

Chapter 2

Problems and Existing Theories

2.1 Introduction

The theoretical approach presented in this work had its genesis in specific morphological problems. In some case these problems are tackled but not entirely resolved by existing theories. Other phenomena remain wholly unaccounted for.

These include certain relationships among the elements of word formation, in paradigms and in groups of words. In some cases there are apparent similarities between entities which not normally considered related. In other cases groups of entities are regarded as closely related, while they demonstrate significant dissimilarities.

This chapter outlines some of these problems, and the major existing approaches to dealing with them.

2.2 Problems

2.2.1 Related Forms

It is not uncommon to observe a group of words which share some element of meaning, and which also share phonological similarities. In some cases these groups of words appear to in fact share variants of a single root. Certain kinds of such relationships have been largely ignored by morphological theory.

Consider, for example, the relationship between *isle* and *island*. Semantically they are very similar, and are often used interchangeably. *Isle* perhaps refers to something smaller than *island*, with *-land* reinforcing the idea of something slightly more substantial. Once *-land* is isolated, the two roots are phonologically identical. This is a single root, meaning a body of land separated from other land by being surrounded by water. Since the earth's surface may be characterised as consisting of either land or water, an *isle* may be regarded as fundamentally a piece of land characterised by being *isolated* from other land by being surrounded by something that is not land.

Isolate itself contains the familiar verb marker *-ate*, which if removed reveals the root *isol-*. An *isolated instance* is one which is not part of a pattern of similar

instances, but a lone occurrence separate from other like occurrences. An *isolated farmhouse* is one which is not part of a larger body of buildings such as a town, or even a scattering of farmhouses. Millions of years ago Tasmania was *isolated* from the mainland by rising sea levels. The meaning is not limited to physical separateness: a role taken on at work which deprives the worker of support from colleagues would *isolate* that worker, even in a crowded office.

So the root *isol-* means separateness, the absence of like with like. *Isle* also means this separateness, but specifically related to land. The occurrence of the root in the phrase 'no man is an island' is metaphorical. But why does the metaphor work? 'No man is an atoll' fails: the semantic focus of *atoll* is not its isolation, but its coral construction. The metaphor works because the semantic focus of *island* is the notion of separateness of like from like. If someone is described as "acting like an island", it is not meant to suggest that they sit in water. Perhaps the fact that *isle* is marginal despite being shorter than *island* indicates that there is some semantic benefit from the presence of the element *-land*.

There is a close semantic relationship between *isle* and *isol-*: they have the same underlying semantic focus. Phonologically the links between [aɪl] and [aɪsəl] are more complex, but nonetheless there are clear links. This relationship may be extended to include *insular*.

The root in *insular* is also present in *insulate* and *insulation*, which have the familiar verb and noun marking suffixes. *Insular* itself has the suffix also found in *popular*, *similar* and so on. This isolates the root *insul-*, which appears to have several closely related meanings. The American Heritage Dictionary defines *insular* as "of or constituting an island" or "characteristic or suggestive of the isolated life of an island." This is especially evident in *peninsula*.¹ To *insulate* is to separate something by surround it with some other material. *Insul-* involves the underlying notion of separation, of a lack of contact between things. This is strikingly similar to the semantic focus of *isle* and *isol-*.

Phonologically there are also similarities between these and *insul-*. There appears to be a progression, a building up of phonological features: [aɪl] to [aɪsəl], then to [ɪnsyəɪl]. While the phonological relationship between *isle* and *insul-* is not particularly strong, *isol-* forms a bridging link in a chain of phonologically linked forms. This relationship is supported by the virtually synonymy of *isle* and *insul-*. There is a relationship between these three entities which needs accounting for.

1. The prefix *pen-*, means "the nearest thing to", as in *penultimate* (the next-to-ultimate). A *peninsula* is a piece of land which projects into a body of water: the nearest thing possible to an island without actually being an island. The relationship between *isle* and the *-insul-* in *peninsular* is one of synonymy.

A more complex example involves *name* and *noun*. There are significant phonological similarities between the two: both are monosyllables starting with [n] and ending with an anterior nasal. There are also significant semantic links. Not all names are nouns (though they must at least function as a Noun Phrase), and *name* is also a speech act; however, all nouns are names. *Name* and *noun* sound similar, and their meanings overlap. This relationship needs accounting for.

Now consider *pronoun*. This *-noun* is the same entity as the word *noun*: a *pronoun* is a type of *noun*, and the two are phonologically identical. Now consider *pronounce*. This *-noun-* is phonologically the same as the word *noun*, but the meanings differ. However, *pronounce* is not semantically unrelated to *noun* in the way that, say, *hippo* or *walking* are. A *noun* is a kind of word. *Pronounce* may be a speech act (to say the words of judgment); or may refer to the manner of saying words. Both *noun* and *pronounce* are about words, and so is *name*, which is either a kind of word or a speech act involving that kind of word.

Now consider *denounce* and *announce*. This *-noun-* is phonologically the same as that in *pronounce*, and, like one sense of *pronounce*, refers to speech acts. From elsewhere in the language we can see that *pro-*, *de-* and *a-* are probably prefixes. These words appear to share a single root, as does *enunciation*. Like one sense of *pronunciation*, *enunciation* refers to manner of speech. However, unlike *pronunciation*, there is no **enounce*: the verb is *enunciate*. That aside, *announce* and *denounce* are speech acts; *enunciate* refers to manner of speech; and *pronounce* shares both related meanings.

Now consider *antonym*, *synonym*, *homonym*, *eponym* and *patronymic*. These all refer to kinds of words. They may be characterised as words whose key characteristic is the relationship they have with certain other words. The first three may be any kind of word. *Eponym* and *patronymic* are specifically kinds of names. A *patronymic* is a name taken from the father. The *patro-* refers to the father component of the meaning, and the *-nym-* is the "name" component.

The phonological relationship between *-nym* and *name* is significant. Indeed, it could be argued that the former is the result of a vowel shortening rule. Semantically there is also a significant relationship. A *name* is a specific kind of word. The entity *-nym* also refers to a specific kind of word, and sometimes to a specific kind of name.

As a verb, *name* has a close relationship with *nominate*. Both are speech acts involving names: one assigns a name, the other proposes a name. If the verb marker *-ate* is removed from *nominate*, the root *nomin-* is revealed, and the same entity occurs elsewhere: *nomenclature* is a system of naming; and in one sense

nominal means "in name". However, in another sense *nominal* is the adjective of *noun*, as *pronominal* makes clear.

This is clearly an extremely complex group of words. All involve roots with the form [n], followed by a vowel, followed by an anterior nasal. All have meanings which relate in some way to specific characteristic of a word or words. Within that, *synonym* is related formally and semantically to *name*, which is semantically related to *noun*, which is related semantically to *nominal*, which in turn is formally identical to *pronominal*, which is related semantically to *pronoun*, which is formally related to *pronounce*, which in turn is formally and semantically related to *enunciation* and so on and so on, backwards and forwards through the set.

This kind of relationship is interestingly illustrated in certain language tests.² One set includes multiple choice tests in which there is a given word, and a set of five answer options. The object is to identify the answer option closest in meaning to the given word. In a number of questions, the correct answer, characterised as being nearly synonymous, bears a striking phonological similarity to the given word. For example:

- (1) 10 slim
 A slender
 B figure
 C weak
 D branch
 E light

Here the correct answer is *slender*. The second syllable could possibly be analysed as a suffix, leaving the root *slend-*. Either way, there is a transparent formal relationship between *slim* and the first syllable of *slender*: [sl], then a [- low front] vowel, followed by an anterior nasal. The presence of the [d] in *slender* may be epenthetic or may be part of an allomorphic phenomenon. Either way there is a significant phonological similarity between this root and the given word. In addition, the semantic relationship between the two is sufficiently strong for the test designers to have identified them as synonymous or nearly synonymous.

- (2) 29 sullen
 A faded
 B supple
 C sulky
 D fierce
 E sleepy

Here both the correct answer, *sulky*, and the given word involve two formal and semantic elements. *Sulky* is the adjectival form of *sulk*, the suffix also found in

2. The examples are from the Test Of Learning Ability tests: TOLA 4 and TOLA 6. These tests were used until recently in NSW for primary school age children, and included a section of language proficiency tests.

angry, hungry etc. In *sullen* the *-en* may be isolated as the form also found in adjectives like *wooden, woollen, golden, earthen, ashen* etc. The isolation of these two suffixes reveals the roots *sull-* and *sulk*. Again there are significant and readily apparent phonological similarities, and again the forms are considered by the test designers to be at least nearly synonymous.

- (3) 30 sieve
 A wash
 B sift
 C dry
 D mix
 E drain

Presumably we are to take *sieve* to be a verb, since the options are all verbs, and that *sift* is the correct answer.³ Again the phonological similarities are readily apparent and the two forms are presented as at least nearly synonymous.

Example (4) provides a particularly interesting illustration of this kind of relationship:

- (4) 6 mimic
 A point out
 B hide
 C scribble
 D fly
 E imitate

The correct answer is *imitate*. By isolating the verb marker *-ate* we are left with *mimic* and *imit-*. Here there is a significant level of phonological similarity: [ImI] followed by a voiceless stop. But the relationship is more complex and interesting than that seen between *sieve* and *sift* because of the initial [m] in *mimic*. *Mimic* and *imitate* are semantically very similar, and in *mimic* the meaning of imitation may be reinforced by reduplication. In any case both words are phonologically very similar and are considered by the test designers as synonymous.

These test examples support the psychological reality of a relationship between formally and semantically similar roots. The degree of formal similarity between words regarded as synonymous by the compilers demonstrates a phenomenon which needs accounting for.

It is easy to ignore the problems posed by the data in this section: all these words could be regarded as unrelated. Alternatively, the complex network of relationships between them can be recognised and an attempt made to account for them.

3. To me *sieve* is simply a noun, while *sift* is the verb, what you do with a *sieve*.

Roots which are similar in meaning, but not in form, have a semantic relationship. Roots which are similar in form, but mean irreconcilably diverse things, have a phonological relationship. When two roots are both semantically and phonologically related, they may be variants of a single entity, and that is of significance to the morphology.

2.2.2 Variation by Ablaut

In the previous section the relationship between semantically and phonologically similar forms was discussed. On the face of it, a similar relationship appears to exist between forms such as:

(5) sing sang sung song

These words demonstrate significant phonological similarity. All consist of [s], followed by a vowel, followed by [ŋ]. Equally there is a great deal of semantic similarity. Indeed, the three are verbs differentiated only by tense. The fourth is a noun which has significant enough semantic similarity for it to be regarded as the noun of *sing*. A relationship exists between these four words which needs accounting for.

With the verb forms the relationship involves variation in vowel quality and commensurate change in tense. There are a number of irregular patterns of tense formation in English, and the data in (5) exemplify one of the most widespread. Here the past participle, past tense, and non-past tense are distinguished by the vowel. This is normally described as derivation by ablaut from the underlying non-past vowel. There are two problems with this. Firstly there is no phonological reason to assume that any one of these vowels is underlying and any other derived from that. The relationship between these vowels is the result of neither a general phonological rule, nor a morphophonemic rule attestable elsewhere. It is arbitrary to regard [æ] as a mutated form of [ɪ]: the reverse is no less likely.

The second problem is that there is nothing about the shape of verbs like *sing* which suggests that the non-past form is formally less marked than the past or past participle. To assume that the non-past form is less marked is to take the markedness status of the regular past tense formation (a completely different system of tense marking), and apply that markedness status to this pattern. Clearly *walk* is formally less marked than *walked*. There are no grounds for making the same claim about *sing*.

Moreover, *sing* exemplifies a regular, if closed, set of forms. This is in contrast with the relationship between, for example, the vowels in *do*, *did* and *done*, which are

entirely idiosyncratic. With *sing* the quality of the non-past vowel is regular. If a variety of vowels were replaced by the past and past participle vowels, there would be some grounds for regarding the non-past vowel as underlying, but that is not the case.

The consequence of this perspective is the statement that there is a closed set of verbs in English in which the non-past form is marked by the vowel [ɪ], the past tense by [æ] and the past participle by [ʌ]. To say any more requires bringing a baggage of assumptions to the data.

Turning to the noun form, the phonological and semantic relationship between *song* and the verb forms is significant. It is readily apparent that the two have a formal relationship not found between, for example, *travel* and *journey*. Again the only distinguishing feature is vowel quality. Again there are no phonological grounds for assuming this is either derived or underlying. In addition, it is not a unique alternation (eg. *lift* and *loft*). This kind of noun/verb alternation is less regular and less widespread than the *swim, swam, swum* pattern, but is nonetheless a readily identifiable phenomenon.

In the data in (5) semantic distinctions are realised by different vowel qualities. The assignment of underlying status to one is arbitrary or the result of assumptions based on other forms in the language. This is apparent if the forms are compared to forms which behave similarly. For example it is apparent from a comparison of *sing* and *sang* that it is the formal element [s-ŋ] that conveys the root meaning, and the components [ɪ] and [æ] which convey non-past and past. Equally, a comparison of *sing* and *begin* indicates that it is the formal component [ɪ] which conveys the non-past tense, and the components [s-ŋ] and [bæg-ŋ] which convey the distinct root meanings. Relationships like these need to be accounted for.

2.2.3 Formally Dissimilar Allomorphy

Phonologically diverse entities are generally regarded as allomorphs if they are structurally identical. It is clear, for example, that *went* is the past tense of *go*, but does that mean they are allomorphs of a single morpheme? There is an extremely close semantic and structural relationship between the two, yet they display total phonological dissimilarity.

Consider English plural markers. The regular productive forms [s~z~əz] have a close phonological relationship with each other. *Geese*, however, distinguishes plurality by vowel quality. A standard structural analysis⁴ would postulate the presence of an

4. For a fuller discussion of this kind of analysis see section 2.3.1.

zero suffix, an allomorph of the plural marker which co-occurs with a mutated vowel. However, no suffix is apparent. To distinguish number the hearer must rely on the vowel. While the vowel in *geese* fulfills a function identical to that of the productive suffix, the two display total phonological dissimilarity. This suggests that although they fulfill the same function, they are different morphemes.

Even where plurality is marked by a suffix, it may not be formally related to the productive suffix. The suffix in *phenomena* and *criteria* bears no phonological relation to the productive form. Indeed, the only formal commonality between the two is that they are both suffixes. It could be argued that in spite of that they are allomorphs, implying that allomorphy has no phonological limitations, and raising significant questions about the definition of 'morpheme'. Alternatively such forms may be regarded as synonymous but distinct morphemes, no more variants of the same morpheme than *transformation* and *metamorphosis* are variants of the same word.

2.2.4 Irregular Paradigms

It will be apparent from the discussion so far that many problems in morphemic analysis involve the presence or absence of relationships between forms, and the semantic and formal links upon which those relationships are based. In 2.2.3 *go* and *went* illustrated the fact that the phonological relationships between forms in an irregular paradigm do not necessarily parallel the paradigmatic structure. The paradigm for the verb *go* is:

(6)	Nonpast -3SG	Nonpast +3SG	Past	Past Part.
	goU	goUz	went	gon

It is immediately apparent that *go*, *goes* and *gone* have a phonological relationship not shared by *went*. Despite the synonymy (apart from tense), it appears that two separate roots are present.

English verbs typically form +3SG by inflecting an unmarked Nonpast, -3SG form, and that is regular here. The Past Participle, however, is not. The suffix, realised as [n] with a vowel-final root, uniquely co-occurs with a shortened vowel (unlike *know* and *known*).

This paradigm involves two suffixes and two separate roots, one of which displays vowel alternation and inflects with a subregularity. This network of morphemic relationships does not parallel the structural relationships within paradigm. While the paradigm provides a structural account of the data, it does not provide a morphological account.

The English pronominal system provides a much more complex illustration of idiosyncratic paradigm-internal morphemic relationships:

(7)	Nom.	Non-Nom.	Attributive Poss.	Absolute Poss. ⁵
1SG	aɪ	mi	maɪ	maɪn
1PL	wɪ	ʌs	a:	a:z
2SG	yu	yu	yɔ	yɔz
2PL	yu	yu	yɔ	yɔz
3SGM	hi	hɪm	hɪz	hɪz
3SGF	ʃi	hɜ	hɜ	hɜz
3SGN	ɪt	ɪt	ɪts	ɪts
3PL	ðeɪ	ðeɪm	ðe:	ðe:z
cat	kæt	kæt	kæts	kæts

A comparison of *cat* and the 3SGN forms reveal that *it* is inflected regularly. The remaining forms display various irregularities. To accurately reflect the distinctions made morphemically it is useful to rewrite the paradigm:

(8)	Nom.	Non-Nom.	Attributive Poss.	Absolute Poss.
1SG	aɪ	mi	maɪ	maɪn
1PL	wɪ	ʌs	a:	a:z
3PL	ðeɪ	ðeɪm	ðe:	ðe:z

(9)	Nom.	Non-nom, Non-Attrib. Poss.	Absolute Poss.
3SGF	ʃi	hɜ	hɜz

(10)	NonPoss.	Attributive Poss.	Absolute Poss.
2	yu	yɔ	yɔz

(11)	Nom.	Non-Nom.	Poss.
3SGM	hi	hɪm	hɪz

None of the forms in (8) to (11) can be generated by regular rules of English. However, within these paradigms there are a number of sub-regularities:

- (12) 3SGM forms the Possessive by the regular suffix. The remaining person/number categories distinguish Absolute Possessive by marking the Attributive Possessive with a suffix, which apart from 1SG is formally identical to the regular Possessive suffix. The 1SG suffix is unique.

5. I have used the term non-nominative to allow for the various uses of the so-called Accusative (such as Dative etc). To distinguish the two Possessive forms I have adopted the terms used by Huddleston (1984:294).

Assuming this suffixing, other subregularities include:

- (13) Second person does not distinguish number, and distinguishes only between Possessive and Non-Possessive by vowel height. Second person and 3PL are characterised by their own initial segment or segments, and that is also the situation with 1SG and 1PL, except for their Nominative forms. 3SG non-Neuter forms share an initial segment, except for 3SGF Nominative, which is distinguished from 3SGM Nominative only by initial segment. 3PL and 3SGM distinguish non-Nominative by a final /m/. Etcetera.

The kinds of relationships raised in (12) and (13) illustrate that while a paradigm such as a pronominal system consists structurally of a grid, the morphemic relationships are not necessarily exactly parallel. *They* is related to *them* in a way that *we* is not related to *us* ; and *them* is related to *him* (phonologically and semantically) in a way that neither are related to *me*. A morphemic analysis of the forms present in these paradigms must account for these kind of morphological facts. An adequate theory of morphemics must be able to at least recognise such relationships.

2.2.5 Semantic Problems

Semantics poses numerous difficulties for morphology. So much so that some morphological theories resort to regarding meaning as incidental. A significant problem in morphology lies in determining whether meaning is an essential characteristic of morphemes, and if so what the nature of the relationship is between morphemes and their meanings. This will be discussed at length in 2.3, particularly in the section on Aronoff. The present work regards the relationship between form and function as crucial to morphology.

Perhaps the classic illustration of the issue of morpheme and meaning is the problems posed by cranberry morphs. Spencer (1991:40) claims that such a morpheme "has neither meaning nor grammatical function, yet it is used to differentiate one word from another." Aronoff (1976) in part bases his theory on the meaninglessness of elements like *cran-*, yet even he observes that this entity "in a sense told us what sort of berry we were dealing with". (1976:11) However meaningless it may appear in isolation, an entity like *cran-* communicates semantic information: that this berry has the real world characteristics XYZ, which distinguish that berry from others. If 'meaning' is the communication of information, then cranberry morphs are meaningful.

If cranberry morphs seem like form without meaning, zero marked changes in word class seem like meaning without form. This is sometimes regarded as the result of a zero morpheme. This is not, however, very useful, since it is impossible to tell whether the word class marker is present except by context, in which case the word class is apparent anyway.

Alternatively it is argued that rules cause words to change class. Syntactically such rules may be useful, but they are semantically irrelevant. In terms of the morphemic representation of meaning there are no grounds for regarding a phonologically unmarked word as a noun, and also separately a verb.⁶ As a morphemic object such a word is a single entity, indistinguishably a member of both word classes.

The relationship between meaning and morpheme is discussed in more detail elsewhere in the present work.

2.2.6 Phonesthemes

Throughout many of the most common English words there are small elements which display a sound-meaning correspondence. Generally referred to as 'phonesthemes', these entities are identifiable by means of recurrent association, and operate below the level of a structurally defined morpheme. Bloomfield (1933:245) exemplifies the phenomenon with the onsets in *flash, flare, flame, flicker* etc ('moving light'); in *flap, fly, flit* etc ('movement through air'); in *glow, glare, gloom, gleam, glint* etc ('unmoving light'); and in *slime, slush, slobber, slip, slide* etc ('smoothly wet'); and the rhymes in *bash, crash, gash, slash, mash* etc ('violent movement'); and in *bump, chump, frump, dump, lump* etc ('clumsy').⁷

Many English phonesthemes correspond to phonologically similar elements found throughout the Germanic family, and the phenomenon is evident throughout

6. Unless the word classes are represented by separate allomorphs for other reasons: viz *knife* (verb) becomes *knives* (3SG) while *knife* (noun) becomes *knives* (plural).

7. Bloomfield (1933:245) regarded phonesthemes as "*root-forming morphemes, of vague signification*" (his italics). This position was subsequently rejected by post-Bloomfield structuralists, exemplified by Nida's Principle 6.3 (1949:61), on the grounds that phonesthemes "do not occur with free forms or with forms which occur in other combinations". The view that phonesthemes are not morphemes because they are not isolatable by paradigmatic contrast is flawed on two counts. First the distinction between recurrent association and paradigmatic contrast is a distinction between types of occurrences, not types of entities. Secondly phonesthemes are partially paradigmatic: with *flow* and *flitter*, for example *flow* suggests a continuous, uninterrupted and directed movement. *Flitter* on the other hand seems to suggest constant change of direction and pace. Now consider *glow* and *glitter*. Both refer to a kind of emission of light. *Glow* suggests an unwavering sort of emission, while *glitter* is quite the reverse. These four words may be placed into a paradigmatic structure:

	'movement'	'emission of light'
'evenness'	flow	glow
'erraticness'	flitter	glitter

Indo-European languages and beyond.

It could be argued that in the total vocabulary of English it would be possible to arbitrarily collect words which share a few phonemes and can be squeezed into a single semantic group. However, the kinds of examples given by Bloomfield are common words, and representative of large groups of words which share sound-meaning associations.

For example, Bloomfield analyses the initial cluster /fl/ as having two meanings: 'moving light' as mentioned already, and 'movement through air' as in *flap, fly, flit* etc. In fact an analysis of the total number of /fl/ initial words reveals an overwhelming presence of a general sense of movement. In an effort to determine how widespread this particular phonestheme is, an analysis was made of a four dictionaries.⁸ In each, the total number of entries with the initial consonant cluster 'fl' was analysed semantically, and divided into three categories: those entries with meanings which clearly involve movement, those which clearly do not involve movement, and those about which there was some uncertainty.

Entries were included in the 'move' category if they referred a) to movement or a type of movement, either as a verb or a noun (*fly, flight*); b) to something which has some element of movement as a key semantic feature (*flipper*); or c) to something which has some sense of movement as a feature of the referent (*flue*).

(14)	Dictionaries	total entries	move	non-move	unsure
	Concise OED	260	53.1%	33.5%	13.4%
	Websters enc'lic	371	60.1%	30.2%	9.7%
	Greens Jargon	203	66.5%	30.1%	3.4%
	Macquarie New Words	39	76.9%	20.5%	2.6%

In each case the entries involving movement account for more than both the non-movement and the doubtful categories combined. This majority ranges from slightly more than half the total entries to slightly more than three quarters. This is clearly not merely a representative cross section of English: it would be ludicrous to suggest that somewhere between a half and three quarters of all English words mean something to do with movement. The figures in (14) suggest that there is more than simply a passing or coincidental correlation of this cluster and its proposed meaning. This phenomenon is widespread in English.

8. The dictionaries used were Webster's Encyclopedic Dictionary (1976), the Concise OED, the Macquarie Dictionary of New Words (1990), and Jonathon Green's Dictionary of Jargon. The analysis excluded proper nouns and their derivatives (such as *Flanders* and *Flemish*). Also excluded were initialisms (as opposed to acronyms): for example from the entries in the Green book *FLK* [ef el kel], from *funny looking kid*, was excluded, while *flip* [flɪp], from *floating instrument panel*, was included.

This preponderance of movement related entries is made more marked by the fact that the 'non-movement' entries include several large groups of words which are related to single non-movement roots. For example in Webster's, out of 112 non-move entries, 54 (48.2%) fall into either the *fleur/flora/flower* group, the group of chemical terms involving *fluor-*, or derivatives of *flat*. A further 28 (25%) fall into one of another three groups. Thus nearly three quarters of the non-movement entries are related to just six roots. This serves to reinforce the view that in a general sense there is an association of the cluster 'fl' and the semantic value move. I suggest that the testing of speakers on /fl/ initial nonsense words would result in a high proportion of suggested definitions involving movement.

This phenomenon is not merely a matter of fossilised hangovers from earlier periods in the language. Blust (1988:67), discussing the phenomenon in Austronesian, asks: "How are (such) sound-meaning associations passed on through the generations...?" This question is both prompted and answered by the productivity of the forms. As a language evolves, among the kinds of words which potentially contain phonesthemes, those which lack phonesthemic strength are more likely to be lost, and less likely to be borrowed than those which conform to a possible phonesthemic analysis.

The Dictionary results support this view. The OED, based on historical principles, and with several anachronisms among its entries, demonstrates the smallest majority of 'fl' initial entries with 'movement' meanings, while the Macquarie Dictionary of New Words has the largest.

The process of language change can be driven by phonesthemic force, either by the phonological shape of a word changing to match a phonestheme relevant to the word's meaning; or by a word's meaning shifting to incorporate the semantic value of a phonestheme which is homophonous with some part of the word.

Table (15), reprinted from McCune (1988:93) (citing Wescott (1979)), exemplifies phonesthemically driven phonological change.

The examples come from at least four separate historical sources, and all 11 originally had distinct phonemic forms. The phonemic shape of the rhyme of 10 of the 11 underwent a change, resulting in the present correspondence of sound and meaning throughout the set.

(15) Sources and Forms of English -ash 'violent motion'

Source Language	Phonemic Form in Source Language	English Word From This Source
Old English	- aesk	rash
Old English	- a:ks	mash
Middle English	- ast	gnash
Middle English	- ais	abash
Middle English	- ars	gash
Middle English	- as	flash
Middle English	- esh	thrash
Old French	- ash	hash
Danish	- ask	bash
Frisian	- ats	clash
Frisian	- atsk	(s)plash

With the second kind of change the apparent presence of a phonestheme can cause speakers to reanalyse meaning. While *flay* appears in the OED as 'to skin', its current Australian use is 'to whip or beat'. The strength of the association between /fl/ and movement has caused a shift from a meaning of 'to skin' to the movement causing skinning, and finally to just that whipping motion.

In addition to language change, phonesthemes play a role in the creation of new words. Acronyms which are formed from words containing initial phonesthemes typically retain the entire phonestheme, rather than just the first letter. The word *flak* is an excellent example of this. Coined in German, the initial consonant cluster was retained from *flieger*, meaning 'flier'. (The /fl/:'movement' phonestheme is also found in German.) *Flak* contains not only this initial phonestheme, but also the rhyme [æk], with its meaning of dangerous impact (as in *smack*, *whack* and so on). It is noteworthy that the word was not only borrowed into English, supplanting locally coined terms, but has undergone a broadening of meaning, now referring to any incoming danger, including criticism.

Phonesthemes also appear to be productive in what are entirely new phonological coinings. Green's Dictionary of Jargon lists three /fl/ initial words which are entirely new to me, all involving movement:

- (16) *flodding n* [Communications] in packet-switching, this is a routing method in which each node reproduces in-coming packets and sends duplicates to its neighbours; this ensures swift and accurate communications, but uses up a very large amount of transmission capacity.
- flugie n* [US] [Politics] a rule that benefits only the creator of that rule, and one that can always be changed as required by the creator to ensure that no-one else is ever able to use it to their own advantage.
- fluther v* [Science fiction] to move around aimlessly.

The semantic strength of phonesthemes is demonstrated by their appearance in language tests. The tests discussed in 2.2.1 exemplify this.

Each question involves a given word, and a set of five options for the answer most nearly the same in meaning. The makers of these tests included among each set of options a foil which resembles the question word phonologically. In every question where a phonestheme appears in the given word, the same phonestheme appears in the answer options, either as a foil, or as the correct answer, or as both. In (17) the given word and the foil share not one, but variants of the same two phonesthemes.

- (17) **sparkle**
 A star
 B bright
 C glitter
 D sprinkle
 E diamond

The initial consonant cluster /spr/ is part of the set of variants *sp-*, *spl-* and *spr-*. This phonestheme may be given the meaning 'the dispersal of small element(s) of a greater whole', and is found in *spatter*, *spit*, *spark* etc; in *splatter*, *split*, *splash* etc; and in *sprawl*, *spray*, *spread* etc. The residues *-arkle* and *-inkle* are variants in a set including clusters found in *speckle* and *spangle*, as well as in *prickle*, *tickle*, *tinkle*, *tingle*, *jingle*, *jangle*, *jungle*, *wrinkle*, *wrangle*, *mangle*, *tangle* and *bungle*. In each case there is a semantic component of 'unordered diversity'. *Sprinkle* shares with *sparkle* variants of their two component phonesthemes.

The fact that phonesthemically related words were chosen as answer options in these tests indicates that although the semantic values of the phonesthemes are more nebulous than simple referents, they lack nothing in strength of semantic association.

Phonesthemes may be characterised as recurrent sound-meaning correspondences: small, meaningful, apparently root forming entities which have a single, if broad, meaning, and which may be isolated by means of recurrent association. They are common, perhaps even pervasive, and are productive to the extent that their meanings are constant enough and strong enough to drive language change.

Appendix II provides additional evidence supporting the linguistic reality of phonesthemes.

2.3 Existing Theories

The present work is concerned with the nature of morphemes. In any attempt to resolve morphemic problems it is crucial to define 'morpheme', and its role in morphology. The varying approaches taken by different morphological theories when faced with morphemic problems depend in many cases on varying notions of

'morpheme', and on differing views on the relationship between the lexicon and word formation. This section examines the major existing approaches to this matter. No comment will be made here on word formation or morphological structure, except where those issues have an impact on defining the concept of morpheme. A general understanding of each theory is assumed.

Broadly speaking, there are two kinds of approaches: those which regard morphemes as the atoms of word formation; and those which regard morphology as processes by which words are derived, in which morphemes are the phonological realisations of rule applications.

The view of morphemes as atoms was held by many within the structuralist school, who were concerned with defining morpheme, but who substantially limited the role of meaning. More recently, morpheme-based approaches have been proposed within a generative framework. This path was commenced upon by Halle, and now finds widespread acceptance in Lexical Phonology.

A widely accepted process-based alternative to LP is Aronoff's generative approach to word formation. This and other process theories regard morphemes as having no independent life. They are listed as part of rules, and surface in words derived or inflected by those rules.

Within the generative morpheme based approaches there is a tendency to assume that the notion of morpheme has been defined, and that morphemes are readily identifiable. Indeed, current morpheme and process based theories both regard morphology as about word formation, not morphemes. Consequently few modern theories deal with morphemics at all.

2.3.1 Structuralist Morphemics

For many in the structuralist school, utterances were analysable on distributional grounds into a series of morphemic entities called 'alternants' or 'morphs'. According to Harris (1951:189) "all that is possible...is to set up the morphemic segments purely on the basis of the relative distribution criteria...". To Harris (1951:207) both *knife* and *knive-* (as in *knives*) are morphs, and are allomorphs. They are conflated into a single morpheme, not because they are virtually identical in form; nor because they are identical in meaning; but because they appear in complementary distribution. *Knife* is not conflated with *wive-* because its total distribution environments are not parallel: only *knive-* occurs in *I'll sharpen my ---s on the whetstone*, paralleling *I'll sharpen my --- on the whetstone*.

This exemplifies a key problem with structuralist approaches to morphemics: the role of semantics is downplayed, while remaining of inescapable significance. While morphemes must be meaningful, their meanings must play no direct part in their analysis.

Clearly *knife* is related to *knive-* in a way that it is not related to *wive-*: *wive-* shares fewer formal features, and has an unrelated meaning. To demonstrate this link without resorting to phonological and semantic considerations, distributional dissimilarities must be demonstrated between *knife* and *wive-*, and distributional parallels demonstrated between *knife* and *knive-*. But the distribution of all three are frequently parallel. It is impossible to distinguish on distributional grounds between *knive-* and *wive-* as the parallel for *knife* in the sentence *their ---s are in the house*. Consequently the analyst must construct sentences like the *whetstone* example to demonstrate the relationship. Yet even in those examples the distinction is actually semantic. *I'll sharpen my wives on the whetstone* is strange, but it is grammatically possible. The sentence is 'wrong', not on distributional grounds, but because under normal circumstances it doesn't make much sense. The distributional parallel actually demonstrates that *knife* and *knive-* mean the same thing. The 'distributional' bar is in fact semantic.

Harris isolates and conflates morphs, ostensibly on distributional grounds, but in fact to do so must covertly resort to meaning.

Hockett pursued a less extreme view than Harris (1951), a view more in keeping with Harris's (1942:115) earlier approach. Here the reference to semantics in morph isolation and conflation is less covert: morphs are isolated as "the smallest parts which recur with the same meaning in different expressions". Hockett (1947b:241) claims morphs may be conflated if they fit three criteria: they have a single meaning; they occur in non-contrastive distribution; and they have a range which is not unique. This prevents the conflation of semantically distinct entities on the basis of accidents of distribution; and recognises, for example, that *knife* and *knive-* are synonymous. However the distribution criteria must still be met.

It is a tenet of structuralism that the range of forms present is dictated by the entities present in the most complex paradigm in the language. For example, a subset of English verbs distinguish the past participle by means of the suffix *-en*. Since these verbs distinguish between the simple past and past participle, there is a distributional requirement that all verbs must do so. Therefore in *walked* the past participle is marked by the morpheme *-en*, realised as its allomorph *-ed*. That allomorph is homophonous with the completely separate morpheme *-ed* which has a different meaning: simple past. Because *prove* ~ *proved* ~ *proven* distinguishes between simple past and the past participle *walk* ~ *walked* ~ *walked* must do the

same, *walked* thus contains one of two indistinguishable but distinct morphemes. The conclusion that *walked* simply does not distinguish between the categories of past and past participle is not possible.

Equally, logic dictates that since the paradigm for the verb *to be* contains *am*, every other English verb must also distinguish first person singular nonpast from all other nonpast forms: *walk* (1SG) and *walk* (1PL) must be homophones. As Bazell (1949:225) noted, the view "that a distinction in however few paradigms implies homonymy in all those in which the distinction is not made, would involve a denial of the privative nature of oppositions expressed by morphemes, and could of course not be carried to its logical conclusion in any grammatical description."

Hockett (1954:389) lists a range of types of morphs as "devices" which may be used "where they prove convenient". This is not an introduction which engenders confidence in the linguistic reality of what is being proposed. Such devices include, among other things, zero, portmanteau and empty morphs. Zero and portmanteau morphs are a direct consequence of the primacy of distributional criteria in the approach, and pose a number of problems. The concept of empty morphs is so fraught with difficulties that it fails to stand even within structuralist assumptions.

The liberal use of zero morphs is a direct consequence of the parallel distribution approach: *cat* contrasts with *cats* by a plural suffix, so there must also be a parallel distinction between the singular and plural *sheep*. The plural *sheep* must be marked by a plural suffix, occurring here as an invisible allomorph. This involves proposing an invisible morph which is not isolatable by contrast and serves to distinguish neither meaning nor form, but simply satisfies the paralleling requirements.⁹

This difficulty is made even more apparent when a zero morph is claimed to be present in a form which already overtly marks the relevant category. The word *took* is dealt with in a number of different ways, several involving zero morphs. One, discussed by Nida (1949:54), involves regarding the vowel as a replacive morpheme marking past tense, occurring along with a second past tense marking morpheme, a suffix consisting of a zero allomorph of the regular *-ed*. The rationale for this is that the past tense form is distinguished by means of the vowel, so that must have some tense marking significance, but equally, the form must contain the tense marking suffix to satisfy paralleling requirements, so that suffix is realised as a zero.

9. This kind of zero is entirely different to a paradigmatic gap (or meaningful absence). In the case of a meaningful absence, the failure to meet verb marking expectations, for example for subject person and number agreement, indicates a specific person/number category: that category is identifiable even when the word is in isolation as it contrasts with overtly marked forms. This is an entirely different phenomenon to the proposed zero plural marker in *sheep*, which is primarily characterised by the absence of any formal contrast with the unmarked form. See 4.12.

Took is marked for tense by two separate morphemes, one serving communicative purposes, and the other satisfying structural requirements. Since there are two past marking morphemes, logic dictates that as a consequence of paralleling, all verbs must take both, with regular verbs containing a zero replacive plural morpheme paralleling that in *took*.

The alternative zero marking approach is to regard *took* as an allomorph of *take*, with the distributional environment of preceding the tense marking suffix. This suffix is realised as a zero allomorph. In other words the vowel does not mark tense, it merely points to the presence of an invisible tense marking suffix. This approach is rejected by Hockett (1954:394) on the grounds that under the rules that would allow this analysis, "it would also be possible to propose that took be interpreted as an allomorph of /ed/ accompanied by a zero allomorph of take." Hockett recognised that this solution ignores the special relationship between the phonological forms *took* and *take*.

Hockett (1954:394) also recognised that an analysis employing zero morphs "is arbitrary because it assigns to something that isn't there phonemically...a meaning which would more naturally be assigned to some location within the overt form..." Nida (1949:46) warns that "One should...avoid the indiscriminate use of morphemic zeros. Otherwise, the description of a language becomes unduly sprinkled with zeros merely for the sake of structural congruence and balance." Despite this warning Nida makes widespread use of zeros, and despite Hockett's attempts to reject the whole notion of zero morphs, they remained an integral part of structuralist practice. Their problematic nature recognised, they remained an inevitable consequence of structural parallelism.

Hockett (1954:393) canvassed several other solutions to the *take~took* problem. In one *took* is a single morpheme, but this violated the underlying principles of structuralism: *took* had to represent both the root and the past tense morphemes. A further solution posited *took* as a portmanteau representation of the two morphemes. This solution gained widespread acceptance although (or, according to Hockett, because) it "avoids" noticing that *took* and *take* are phonologically similar.

Portmanteau morphs posed similar problems to zero morphs. If two categories are represented by separate morphemes, but in one paradigm are represented by a single morph (perhaps a single phoneme), that morph belongs simultaneously to both the separate morphemes. Hockett (1947b:241) analyses *mine* as a single morph representing the morphemes *I* and *-s*. The pronoun *my* is analysed in an identical manner. The rationale for this is a paralleling of *John's book* and *the book is John's* with *my book* and *the book is mine*. Thus while *my* and *mine* are separate morphs, they are morphemically identical, both uniquely representing the same two

morphemes in a portmanteau manner. This analysis ignores the fact that *my* and *mine* are neither formally identical nor overtly divisible. The distributional parallel requirement precludes analysing the forms as not parallel to other forms, which is overtly the situation.¹⁰

The reason for this highly counter-intuitive analysis is clear when one considers the enormous number of zero marked case forms of nouns required to parallel the English pronominal system. This analysis does violence to underlying structuralist assumptions: such paralleling is obligatory, and the only way an absurdly complex system of identical noun forms can be avoided is by using portmanteaux to cause the pronominal system to make fewer distinctions.

Hockett (1954:394) identifies a further possible solution to the *take~took* problem: a replacive morph /u/ <- /ey/. This has the advantage of recognising both the phonological relationship between *take* and *took*, and the role of the vowel in conveying the change of tense. It also avoids attributing tense marking to a zero while the forms display an overt difference. Hockett, however, rejects this, saying that a replacive "is not by any stretch of the imagination composed of phonemic material."

The solution favoured by Hockett involved an allomorph of *take* with the form *t...k*, and an infixing allomorph of *-ed* with the form *-oo-*. This has some of the advantages of the replacive solution, but involves regarding the relationship between the suffix /ed/ and the infix /u/ as the same as the relationship between the suffixes /ed/ and /d/. While all three have an identical function, this solution does not recognise that the suffixes have a formal relationship not shared by the infix. A second problem is that the solution is not meant to imply that the vowel in *take* also marks tense. However, for /u/ to be tense contrastive, logical demands that /ey/ must also be contrastive: *take* is recognisably not past tense because of the vowel. The solution accounts for *took*, but fails to explain how *take* marks nonpast.

Phonesthemes were particularly problematic for structuralist models. Bloomfield (1933:245) noted the recurrently associative meanings of these forms, describing them as "*root forming morphemes, of vague signification*". Later structuralists rejected this, phonesthemes not meeting criteria of parallel distribution. They are specifically excluded by Nida's (1949:61) Principle 6.3, as "they do not occur with free forms or with forms which occur in other combinations". However, phonesthemes do occur with forms found in other combinations.¹¹ Harris (1951:193) claims that

10. Similar problems apply to *I* and *me*: Hockett (1947b:242) claims both represent a single morpheme which does not distinguish case, and this provides evidence for the complete absence of any morphemic case marking in English. This ignores the fact that *I* and *me* do indeed distinguish case (or grammatical function).

11. They are partially paradigmatic. See 2.2.6.

phonesthemes display "a correlation between meaning and phonemic form, of the type which is also true for most...distributionally separable morphemes..." He goes on to describe them as productive. Despite all this, since "no adequate distributional basis can be found for supporting this segmentation" a phonesthemic sequence "is not a distributionally separable element; therefore it is not a morpheme..." He is not entirely at ease with this rejection, saying it is important not to leave these morpheme like characteristics unstated.

It is not necessary to discuss the notion of empty morph in detail here. The idea was proposed by Hockett (1947b) and was so problematic it was dropped almost immediately, except occasionally in name. Moreover, contrary to a widespread current misapprehension, it did not refer to cranberry morphs. The term referred to small elements of phonemic material present between two attestable morphemes, and cannot be described as a feature of structuralist theory.¹² This is as well, since no discussion was undertaken (to my knowledge) of its effect on the distinction between morphology and phonology.

Despite this rejection of meaningless morphs, semantics remained the main sticking point in structuralist theory. Meaning was regarded as entirely subservient to distribution. According to Hockett (1958:137-142) language consists of three central and two peripheral subsystems. The central ones are morphemes and their arrangement; phonology; and morphophonemics. They are central because "they have nothing to do, directly, with the nonspeech world in which speaking takes place." The two peripheral subsystems are phonetics and semantics. Utterances have no semantic structure, semantics residing in "habits". It is "made manifest...by the relationship of the grammatical structure of an utterance to the context in which the utterance occurs." Harris (1951:189) clearly limits the role of semantics: "linguists often use apparent differences or identities of meaning...as hints in their

12. The term empty morph was never said to apply to cranberry morphs, which were never claimed to be meaningless. Harris (1951:192) states that, in *boysenberry*, *berry* is a morpheme. "Therefore, *boysen* is also a morpheme, having as its meaning the differentia between boysenberries and other berries." Hockett (1958:126) echoes this, adding that "it might be hard to describe this meaning, but it is easily demonstrated in a fruit market." Empty morphs were proposed by Hockett (1947b:235-236) to account for isolated phonemic elements left by morphemic analysis. He gives as an example the /r/ in *children*. Both *child* and the plural *-en* are independently attestable (allowing for vowel variation in the root). Isolating these two morphemes leaves the single consonant residue which, if attributed to either morpheme, would generate a unique allomorph. The proposal is highly problematic, particularly in the structuralist framework, in which all morphemes must be meaningful, and all utterances must be composed entirely of morphemes. (cf Harris 1942, 1951; Hockett 1954; 1958) In Hockett (1950) he makes no use of the idea of empty morphs, saying instead that the phonemes are non-morphemic. In Hockett (1954:387) he states that "one assumes that any utterance in a given language consists wholly of...morphemes". He goes on to say (p389) that all phonemic material in an utterance is accountable for in terms of the morphemes present. On the same page, however, he pays lip service to the notion of empty morph as a possible solution to the problem of these feral phonemes. He states (p387) that "we may recognise a morph which belongs to no morpheme... Or, instead of this, we may simply classify some of the phonemic material in some utterances as nonmorphemic..." This is at odds with the underlying tenets of structuralism: a throwaway line contradicting every other facet of the approach presented in that paper. He makes no further use of the idea, and it is maintained elsewhere in the body of structuralist literature that all morphemes are meaningful and all utterances consist entirely of morphemes.

search for morphemic segments. However these hints must always be checked with the operations of 12.2 [the distributional criteria] if the resulting segment is to satisfy the purposes of our procedures, so that meaning never functions as a fully-fledged criterion for morpheme segmentation..."

Structuralist theory was driven by the need for practical tools for the analysis of exotic languages, and its difficulties with semantics stem from the lack of an effective means of analysing such languages semantically. Structuralism provided a number of useful insights into morphemics, principally its recognition that in every language there exists a set of entities which are functionally (viz meaningfully) undecomposable. It devoted considerable effort to describing and accounting for these entities. Consequently the approach recognised the role of morphology in grammar.¹³ Structuralism also had numerous failings, proceeding largely from its inability to recognise and account for similarities or dissimilarities in form or meaning between entities, other than those defined by their distributions. This failing was the inevitable consequence of an approach which regarded semantic and even phonological issues to be subservient to rigorous structural parallelism.

2.3.2 Halle's Prolegomena

"In the early days of generative phonology and syntax, the assumptions of structuralists about the nature of words and their component parts were not so much attacked as ignored; and when morphology came to be studied again, these basic notions were simply taken over unexamined as defining the subject matter."
(Anderson 1988:151)

The early generative approach (Chomsky 1965; SPE) divided the grammar into the phonology and the syntax. Apparent morphological phenomena were explained phonologically or by transformations. To a large extent the rules proposed in SPE account for phonologically motivated morphological alternation, and provide a means of removing such alternations from the lexicon. The transformational resolution of morphological problems was less successful. When Chomsky (1970) applied the approach to certain phenomena, notably English derived nominalizations, it became apparent that there were idiosyncrasies associated with derivation at least, which could only with difficulty be accounted for transformationally. Chomsky consequently claimed that derivation took place in the lexicon not the syntax. Thus the grammar needed a morphological component.

13. If anything it placed too much emphasis on morphology. According to Hockett (1958:129) the grammar of a language consists of its morphemes and their arrangements.

This problem was taken up initially by Halle (1973), who, proceeding from the perspective of generative phonology, identified three questions he believed the morphological component needed to address: how the grammar codes the inventory of actual as opposed to possible words; morpheme order; and the idiosyncratic features of words.

Halle (1973:5-6) proposes a model of morphology which "consists of three distinct components: a list of morphemes, rules of word formation, and a filter containing the idiosyncratic properties of words." The morpheme list (lexicon) and the WFRs generate every possible word in the language. The filter then operates on this overgenerated list to cull it to the list of actual words, called by Halle the 'dictionary'.

Halle (1973:6) regards possible words which do not actually occur as existing, but "incapable of appearing in any actual sentence of the language." Generated by the lexicon and WFRs, the filter records their idiosyncratic features. Halle's approach is concerned with lexical and phonological idiosyncrasy, and those idiosyncrasies are manifest on words after the word formation process has taken place. This involves a complex model with a considerable number of operations, many of which reverse previous overgenerations. The lexicon and WFRs generate *arrival*, also **arrivation*, and presumably **arriveness*, **arrivesis*, **arrivance*, **arrivement*, **arrivery*, **arrivage*, **arrivitude*, **arrivehood* etc, all constrained by the filter as [-Lexical Insertion]. Similarly *obesity* is generated with a short second vowel, which the filter then undoes by the constraint [-Trisyllabic Shortening].

This complication is avoided by preventing such overgeneration. To provide constraints prior to word-formation involves having the constraints operate on morphemes not words. This involves regarding it as an idiosyncrasy of *arrive* that it forms a nominal by co-occurring with the suffix *-al*.

Equally, it complicates the analysis of *obesity* to say that the absence of trisyllabic shortening is a feature of the collocation of *obese* and *-ity*. It is considerably less complex to say that *obese* does not undergo trisyllabic shortening. This would then prevent the generation of the shortened vowel, rather than generate it and then undo the effect of the rule. The number of rule operations relating to trisyllabic shortening occurring with *obese* would be nil, rather than the two required by Halle.

A major area of difficulty in Halle's approach lies in the location of the storage of idiosyncrasies. For Halle's filter to reduce the set of possible words to a dictionary of actual words it must store every single possible word in the language, actual or not, to the *n*th derivation, and every idiosyncratic feature of every one of those words. This list would have to be infinite. This also means that Halle's model requires the morphology to contain three lists: a list of morphemes, a list of actual words, and

the filter.

In addition, Halle has WFRs recording idiosyncratic information. He proposes (1973:10-11) that changes in syntactic category be recorded among the WFRs with a form such as [STEM + *al*]A. This template allows the suffixing in *arrival*, and causes that word to behave as an adjective. However, this is not really a rule, but rather a statement of certain features of the suffix *-al*. This information could just as easily form part of the lexical entry for that morpheme, removing the duplication of listing idiosyncratic information about *-al* in both the morpheme list and the WFRs.

The multiple location of idiosyncratic information is not, however, the approach's principal failing. Halle states explicitly that words are decomposable into small units of sound and meaning, and that the behaviour of these morphemes must be accounted for. He assumed (1973:3) that "a grammar must include a list of morphemes", but his concern is only with "the character of [word formation] rules and their relationship to other parts of the grammar." Having assumed that there is an isolatable set of morphemes in any language, the matter of defining or characterising 'morpheme' was not considered an issue.

Halle demonstrated that there are morphemes, and that the grammar has a morphological component, and that this is compatible with generative linguistics. However, since the generative approach rejected the tenets of structuralism, and Halle offers no replacement definition of morpheme; and since in his model many morphological idiosyncrasies are located in the filter and the WFRs, Halle's reader is left wondering exactly what is contained in the morpheme list he proposes.

2.3.3 Lexical Phonology

"According to a widespread view the lexicon is a kind of appendix to the grammar, whose function is to list what is unpredictable and irregular about the words of a language. In more recent studies it has been acquiring a rich internal organization of its own and is becoming recognised as the site of pervasive grammatical regularities." (Kiparsky 1983:3)

The theory of Lexical Phonology proceeds from a recognition that elements of the phonology interact with elements of the morphology in a way which is regular and predictable. It also assumes, like Halle, that that which is "unpredictable and irregular about words" includes a list of morphemes. These two factors led to an approach which is primarily a theory of phonology, but which has an impact on other domains.

Initially a set of observations about the interaction between the phonology and the morphology, LP is now considered by many of its proponents to be applicable to a theory of morphology, and even to syntax.¹⁴

LP was intended to account for the way phonological processes apply to morphological data. A strata based structure was proposed within the morphology, on the basis that it would facilitate the observed phenomena of morphological/phonological interface. That structure was then assumed to be inherent to the domain of morphology beyond its phonological requirements. According to Kiparsky (1982a:21) "Even if the phonological evidence were set aside we should still have to adopt [the proposed structure] on morphological grounds alone. The fact that the same disposition of levels is also necessary for the phonology supports the proposed theory of the lexicon." The first part of this claim is based on the fact that it is possible to analyse certain morphological processes (such as compounding and some affixation) within the phonologically driven strata structure. The second part is the claim is circular: it justifies this phonologically driven approach to morphology on the grounds that it is also applicable to phonology.

Despite being "the dominant morphological theory in North America" (Hammond and Noonan 1988:6), LP is not a theory of morphology so much as a theory of the role of phonological rules and processes in word formation. It proceeds from a concern with the manner in which phonological rules operate upon morphemes, rather than the nature of morphemes themselves. The structure of the module is predicated on what is most suitable for dealing with this interface of phonology and morphology. While LP may be effective in that regard, it is incapable of dealing with many of the morphemic problems raised in 2.2.

LP is concerned with the relationship between forms only in terms of phonological derivation, and where no such relationship can be demonstrated then no relationship is deemed to exist. Kiparsky (1982a:31) states that "we would want the theory to be able to rule out a lexical relationship between such pairs of words as *father* and *paternal*, *ear* and *hear*, and *semen* and *seminary*, in these cases presumably on the grounds of phonological, morphological and semantic complexity respectively."

In the case of 'semantic complexity' it is certainly justifiable to regard the purely homophonous relationship between the roots in *semen* and *seminary* as of no lexical significance. The same cannot be said for the pairs in *father* and *paternal* and *ear* and *hear*. Both pairs share significant phonological and semantic features, and

14. Mohanan 1986:ChVI proposed that LP could readily be applied to the interface between phonology and syntax, and proposed a model of the grammar which applied the approach throughout.

have the kind of relationship discussed in 2.2.1.

Clearly *father* has a significant semantic relationship with *paternal*. The roots shares canonical structure, and words demonstrates phonological similarities paralleling those found in *brother~fraternal* and *mother~maternal*. Kiparsky rejects these formal links on the grounds of unproductivity. Nonetheless, links exist, and these need accounting for.¹⁵ Equally *ear* and *hear* have a readily apparent semantic and phonological relationship. LP fails to account for relationships such as these.

LP is capable of dealing only with certain kinds of morphological phenomena. This, however, is not its principal failing. Its underlying flaw is a failure to address the notion of 'morpheme'. Kiparsky (1982a:3) describes LP as an "approach to word structure", and Mohanan (1986:55) defines morphology as "the study of the structure of words". LP recognises the need for a morpheme list, but employs it merely to list the roles of morphemes in certain derivational processes. Like Halle, LP assumes that 'morpheme' is defined; that morphemes are preexisting entities, the characteristics of which are known. As well as dealing with word formation, any adequate morpheme-based theory of morphology must also address the notion of 'morpheme'. LP does not do so, and consequently is not a general theory of morphology: it is a theory of morphological structure, but by no means a theory of morphemics.

An additional and less significant failing is the virtually interchangeable use of the terms 'morphology' and 'lexicon'. In LP the lexicon is the site of the operation of rules and processes, and the output of the lexicon equals the output of the morphology. The lexicon is often regarded as a component of the morphology in which idiosyncratic information about morphemes is stored. The blurred line between the terms is a consequence of the disregard for non-phonological idiosyncrasies.

Kiparsky (1982a) initially proposed that LP include the operation of Word Formation Rules. In Kiparsky (1983) this position was rejected in favour of Lexical Insertion. This is a strength of LP. If a theory is to take the view that morphemes are entities, then those entities must be stored along with their idiosyncratic features. A theory of word formation in which affixes are stored as features of Word Formation Rules, rather than the reverse, fails to recognise and account for affixes as morphemes, thus precludes the listing of affixing morphemes in a manner consistent with that of other kinds of morphemes, in which their various formal and semantic features are stored, and not merely their roles in word formation. The development of the LP model to involve Lexical Insertion renders the model potentially compatible with a

15. It will be noted that *father* and *paternal* have different etymologies. This fact is of no significance in synchronic analysis. See the discussion in 3.14.

theory of morphemics.

LP attempts to account for phonological and morphophonemic alternation, but ignores non-phonological variation. An adequate theory of morphology must address the notion of morpheme, and must be capable of at least attempting to deal with all manner of morphological phenomena. In an admirable effort to account for the "pervasive grammatical regularities" in morphology, Lexical Phonology forgot the irregular and idiosyncratic. LP includes significant insights into the interface of phonology and morphology, it is not, however, an adequate theory of the entire domain of morphology.

2.3.4 Aronoff's Process Morphology

Aronoff (1976) proposes a word-based theory of morphology. In doing so he explicitly tackles the problem of defining 'morpheme', at that time a task ignored since the generativist revolution. The result was a view of the nature of morphemes that, despite the radically different approaches, shares a key feature with structuralist morphemics: that semantics is incidental to the operation of morphemes.

Aronoff proposes a word-based approach, not because there are no morphemes, but because morphemes, "while they must be assumed to be real linguistic elements, have no meaning which can be assigned independently of each of the individual words in which they occur." (1976:9-10)

This position is justified by the claim that there are elements below the level of word which must be analysed as morphemes, yet cannot be assigned any meaning. Aronoff exemplifies this claim with *-berry* nouns; the various independent and compound occurrences of irregular verbs such as *stand* and *take*; and Latinate stems. It is claimed that each example contains entities which must be morphemes on the grounds of unitary behaviour in relation to various features of irregular word formation or morphophonemic rules, but that no unitary meaning may be assigned. In addition he claims that there is an inherent circularity to the assigning of meaning to entities smaller than words, a circularity not encountered in assigning meanings to words.

With the example of *cranberry*, the element *berry*, being a word, has a meaning. The element *cran-* is a morpheme by virtue of the status of *berry*, but unlike *berry*, is meaningless on its own. To attempt to assign a meaning would be to face a problem of circularity: "is it the word or the morpheme which specifies the meaning?" (Aronoff 1976:11) Because cranberry morphs are unique there is "no noncircular way of assigning meaning to the morphemes...Their meanings are intimately

connected with those of the individual words in which they occur." (Aronoff 1976:10) The meaning of *cranberry* is thus a single meaning associated with the unanalysed word.

However, unless it can be argued that there is something inherently unique about *cran-*, rather than that its uniqueness is an accidental fact (and Aronoff makes no such claim), then *cranberry* must be dealt with in the same way as any other compound. This is problematic since Aronoff explicitly states his approach is not based on a consideration of compounds. However consider *elderberry*, in which both elements are also words and consequently have meanings, *elder* being the name of the tree on which the berries grow. The meaning of *elderberry* is exactly "berries from the elder tree". There may, however, be speakers who are unaware of the existence of the plant *elder*. To such speakers the first element of the compound will serve the semantic function of differentiating this from other berries. Should further investigation reveal the meaning of *elder*, then its meaning in *elderberry* will be narrowed to the differentia of berry source. Such a speaker will have no reason to analyse *cranberry* any differently. The meaning of *cranberry* proceeds from an understanding of the meaning of *berry*, and an assumption about the meaning of *cran-*. This occurs in precisely the same way that a speaker learns the meanings of words. If a speaker read in a book the sentence *they all piled into the charabanc and drove off*, they would assign a meaning to *charabanc*: something like "a means of transport requiring driving and capable of holding several people". Later investigation may reveal a more detailed definition, but even if the speaker never encountered the word elsewhere, and knew no more information about it, if they read that line again they would assign the same meaning to the word, it having entered their lexicon. The same principle applies to *cat*. Any speaker will quickly encounter the word many more times, and see actual cats, and thus form a complex, detailed and subtle definition. The difference in this sense between *charabanc* and *cat* is not a principled one, but merely a matter of exposure to the word. I, as it happens, have never seen a cranberry. All I know about *cran-* is that it distinguishes linguistically a particular kind of berry. Other speakers (such as a North American fruit vendor) may have a much more explicit definition stored in their lexicon.

When an analyst assigns meaning in an apparently circular manner, the process is similar to that of the language learner. Each differentiating element is assumed to have a meaning, and, subject to further evidence, a meaning is assigned on the basis of context. For the speaker, further evidence may take the form of additional occurrences of the linguistic item, or of the element's association with some real world phenomenon. The assignment of meaning is the result of the analysis of utterances in whole or in part. *Charabanc* and *cat* are assigned meanings on the basis of their occurrence in utterances, and that does not prejudice their status as meaningful entities. The same can be said for *cran-*. Once the word *cranberry* has

been learnt, the speaker's morpheme lexicon will contain the meaningful morpheme *cran-*. To the analyst the assignment of meaning appears to involve basing the meaning of the part on the meaning of the whole. In fact the meaning of the whole is merely the key to unlock that of the part. In the case of cranberry morphemes the relationship between the meaning of the part and that key appears particularly intimate because of the accident of uniqueness.

The question of circularity is intimately connected with the notion of morphemes as "the smallest **individually meaningful** elements in the utterances of a language" (my bold), part of Hockett's (1958:123) definition discussed and rejected by Aronoff (1976:7). There are two possible interpretations of "individually meaningful". One is that each element must be meaningful as a lone individual, and thus elements which cannot meaningfully appear by themselves, must be meaningless. That perspective favours a process-based approach.

The alternative is to regard "individually meaningful" as referring to individual components of words, which have meanings as components; it being of no consequence that some individual components may form words on their own while others may not. This view involves regarding morphemes as elements which, once learnt, are stored with their meanings, and involves regarding *cran-* as an individually meaningful component of *cranberry*. This perspective favours a morpheme-based approach, and explains how it is that morphemes serve to distinguish meaning between words, a test the process-based approach fails.

Turning to the notion of meaningless morpheme, Aronoff exemplifies this by several sets of data. However, by identifying semantic commonalities in some cases, and allowing for homophony in others, all the data may be accounted for as consisting entirely of meaningful morphemes.

Aronoff (1976:10-11) claims that the initial elements of *strawberry*, *gooseberry*, *blackberry* and *blueberry* also occur as independent words. He argues that the difference in meaning between their appearance as berries and their independent form are so great that they must either have no meaning as morphemes, or else their meanings must be so underspecified as to amount to a zero meaning.

However, there is no apparent reason to assume that *strawberry* and *gooseberry* involve anything other than cranberry morphemes. Aronoff nowhere precludes the possibility of homophonous morphemes. Indeed, it would be hard to justify such a position. Assuming homophony is possible, then there is no reason to imagine that *straw* and *strawberry* are in any way connected other than phonologically. The same may be said for *goose* and *gooseberry*, which are in fact

not even homophonous.¹⁶

Blackberry and *blueberry* pose a different problem: not all black berries are blackberries, and not all blackberries are black.

"There is therefore no way to assign a meaning to the item *black* which will be valid both when it occurs as an independent word and when it occurs in the word *blackberry*. [We cannot] resort to the simple ruse of assigning...constant meanings, for [such morphemes] do occur elsewhere than in the ['berry'] words..., but with meanings which are totally incompatible with those we would like to assign to them on the basis of the meaning of...[the 'blackberry' word]." (Aronoff 1976:10-11)

Aronoff notes that it is possible get around the problem of one morpheme with different meanings in different words by underspecifying their meanings, with contextually determined "allo-meanings". This "serves to obscure the truth, that it is the words which are idiosyncratic. Though this system may allow us to preserve the idea that morphemes are meaningful, it is only at the level of the word that the meaning can be fully specified." (Aronoff 1976:11) Since some blackberries are actually red the meaning of black will have be underspecified for colour. The alternative is to say that there are several unrelated morphemes *black*, one of which is a cranberry morpheme. Aronoff's solution is that a single morpheme with no meaning is present.

However, the relationship between *black* and *blackberry* is not merely homophony. It is considerably overstating the case to say the two meanings of *black* are "incompatible". It is true that not all black berries are *blackberries*, and not all *blackberries* are black. It is not true, however, that *blackberries* have nothing whatsoever to do with blackness: there is a relationship between the berry and the colour. This relationship may be illustrated by *Easter egg*. Not all Easter eggs are made, or bought, or eaten at Eastertime. Not all eggs eaten at Easter are Easter eggs. It would, however, be laughable to suggest that there was no semantic relationship between *Easter* and *Easter eggs*. The season is a prototypical feature of the object: whenever in the year one may be seen, it is characteristically associated with Easter. The same principle applies to *blackberry*. Whatever the colour of a few individual examples, *blackberries* are characterised as prototypically black.¹⁷ Since in both environments *black* refers to the colour, it would be incorrect to regard the two either as homophonous, or as underspecified for colour.

16. While the two are graphologically identical, in my dialect (at least) the berry is pronounced [gUz], while the bird is pronounced [gus], [gis] or [goz] (as in *gosling*).

17. If this was not a possible naming strategy, *blackberries* would have to be called something like *usually-black-though-occasionally-green-or-red-berries*.

For Aronoff (1976:14), the various occurrences of *stand* in (18) - (22) exemplify the notion of a single morpheme with irreconcilable meanings.

- (18) We stood there for a while.
- (19) We stood the chairs in the corner.
- (20) I stood it for as long as I could.
- (21) understood
- (22) withstood

The subregularity of past tense formation present must be accounted for by regarding these as occurrences of a single morpheme.

Aronoff considers it possible to attempt to relate the meanings of (18) and (19), but cannot see how this could also be done for (20). The meanings of (21) and (22) cannot be related to each other, or to those of (18) - (20). Each has a different subcategorisation and hence a different meaning. Each example must have a separate lexical entry and thus constitute separate words, while the identical coding of PAST "must be conditioned by some abstract property which is common to all occurrences of the meaningless entity..." (Aronoff 1976:15)

This view requires a literal and narrow approach to semantics. It is frequently possible to identify some underlying concept or experiential feature present in a range of occurrences of what appears to be an isolatable entity. In this case the difference between (18) and (19) is merely one of transitivity, and (20) is related almost as directly as the maintenance of a standing position, either literally (*the tree stood the blast*) or metaphorically (*I stood the embarrassment*). *Withstand* and the use of *stand* in (20) have frequently interchangeable environments (*It stood the blast* and *It withstood the blast*). *Understand* is a little more problematic, but could be seen as metaphorical. In English at least, it is common for terms of physical position or process to be used as metaphors for non-physical positions or processes (*he sat out the war; he scoured every avenue*). If we accept that view the *stand* in (21) is semantically related to all the other *stands*. If that view is rejected, and the *stand* in (21) cannot be related semantically to the others, then need it be regarded as an occurrence of the same morpheme?

It is true that all show the same past tense formation. If that alone is grounds for assuming a single morpheme, then *wring* and *ring* must be considered a single morpheme on the strength of *wrung* and *rung*. Further, since Aronoff makes no claim that this conflation of forms into single morphemes applies only to irregular formations, *need* and *knead* must be regarded as a single morpheme on the strength of *needed* and *kneaded*. The only way to prevent such results is to permit

homophonous morphemes which behave similarly.¹⁸

Aronoff's discussion of *stand* highlights a significant failing in his approach. Each of the five occurrences of the form are listed separately in his dictionary as full words. This precludes the possibility of capturing the fact that this is a set of forms which are phonologically identical, and have a kind of special semantic relationship. Aronoff's approach is incapable of recognising this congruence of identical phonological form and similar meaning. Needless to say the approach is entirely incapable of addressing the complex lexical relationships posited in 2.2.1. The semantic results achieved by Aronoff reveal the need to avoid a narrow literal approach to meaning, a need illustrated by the problems discussed in 2.2.1.

The most persuasive evidence Aronoff presents in favour of delinking meaning and morpheme comes from English Latinate stems. He lists such sets of data as "X-fer" (*refer, defer, prefer, infer, confer* and *transfer*), and "re-X" (*repel, remit, refer, resume, receive* and *reduce*). (1976:13-14) Unlike *cranberry*, where at least *berry* has a "fixed meaning", these Latinate words each consist of two morphemes, neither of which have a fixed meaning.

These forms are members of a limited, if fairly large, set of words which are frozen from some earlier period in the language. Aronoff, however, quite rightly demands a synchronic analysis. Synchronically the words intuitively appear dimorphemic, and Aronoff claims that neither the stems nor the prefixes have any fixed meaning. That being so, "how are we to segment the meaning of *reduce* into two parts...in a principled manner? We can't." (Aronoff 1976:14) The only solution is that the two morphemes present are meaningless.

However, as with the berries, a combination of semantic analysis and the acceptance of homophony provides a solution involving meaningful morphemes. In many such cases semantics unifies each morpheme as a single meaningful entity. With the prefixes there is elsewhere a readily apparent meaning: the *re-* in *reinvent*; the *dis-* in *disinterested* etcetera. Aronoff suggests that in most cases these meanings are not apparent as elements of meaning in the examples given. The stems pose greater problems. At least in *reinvent* the root has a fixed meaning; *-sume* appears nowhere with an immediately apparent meaning. It is, however, often possible to detect an underlying meaning which is present throughout a set and combines with the expected prefix meanings.

18. While I regard forms like *understand* and *undertake* as metaphorical, it would be possible to analyse them as containing morphemes homophonous with the free forms, despite the concord of past tense formation. For example, just as *undertake* follows the same past and past participle pattern as *take*, so do *shake* and *forsake*.

It can be argued, for example, that *-sume* means "take", in various shades of meaning, while *-cord* means "agree", somewhat in its linguistic sense:

(23) *-sume*

resume (land):	back/again-take.	"Take back title over the land."
resume (work):	back/again-take.	"Take up work again."
assume (a fact):	at-take.	"I take it you're going."
assume (a position):	at-take.	"Take a defensive posture."
presume:	before-take.	"Dr Livingstone, I take it?"
consume:	with-take.	"Take nourishment"; "partake"
subsume:	under-take.	"The smaller department was taken under the control of the larger."

(24) *-cord*

accord:	at-agree	"All parties reached an agreement."; according to John...("these remarks agree exactly with what John said.")
concord:	with-agree	"They found a point of agreement."
discord:	absence-agree	"They overcame their disagreement." ¹⁹

This kind of semantic analysis overcomes many of Aronoff's problems. However, not all such data may be so readily analysed. It may be that in other examples there is no discernible commonality of meaning. But if that is so, then why not regard the several forms as homophonous? Aronoff argues against this on grounds similar to those used with *stand*. This time the link is not a process like PAST marking, but the operation of a morphophonemic rule. When the stem *-mit* undergoes derivational affixing it appears with final consonant alternation: *permit*, *permission*, *permissive*. This is not the result of a general phonological rule. Nor is it the result of a morphophonemic rule applying to either all Latinate stems (*assertive* not **assersive*), or to all *-mit* final words (*vomitory* not **vomissory*). The only explanation is that the alternation is the result of a morphophonemic rule applying to a single morpheme: "all the items...*mit*₁,...*mit*_n, must be at some level instances of the same thing. Otherwise there is no way to express the fact that all occurrences of *mit* exhibit the same allomorphy." (1976:13)

This involves the same assumption made regarding *understand*: that forms cannot be regarded as merely homophonous if they behave the same way in word formation. But if all the occurrences of *mit* are to be regarded as single morpheme solely on the strength of the operation of a morphophonemic rule throughout the

19. This could possibly be extended to include *record* (back/again-agree) in the sense that a record is prototypically a copy which agrees exactly with the original, be it a disc of music or notes of a meeting. A record is not a general impression or rough sketch, but a faithful depiction.

set, then presumably the English plural, genitive, and third person singular suffixes [-S] must be regarded as a single morpheme on the strength of the morphophonemic rule of schwa insertion applying uniformly to all three.

Aronoff explicitly sets out to define 'morpheme', and proposes that "a morpheme is a phonetic string which can be connected to a linguistic entity outside that string. What is important is not its meaning, but its arbitrariness. This is close to the position of Harris (1951)." (Aronoff 1976:15) The word based nature of Aronoff's morphology proceeds from "freeing the morpheme from the requirement that it be meaningful". (1976:17) However, by conducting semantic analysis in a manner which does not expect an identical semantic reading in every occurrence of each morpheme (an impossible expectation for any word or morpheme, I suggest), and by allowing homophony, I propose that none of the evidence given justifies Aronoff's position. The justification for abandoning the notion of a morpheme based morphology is the proposed need to delink morpheme and meaning. Aronoff's evidence, on close scrutiny, fails to maintain that premise.

2.3.5 Extended Word and Paradigm

EWP does not purport to be a unified theory of morphology: it is a theory of inflection, that is, "exactly a theory of just how much interpenetration of the morphology by the syntax (and vice versa) there is." (Anderson 1988:23) Anderson (1982:587) observes that "inflectional differences correspond to differences with in *paradigm* of a single lexical item, while derivational processes give new lexical items from old." (His italics)

According to Anderson inflection and derivation constitute two entirely separate submodules, each with their own internal structure. Rejecting the Lexicalist Hypothesis, Anderson makes the inflectional component of the morphology interactive with the syntax. The derivational component takes place in the lexicon, opaque to the syntax, involving a structured lexical list and derivational word formation rules. "The lexicon supplies a comprehensive set of well-formed stems [which] represent complete words, with the exception of inflectional material." (Anderson 1982:592)

Unlike other theories of morphology, EWP neither attempts to define the notion of morpheme, nor assumes there is a definition. Indeed, it is concerned with the notion of morpheme only to the extent of proposing that phonological realisations of syntactic categories are not morphemes: they are 'morphosyntactic realisations'. The validity or otherwise of this assertion is the limit to the morphemic relevance of EWP.

Underlying EWP is the view that inflectional marking frequently fails to divide into separate forms each with distinct meanings. "There is no one-to-one correspondence here between the formal markers and the categories they express." Instead there is an "association between the entire form...and the entire set of categories...The association...is not one-to-one but many-to-many". (Anderson 1985:160) It is consequently argued that inflectional forms are undecomposable.

However, if a listener distinguishes between an uninflected and an inflected form, then there is something about the formal quality of the inflected form which conveys elements of information not conveyed by the uninflected form. Equally, two differently inflected forms convey two separate and distinct sets of meanings (including grammatical category). To a speaker of Latin *amo* conveys one thing, while *amas* conveys something else. But the two also have common elements of meaning. It is the *-o* in *amo* which conveys 1SG. If the *-o* is absent, another form will be present (eg. *-as*) conveying another set of information. Equally, *amabo* conveys something different again, but not entirely different. It conveys some information also found in *amo* and *amas*, some also found in *amo* but not in *amas*, and some found in neither. The listener can only distinguish the meaning of one from the meaning of another by formal features. These words contain formal components, the nature of which allows the listener to receive elements of information. This relationship between the forms and the meanings demands the decomposition of the words. This is a simple example, but the principle may be applied to more complex forms.

There is no reason to assume that a morpheme based approach to inflection poses problems within the generative framework. Using Lieber's model, Jensen and Stong-Jensen (1984) reanalyse Anderson's examples and demonstrate that by proceeding from the morphology, morphosyntactic representations may in fact be analysed as morphemes.²⁰

Anderson, however, proceeds from the syntax, examining inflection for the representations of each grammatical category. The result is often complex: a formally indivisible entity may represent more than one category, or a single category may be represented by a formally complex entity. Since there can frequently be no simple one-to-one relationship between a form and a grammatical category, it must not be possible to decompose inflectional forms. However, while a form may be complex it is none-the-less distinct from other forms and the distinguishing formal features constitute an element which is inherently associated with the corresponding distinguishing set of features of meaning or function.

20. Spencer (1991:221-223) provides a brief overview of Jensen and Stong-Jensen's reanalysis of Anderson's data.

In the case of portmanteaux, the representation of several syntactic categories constitutes the single communicative function of an entity. A portmanteau may have several syntactic functions, but as a morphological object, it has a single complex meaning. These two concurrent facts reflect the roles the form plays in separate domains. The English pronoun *I* is formally undecomposable. As a morphological object it represents the single set of morphosyntactic features: 'first person singular nominative'. This is the meaning set a speaker will intend and understand. This single meaningful entity then performs three separate functions in the syntax. The same linguistic item is behaving differently in different domains. The fact that some morphological objects perform several syntactic functions does not indicate that all inflectional forms are undecomposable.

As a further problem, EWP has an internal inconsistency: inflection occurs postlexically, but irregularly inflected forms are listed. This gives the lexicon a significant element of duplication. Base forms are listed, while derived forms, being generated by rules, are not. Irregularly inflected (and thus potentially derived) forms are then also listed alongside the underived. The only derived forms to make it into the lexicon are thus those which accidentally inflect irregularly.²¹ Anderson proposes the listing of irregularly inflected forms despite accepting that such forms may demonstrate significant subregularities. A consequence of this is that the model cannot capture the phonological and semantic relationship between the underived base and the fully derived and irregularly inflected form of the same base.

EWP is a theory of inflection. It does not deal with the notion of morpheme in derivation, and consequently is unable to deal with many of the problems raised in 2.2. EWP is relevant to morphemics only in its rejection of the role of morphemes in inflection, and as Jensen and Stong-Jensen demonstrate, that rejection may be challenged.

2.3.6 Summary

Throughout morphological literature there is a considerable lack of attention given to defining 'morpheme'. Following the views of Bloomfield (1933) there was considerable debate among structuralists about the precise nature and characteristics of morphemes. This debate was swept away by the generativist revolution, and since then, the only major theorist to deal with the issue is Aronoff. Elsewhere in generative linguistics morphological theory equals theory of word formation, with the existence and definition of 'morpheme' tacitly assumed. Other theories, such as EWP, do not concern themselves with the notion of morpheme at all.

21. If inflection and derivation are distinct kinds of operations, the inflectional process applying to a derived form can only be incidental to any derivational process that generated that form.

Both the major approaches which deal explicitly with the notion of morpheme reject the idea that meaning is a key feature. Despite a theoretical gulf, the structuralist school and Aronoff reject the morphemic significance of semantics for very similar reasons. A concern with readily applicable isolation and description techniques, coupled with the absence of any effective system of rigorous semantic analysis, led the structuralists to regard distributional paralleling as the primary factor in characterising morphemes. Since semantic relationships exist which are not definable distributionally, those relationships were relegated to some future semantic analysis and not regarded as a feature of the morphology.

Aronoff also delinked meaning and morpheme on the grounds that while semantic analysis is possible, it reveals the presence and absence of relationships between forms which do not parallel their relationships in terms of syntactic behaviour. Both Aronoff and the structuralists reject an intrinsic link between semantics and morphology on the basis of the behaviour of morphological objects in larger structures: for the structuralists the behaviour of morphemes in words and sentences, and for Aronoff the behaviour of words in sentences.

Whether explicitly dealt with or assumed, morphemes are regarded by theorists as either objects which form the atoms of morphology, or as the manifestations of processes. The morpheme-as-thing approach, currently represented by Lexical Phonology, regards word formation as morpheme based, while process theorists such as Aronoff propose a word based model.

Many theories which regard the morpheme as the basis of word formation, regard the morpheme as an inherently meaningful entity. Unfortunately these theories are the ones which assume a definition of morpheme and concern themselves with other matters. What the literature lacks is a theory which regards the morpheme as the basis of word formation, deals explicitly with defining 'morpheme', and regards meaning as an intrinsic characteristic of morphemes. This is a significant absence.

2.4 Conclusion

The first part of this chapter examined a number of morphological problems which all involve the status of morphemic entities, and the relationships between them. Examples were given of forms which have a great deal in common both phonologically and semantically. Others demonstrated variety in meaning or function co-occurring with phonological variation. Yet others involved the reoccurrence of certain elements of meaning coinciding with recurring formal features. It was proposed that an adequate theory of morphology must account for these relationships.

The second part of the chapter examined the ways in which existing theories deal with morphological problems. None of these theories proved capable of dealing with many of the kinds of problems discussed in 2.2. Several were incapable of even recognising many such problems.

The difficulties existing morphological theories have when faced with many phenomena stem from an unwillingness to recognise the role of semantics in linking entities in speakers minds, and the intrinsic relationship between formal elements of utterances and their several semantic elements conveyed by those utterances. There is a need for a theory of morphemics which concerns itself primarily with the morpheme as a semantically functional and linguistically real entity.

Chapter 3

The Theory

3.1 Introduction

It was noted in 2.3 that many theories of morphology are concerned with word formation and assume a notion of 'morpheme'. In contrast, the present work proposes that a theory of morphological structure must proceed from a definition of morpheme. The present work is specifically concerned with the notion of morpheme, and seeks to define and characterise 'morpheme' from a theoretical perspective. This work assumes a theory of morphological structure exists, or can be developed, which will operate in conjunction with the proposed theory of morphemics.

3.2 The Relationship between Form and Meaning

This approach is predicated on the view that language is a phenomenon of form and function in which they are entirely inter-dependent. The form exists to provide a vehicle for the function and the function depends on the form for realisation. Without the form, communication would be impossible, therefore meaning, being the function of language, depends on the form for its existence. And without meaning the form would not be language. The two are essential to each other. The present approach assumes that since form and meaning are symbiotic they have equal significance. Consequently the study of morphemics must give equal regard to formal and semantic issues. This view may be codified:

- (1) If X is the semantic value of formal entity Y, then Y is the formal realisation of semantic value X.

In this work the term 'meaning' refers to communicated information of any kind. It may be referential, or perform a syntactic function. The distinction between semantic and syntactic values is not an issue here, and the term 'semantic value' will be used to refer to meaning of any kind. The term 'form' will be used to refer to any linguistic vehicle for meaning.

The effect of (1) is that the hearer of an utterance may only distinguish between forms by means of differences in their formal features. The distinction between the meanings of forms is conveyed by those differences.

A speaker of English, hearing *rise* and *fall*, will know that movement, and a consequent altitude change, are features of both. They will also know that while *rise* involves 'up', *fall* involves 'down'. This information is conveyed by the forms. The shared elements of meaning do not correspond to any significant shared formal features. The shared meaning is therefore not carried specifically by any part of either entity, but by each entity separately. The entities are therefore entirely discrete.

In contrast, both *fall* and *fell* will convey 'movement', 'altitude change', and 'down'. The only meaning not shared is when the action occurred. There are also many formal similarities, to the extent that a speaker may only distinguish between them by means of the vowel. The phonemes [f] and [l], separated by a vowel which is either [ɔ] or [ɛ], convey downward movement. That meaning is conveyed by that set of phonological features. The tense is not carried by that set of features in general, but solely by the choice of vowel. It follows that in these two words the form [f - l], separated by [ɔ] or [ɛ], means downward movement, without any indication of when it happened; and that the form [ɔ] means non-past, and the form [ɛ] means past.

3.3 The Parameters of Morphemics

In chapter 2 it was noted that every language contains a set of entities which have both form and meaning and cannot be meaningfully subdivided: a set of semantically undecomposable sound-meaning associations. An utterance comprised of meaningful elements may be broken down into those elements, and they, if comprised of even smaller meaningful elements, may also be broken down. Ultimately there is a set of semantically undecomposable meaningful forms. It may be argued that in any language some members of this set will also be words, and as such their behaviour will be accounted for by syntax. However, in almost every language, at least some members of the set combine to form words. Words are, therefore, combinations of one or more of these entities.

So we have two entities: undecomposable sound-meaning associations; and words. Di Sciullo and Williams (1987) argue that 'word' is both a morphological object, and a syntactic atom. Syntactic atoms are the smallest objects over which the syntax operates, and are opaque to it. The word as morphological object is the output of the morphology: an entity generated by morphemes and morphological rules.

Just as words are syntactic atoms, the semantically undecomposable meaningful forms are the atoms of word formation. The formation of words out of these atoms is 'morphology', and the atoms 'morphemes'.

The domain of morphology therefore consists of a set of morphemes, and the manner in which they form words. This parallels words and syntax, but the morphology has its own set of atoms and its own set of rules controlling the combination of atoms. The study of morphology involves two issues: the study of morphemes, and the study of word formation. The former will be termed 'morphemics', and the latter 'morphological structure'.

3.4 Defining Morpheme

3.4.1 Principle 1

A morpheme must have a form and a meaning. A 'form' may consist of one or more segments, or of suprasegmental features, or of a meaningful absence, or of one or more phonological features within a segment. In the present work the term 'phonological' will encompass any of these characteristics. Any part of an utterance, no matter how simple or complex, which conveys any element of meaning, is a form-meaning correspondence. If it is meaningfully undecomposable, it has some morphemic status. Such a correspondence will be referred to as a 'morphemic entity'. The term 'entity' will only be used in that context.

The above may be codified by a morphemic principle:

- (2) Principle 1 Any phonological feature or set of features which has any semantic value and is semantically undecomposable has morphemic status.

This does not mean that anything which fits this description is a discrete morpheme, merely that it has some status in the language's morphemics, that it is a morphemic entity.

This Principle resolves many of the problems raised in chapter 2. Phonesthemes, for example, are revealed to be a type of morpheme. They have something in common with even the most clear cut morphemes like *walk* or *cat* that they don't share with phonemes: they are meaningful.

3.4.2 Principle 2

Principle 1 reveals what has morphemic status. However, in any language not all morphemic entities isolated by this Principle will be separate morphemes. Some will be related to each other. Chapter 2 raised problems involving the relationships between morphemic entities. Some pairs or groups of entities demonstrate

significant phonological and semantic similarities (as in *isle*, *isol-* and *insul-*).¹ On other occasions functionally similar entities are phonologically disparate, paralleling homophonous but semantically distinct entities (as the plural suffixes in *cats* and *criteria*).²

A second morphemic principle may be proposed which accounts for and defines the relationship between similar entities, and the absence of a relationship between disparate entities:

- (3) Principle 2. Any group of morphemic entities which have both formal and semantic links constitute separate realisations of a single morpheme.

This means that any two morphemic entities, that is, any two semantically undecomposable form-meaning associations, which are significantly similar to each other in form and in meaning, have a relationship. That relationship is one of allomorphs within a single morpheme. It is implicit in this Principle that forms which fail to demonstrate both formal and semantic links constitute separate morphemes, and that merely homophonous or merely synonymous entities are also separate morphemes.

This is in keeping with the view of linguistic form and function expressed in 3.2: entities which are not both formally and functionally related cannot be considered parts of the same thing.

Principle 2 raises a terminological issue. The relationship between separate parts of a single morpheme is conventionally termed 'allomorphy', and the entities 'allomorphs'. It is, however, an inherent implication of these terms that the only acceptable diversity is formal. Parallel semantically related entities have sometimes been termed 'allosemes'. In contrast, the present approach makes no distinction between formal or semantic differences. A morpheme may have several separate surface realisations, and they may be semantically identical, and have slight formal differences. Equally they may be formally identical but display slight semantic variation. The term 'allomorph' is inappropriate in this context. Consequently, separate parts of a single morpheme will be termed 'Separate Surface Realisations' (SSRs).

1. For a full discussion of these entities see 2.2.1.

2. See 2.2.3.

3.5 The Typology of Morpheme

Since a morpheme may consist of several slightly different Separate Surface Realisations, 'morpheme' must have an internal structure.

3.5.1 The Relationship Between the Underlying Representation and Surface Realisations

In a conventional operational approach, where a morpheme has several allomorphs one is nominated as underlying, and from this all others are derived. This assumes that one surface form is also the underlying representation, often involving an arbitrary assignment of underlying status. In contrast, I propose that what is underlying in a morpheme is that which unites its surface realisations: its formal core consists of those formal features which are common to all its disparate SSRs, and its semantic core those semantic features which are common. Any features which apply to only one or a subset of the SSRs are additional to those key features which unify the morpheme.

This view regards no one SSR as underlying. At the core of any morpheme is an Underlying Representation (UR), consisting of those formal and semantic features common to all the SSRs, and distinct from any surface realisation. The UR is the distillation of the morpheme's realisations. The SSRs are consequently the surface realisations of the UR. Each surface realisation consists of those features specified in the UR plus additional features specific to that realisation.

3.5.2 Principle 3

The relationship between the surface and the underlying may be represented by a further morphemic principle:

- (4) Principle 3. A morpheme consists of an Underlying Representation and its realisations. The UR is the semantic and formal distillation of the morpheme's various semantic and formal realisations, and those various realisations are the surface realisations of the UR.

3.5.3 Subfamily Groups

An effect of Principle 2 is that a morpheme may consist of a set of SSRs which, while clearly related, are also varied. This variation is not necessarily even. Within a single morpheme some SSRs may share more features with each other than with others. In a morpheme consisting of many SSRs, they may appear to form groups,

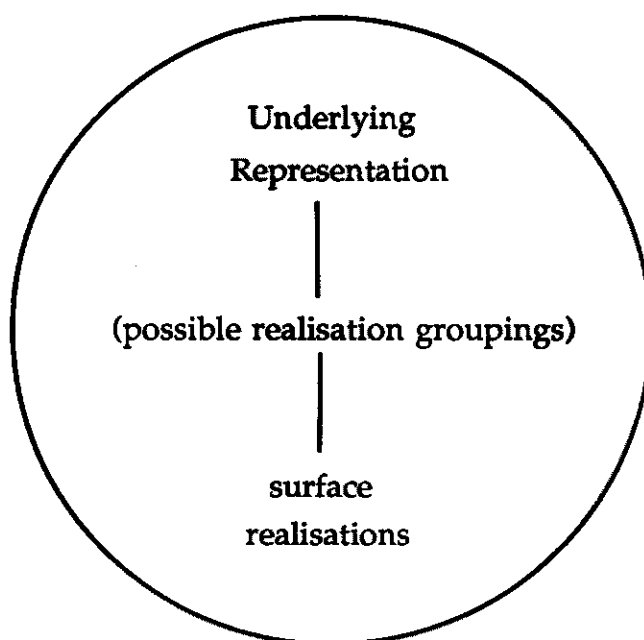
and within these there may be further divisions into smaller groups. Such complex morphemes may be termed 'family morphemes'. The subsets of SSRs within such morphemes may be termed 'subfamily groups'.

A family morpheme may, for example, have 16 SSRs, and divide into two subfamily groups each of eight particularly closely related SSRs, and those two groups may themselves consist each of two further subfamily groups, all of four SSRs, and so on. There is theoretically no limit to the number of subdivisions possible, although more than two or three is unusual. There is also no expectation that such structures will be symmetrical: a morpheme may consist of one subfamily group, and outside that one slightly less similar SSR.

3.5.4 The Typology Schematised

Implicit in Principle 3 and in the concept of subfamily groups is the following typology of the structure of 'morpheme':

(5)



3.6 The Role of Phonology

The "surface realisations" level in (5) does not include every observable realisation. The typology is that of the stored morpheme, and as such contains exactly that which is idiosyncratic and not the result of any rule. This specifically excludes phonologically and morphophonemically generated variants, which are generated outside the lexicon.

Some features of surface realisations are generated by phonological rules general to the language. In such cases the phonology operates regularly upon the morpheme, and the result need not be stored.

Other alternations are the result of phonological rules applying idiosyncratically. These will be referred to here as morphophonemic rules. The invocation of a morphophonemic rule is an idiosyncratic feature of the morpheme which invokes that rule. As such it is stored. The resulting realisation is not stored since it is the fact that the rule applies which is idiosyncratic, not the rule itself.

Phonologically and morphophonemically generated features represent true allomorphy, and need not and must not be entered in the lexicon.

The English third person singular suffix, for example, has three realisations: [-s], [-z] and [-əz]. None of these are part of the morpheme itself. The schwa is inserted in a post-sibilant environment as the result of a syllable structure rule preventing the collocation of like with like; and English has a morphophonemic rule of voice assimilation which applies to some single consonant suffixes. This morpheme is stored with only the following phonological details: it is a single consonant, which is an alveolar fricative, and to which the morphophonemic rule of voice assimilation applies. This is sufficient to generate all three realisations.

During the analysis in this work the expression 'rule referral' will be used. This is meant only as a shorthand. It indicates that the entity with which it occurs has as a feature the invocation of a rule. With the 3SG suffix example, the schwa insertion, being a phonological rule, applies automatically and is not listed at all. The voice assimilation, on the other hand, is morphophonemic, and so does not apply automatically. For it to apply the rule must be invoked somehow. This is done here by the morpheme (or appropriate realisation) being referred to the appropriate rule. This process is called here a 'rule referral'.

It should also be noted that to avoid orthographic confusion IPA representations are used in analysis. These will always be broad transcriptions of my own dialect of Australian English.

3.7 Semantics

In the discussion of form and meaning in 3.2 an argument was made for the joint consideration of formal entities and their meanings. This is not easy: the nature of tangible formal entities and of meanings are very different, as are the means by which the two may be analysed. The identification and characterisation of formal

features is often controversial, but is considerably easier than the analysis of meaning, a very much more elusive thing.

Many approaches respond to these difficulties by removing semantics from morphemics. As discussed in 2.3.4, Aronoff (1976) regards the meanings of morphemes as arbitrary and irrelevant. This ignores significant semantic regularities among morphemes. It also belies the communicative function, manifest in meaning, as the *raison d'être* of morphemes.

This disinclination to give due regard to semantics has been noted by Wierzbicka (1986:368): "The dangers of subjectivism and arbitrariness involved in a search for such correlations [between language and culture] are no doubt real enough. But to abandon the search because of these dangers is, in my mind, analogous to saying, as Bloomfield did, that linguistics should stay clear of meaning because all attempts to study meaning are fraught with dangers of subjectivism and arbitrariness."

The key semantic problem for morphology lies in characterising elements of meaning. If the meaning of a form is significant to the discussion of that form, then a statement must be made about that meaning. This usually takes the form of a definition, in which, as Wierzbicka (1987:6) observes, "the meaning is **alluded** to, but it is not spelled out explicitly" (her bold).

Such allusions cannot form the basis for a rigorous system of analysis. Wierzbicka (1987:11) responded to this problem by attempting to develop a model in which semantics may be analysed and meanings represented. A model in which, "...instead of comparing an unidentified meaning with various other meanings, one could simply identify it, by enumerating all its constituents." She does this by means of "reductive paraphrases" involving a metalanguage with a limited vocabulary. The effect of this approach is to not define meanings, but to systematically characterise them.

This approach has the potential to allow a semantic analysis capable of occurring alongside formal analysis, allowing results which reflect the symbiotic relationship between form and meaning. This would preclude the need to delink morphemes and meanings on the grounds of semantic difficulties.

At this stage, however, Wierzbicka's metalanguage involves not a few words, but a few sentences. These "explications" effectively characterise meaning, and are suitable for lexicography. However, they are unwieldy for the purposes of providing the semantic component of an analytic tool.

This, for example, is her meaning of *apologize* (1987:215):

- (6) I know that I caused something to happen that was bad for you
 I think that you may think something bad about me because of that
 and feel something bad towards me because of that
 I say: I feel something bad because of that
 I don't want you to think something bad about me because of that
 and to feel something bad towards me because of that
 I say this because I think I should say it to you

In the present work, an attempt has been made to go beyond definitions, and to provide an idea of the key semantic characteristics of elements of data. However, no attempt has been made to provide full explanations of meanings. The 'semantic values' proposed in this work are not intended to be either definitions, or explications of the full meanings of the analysed forms. They are a shorthand representation of the key semantic features relevant to the analysis undertaken, and for that reason they will be inadequate as explications. If semantic features are represented in analysis, it will be because they are shared or are contrastive with the features of another form. For example, in a comparison of *flow* and *flitter* a key semantic feature of both is that they are ways of moving. This will be represented as the shorthand semantic value <movement>, and will be represented as a semantic feature the two words share. The differences in meaning between the two relate to the manner of the movement. *Flow* will be characterised as an even kind of movement, while *flitter* is an erratic kind of movement. These key semantic characteristics are contrastive, and will be represented by the shorthand semantic values <evenness> and <erraticness>. *Flow* will thus be represented as having the semantic features <movement, evenness>.

In the absence of a rigorous yet not unwieldy system, the semantic features assigned in this work to English entities are based on a combination of introspective data and meanings suggested by dictionary definitions. This is potentially problematic, and a conscious effort was made to avoid having the explications fit the example.

In the case of some examples, semantic features are not represented in detail at all. In these instances the semantic value of the forms is not the issue of the example, and it is not necessary for the purposes of the example to attempt to represent the entity's meaning. For example, when comparing *go* and *went* the only difference in meaning lies in tense, and it is unnecessary for the example to characterise the common semantic value. This will simply be represented by the shorthand "assume the semantic value <go>". As another example, in a comparison of the suffixes in *hypnosis* and *hypnotic*, the meaning of the root is incidental to the example, and will be represented as "assume the semantic value <hypno>".

To some extent the approach to morphemics presented in this work depends on the existence of a semantic system which combines the explanatory strengths of the Wierzbicka approach with a practically applicable form. The development of such a system is, of course, beyond the scope of the present work. The absence of rigorous criteria for the semantic analysis carried out in this work is a limitation on the application of the approach proposed in this work. I believe, however, that it does not prejudice the validity of the underlying approach proposed.

3.8 Commonality and Distinctiveness

3.8.1 The Analytic Procedure

Sections 3.2 to 3.5 propose an approach to defining and characterising 'morpheme'. This requires a corresponding means of analysing and representing morphemic data.

The process of analysis involves four stages:

- (7) (7.1) The isolation of entities: The corpus of the language must be examined in a way which will reveal the set of morphemic entities, the set of undecomposable correspondences of form and meaning.
- (7.2) Relating entities: The forms revealed by 7.1 must be examined to determine what relationships they have with each other, which are SSRs of single morphemes, etc.
- (7.3) The proposal of morphemes: Related entities must be conflated into single morphemes and the relationship between each SSR and the UR must be determined. This stage also involves the elimination of phonologically and morphophonemically generated alternations from the morphemics.
- (7.4) The listing of morphemes: The results of the analysis must be entered in the lexicon.

Stage (7.4) involves proposing a system of lexical representation. This will be dealt with in 3.9.

The analysis in (7.1) to (7.3) may be conducted using a single tool: the Principle of Commonality and Distinctiveness (PCD). This tool is the means by which the theory is applied to data.

3.8.2 The Principle of Commonality and Distinctiveness

In accordance with the relationship between form and meaning proposed in 3.2, the Principle of Commonality and Distinctiveness isolates entities, then conflates or separates them, on the basis of the correspondence of elements of form and elements of meaning. This involves associating forms and meanings, and testing for undecomposability.

PCD compares entities by revealing shared formal and semantic features, identifying these features as 'commonalities'.

Correspondences of formal and semantic features which are not shared, which distinguish the forms, are also revealed. These distinguishing features are 'distinctivenesses'.

PCD usefully takes the form of a table, upon which elements of data may be represented. This table isolates morphemic entities by making transparent the correspondence of form and meaning which distinguish elements of the data.

(8) The form of the PCD table:

Features	Commonality	Distinctiveness
formal		/
semantic		/

The table compares two elements of data. Formal features of the compared elements are represented on one line, and semantic features on the next.

These lines of feature types are divided into columns. Features which are common to both compared entities are represented in the 'commonality' column. Features which distinguish the compared entities from each other are represented in the 'distinctiveness' column. This column is divided by diagonal slashes. The distinguishing features of one entity are represented on the left of the slashes, while those of the other entity are represented on the right.

This arrangement of elements of data clearly reveals correspondences between the formal and semantic commonalities, and between the two formal and semantic distinctivenesses.

3.8.3 Implementing Principle 1

PCD serves to isolate and identify a language's set of morphemic entities, thus implementing Principle 1.

3.8.3.1 The Isolation of Entities

To demonstrate the PCD table, consider *walk* and *hiss*. It is important when applying data to the table to represent formal features in a way which avoids misleading orthographic similarities or dissimilarities. Consequently henceforth all forms will be represented on the table by IPA symbols. Semantically *walk* may be characterised as self-propulsion by means of legs. *Hiss* may be characterised as the production of a sibilant sound. They appear on the PCD table as follows:

(9) Comparing *walk* and *hiss*:

Features	Commonality	Distinctiveness
formal		wɔk / his
semantic		propel self on legs/produce sibilant sound

In (9) the two words share no formal or semantic features. Consequently they may be regarded as morphemically unrelated. The table revealed that the formal distinctiveness [wɔk] corresponds with the semantic distinctiveness <propel self on legs>, and that the formal distinctiveness [his] corresponds to the semantic distinctiveness <produce sibilant sound>. This table reveals nothing about the internal structure of the two words, such as how many morphemes are present. It simply reveals that there are no morphemes which are present in both. In contrast compare *walked* and *hissed* :

(10) Comparing *walked* and *hissed*:

Features	Commonality	Distinctiveness
formal	t	wɔk / his
semantic	past	propel self on legs/produce sibilant sound

These words are revealed to contain at least two morphemes: the distinctivenesses revealed in (9), and also the shared features of the word-final segment [t], and past tense. This is a correspondence of form and meaning and thus a morphemic entity. From (10) it is possible to make a tentative postulation: English includes the morphemic entity [t]:<past>. The table does not reveal whether this is the only English past marker or one of a set of SSRs; or whether it is regular and productive or idiosyncratic. It simply reveals that it is a morphemic entity.

In 2.2.6 phonesthemes were proposed to have morphemic status, and the PCD table effectively reveals the morphemic construction of phonestheme-based words. Consider *flow* and *flutter*. They may be characterised as meaning <even movement> and <erratic movement> respectively.³

3. See 2.2.6, footnote 6.

(11) Comparing *flow* and *flutter*:

Features	Commonality	Distinctiveness
formal	fɪ	oʊ / Itə
semantic	movement	evenness / erraticness

The formal commonality [fɪ] corresponds to the semantic commonality <movement>. The formal distinctiveness [oʊ] corresponds to the semantic distinctiveness <evenness>; and the formal distinctiveness [Itə] corresponds with the semantic distinctiveness <erraticness>. Table (11) allows three distinct morphemic entities to be tentatively postulated: [fɪ]:<movement>; [oʊ]:<evenness>; and [Itə]:<erraticness>.

3.8.3.2 The Resolution of Non-corresponding Features

Examples (9), (10) and (11) are straightforward: in each all formal and semantic features are corresponding. This will often not be the case. The PCD table frequently reveals formal commonalities with no corresponding semantic commonalities, or vice versa; or formal distinctivenesses with no corresponding semantic distinctivenesses, and vice versa.

Consider *walk* and *fork*. These have the formal features [wɔk] and [fɔk].⁴ For our purposes *fork* may be characterised as a pronged implement. An initial PCD comparison of the two looks like this:

(12) Comparing *walk* and *fork*:

Features	Commonality	Distinctiveness
formal	ɔk	w / f
semantic		propel self on legs/pronged implement

Although there is a formal commonality in (12), it has no corresponding semantic value. This commonality is thus revealed to have no morphemic status, merely reflecting the partial homophony of separate entities. To accurately account for all morphemic entities in data, a process of analysis and reanalysis must continue until all non-corresponding features are resolved. Table (12) must be revised to reflect the non-morphemic status of the formal commonality:

(13) Revised comparison of *walk* and *fork*:

Features	Commonality	Distinctiveness
formal		wɔk / fɔk
semantic		propel self on legs/pronged implement

4. These are broad transcriptions of the words in my dialect of English.

Just as (13) revealed a formal commonality with no semantic correspondence, the opposite is possible. Table (14) compares *walk* and *run*. Like *walk*, *run* may be characterised as <propel self on legs>, but raises the issue of speed. *Walk* may be characterised as a slow pace, *run* as rapid.

(14) Comparing *walk* and *run*:

Features	Commonality	Distinctiveness
formal		wɔk / rʌn
semantic	propel self on legs	slow / rapid

It is clear from (14) that although both words refer to self-propulsion on legs, there is no corresponding formal commonality: *walk* and *run* contain no formal features which convey the semantic commonality. That commonality has no morphemic status, merely reflecting partial synonymy. As in (13), the non-correspondence must be resolved:

(15) Revised comparison of *walk* and *run*:

Features	Commonality	Distinctiveness
formal		wɔk / rʌn
semantic	propel self on legs, slow / propel self on legs, rapid	

Where two forms demonstrate a semantic or a formal commonality, but that commonality has no correspondence it has no significance for morphemic analysis.

This principle also applies to a lack of correspondence between distinctive features. Consider *know* and *knowledge*.⁵ *Knowledge* contains two morphemes: the noun marker *-edge*; and the residue *knowl-*, which may be compared with *know*. *Know* may be characterised at its most fundamental as something like 'the mental presence of information and/or understanding'. For our purposes let us assume the semantic value <know>.

(16) Comparing *know* and *knowl-*:

Features	Commonality	Distinctiveness
formal	no	u / l
semantic	know	/

The formal distinctivenesses have no corresponding semantic distinctivenesses, and consequently are not morphemic entities.⁶ That these features have no morphemic

5. A full analysis of this data may be found in 4.6.

6. Features which have no morphemic status but form part of the representation of a morphemic entity are called 'Realisation Features', and are discussed in detail in 3.11.

status does not mean they are of no significance in the representation of the morpheme. What it does mean is that they are SSRs within a single morphemic entity. Their significance lies in morphological structure: one is free, the other bound.

3.8.3.3 The Status of Non-comparative Distinctivenesses

A further issue requiring resolution is that of a distinctiveness correspondence which contrasts only with the commonality. For example *walk* and *walked*:

(17) Comparing *walk* and *walked*:

Features	Commonality	Distinctiveness
formal	wɔk	/ t
semantic	propel self on legs	/ past

Being a correspondence, the distinctivenesses constitute a morphemic entity, but with what are they contrastive? The immediate answer is the unmarked non-past word, but this is problematic: it suggests that PAST is marked by adding a past tense marker to a non-past word; and more significantly, it implies there is something inherently non-past about *walk*. What (17) actually reveals is that *walk* is underspecified for tense. While *walk* by itself refers to any tense other than PAST, those tenses are not represented morphologically. Table (17) reveals that for *walk* at least, the only tense marking morpheme possible marks past.

3.8.3.4 Testing Postulated Entities

Having completed the analytic procedures outlined so far, various morphemic entities will be tentatively postulated. These must be tested against other data. The analysis of a single pair may give a spurious result, leading to the postulation of a non-existent entity. Such errors will become apparent by the application of PCD across a body of data. A comparison of the verbs in *yesterday he walked* and *yesterday he hit him* would give this result:

(18) Comparing *hit* and *walked*:

Features	Commonality	Distinctiveness
formal	t	hɪ / wɔk
semantic	past	strike/propel self on legs

This would lead to the tentative postulation of the morphemic entities [t]:<past>; [wɔk]:<propel self on legs>; and [hɪ]:<strike>. The latter is counter-intuitive. However, an adequate analytic procedure must not require intuitive judgments. The counter-intuitive analysis of *hit* in (18) must be disproven by procedure.

The rigorous application of PCD across the corpus of data exposes such spurious analyses. Each postulated entity must be tested against other forms. As demonstrated earlier, both *walk* and the tense marker are attestable across the language. With the postulated *hi-* on the other hand, a comparison with other forms fails to support the analysis. Two avenues may be pursued for evidence supporting *hi-*: a comparison may be made with other vowel-final roots in their past tense forms, as with *bored* in (19); and a comparison may be made with non-past occurrences of this postulated root, as of *he will hit him* in (20).

(19) Comparing *hit* and *bored*:

Features	Commonality	Distinctiveness
formal	C 	CV C / CVC /
	+alveolar,+stop	h I -voice / b ɔ +voice
semantic	past	strike / bore

A comparison of *bore* and *bored* will reveal that *-d* is the tense suffix, and further comparisons across the language will confirm that in a post-vowel environment a voiced past suffix is expected. Consequently the past tense of *hi-* would be expected as **hid*.

(20) Comparing the past and nonpast of *hit*:

Features	Commonality	Distinctiveness
formal	hit	/
semantic	strike	past / non-past

This reveals that there is no morphologically marked past tense version of *hit*. In accordance with the approach to non-corresponding distinctivenesses discussed above, *hit* will be postulated as underspecified for tense, that it does not mark tense morphologically.

Tables (19) and (20) disprove the analysis in (18). As a result of this testing process (18) must be revised:

(21) Revised comparison of *hit* and *walked*:

Features	Commonality	Distinctiveness
formal		hit / wɔkt
semantic		strike, past/propel self on legs, past

This indicates that in *walked* and *hit* the shared phonological material and shared tense do not constitute a commonality; and thus the tense-underspecified morpheme [hit]:<strike> may be proposed.

This analysis and reanalysis demonstrates that initially entities must be postulated only tentatively, and a rigorous testing process must follow. Although a PCD comparison of a single pair may give spurious results, the application of PCD across a body of data invariably gives accurate results. This is one of the strengths of PCD: it is inherently self-correcting.

3.8.4 Implementing Principle 2

Morphemic Principle 1 is concerned with identifying morphemic entities. The application of PCD so far has involved implementing Principle 1. Having established a set of morphemic entities, it remains to implement Principle 2: to determine the relationship between the isolated entities. As noted in 3.5, not all entities will be discrete morphemes. Many will be SSRs of one morpheme. Determining the relationships between entities involves determining the boundaries of individual morphemes. A process must ensue turning the set of morphemic entities into a set of morphemes, each with their own SSRs. This process will identify some entities as the sole realisations of discrete morphemes; and conflate others into single morphemes. For this process the tool is again the Principle of Commonality and Distinctiveness.

3.8.4.1 Separating Entities

PCD may be used to separate entities which are realisations of discrete morphemes. Consider *go* and *went*, assuming for our purposes the semantic value <go>.7 The comparison in (22) of these two words contrasts strikingly with the comparison of *walk* and *walked* in (17).

(22) Comparing *go* and *went*:

Features	Commonality	Distinctiveness
formal		goU / went
semantic	go	non-past/past

From (22) it is apparent that although these words differ semantically only in tense, they have no formal commonality whatsoever which corresponds to the semantic commonality <go>. Principle 2 precludes the conflation of entities which do not share both formal and semantic links. Here two entities share a significant proportion of their semantic features, but, failing to demonstrate any corresponding formal commonality, demonstrably belong to separate morphemes.

7. A full analysis of the *go* paradigm may be found in 4.4.

3.8.4.2 Conflating Entities

The PCD table also determines when two entities are realisations of a single morpheme. Assuming the semantic value <do>, comparing *did* with other past tense verbs allows the isolation of the entity [dɪ]:<do>.⁸ This may be compared with the non-past entity:

(23) Comparing *do* and *di-*:

Features	Commonality	Distinctiveness
formal	d	u / ɪ
semantic	do	/

The vowel variation does not correspond to any semantic variation and consequently has no morphemic status. The two entities are semantically identical realisations of a single morpheme: SSRs which differ in some phonological features, and of which one is free, while the other is bound and occurs uniquely with the past tense suffix.

3.9 The Lexicon

The Principle of Commonality and Distinctiveness implements Principles 1 and 2. The lexicon is the implementation of Principle 3. The typology of morpheme implicit in Principle 3 captures the relationship between the internal parts of a morpheme: the UR, the Separate Surface Realisations, and any subfamily groups. This information, identified by the analysis in 3.8, is entered in the lexicon.

3.9.1 Defining the Lexicon

In 3.3 morphology was characterised as consisting of morphemics and morphological structure. While the latter consists of rules, morphemics is located in the lexicon.

In some lexicalist theories such as LP, 'lexicon' refers to the domain of morphology, and the output of the lexicon equals the output of the morphology. The formation of morphemes into words, however, involves significant regularities, representable as rules. It would be inefficient to list such regularities. The present approach regards the lexicon as a list of that which cannot be represented by rules, as a repository of that which is idiosyncratic. The lexicon lists all and only the idiosyncratic features of all morphemes.

8. A full analysis of the *do* paradigm may also be found in 4.4.

In 3.6 phonological and morphophonemic alternations were excluded from the lexicon. Equally, predictable semantic information, such as that a free morpheme which is underspecified for word-class is a noun in a post-article environment, need not be listed. Structural idiosyncrasies, such as the possible collocation of other morphemes, are listed, but the resulting forms need not be as they will then be predictable.

By this definition the lexicon cannot list idiosyncratic polymorphemic forms. It cannot list non-compositional compounds or idiomatic phrases, or predict the idiosyncratic syntactic behaviour of polymorphemic words. The lexicon defined above must contain as much information as possible to allow the prediction of the characteristics of polymorphemic forms, but this is unlikely to be exhaustive.

It is beyond the scope of the present work to deal with the storage of the output of the morphology. Indeed, it is beyond the scope of the morphemics to do so. It is, however, consistent with the present approach that there are two lexica, one listing morphemes, the other more complex forms.⁹ In the present work, however, our concern is with the morphemic lexicon.

3.9.2 The Structure of the Lexicon

Every morpheme has a lexical entry, and that entry has an internal structure in accordance with the typology in Principle 3. Each entry consists of the UR; any subfamily groups; and all SSRs. The structure typically looks something like this:

- (24) Underlying Representation
1. Subfamily Group
 - 1.1 SSR
 - 1.2 SSR
 2. Subfamily Group
 - 2.1 SSR
 - 2.2 SSR

Of course there may be no subfamily groups, or several levels of them, and any number of SSRs in any group. The representation at (24) merely indicates the structure of lexical entries.

9. It may be argued that these lexica operate at different levels. The second lexicon consists entirely of consciously defined entities, which the speaker will say "means something". The first lexicon contains much material definable by speakers, but this is not a necessary feature. For example, as has been demonstrated, phonesthemes have a semantic strength which allows them to be learned and stored, and generate new words in a language. No speaker would associate a meaning with a phonestheme in isolation, yet words containing phonesthemes "sound right".

Table (25) below exemplifies the structure of a line in a lexical entry, using the entity *di-*, as analysed in (23). In each line in an entry there are three sets of brackets: square brackets containing formal features; angle brackets containing semantic features; and curled brackets containing structural information. The formal and semantic brackets contain the features which characterise the morpheme. The structural brackets supplement this by identifying any idiosyncratic structural matters.

(25) $\left[\begin{array}{ccc} C_1 & V_1 & -> \\ | & | & \\ d & i & \end{array} \right]$ < do > { Specification for collocation with PAST suffix.}

Note that (25) contains an arrow. This is an 'Other Morpheme Requirement' (OMR). These will be discussed in 3.10. For the present purposes it simply means the entity is bound.

3.9.2.1 Formal Feature Brackets

These appear at the beginning of each line, and list the morpheme's formal features. Segmental features are listed as Cs and Vs, and numbered. These canonical segments are linked by association lines to sets of phonological features. For the purposes of simplicity, where these features constitute the full set of features of a phoneme, that phoneme is listed by its IPA symbol. For example, in (23) the morphemic entity *di-* was analysed. This entity's formal features would be represented as follows:

(26) $\left[\begin{array}{ccc} C_1 & V_1 & -> \\ | & | & \\ d & i & \end{array} \right]$

3.9.2.2 Semantic Feature Brackets

These list all semantic features. Syntactically functional meanings are represented in small letters, rather than as syntactic categories in capitals. For *di-* the semantic brackets appear as follows:

(27) < do >

3.9.2.3 Structural Details Brackets

The structural brackets contain a range of details. Any morphophonemic or word formation rules which apply to the entity are invoked by rule referrals. Restrictions on what other morphemes or kinds of morphemes may be collocated with the entity are listed.

As the present work assumes no specific theory of morphological structure, it describes information given in the structural brackets as simply as possible, intending to avoid any theoretical formalism. The manner in which the information is listed is meant to represent a kind of shorthand, not a formal representation. Where no structural information has been discussed in relation to an entity, its structural brackets will contain 'X', representing the set of applicable structural details. Where a structural issue is discussed it will be listed in the structural brackets, but this does not imply that it is the only structural information relevant to that entity.

The entity *di-* occurs uniquely with the past tense suffix. For the present purposes that will appear in the structural brackets as:

(28) { Specification for collocation with PAST suffix.}

In the structural brackets the terms 'subcategorisation' and 'specification' are used. In this work where an entity co-occurs with a particular entity, for example, it is said to 'subcategorise' for that entity. Where some information about an entity is relevant to its role in morphological structure, it is said to 'specify' that role. For example *ox* will subcategorise for its irregular plural suffix, while the suffix *-en* will specify as PLURAL.

3.9.3 The Underlying Representation

The UR is the distillation of all surface realisations of the morpheme. It consists of features present in every realisation of the morpheme. It may also contain optional features, which are listed in parentheses, indicating their optional status, and must be invoked by the realisations to which they apply. Optionality will be discussed in detail later.

As (23) indicated, *do* and *di-* are SSRs within a single morpheme. Consequently the UR will contain the formal and semantic features present in both, as (29) illustrates.

(29)
$$\left[\begin{array}{l} C_1 \quad V_1 \text{ (->)} \\ \left[\begin{array}{l} | \\ d \quad +\text{high} \\ -\text{back} \end{array} \right] \end{array} \right]$$
 < do > { X }

The UR in (29) fully specifies everything which is general to both realisations. The two SSRs only differ in the frontness of the vowel. That will be specified in the lines for each SSR. The structural brackets contain 'X', representing for our purpose all structural details relevant to the whole morpheme. It does not include the

collocation restrictions for the bound SSR because that applies to only one SSR.

Note that the OMR arrow is in parentheses, indicating it is optional. It is optional in the UR because it applies to only one SSRs.

3.9.4 Separate Surface Realisations

The entry for a morpheme's SSRs do not list features specified as obligatory in the UR, as these automatically apply to every realisation.

In the UR in (29) only two issues are unresolved: the frontness of the vowel, and the presence or absence of the Other Morpheme Requirement. Only these issues are dealt with by the SSR entries. The entry for the SSR *do* is:

- (30) 1. $\left[\begin{array}{c} V_1 \\ | \\ [-\text{front}] \\ [+long] \end{array} \right]$ < > {}

The empty semantic and structural brackets indicate there are no semantic or structural features specific to this SSR: only and all the features specified in the UR apply. The formal brackets contain only those phonological features remaining for the vowel to be realised. The Other Morpheme Requirement, optional in the UR, is not invoked by the SSR in (30), so does not apply. This SSR, in conjunction with the UR, generates *do* :<do>.

The second SSR of this morpheme is that occurring in *did*. This appears in the lexicon as:

- (31) 2. $\left[\begin{array}{c} V_1 \text{ ->} \\ | \\ [+front] \\ [-long] \end{array} \right]$ < > { Specification for collocation with PAST suffix.}

This entry specifies no additional semantic features, and specifies the remaining phonological features of the vowel. In addition it invokes the Other Morpheme Requirement, indicating the SSR cannot occur in isolation. The structural brackets specify constraints on what morphemes may co-occur. This SSR, coupled with the UR, generates *di-* :<do>

The UR and SSR entries form a single lexical entry for the morpheme:

(32)	$\left[\begin{array}{c} C_1 \quad V_1 \quad (->) \\ \quad \\ d \quad +high \\ -back \end{array} \right]$	< do >	{ X }	
	1.	$\left[\begin{array}{c} V_1 \\ \\ -front \\ +long \end{array} \right]$	< >	{ }
	2.	$\left[\begin{array}{c} V_1 \quad -> \\ \\ +front \\ -long \end{array} \right]$	< >	{ Specification for collocation with regular PAST suffix. }

3.9.5 Default Values

Structural details brackets do not contain rules of word formation. Instead they refer the morpheme to rules which apply to that morpheme. This, however, causes the potential for duplication between the lexicon and morphological structure, duplication which is avoided by making only unpredictable rule referrals. When a morpheme behaves regularly in word formation, relevant rules are not listed. Idiosyncratic behaviour is dealt with in the lexicon, not by accounting for the behaviour, but by referring the morpheme to the appropriate rule.

Cat, for example, forms *cats* by the regular, productive plural marker. Consequently its lexical entry does not include a rule referral for plural. English nouns must mark plurality unless otherwise specified, so *cat* must take a plural marker, and since which marker it takes is not specified, it must take the regular marker by default. To enable this, the regular suffix, in its own lexical entry, is specified as the default marker.

In contrast the lexical entry for *ox* contains a rule referral for the irregular marker *-en*. *Sheep*, on the other hand, being specified in the lexicon as being both singular and plural, neither refers to a plural marker nor attracts the default form.

This default system extends to broader issues like word position. In a language where subject agreement markers always occur immediately preceding the root, no word position is listed in the lexical entry for such entities, that being dictated by rule.

3.10 Other Morpheme Requirements

The presence of an Other Morpheme Requirement (OMR) in a lexical entry indicates that the morpheme cannot be realised without the presence of another morphemic entity. OMRs are represented as arrows:

- (33) [**<-**] This indicates that the other morpheme must precede the listed entity. This may mean that the required entity is a prefix, or that the listed entity is itself a suffix. For example the root in *resume* would be [**<-zyum**], and the irregular plural in *oxen* would be [**<-ən**].
- [**->**] This indicates that the other morpheme must follow the listed entity. This may mean that the required entity is a suffix, or that the listed entity is a prefix. For example the *knowl-* from *knowledge* discussed in (16) would be [**no1 ->**], and the prefix in *prefix* would be [**pri ->**].
- [**<->**] This indicates that the required morpheme is an infix. This arrow may appear in a discontinuous form. For example in the Hebrew *kitab* 'write', in which the consonants are a root morpheme and the vowels a tense/aspect morpheme, the consonant morpheme would be represented as [**k<-t->b**], indicating that a single discontinuous two part infix occurs in the slots indicated by the arrows.¹⁰ This OMR example does not have the form [**k <-> t <-> b**] because that would indicate that two separate infixes were required.
- [**<<- ->>**] This indicates a circumfix is required. The vocalic morpheme in *kitab* would be [**<<- i - a ->>**], indicating that the required morpheme is a discontinuous three part entity. The middle dash is of the circumfix. If that dash had arrowheads of its own it would indicate it was an OMR in its own right, separate from the circumfix. This OMR arrow has double arrowheads to distinguish it from merely a prefix and a suffix OMR: [**<- X ->**] merely indicates that entity X requires both a preceding and a following morpheme, while [**<<- X ->>**] indicates that a circumfix is required.

An OMR is a feature of the morpheme. OMRs restrict the realisation of bound entities, and indicate how many other morphemes are required and what kind of morphemes they must be. This allowing a mapping of appropriate morphemes: in the Arabic illustration the OMR arrows indicate that *k-t-b* and *-i-a-* may co-occur.

10. An analysis of semitic discontinuous morphemes may be found in 4.10.

3.11 Realisation Features

Realisation Features (RFs) are not morphemic entities. They are unique features of a single morpheme which are present in one or more surface realisations, but are not in all. They have three key characteristics:

- (34) (34.1) An RF may be a formal or a semantic feature, but, having no independent morphemic status, may not be a formal-semantic correspondence.
- (34.2) RFs must not be attributable to any rule. Phonologically or semantically predictable features are not listed in the lexicon. Being features unique to realisation(s) of a single morpheme, and consequently not part of any sub-regularity, RFs may not be attributed to a morphophonemic rule.
- (34.3) RFs are optional. Not being present in every realisation, RFs are not underlying obligatory features of their morphemes. They are only obligatory for the appropriate SSR.

Two RFs have been discussed already: the vowel frontness in *do* and *did*; and the [l] in *knowledge*. Neither carry meaning or can be attributed to a morphophonemic rule. As such they are stored features of their respective morphemes.

RFs are listed in one of two ways. Where the features are smaller than a single segment that segment is underspecified in the UR for those features, which are then specified in the appropriate SSR(s), enabling realisation of that segment. This applies with the *do/did* vowel.

Where RFs consist of one or more segments, those segments are listed as optional in the UR, and then invoked in the appropriate SSR(s). This applies to the [l] in *knowledge*.¹¹

3.12 Submorphemes

A submorpheme is a morphemic entity which occurs uniquely, in a symbiotic relationship with a host morpheme, and as such does not have the status of morpheme.

11. *Do* and *did* are analysed in detail in 4.4. *Knowledge* is discussed in 4.6.

3.12.1 Unique Residues as Morphemic Entities

As discussed in 2.3.4, allegedly empty morphemes such as *cran-* are meaningful. PCD readily demonstrated this, as a comparison of *berry* and *cranberry* reveals:

(35) Comparing *berry* and *cranberry*:

Features	Commonality	Distinctiveness
formal	beri	/kræn
semantic	small fruit etc	/those real world characteristics which distinguish cranberries from all other berries

An apparently meaningless residue may be demonstrated to be meaningful if it is not unique. With the phonestheme based words discussed in 3.8, the isolation of [fl] from *flow* leaves the residue [oʊ], which also occurs in *glow* and is attestably meaningful:

(36) Comparing *flow* and *glow*:

Features	Commonality	Distinctiveness
formal	oʊ	fl / gl
semantic	evenness	movement / emit light

Equally, isolating [oʊ] from *flow* leaves the residue [fl], also occurring in *flutter* and also attestably meaningful.¹² Being non-unique sound-meaning correspondences [fl] and [oʊ] are independent morphemic entities.

Other phonesthemic words, however, are like *cranberry*, and leave a unique residue. Consider the isolation of [fi] from *flee*, by comparing *flee* and *flow* :

(37)

Features	Commonality	Distinctiveness
formal	fi	i / oʊ
semantic	movement	rapidity, away / evenness

As we saw in (36), [oʊ]:<evenness> is a non-unique entity. In (37) the semantic distinctiveness corresponds with the formal distinctiveness [i]. Being a sound-meaning correspondence, [i] is a morphemic entity, however it is unique.

12. A full analysis of this data may be found in 4.8.

3.12.2 Unique Entities: Morpheme or Submorpheme?

Like *cran-*, [i]:<rapidity, away from> is a unique morphemic entity. However, the relationship between *cran-* and *berry* is different to the relationship between [fl] and [i]. *Cran-* is overtly isolatable because it combines with a semantically independent entity, *berry* needing no additional information to be realised semantically. In contrast, [fl] has a semantic value which, must be activated by the presence of an additional morphemic entity. This is manifest by the Other Morpheme Requirement of [fl]. This OMR may be satisfied by another, separate, morpheme, or as in the case of *flee*, by a unique morphemic entity. Since the host morpheme in *flee* depends on the presence of another entity, where that entity is unique the two are in a particular kind of relationship: the host is an independent morpheme in a symbiotic semantic relationship with the non-independent unique entity.

These dependent unique entities are morphemic entities, but do not have the status of a distinct morpheme. They are 'submorphemes'. A submorpheme functions to generate a surface realisation of the morpheme which is semantically dependent on it. They associate with their host morpheme precisely as another morpheme would, adding semantic information and formal features; but since they are unique, and satisfy the semantic dependency (represented as the OMR) of the host morpheme, they operate entirely within the structure of the host.

Submorphemes have no life independent of their host. The host in turn depends on them to satisfy its own realisation requirements. They do not combine with the host as equal partners, but are invoked from within the host morpheme's structure.

3.12.3 Principle 4

This defines 'submorpheme':

- (38) Principle 4. A unique morphemic entity which satisfies an "Other Morpheme Requirement" of the host morpheme has the status of submorpheme and operates as part of the host morpheme. A unique morphemic entity which attaches to its host morpheme without satisfying an "Other Morpheme Requirement" is a distinct morpheme and operates from outside the host.

3.12.4 The Lexical Representation of Submorphemes

Submorphemes appear in the lexicon within the host's entry, as a sub-entry with its own set of features. As submorphemes have their own morphemic status their

features remain distinct from those of the host, and do not appear in the host's UR. Their features combine with those in the host's UR to generate surface realisations, and they distinguish the realisations in which they occur from other realisations.

Flee is an SSR of the [fl]:<movement> morpheme, generated by the submorpheme [i]:<rapidity, away from>. In the lexicon the submorpheme generates an SSR which contrasts with the SSR isolated from *flow* :

(39)	$\begin{bmatrix} C_1 & C_2 & -> \\ & & \\ f & 1 & \end{bmatrix}$	< movement >	{ Specification for collocation with another phonestheme.}
	1. []	< >	{ }
	2. [[V ₁]]	<< rapidity, away >>	{ }
	[[]]		
	[[i]]		

The UR is the distillation of features found throughout the morphemic entity [fl], but does not including those found in the submorpheme, which is separate entity.

There are two SSRs, each containing all features the morpheme requires for realisation. SSR1 is that isolated from *flow*. It has no submorpheme to satisfy its OMR, so that remains, allowing the SSR to combine with appropriate morphemes like [oU] and [itə].

SSR2 includes the submorpheme, listed with its formal and semantic features. These features are contained in their own sets of feature brackets, within the feature brackets of the host. This distinguishes the features of submorpheme from the features of the host morpheme. The features listed for the submorpheme combine with those in the host's UR, mapping on to and satisfying the host's OMR, and generating the free form *flee*.

3.12.5 Submorpheme or Realisation Feature?

Any unique residue which cannot be attributed to a rule must be examined to determine whether it is part of a sound-meaning correspondence. If it is not, then it has no morphemic status and is a Realisation Feature of one of the morphemes present. If it is part of a correspondence, then it is a submorpheme. The [l] in *knowledge*, apparently left as a residue by the isolation of the suffix and the root, has no meaning. Consequently, having no morphemic status, it must be an RF of an adjacent morpheme, in this case the root. If *knowl-* had been semantically distinct from *know*, then the [l] would have been a submorpheme.

3.13 Recapitulation of Morphemic Principles

- (40) Principle 1 Any phonological feature or set of features which has any semantic value and is semantically undecomposable has morphemic status.
- Principle 2. Any group of morphemic entities which have both formal and semantic links constitute separate realisations of a single morpheme.
- Principle 3. A morpheme consists of an Underlying Representation and its realisations. The UR is the semantic and formal distillation of the morpheme's various semantic and formal realisations, and those various realisations are the surface realisations of the UR.
- Principle 4. A unique morphemic entity which satisfies an "Other Morpheme Requirement" of the host morpheme has the status of submorpheme and operates as part of the host morpheme. A unique morphemic entity which attaches to its host morpheme without satisfying an "Other Morpheme Requirement" is a distinct morpheme and operates from outside the host.

These principles provide a theoretical framework for morphemics. Principle 1 isolates morphemic entities; Principle 2 turns these entities into a set of morphemes and their surface realisations; Principle 3 outlines the internal structure of morphemes; and Principle 4 defines submorphemes.

3.14 The Irrelevance of Etymology in Synchronic Analysis

It remains to make a few remarks about the kind of information which may be considered when conducting synchronic analysis.

If an approach to analysis is to be adequate, it must be applicable to any language. Consequently, it must only rely on the kind of data which is available in any language, and must be capable of accounting for all data extant at the time of analysis. Synchronic analysis must exclude any information which relies on historical data, and must consider actual usage, even if that offends the sensibilities of the etymologically informed.

3.14.1 Historical Data

Etymological and other historical data is irrelevant to synchronic analysis. To use such data in analysis reduces the adequacy of a theory in two ways: in terms of psychological reality; and in the practical consistency of the theory's application.

If analysis is to accurately account for a body of data in a way which reflects psychological reality, it must limit its source of data to all and only the forms present in the language. Few speakers of a language know its history, and those that do learn that long after the language itself.

Secondly, if a theory is to be adequate it must be able to be applied consistently to any language. To regard historical material as an appropriate factor in analysis has two possible effects: either the theory must be applied inconsistently (it will deal with languages whose history is known differently to those whose history is unknown); or if applied absolutely consistently, it will be unable to deal with a great many languages. Inconsistent or widely inapplicable, either way the adequacy of the theory is severely impaired.

Having said that, it will be apparent that the PCD approach, which is concerned with accounting for the relatedness of forms, will reflect elements of a language's history. This is, however, an incidental byproduct of the approach.

Leaving theoretical adequacy aside, historical fact may be misleading synchronically. There is no reason to assume that contemporary relationships between forms will correspond with their historical relationships. A form may change meaning or sound, and be regarded by speakers as related to an etymologically distinct form. For example *leave*, as in *without pay*, has a widely regarded connection with *leave*, as in 'depart': you *leave* work for a couple of weeks. However while *leave*, as in 'depart' is from the Germanic *laibhjan*, meaning 'remainder', via the Old English *læfen*; the *leave* we go on from work comes from the West Germanic *laubha*, meaning 'love', via the Old English *leaf*, meaning 'permission'. The etymologies are separate, yet the forms have a relationship in contemporary English. This suggests there is a pressure for speakers to semantically reanalyse and attempt to link homophonous forms. Accepting the existence of that pressure allows the analyst to accept relationships between forms despite divergent etymologies.

Equally a single etymological root may generate forms which diverge to the point where there is no psychological link. *Brochure* and *brooch* are derived from the Old French *broche*, meaning 'pointed tool', but would be extremely difficult to argue that the two have any current connection.

The irrelevance of etymology to the psychological reality of a language, and the total absence of such information for many languages, indicates that such information has no place in an adequate synchronic theory.

3.14.2 The Effects of "Folk Etymology"

"Folk etymologising" is a term describing speaker reanalysis of semantic values. It is misleading, suggesting that there is 'proper' language change, and change motivated by speaker ignorance. This is an inappropriate attitude for the linguist. Folk etymologising is a process of speakers identifying links between forms. It involves assumptions on the part of speakers that a relationship exists between particular forms. It is irrelevant to language change whether an historical relationship exists between such forms, and it is certainly irrelevant to synchronic analysis.

As indicated by *leave*, and by phonesthemically driven phonological and semantic change,¹³ if the view that particular forms have a relationship is prevalent among speakers, then that relationship exists, and will influence semantic and formal change. Lyons (1977:4) observes that "the meaning of words and sentences is learned and maintained by the use to which language is put in communicative situations." If that use involves an assumed link with another form, and the related meaning is both intended and understood, then the associated meaning becomes the meaning.

An understanding of this speaker reanalysis allows the analyst of languages with a wealth of historical data, like English, to conduct truly synchronic analysis. As an example of this kind of reanalysis and its effects, consider *hamburger*, which underwent a striking series of speaker reanalyses, striking because the element initially responsible for generating change then became the subject of further change, generated by the element it had initially caused to change.

Originally the homophony of the word *ham* and the first syllable of *hamburger* gave the latter the appearance of a compound. The apparent presence of the nominalising suffix made *burger* itself sound like a noun.¹⁴ A *hamburger* is an object with the key ingredient of meat, and *ham* is a meat. The environment for reanalysis was perfect, making it almost inevitable. Consequently the word *burger* was created, referring to a specific kind of food item.

This item, being flexible, was developed with a range of central ingredients: minced beef; cheese; minced vegetarian foodstuffs etc. *Burger*, having become the term for

13. Phonesthemes were discussed in detail in 2.2.6.

14. The *-er* was historically a suffix attached to the town name *Hamburg*, the whole word having been borrowed as a single unit from German.

that kind of object, could then form additional compounds and become *beefburger*; *cheeseburger*; *vegeb主rger* etc. The original component *ham-*, which had caused the reanalysis of *burger*, then itself came under semantic scrutiny.

Burger was defined, so *ham*, appearing in a compound with *burger*, had to mean something. However, since *hamburgers* are made with beef, that *ham-* was not synonymous with the meat *ham*. The notion of *hamburger* was distinct from *burger* in that the former was the 'traditional' *burger*, as available before the advent of *cheeseburgers* etc. It has a certain set of traditional ingredients and is available from a traditional takeaway. *Hamburger* denotes a particular kind of *burger*, and carries with it a certain amount of cultural baggage.

The *ham-* in *hamburger* was analysed by speakers as the meat, inspiring speaker reanalysis of *burger*. Once established this itself inspired reanalysis of *ham-*.

As a consequence of this, a synchronic analysis of English must ascribe the status of morpheme to *burger* (or perhaps to *burg-* and *-er*), and to *ham-*. The latter being isolated and defined precisely as *cran-* was in 3.12.1.

Had this semantic reanalysis occurred six hundred years ago, there would be no question about the validity of this analysis. As the reanalysis occurred sufficiently recently for some speakers, and certainly any analyst, to be aware of it, the veracity of the analysis is questioned. However, if a truly synchronic analysis is the goal, then six hundred years or sixteen makes no difference: this analysis of *ham-* is correct for contemporary English. It is in this spirit that all synchronic analysis should take place.

The phenomenon of folk etymology is a feature of the normal process of language change: speakers tend to look for a link in meaning between homophones, and to attempt to unify the sound shape of phonologically similar and semantically related items. These processes may or may not reflect historical derivation, and that, for the purposes of synchronic analysis, is irrelevant.

The PCD approach regards language as involving a constant process of speaker reanalysis, and allows an analysis which recognises this and provides a truly synchronic account.