cut', and *kùsazà* means 'to cut with'; *kùleèta* means 'to bring' and *kùleèsa* means 'to make to bring, or bring by means of' and *kùgoberera* means 'to follow' and *kùgobereza* means 'to cause to follow'. Even though the addition of the causative suffix to the verb always produces a predictable meaning, there are many verb plus causative combinations that now have in addition an idiosyncratic meaning. For example, the causative of *kùbala* which means 'to count or calculate' is *kùbaza* and means 'to multiply'. The causative of intransitive *kùkyûka* which means 'to turn around, change or be converted' is transitive *kùkyûsa* meaning 'to turn, change or convert (transitive)' or 'to retrace steps' or 'to translate'.

Because the semantic element *causative* makes a considerable semantic change in the event or state being described by the verb, and because the resulting meaning varies considerably according to the verb stem it is combined with, a morphological causative may be very general, but it is prone to lexicalization. In some languages it is totally lexical (and syntactic) as in English, and in others, such as Luganda, it is morphological with many cases of lexicalization.

Now compare the *causative* notion with *tense*. Tense combines quite acceptably with the notions expressed by verbs, in that the verb describes a situation and the tense fixes it in time with respect to the moment of speech. Tense does not, however, alter the situation described by the verb at all. It is the same event no matter when it happens. Thus, while we expect the

expression of tense as a verbal inflection, we also expect that tense distinctions would rarely lead to lexical distinctions. In English we have a fair number of verbs with unpredictable phonological expression in the Past Tense, but there are few tense distinctions that lead to unpredictable semantic representations. An example of the lexicalization of a former tense distinction that is in progress in American English is in the verb *get, got*. *Get* means 'acquire'. *Got* means 'acquired', but because it is past, implies possession in the present: 'I got it, now I have it.' For a substantial group of American children I have observed, *got* can be used in the present tense: *I got; you got; he, she, it gots; we got; I don't got; he doesn't got; do you got? does he got?* are all frequently heard. It is not yet used after modals, **I'll got it; *I can got it*, but its use in the Present Tense shows lexicalization. Note, however, that once this change occurs, the difference between *get* and *got* is not one of tense, but rather of stativity. I know of no examples of the true lexicalization of tense.

The reason that causative is often lexicalized and tense never is, is that the combination of the causative meaning with a verb stem has a radical effect on the meaning of the resulting verb, while the combination of tense with a verb stem does not affect the inherent meaning of the verb. The consequence of this is that even productive morphological causative processes will produce many verbs that will be lexicalized, and it will be difficult for such a process to remain productive. Since tense does not change meaning, and does not give rise to lexicalizations, there is no problem with tense inflections remaining productive.

To summarize the hypothesis: there are two factors that determine the likelihood that a semantic notion will be encoded as an inflectional category. First, the semantic notion must be highly relevant to the meaning of the stem to which it attaches. Second, it must be a very generally applicable semantic notion, or it simply will not apply to enough different items to be inflectional. The interaction of these factors is such that the semantic notions with the highest relevance tend not to be generally applicable, for the reasons outlined above. Thus if morphological categories are graded in terms of relevance, we will find that the most common categories to be inflectional are those that appear in the center of the scale, where relevance is sufficient, but not so high that the meaning changes produced by the process tend toward lexicalization.

3. *Inflectional categories*

The remainder of this chapter concentrates on elaborating and testing the hypothesis that the relevance of one semantic concept to another in part determines the type of expression the concepts will take. The present section discusses some inflectional categories commonly occurring with verbs, and specifies how the hypothesis presented in the preceding section applies to these categories. Deriving from the discussion are certain predictions concerning the relative frequency with which these categories occur as inflectional categories of verbs. These predictions are tested in section 5, where the results of an extensive cross-linguistic survey are reported. Furthermore, this section lays out certain predictions concerning which categories may have lexical expression. No systematic test of these predictions has been carried out, but the reader is invited to apply his or her knowledge of the languages of the world to test these predictions. The categories to be discussed here are valence, voice, aspect, tense, mood and agreement.

Valence-changing categories such as transitive, intransitive and causative are highly relevant to the situation described in the verb stem, since the situation expressed by the verb stem changes according to the number and role of the participants in the situation.[3] The change is sometimes dramatic, as in the case of the causatives discussed above, predicting a tendency toward lexical expression of valence categories, such as English *lay*, *lie*; *sit*, *set*; *go*, *send*, etc. But in other languages the transitive/intransitive distinction figures as 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 in Hebrew (Berman 1978): *avad* 'work' vs. *ibed* 'cultivate'; *yaca* 'go out' vs. *yice* 'export'; and *paxat* 'lessen' vs. *pixet* 'devaluate'.

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 English verbs *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 (a)* '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.

As we mentioned earlier, 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.[4] 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 the Latin example given in section 2.1 of this chapter. These usually express more specific meanings, such as inchoative, as in the Latin example, or completive as in Russian *usinat'* 'have supper', which contrasts with *otusinat'*, which means 'finish supper'.

When *aspect* is an inflectional category, the meaning change effected by it tends, as predicted, to be small. Hopper (1977) 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. 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. Since tense is not exclusively relevant to the verb as aspect is, it is less relevant to the verb than aspect. On the other hand, tense is somewhat more 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. Consider also the case of *get* and *got* discussed above. Here a tense distinction is becoming lexicalized, but note that when *got* is lexicalized it differs from *get* in stativity, not in tense.

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 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 and tense are. Thus we might expect mood to occur less frequently as an inflectional category of verbs than aspect and tense. Since mood does not affect the meaning of a verb, examples of lexical expression of mood-like distinctions are rare or non-existent.[5]

Agreement categories in verbal inflection refer not to the situation

described by the verb, but rather to the participants in the situation, as we mentioned before. Thus we expect subject and object agreement categories to be less frequent than categories that more directly affect 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 different. The number of participants in a situation, whether agents or recipients of an action, can affect the situation profoundly. Thus lexicalized distinctions based at least in part the number of participants do exist, e.g. English *run* vs. *stampede*, *murder* vs. *massacre*. Ainu, which has no inflectional agreement categories, nonetheless has a number of irregular and suppletive verb stem pairs that are distinguished semantically solely on the basis of the number (singular vs. plural) of the object (Batchelor 1938). We will return to a discussion of examples such as these in Chapter 4.

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. The diagram, then, represents only a gross overall pattern of tendencies to be tested in the next section.

This scale 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 and lexical expression.

The relevance, generality and semantic change criteria have yielded a testable hypothesis concerning the relative frequency of certain verbal categories. In addition, two other testable predictions emerge. The first con-

cerns the order in which affixes appear in relation to the stem of the verb. If linguistic expression is iconic, then we would predict that the categories that are more relevant to the verb will occur closer to the stem than those that are less relevant. The second prediction is related. It is that the categories that are more relevant will have a greater morpho-phonemic effect on the stem than the less relevant categories. Thus the greater semantic cohesion of concepts is reflected in a higher degree of fusion in their corresponding expression units. These predictions concerning the frequency of inflectional categories in the languages of the world, their order with respect to the stem, and their degree of fusion with the stem were tested on a stratified probability sample of 50 languages of the world. The next section describes the methodology employed in carrying out these tests, and the following sections present the results.

Expression type

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

4. *Cross-linguistic survey*

Most previous work in cross-linguistic typology and univerals research has paid very little attention to the problem of selecting an appropriate sample for the testing of hypotheses. Often, hypotheses are formulated and "tested" taking too few languages into consideration. Even when a large number of languages is used, little attention is given to the problem of making the sample representative of the languages of the world. Instead, samples are chosen for convenience — the main criterion used is the availability of information about the language (Bell 1978). This is understandable, but it introduces a bias into the data, because languages in certain parts of the world are better described than languages in other parts. Thus a sample chosen for convenience may have in it too many languages that are genetically related,

or too many languages that are in areal contact. Convenience may also weight a sample heavily in favor of languages that are written and standardized, which may also introduce an unwanted bias (Perkins 1980).[6]

The only way to make a fair test of a hypothesis about the languages of the world is to use a sample of languages that is truly representative. The sample must attempt to be representative of *all* the languages of the world and at the same time be as free as possible of genetic or areal bias. The sample used in the present study has these properties. It is a sample of 50 languages chosen by Revere Perkins (1980) in such a way that no two languages are from the same phylum (according to Voegelin and Voegelin 1966) and no two are from the same cultural or geographic area. The division of the world into geographic and cultural areas that was used in the sampling procedure was made by Kenny (1974) based on the listing of cultures in Murdock (1967). The full procedure is outlined in Perkins 1980.

This method ensures a representative sample free of known biases. There are, however, a few problems associated with the sampling procedure. One problem is that some language phyla will not be represented in the sample at all. The reason is that some phyla occur only in the same cultural area as another language phylum, and only one of these can be included in the sample. The difficulty presented by this problem is small in a sample of 50 languages. Bell 1978 has calculated that in a sample of 50 languages, a type that occurs in only 10% of the world's languages will have 999 chances out of 1000 of occurring in the sample. This means that only a rather rare type of language would be totally absent from the sample.

A second problem is associated with choosing languages at the level of the *phyla* as established by Voegelin and Voegelin, because language isolates (individual languages with no known genetic affiliations) are treated as phyla. This means that there are probably too many language isolates in the sample, and that some areas, such as North American are over-represented because more language isolates occur there than elsewhere. Once this source of bias is recognized, it is easily controlled, by checking whether frequent phenomena are concentrated in one area or have broad distribution.

A third problem with this sampling procedure is inconvenience. Perkins selected the languages from each cultural group randomly (by using a table of random numbers), and then attempted to find information about the language. In some cases, no adequate documentation exists, and Perkins then randomly chose another language. Inevitably, in some cases the information about some languages in the sample is better than the information

about others. In four cases, I found the information Perkins had used about a language inadequate for my needs, and was forced to substitute another closely related language. Perkins' original sample contained Haisla, a Wakashan language, for which I substituted Kwakiutl; Ingassana, a Nilo-Saharan language, for which I substituted Logbara; Ket, for which I substituted Gilyak; and Yuki, for which I substituted Wappo. (The languages used in the survey are listed in the Appendix.)

The advantages of using the Perkins sample far outweigh the disadvantages. This sample was drawn up independently of the hypotheses to be tested, so there is no chance of bias introduced by the investigator consciously or unconsciously choosing languages that are known in advance to fit the hypothesis. Since it is a representative sample, it allows the possibility of quantifiable results that are not only valid within the sample, but extendable to the universe from which the sample was drawn. That is, if 52% of the languages in the sample have inflectional aspect, then we can predict that inflectional aspect occurs in approximately 52% of the languages of the world. A consequence of using this type of sample is that there are languages included that do not have any inflectional morphology at all. In many respects these languages are irrelevant to the hypotheses to be tested. It is necessary to include these languages in the investigation, at least at first, to ensure that the hypothesis is tested on the full range of morphological types. Moreover, the inclusion of non-inflectional languages allows inferences about the relative frequency of inflection in the languages of the world, and the possibility of attempting correlations of inflection with basic word order type (see section 9 of this chapter).

In order to test the current hypotheses, the information about each of the 50 languages was examined and certain facts about the inflectional morphology of verbs was extracted. A list was made of the morphological categories that appear inflectionally marked on verbs in each language. For each language and for each category a sublist of the meanings distinguished in the categories was made. And for each meaning, the associated mode of expression (suffix, prefix, zero marker, stem change, etc.) was recorded. In addition, information about the order of morphemes, the conditioning of allomorphy, irregular and suppletive verbs, and basic word order were recorded. The criteria applied in coding each type of information will be discussed in connection with the particular hypothesis being tested, throughout this chapter and following chapters. In this section, the criteria for coding a morphological category as inflectional are discussed.

Inflectional expression must be distinguished from *syntactic* expression on the one hand and *derivational* on the other. *Boundedness* is the primary criterion for distinguishing inflectional from syntactic expression. If a morpheme is inseparable from the stem, and/or occurs in a fixed order contiguous to the stem, or with only closed class items intervening between it and the stem, it is considered bound. If in the language descriptions this information is not given explicitly, then the author's decision to write the morphemes as separate or bound is taken as an adequate indication of their status. Of course certain unclear cases were encountered and decisions had to be made that might be controversial. In some cases, the second major criterion had to be relied upon.

The second major criterion distinguishes inflectional from derivational morphology, and in some cases, inflectional from syntactic expression. This criterion is *obligatoriness*. An inflectional category is obligatorily marked every time a stem category to which it applies appears in a finite clause.[7] The consequences of this are that there must be some means of expression for the category with every stem. A statement in a grammar to the effect that a certain arbitrary set of stems does not appear in the Causative, for example, would mean that the Causative is derivational, not inflectional.

A third criterion associated with the previous one is *predictability of meaning*. The meaning of the category must be predictable with every verb. If the grammar lists the meanings of the combinations with different stems, then the category is not inflectional.

Further, in inflectional expression, the lack of a marker must be interpreted as meaningful (as zero expression) rather than as the absence of the category. Thus in Tiwi verbs are marked for time of day. There is a prefix for morning and one for evening. However, the author notes that the absence of a mark for time of day does not mean "mid-day", but rather that time of day is not indicated in the sentence (Osborne 1974). Thus time of day is not an obligatory category.

In general, in examining the grammars it was possible to follow the author's division of categories into derivational and inflectional, even if explicit criteria were not given. It was assumed that most linguists' intuitive understanding of the inflectional/derivational distinction coincides with the criteria made explicit here, and that the author, who is the expert on the language, should be given the last word.

For certain categories, more specific criteria could be used. For instance, person and number categories are not obligatory unless they are required in

clauses where the subject (or object) with which they agree also appears.

Morphological categories were assigned *labels* based on a list of definitions drawn up before the coding began. With few exceptions, which will be mentioned below, the categories found fit these definitions. When determining how to label the inflectional categories for a language, the labels used by the authors were not relied upon as heavily as the description the author gave of the *function* of the morphemes. For instance, if a grammar characterizes a "Present Tense" as used in the present and past for incomplete action, and a "Past Tense" as used for completed action in the past, these "Tenses" are coded as *aspect*. The following is the list of definitions of the categories used in coding the data:

Valence refers to differences in the number or role of argument that the verb stem can take.

Voice indicates the perspective from which the situation described by the verb stem is viewed.

Aspect refers to the way the internal temporal constituency of the situation is viewed.

Tense places the situation in time with respect to an established point in time, either the moment of speech, or some other point in time.

Mood refers to the way the speaker presents the truth of the proposition in the discourse and real-world context. Included here are expressions of probability, possibility and certainty. *Evidentials* indicate the source of the information expressed by the proposition and were included under mood.

Number agreement is concord with one or more of the arguments of the verb. Subject, object and indirect object agreement are distinguished.

Person agreement is concord by person with one or more of the arguments of the verb. Subject, object and indirect object agreement are distinguished.

Gender agreement is concord with one or more arguments of the verb according to their assignment to lexical classes either arbitrarily or based on inherent qualities of the entity referred to.

While there is a rich enough variety of meanings distinguished under these categories, it was found that by and large these categories and definitions were adequate to cover the inflectional categories described for the languages. The category *mood* contains the widest variety of meanings, and occasionally its boundaries fade off into other categories, as for instance, in

one case where it is difficult to distinguish an intentional mood from a future tense. Only two new categories had to be created during the examination of data. One was *purposive*, to use the terminology of Andrews 1975 in his description of Classical Nahuatl. This category usually distinguishes two meanings, one is "going to" and the other is "coming to". This category was found to occur in three languages. In Zapotec, its meaning is close to that of an intentional and even a future. The other was *status* which refers to markers of the relative social standing of the speech act participants. This was found as a verbal inflection only in Korean.

Once the information from each language was extracted from the grammars, it was entered on a computer so that various correlations could be attempted. The most basic of these was the determination of the relative frequency of these inflectional categories for verbs. We turn now to an examination of these data.

5. *The distribution of verbal categories*

Figures 1 and 2 show the frequency of morphological categories for verbs in the 50-language sample. Figure 1 shows those categories that were considered to be inflectional (obligatory and bound), and Figure 2 shows the occurrence of the same categories counting both inflectional and derivational affixes.[8] Note that an exhaustive survey of derivational morphology was not undertaken — there are many more possible derivational categories than the ones mentioned here. Rather, the presence or absence of derivational morphology with meanings related to those expressed by inflectional categories was noted. The direct test of the *relevance* principle is in Figure 2, where both types of morphological categories are considered. Figure 1 is the test of the relevance principle in conjunction with the effects of *generality*, which predicts fewer inflections among the most highly relevant categories. In general, both sets of results conform to the predictions made by the hypothesis, although neither conform perfectly. This indicates, of course, the need to investigate additional factors that might determine morphological status. Let us discuss briefly the specific findings of the survey.

Perhaps the most striking finding in this survey is the near universality of the morphological expression of *valence*. In 90% of the grammars consulted there was mention of an affix or stem change which could be applied to a verb to change the number of arguments required by the verb. The most frequently mentioned morpheme of this sort was a causative morpheme. The

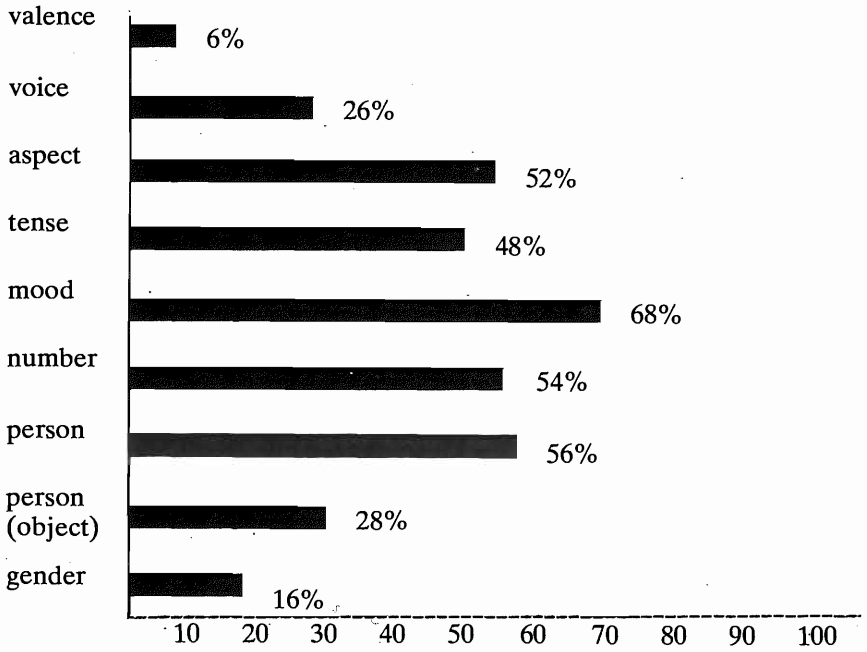

Figure 1: Inflectional categories for verbs

languages for which I found no mention of valence-changing morphology are Haitian, Karankawa, Navaho, Serbo-Croatian and Vietnamese. All of these languages do have other morphology, so it is possible that valence-changing morphology was just not mentioned in the sources.[9]

As might be expected from the high frequency of valence morphology, there are languages which have no other verbal morphology at all, and yet have affixes which change the valence of the verb. Such is the case, for example, in Ainu, Khasi, !Kung and Palaung. The basicness of valence morphology is also seen in developing creoles, such as Tok Pisin, where the first verbal morphology to develop, after the general predicate marker *i-*, is the causative suffix *-im*. Tok Pisin also shows the development of aspect and number markers, but these are not bound to the verb (Mülhäusler 1980).

Figure 1 indicates a very low frequency of valence-changing morphology as inflection. The reason for this is that it is rare to find a case where valence could be considered obligatory in the sense that every finite clause contains a morphological indicator of the number and role of the arguments. All

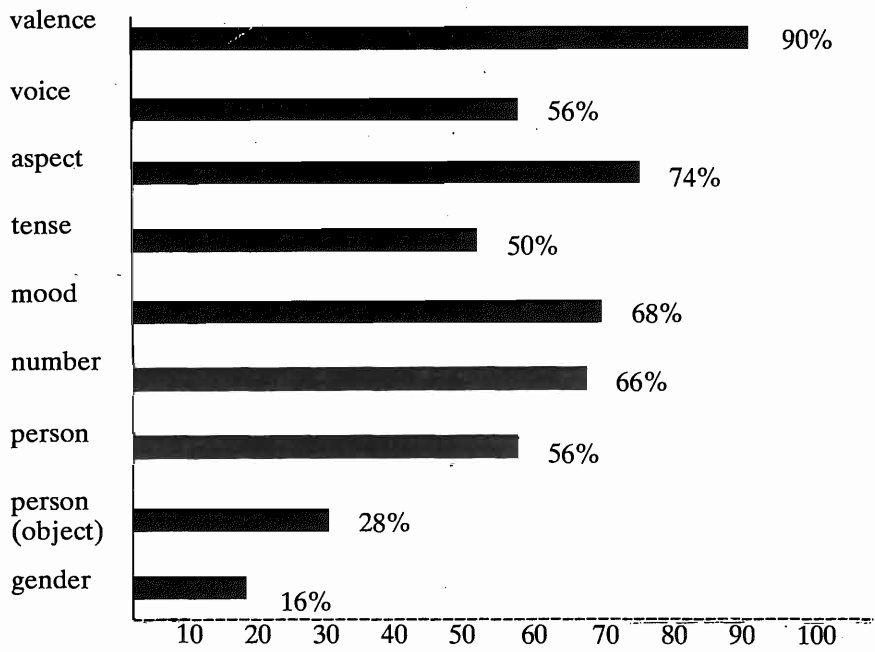

Figure 2: Derivational and inflectional categories marked on verbs

languages seem to have verbs that are inherently transitive and verbs that are inherently 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 sometimes gender as well. Thus object agreement was not counted as an obligatory expression of *valence*. There were, however, three languages in the sample that could possibly be considered to obligatorily express valence. Consider first Kutenai and Maasai: both have object agreement markers, and both have instrumental and benefactive markers on 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. Furthermore, in Maasai if object agreement is not present on a transitive verb, the verb 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", *árányɪshɔ* "I sing", and *adɔl* "I see it or them", *adɔ lɪshɔ* "I see" (Tucker and Mpaayei 1955). The third example of possible obligatory valence is in Gilyak, where a stem initial consonant alternation coincides with the transitivity of the verb. All transitive verbs have consonants from one series, while intransitive verbs have consonants from another series (Jakobson 1957).

The lower frequency of *voice* categories (as compared to *aspect*) is due to the fact that there are other ways of signalling changes in the perspective of the sentence: sometimes the marking occurs on the noun phrases, and sometimes changes in perspective are accomplished solely by changes in word order. However, *voice* can produce a substantial meaning change in the verb, as evidenced by the occurrence of voice as a derivational category. *Voice* morphemes were considered derivational in cases such as Diegueño, where two suffixes were described as having passive-like meaning (be in a state resulting from an action), but were described under the heading of "stem-formation". 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).

Morphological aspectual categories followed valence as the most frequently described verbal categories. The meanings most often recorded in the sources were *perfective, imperfective, habitual, continuous, iterative* and *inceptive*. Again it is at times difficult to distinguish obligatory from non-obligatory aspect. This issue is discussed in detail in Chapter 4, and again in Chapter 6. A rough guess is that a little more than half the languages of the world have inflectional aspect.

Fewer languages have *tense* as an obligatory category, and only one language (Kwakiutl) was found to have *tense* as a derivational category. The lower frequency of *tense* as compared to *mood* is due almost entirely to the near ubiquity of an *indicative / imperative* distinction in languages that have any verbal inflection at all. Here we must invoke some notion of communicative importance, for the distinction between a command and an assertion is a very basic distinction in natural language. Details of the distinctions expressed in the *mood* category, as well as a critical examination of the coherence of this category are found in Chapters 8 and 9.

Person and *number* as inflectional categories tend to co-occur. Diegueño is the only example in the sample of a language with inflectional person agreement that lacks inflectional number agreement. However, Diegueño has a derivational category of number. In fact, several languages in the sample have derivational expression of "number" in the verb, and some even have lexical expression of "number". This phenomenon is discussed in Chapter 4.5.2.

The differences between the frequency of inflectional *aspect, tense, mood* and *subject agreement* categories is probably not highly significant. However, if *aspect, tense* and *mood* are grouped together and compared to *subject agreement*, then a greater difference emerges: 72% of the languages of the sample have *aspect, tense* or *mood* inflections, while only 56% have subject agreement. Further, in accordance with Greenberg's 1964 finding, no language in this sample has subject agreement that does not also have either aspect, tense or mood.

It appears, further, that the presence of inflectional *aspect* or *tense* in a language implies the presence of inflectional *mood*. The only exception in the sample is the case of Susu, where the unmarked form of the verb may serve as both an Aorist and an Imperative, so that there is no inflectional marking of mood. On the other hand, *mood* inflections occur in languages which do not have inflectional aspect or tense, according to the descriptions consulted. The four languages with this arrangement, Keresan (Acoma), Quileute, Kwakiutl, and Yupik, are all North American languages, and all four of them have considerable derivational morphology, which includes derivational *aspect* in the case of Keresan, Quileute and Kwakiutl. Sufficient data on Yupik are not available to determine the existence of derivational aspect. It appears, then, that a typical inflectional system for verbs will express *mood*, but almost always accompanied by the morphological (usually inflectional) expression of *aspect* or *tense* or both.

Finally, it should be mentioned that derivational expression of *agreement* categories other than *number* was not encountered.

6. *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 elements 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.[10]

Tense markers occur closer to the stem than *mood* markers in 8 languages out of 20 that have both tense and mood. In one language, Ojibwa, the mood marker occurs closer to the stem than the tense marker.

Tense markers occur closer to the stem than *person* markers in 8 languages out of the 17 that have both tense and mood. In one language, Navaho, the person markers occur closer to the stem than the tense markers.

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 in a verb form.

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.

7. *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. Examples of cases where the inflectional category has an effect on the verb stem are 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 and Zapotec 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 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 Chapter 4.5.2.

Stem changes with *person* categories are even rarer (Hooper 1979a). 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, support the relevance hypothesis 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.

8. *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 in this chapter, 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 hypothesis 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 main 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. 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 and the closeness of constituents to each other in the surface string. (Vennemann 1973:41)

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.[11] 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 hurtanas
huṛtanay	huṛtanader
huṛtanan	huṛtanar
huṛtanat	huṛtanik huṛtiniŋ

These example show that the order of morphemes need not necessarily reflect an earlier order of words, nor the chronological order in which inflectional morphemes develop. 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 I 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. I 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 entirely 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 in both domains.

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 pronominal meanings with 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 spell *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.

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.

The application of the relevance principle to the degree of fusion of the inflectional morpheme with the verb stem will be treated in detail in the next chapter. To a large extent the degree of fusion is determined rather mechanically by how long and in what order the inflectional morphemes have been attached to the stem. But this is not entirely the case with stem changes that co-occur with inflectional categories, because these can be affected by morpho-phonemic changes. These changes are also governed by the relevance principle, and will be treated along with other matters relating to the organization of verbal paradigms.

9. *Morphology and word order*

In the survey of 50 languages, information about the basic word order typology was recorded along with the information about verbal inflection. While correlations with word order do not bear on the central hypothesis of this chapter, it is interesting to consider briefly the frequency of occurrence of inflectional languages among the different word order types. It is also interesting to compare Greenberg's (1963) findings on the frequency of the

different word order types with the present survey. There are some significant differences, which are probably due to differences in sampling techniques. We take up this matter first.

Greenberg used 30 languages from a wide range of families and areas, but the languages were largely selected for convenience, as Greenberg notes (1963:75). The distribution of the three main types of word order in his sample is shown in Figure 3, followed by the distribution in the Perkins sample, which was used here. For Perkins' sample, the total is only 40 languages because there was one VOS language (the Car dialect of Nicoborese), one langauge, Logbara, with the contrastive use of SOV and SVO order for imperfective vs. perfective sentences, five languages with reportedly free or variable word order (Nahuatl, Navaho, Sierra-Miwok, Tarascan, and Ojibwa, all North American Indian languages), and three in which information about word order was not available (Karankawa, Timucua and Yupik).

The interesting point to note is that in Greenberg's sample there are slightly more SVO languages than SOV, while in Perkins' sample there are nearly twice as many SOV languages as SVO. This is a rather striking difference, and an interesting indication of the importance of sampling techniques. Greenberg's sample is rather heavily weighted in favor of languages in certain areas — there are 7 European languages, 5 of them Indo-European, 7 African languages, and no languages from Australian or North America. I estimate that about one-third of the languages in his sample are closely related genetically to another language in the sample. This leads us to suspect that the distribution of word order types in his sample is not representative of the languages of the world.

Another point of discrepancy between the results from the two samples is that Greenberg found that 10 out of 11 SOV languages were exclusively suffixing (1963:92), while in the Perkins sample only 6 of the 21 SOV languages are exclusively suffixing, while 11 have both suffixes and prefixes.

With the Perkins sample we can also determine the frequency of the different morphological types among the languages of the world, and correlate this typology with the basic word order typology. The basis of morphological typology, as established originally in the nineteen century, and further developed in our own, most notably by Sapir, is the extent of fusion of morphemes. An *analytic* language does not have fusion of morphemes, rather each morpheme is a separate word, while a *synthetic* language does have fusion of morphemes, and consequently morphologically-complex words. This typology suggests that there is no necessity that a language have morphological expression. Clearly all languages have the possibility of syntactic expression, but not all languages use morphological expression. But just how common are purely analytic languages? In the Perkins sample 28% of the languages have no *inflectional* morphology on verbs. But some of these languages, for example the Car dialect of Nicoborese, have extensive *derivational* or non-obligatory morphology. Others, e.g. Vietnamese, make extensive use of compounding and reduplication. In fact, languages with absolutely no morphologically-complex words do not occur in this sample. The closest case is Palaung, which, according to the description by Milne 1921, has only two prefixes, both of which attach to verbs to give a causative meaning. In addition, Khasi and !Kung have very little affixation, but both of these languages allow the formation of compounds.

Another distinction made in morphological typology distinguishes degrees of fusion in synthetic languages. In *agglutinative* languages, morphological boundaries coincide with phonological boundaries (especially syllable boundaries) to an extent that makes segmentation of morphemes transparent. In *fusional* languages, there is greater fusion of morphemes characterized by sandhi at boundaries, allomorphy, and simultaneous expression, all of which make morphological segmentation more opaque. This can be illustrated with the following two extreme examples:

Agglutinative: Garo
/sok + ba + ku + ja + ma/ = /sokbakujama/
arrive + toward + yet + neg + int
"has he not yet arrived?"

Fusional: Hebrew (the verb "to write")
Past 3s masc: katav
Present, 3s masc: kotev
Future, 3s masc: yixtov
Imperative, masc. sg: ktov

In the case of Garo, the morphemes occur sequentially, one per syllable, but in the case of Hebrew, the consonants of the stem interdigitate with the vowels that signal the inflectional categories, so that no clear segmentation is possible. Further, there are alternations in the consonants of the stem in the case of Hebrew, illustrated here by the /k/ and /x/ alternation, while Garo has virtually no allomorphy.

These are extreme cases. Most languages utilize a combination of agglutination and fusion, so that a clear classification of synthetic languages into

the two types would be difficult. An impressionistic classification of the languages of Perkins' sample into agglutinating and fusional, however, shows that pure agglutinating languages are much rarer than languages with greater fusion. The mechanism behind this distinction may be phonological. Most languages have reductive phonological processes that obscure the boundaries between morphemes and create allomorphy. Phonological processes in agglutinating languages apparently do not have this effect, or alternatively, have not had this effect yet, assuming that agglutinating languages can become more fusional. If the mechanism is primarily phonological, however, it is difficult to explain the observation, first advanced by Lehmann 1973, and confirmed impressionistically in the current sample, that agglutinating languages tend to have SOV as their basic word order.

Another parameter that might yield a morphological typology is the number of categories expressed inflectionally: some languages have no inflections, some have a few, and some have a great many. In section 5 it was pointed out that some languages have only *aspect, tense* or *mood* inflections, while others have one or more of these in addition to inflections to mark agreement in *person* and *number* with the subject. If the languages are divided according to whether they have no inflection (I), *aspect, tense* or *mood* (II), and one of the preceding plus *person* and *number* (III), then the correlations shown in Figure 4 result. The figure shows that SOV languages may be of any inflectional type, but OV languages are evenly divided between those with no inflection and those with the full array of inflections. There are no VO languages with *aspect, tense* or *mood* that do not also have *person* and *number* agreement. Furthermore, while SVO languages are divided between the no-inflection and full-inflection types, VSO languages tend to have full inflection in this sample. At the present moment I do no know the significance of these correlations between inflection and word order.

	SOV	SVO	VSO
Greenberg's sample	37% (11)	43% (13)	20% (6)
Perkins' sample	42% (21)	22% (11)	16% (8)

Figure 3. Comparison of the distribution of basic word order types in two samples.

	I	II	III
SOV	19% (4)	33% (7)	48% (10)
SVO	64% (7)	0	36% (4)
VSO	12% (1)	0	88% (7)
VOS	100% (1)	0	0
all VO	45% (9)	0	55% (11)
all languages	32% (13)	17% (7)	51% (21)

Figure 4. Distribution of inflection in languages with different basic word order.

NOTES

1. Greenberg does not explain how the term "mode" is defined in his study.

2. Haiman 1983 points out that the difference in meaning between a periphrastic and a lexical causative is reflected in the expression type. For a lexical causative in which cause and result are expressed in the same stem, the cause and result are at the same time and the same place. This is not necessarily the case for a periphrastic causative.

3. The following conventions are followed throughout the book: the names of categories and members of categories that are defined here and intended as universal labels are printed in italics with lower case letters. The names given categories and their members by authors of language-specific grammars are printed with initial capital letters.

4. An interesting exception is the Finnish genitive vs. partitive distinction. An object marked in the partitive gives an imperfective reading to the 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).

5. Lexical items such as *think, know, believe*, or *doubt* do not represent the lexical expression of mood, but rather periphrastic or syntactic expression. Lexical expression would involve a basic verb meaning, such as "go" appearing in two contrasting lexical items where the only difference between them is a mood difference. Two examples in English that come close to meeting the criterion are the verbs *scram* and *scat*, which are only used in the imperative. What English lacks is a true complement to these verbs: a verb for "go away" that is never used in the imperative.

6. The difference in results obtainable using different sampling techniques is demonstrated in section 9, where the distribution of word order types in Greenberg's 1963 sample is compared with the distribution in the sample used here.

7. Non-finite forms were excluded from the study.

8. Categories that occurred in four or fewer languages were not included in Figure 1. Missing from the table are (1) evidentials, which are included under mood, since all the languages that have evidentials also have mood, (2) purposives and (3) status markers. Status markers code the

social status of the speech act participants and are even less relevant to the verb than agreement markers that refer to participant in the situation.

9. These figures differ slightly from those reported in Bybee 1984 since they are based on a re-examination of a portion of the data.

10. In Georgian the marker for 1st person occurs as a prefix following the aspectual preverb. Other indicators of person and number are suffixes.

11. Certain disagreements with the ordering of morphemes proposed here and those proposed in Foley and Van Valin 1984 will be discussed in Chapter 9.2.

CHAPTER 3: THE ORGANIZATION OF PARADIGMS

A paradigm is a group of inflectionally related words with a common lexical stem. Such a group of words is not comparable to an unstructured list, in which each word bears an equal relation to every other word. Rather, a paradigm has internal structure: there are relations among words that are not symmetrical, and some relations are stronger than others. Descriptive morphology, whether Item and Arrangement, Item and Process, or Word and Paradigm, has not concerned itself with these relations directly, but has rather concentrated on describing the internal structure of each individual word of a paradigm. This chapter will not be concerned so much with descriptions of how to combine morphemes into words, but will concentrate on discovering, through the use of various kinds of evidence, what types of relations obtain among the words of a paradigm.

In all models of morphological description it is recognized to some extent that not all forms of a paradigm have the same status. In Item and Arrangement grammar the fact that some allomorphs have a wider distribution than others is highlighted. In Item and Process grammar there is an underlying form for each paradigm, which does not necessarily correspond directly to any form that actually occurs on the surface. Relations among surface forms are not described, rather all surface forms bear a similar relation to the underlying form, in that they are derived by rule from it. The only difference among these relations is that it might take more rules to derive one surface form than it takes to derive another.

The only descriptive framework that confronts directly the relations among surface forms of a paradigm is traditional grammar, where paradigms are described by structured lists that group forms together according to how closely related they are to one another. While this chapter is not concerned with advocating one descriptive model over another, we will see that the structure implicit in the traditional representation of paradigms is often supported by data from child language, experimentation, historical change and universals.

In this chapter it is proposed that paradigms consist of clusters of closely

TYPOLOGICAL STUDIES IN LANGUAGE (TSL)

A companion series to the journal "STUDIES IN LANGUAGE"

Honorary Editor: Joseph H. Greenberg
General Editor: T. Givón

Editorial Board:

Alton Becker (Michigan)
Wallace Chafe (Berkeley)
Bernard Comrie (Los Angeles)
Gerard Diffloth (Chicago)
R.M.W. Dixon (Canberra)
John Haiman (Winnipeg)
Paul Hopper (Binghamton)
Margaret Langdon (San Diego)
Charles Li (Santa Barbara)
Johanna Nichols (Berkeley)
Andrew Pawley (Auckland)
Frans Plank (Konstanz)
Gillian Sankoff (Philadelphia)
Dan Slobin (Berkeley)
Sandra Thompson (Los Angeles)

Volumes in this series will be functionally and typologically oriented, covering specific topics in language by collecting together data from a wide variety of languages and language typologies. The orientation of the volumes will be substantive rather than formal, with the aim of investigating universals of human language via as broadly defined a data base as possible, leaning toward cross-linguistic, diachronic, developmental and live-discourse data. The series is, in spirit as well as in fact, a continuation of the tradition initiated by C. Li *(Word Order and Word Order Change, Subject and Topic, Mechanisms for Syntactic Change)* and continued by T. Givón *(Discourse and Syntax)* and P. Hopper *(Tense and Aspect: Between Semantics and Pragmatics)*.

Volume 9

Joan L. Bybee

Morphology
A Study of the Relation between Meaning and Form

MORPHOLOGY

A STUDY OF THE RELATION BETWEEN MEANING AND FORM

JOAN L. BYBEE
SUNY at Buffalo

JOHN BENJAMINS PUBLISHING COMPANY
Amsterdam/Philadelphia

1985 911204-48

	lexical	derivational	inflectional
telic/atelic	X	X	
semelfactive	X	X	
stative/active	X	X	
inceptive	X	X	
iterative		X	
perf/imperf			X
habitual/cont			X

Table 6: Expression types for aspectual notions.

Not only do generality and predictability of meaning condition possible morphological expression types, but even within the conceptual domain of aspect, there are differences in the relevance of aspectual notions to the verb, in that there are differences in the extent to which an aspectual modification of an activity or event produces a distinct activity or event.

5.2. *Number in verbs*

As mentioned several times, agreement categories have less relevance for a verb stem than any other inflectional categories, and person-agreement categories do not have lexical or derivational expression, except in the rarest of cases. However, agreement for *number* of the subject or object appears to present quite a different case.[12] There are some languages in which number distinctions are lexicalized in verb stems. A very clear example is !Kung, a language of Southern Africa, which has no inflectional agreement categories, but does have a small but central group of verbs which has one form in a sentence in which the absolutive noun phrase (the object of a transitive or subject of an intransitive) is singular and another form when that noun phrase is plural. When the verb has a plural form, the plural suffix *-si* on the noun is optional (Snyman 1970: 124, 131-132): ([!], [|], [ǂ] and [||] are ingressive consonants)

mi gu n!ao	'I take the bow'
mi n|'hwi n!aosi	'I take the bows'
n!eu !o'a g!heĩ	'The elder breaks the stick'
n!eu kx'oma g!heisi	'The elder breaks the sticks'
kx'ao ||u hema	'Kx'ao hangs up the shirt'
kx'ao g||ao hemasi	'Kx'ao hangs up the shirts'
n|wa !ei	'The cat dies'

n|wa !ao	'The cats die'
n!eu g|i	'The elder goes out'
n!eu g!e'i	'The elders go out'
xei šu	'The loaf of bread lies flat'
xei gǂa	'The loaves of bread lie flat'

Another language with a similar phenomenon is Ainu, which, like !Kung, does not have inflectional expression of agreement on verbs, but does have lexical and derivational expression for some verbs of the number of the absolutive noun phrase. The following lexical pairs of verb forms are given in Batchelor 1938: 121-122:

singular	*plural*	
a	at	'to be'
a	rok	'to sit'
ani	amba	'to carry'
raige	ronnu	'to kill'
ek	araki	'to come'
arapa	paye	'to go'

The majority of plural verbs are formed, however, by the addition of the suffix *-pa* to the singular form:

ama	amapa	'to put or place'
aship	ashippa	'to flower'
heashi	heashpa	'to begin'
hekatu	hekatpa	'to be born'
oboso	oboshpa	'to pass through'
rai	raipa	'to die'

In Diegueño and Kwakiutl plural verbs are derivationally related to singulars by a variety of irregular reduplicative, stem-changing and affixation processes. For these languages, the analysis of meaning makes it clear how plurality of subject or object can affect the meaning of the verb stem. Boas (1947: 246) explains that in Kwakiutl there are three distinct types of plurality for verbs: "one indicating several subjects; a second indicating an action occurring at the same time in different parts of a unit; and a third, expressing repeated action."

mɛdɛ'lqwɛla	'it is boiling'
meʔmɛdɛ'lqwɛla	'many are boiling'
maʔɛ'mdɛlqwɛla	'is boiling in all of its parts'

mɛdɛ'lxumɛdɛ'lqwɛla	'it is boiling repeatedly'
tɛ'nk˙ɛla	'it is sizzling'
te'ʔtɛ'nk˙ɛla	'many are sizzling'
tɛ'ntɛnk˙ɛla	'it is sizzling in all its parts'
tɛnx˙tɛnk˙ɛla	'it is sizzling repeatedly'

These examples show that plurality in a verb may express more than the number of the subject or object. Plurality of action may involve either distribution or iteration of the action. These various plural notions may be expressed by a single morpheme, as in Pawnee, where the prefix *wa:* may signal a distributed activity or state, an iterative action, a distributive plural object, or a dual and/or plural subject (Parks 1976: 279). Consider these examples:

/wa: + wiu:s/ wa:wiua	'to defecate here and there'
/ra + wa: + hak/ rawa:hat	'to pass to (various people)'
/wa: + u/ wa:ʔu	'to give (various things)'

Freeland 1951 observes a similar phenomenon in Sierra Miwok. She says (Freeland 1951: 112):

> 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.

Her examples show the distributive suffix + *i:* +:

poʔa:l + 'to slit open';	poʔ:al + i: + 'to slit open several'
maʔta + 'to kill';	maʔ:at + i: + 'to kill several'
haʔta + 'to toss';	haʔ:at + i: + 'to toss away repeatedly or several'

Detailed semantic information about !Kung and Ainu are not available, but the Kwakiutl, Diegueño, Pawnee and Miwok data show that number can be much more than just an agreement category when it is expressed derivationally. In fact, this type of verbal plurality is more like an aspect, in that it has an effect on the inherent meaning of the verb. It involves meaning components other than aspectual ones, however, implying as it does, particular arrangements of the absolutive argument. This is another case, then, that shows clearly the difference between lexical and derivational meaning on the one hand, and inflectional meaning on the other, for when number is expressed inflectionally, it never has such an effect on the meaning of the verb.

Observe also that inflectional number agreement is often redundant even in the context of the clause, since the verbal agreement is required even when the number is given in the noun phrases. This is not so with derivational or lexical number. Snyman notes for !Kung that the pluralizing suffix for nouns, *-si*, is not necessary in the examples given above, since the verb indicates plurality. Boas says about Kwakiutl, that plural verbs are not used if plurality of noun phrases is indicated by a numeral or other quantifier. Unlike inflectional number, lexical and derivational plurality is not used redundantly. This indicates that the amount of meaning supplied in this way is much greater than that supplied by inflections. I conclude that even though there are cases of what seem to be the same concepts having derivational and inflectional expression, it turns out that they are not precisely the same, but only related concepts.

6. *Compounding and incorporation*

Two morphological expression types related to derivation are *compounding* and *incorporation*. Morphologically-complex words created by these processes differ from those created by derivation and inflection in that such words cannot be analyzed as consisting of a stem or root plus affixes, rather they contain more than one stem or root. That is, the elements combined in these formations are not lexical plus grammatical, but rather two or more lexical elements. *Compounding* can be easily illustrated in English, where noun - noun, adjective - noun and noun - verb combinations occur. For example: *school bus, black board, babysit.* Noun - noun and adjective - noun combinations differ from regular noun phrases but resemble monomorphemic nouns in that the primary stress is on the first element of the compound, while for phrases the main stress is on the last lexical element. Noun - verb compounds differ from verb - object phrases in the position of the noun.

The term "incorporation" has been used to cover a variety of verb-formation phenomena ranging from processes very similar to compounding to processes that are much like derivational morphology (Sapir 1911, Woodbury 1975, Sadock 1980, Rudes 1984). Typically, "incorporation" refers to the fusion of the nominal patient of the verb with the verb, but often two verb stems can be fused as well. Both of these situations can be illustrated with examples from Tiwi, an Australian language (Osborne 1974: 46-48). Consider these examples:

Library of Congress Cataloging in Publication Data

Bybee, Joan L.
Morphology.

(Typological studies in language, ISSN 0167-7373; v. 9)
Bibliography: p. 217
1. Grammar, Comparative and general -- Morphology. 2. Grammar Comparative and general -- Grammatical categories. I. Title. II. Series.
P241.B9 1985 415 85-9201
ISBN 0-915027-38-0 (pb.; U.S.; alk. paper)
ISBN 0-915027-37-2 (hb.; U.S.; alk. paper)
ISBN 90-272-2877-9 (pb.; European; alk. paper)
ISBN 90-272-2878-7 (hb.; European; alk. paper)

PREFACE

This work is concerned with both morphology and morpho-phonemics. The study of morphology approaches morphemes as the (minimal) linguistic units with semantic content, and studies the relations among them. In contrast, morpho-phonemics, as classically defined, studies the relations among allomorphs — the variant phonological representations of a single morpheme. The latter study, it would appear, is a study of form only, and has often been approached that way. The present work proposes to demonstrate that in morphology form should not be studied independently of meaning. The meaning of a morpheme and the meaning of the context in which an allomorph occurs determine many properties of their formal expression.

Some readers of this book may be more interested in morphology, while others may prefer to read about morpho-phonemics. Throughout most of the book, the discussion of both subjects is interwoven, but Chapters 3 and 5 are more relevant to those interested in morpho-phonemics, while Part II is less about morpho-phonemics and more about the content of morphological categories of the verb.

Perhaps some readers will wonder how this book relates to my earlier book on Natural generative phonology, since it might appear at first that the current work deals with a different set of issues. Actually, the present work has grown up spontaneously from this earlier work. Natural generative phonology is an attempt at a formal theory which addresses the problem of arbitrariness in morpho-phonemics. Abstract generative phonology approaches this problem by trying to equate morpho-phonemics with phonology, and thereby reduce its arbitrariness. Natural generative phonology responds by treating morpho-phonemics as part of morphology, emphasizing the notion of morphological conditioning. This approach has turned out to be productive, in the sense that it raises questions which seem to me to be more interesting and important than the formal issues of rule order and levels or means of representation that are debated in generative phonology. They are questions of how morphology is acquired, how morphological classes are structured, what determines productivity, what are possible morpho-