This is to certify that the

thesis entitled

Lexeme Formation Tendencies in Modernising Languages

presented by

J. Barrie Evans

has been accepted towards fulfillment
of the requirements for

Masters degree in Linguistics

Carol Myers Scotton

Major Professor

Date 8/12/85

LEXEME FORMATION TENDENCIES

IN

MODERNISING LANGUAGES

By

James Barrie Evans

A THESIS

Submitted to

Michigan State University

in partial fulfillment of the requirements

for the degree of

MASTER OF ARTS

Department of Linguistics

1985

ABSTRACT

LEXEME FORMATION TENDENCIES
IN
MODERNISING LANGUAGES

By

James Barrie Evans

The thesis examines the phenomenon of lexical development in modernising languages. In so doing it puts forward a hypothesis that planned or conscious lexical development can be distinguished from spontaneous lexical development in terms of the composition of the lexemes that are formed. The thesis also examines the terminology for describing lexical composition and proposes an extended terminology. The hypothesis is tested by examining data from Welsh and Swahili, and by reviewing articles on lexical development in a few other languages. As a justification for the thesis it is claimed that lexical development, particularly in Third World countries, can be related to economic development, since lexical development affects the extent to which indigenous languages can be used in education and training.

ACKNOWLEDGEMENTS

The preparation of this thesis has been greatly facilitated by the assistance and encouragement of Professors Carol Scotton, David Lockwood and David Dwyer.

Professor Carol Scotton, who also kindly acted as the chair of the thesis committee, provided an overall background and orientation for the thesis and offered useful criticism on its structure and rationale.

Professor David Dwyer provided useful and encouraging discussions on the morphology and structure of languages.

Professor David Lockwood provided guidance to reference works on lexical innovation and lexical structure in the English language, and a yardstick of systematic thinking to bear in mind throughout the thesis.

Particular gratitude must be expressed to Yunus Rubanza, Swahili instructor at Michigan State University, who spent many hours patiently explaining the morphological structure of Swahili and the clarifying the word lists that were used in the analysis.

TABLE OF CONTENTS

LIST OF TABLES

1. OVERVIEW

1.0   The intention of this study is to examine some aspects of the phenomenon of lexical development in languages which are having to increase their vocabulary very greatly in order to cope with the demands of the 20th century. Such languages often resort to planned or conscious means of lexical development in order to achieve the expansion that they view as necessary.

The thesis assumes that this process is an important factor in the socio-economic development of a country. The means by which languages expand their vocabulary therefore merits study, in the same way that a study of agriculture or engineering construction techniques would be justified. The thesis proposes that this study is particularly relevant to Third World countries.

The thesis also recognizes the theoretical value of studies in lexical composition, and in section 2 proposes a framework by which lexical composition can be classified, and a hypothesis which distinguishes planned and spontaneous development based on this framework. In the final part of the thesis, data from a few languages which are undergoing planned lexical development are examined and compared to the framework and hypothesis developed in section 2.

1.1 Topic of Study

The topic of this thesis is lexeme formation tendencies in modernising languages. By the term modernising languages is meant languages that are undergoing vocabulary expansion, normally both planned and spontaneous, in order to extend the societal usage of the language because it has become, or is felt to be in danger of becoming, restricted. This definition includes true minority languages, such as Welsh or Breton, and not-so-minority languages,

1

such as Swahili or any other African or Asian language, which may be spoken by many millions of speakers over a wide area and be major languages in some of the areas where they are spoken.

In considering the modernisation of languages, the modernisation of their vocabulary is of prime importance. Consequently a study of the phenomena involved in lexical development is of relevance to the general process of language modernisation. Lexicology or lexical semantics is not an area of linguistic theory which receives very much attention at the moment - certainly not as much as syntax or phonology - yet it would seem a principled approach could be an important contribution to the subject of vocabulary modernisation.

The topic of this thesis is, therefore, to consider the phenomenon of lexical development, and to bring to bear a principled framework as far as is possible. A range of lexeme formation possibilities will be considered, from single-morpheme lexemes through to multiple-word lexemes.

1.2 Practical Relevance of Study

The underlying justification for the thesis is that the use of indigenous languages, particularly for education and training in Third World countries, can be positively related to economic development.

An important document advocating the value of the mother tongue in education is the 1951 UNESCO report on The Use of Vernacular Languages in Education. However, in the scramble of newly independent colonial countries in the 1950's and 60's to achieve a reasonable techno-economic standard of living, the indigenous languages were often considered hinderances and the Western languages of wider communication a gateway to success.

Ikwue (1984) and Chishimba (1984) claim to document this for Nigeria and Zambia respectively. The result in both these countries was an emphasis on the use of English in education, even in the early years at school. In both these countries, however, these papers point out that achievements in education have been low. Ikwue states: 'the obvious disadvantage of this policy has been that the education system has been producing a large population of, generally speaking, semi-literate people highly incapable of using their school knowledge functionally' (1984:40). Chishimba states: 'it has been noted that children entering secondary school come with low competence in speaking, writing, reading and understanding English' (1984:165). Ikwue describes the situation in Nigeria where regional languages are now given greater emphasis in schooling: Chishimba puts forward pedagogical arguments for the greater use of Zambian languages in education, particularly early school education.

Ansre (1977) and Del Rosario (1968) pick out some specific aspects of the use of indigenous languages in education. Ansre considers four arguments put forward by those who favour languages of wider communication for education, one of these being 'that . . . African languages are not developed enough in scientific and technical terminology . . . and since the effort and cost of evolving such terms for completely new concepts are prohibitive, therefore we should be content, at least for now, to use languages in which these terms have already been well developed and whose speakers are advanced in modern science, technology and the humanities in educating our children' (1977:60). Ansre goes on to argue against this position, by first claiming that it is quite possible to develop the vocabulary corpus of a language, and second, by pointing out that '. . . the more developed communities use their own languages in education and technical training' (1977:61). He believes this to be one of the reasons for

the economic and technological advancement of Japan. Consequently he implies that the use of a Western language can rather be detrimental to the acquisition of technological concepts by children.

Del Rosario (1968) follows a similar line of thinking for south-east Asian languages. He advocates the development of the these languages, using indigenous morphemes wherever possible, in order to create languages which will be as understandable as possible to school children. '[A national language] should enable the school system to teach knowledge in general, and science in particular, very quickly to school children' (Del Rosario 1968:3).

Undoubtedly it would be possible to argue the pedagogical advantages of indigenous languages in education in greater detail, however these quotations have highlighted the possibility that their use could be pedagogically advantageous. If this is true of elementary school education then surely it would inevitably be true, at least for the majority of the population, of higher school education and subsequent technical training as well.

Now it is obvious that there are some considerable problems in the use of indigenous languages. The two most important problems appear to be: (i) the cost of developing the languages and providing pedagogical material in them; and (ii) which languages to promote, at least for the vast majority of Third World countries in which more than a few languages are spoken.

Nevertheless there are many countries which are taking this path of modernisation of indigenous languages and providing pedagogical material in them. This is not restricted to Third World languages; it occurs in the United States with American Indian languages, and also with the minority languages of Europe.

Consequently a study which clarifies the phenomena involved in lexical development can be of value to language modernisation, which in turn, as argued above, can potentially be positively linked to economic development.

1.3 The Theoretical Relevance of the Study

In a review of Lyons (1970), in which Lyons reviews the work of Noam Chomsky, Dell Hymes states: 'to say that every language has a sufficiently rich vocabulary for the expression of all distinctions that are important in the society using it is to beg a host of questions; it is a form of 'functional optimism' or Panglossia, that would not pass muster for a moment in the political circles in which Noam Chomsky figures, if seriously considered' (1972:417). Dell Hymes goes on to state: 'if linguists were to cease treating vocabularies as functionally perfect by definition, returning to the study of lexical creation and change, and the more general question of a theory of vocabularies, a great deal of good might result' (1972:417).

The study of semantics and lexicology, and consequently also lexical semantics, has suffered considerable lack of emphasis in this century, particularly in the United States (see Weinreich, 1963). It is perhaps natural that while a considerable amount of progressive study was being done first in phonology and then (and still to this present day) in syntax, the study of lexical semantics should suffer. At the moment there appear to be signs of a renewed interest in lexical semantics (e.g. Fillmore 1971,1978, Lakoff and Johnson, 1980, Downing 1977).

The study of lexical semantics has been kept much more alive in Europe, and has generally been considered a more central component of linguistic theory. Ullmann is one of the scholars who has kept this

tradition alive in the English speaking world and his two semantic studies with a large lexical component — Semantics: An Introduction to the Science of Meaning, and The Principles of Semantics — have been of great value in preparing this thesis.

If three foci of interest can be identified in lexical semantics they are:

(i) the meaning of 'meaning';
(ii) the structure of words;
(iii) change of meaning.

All of these are highly relevant to a consideration of lexical expansion.

Nevertheless few studies seem to attempt to apply formal lexical–semantic considerations to the lexical expansion of Third World languages. Tumbo (1982:90), for instance, in considering the development of Swahili terminology, goes so far as to state that the use of general linguists should be avoided. Ohly (1977, 1982), of all the scholars referred to, seems to have a formal understanding of lexical semantics. On the other hand, scholars of lexical semantics (e.g. Ullmann 1957,1962, Coseriu 1981, Coseriu and Geckeler 1981) do not consider the application of their work to the conscious modernisation of languages.

In addition, most of the studies that have been referred to in preparing this thesis only consider one type of lexical innovation (e.g. Del Rosario (1968) on affixes, Opalka and Passierbsky (1977) on noun compounding, Sharma (1968) on affixing, Epstein (1959) on metaphor and borrowing). In this thesis an attempt has been made to provide a broader classification of lexical structure as the basis for a characterisation of lexical change in modernising languages.

1.4 Theoretical Orientation

Although little generally accepted theory exists within lexical semantics it is nevertheless possible that an underlying theoretical orientation could be used. Within Michigan State University three such 'orientations' are taught - Stratificational Grammar, Tagmemics and Transformational Grammar - having various degrees of compatiblity, and any of these could be used as a basic starting point.

The approach taken in this study is believed to be generally compatible with both Stratificational Grammar and Tagmemics, and terminology from both these theoretical approaches may be found in the study.

However, in order to provide a broader orientation, a broadly functional approach to grammar has been taken as more fundamental. The scholars that have most been referred to in this respect are Ullmann (1957, 1969), Lyons (1977), Coseriu (1962,1981), Coseriu and Geckeler (1981), Dik (1981) and Comrie (1981).

Ullmann has provided a general appreciation of lexical semantic phenomena and terminology. Lyons (1977), in particular his discussion on the lexicon, has provided a starting point for a classification and terminology of lexemes.

The influence of Coseriu is less easy to state. The four aspects of his work which probably have been most influential have been: his categorisation and description of lexical fields; the distinction he makes between the linguistic and the extralinguistic; his distinction between system, norm, and speech (usage); and the kaleidoscopic and continually shifting nature of language.

Dik (1981) has provided an overall paradigm of a Functional approach to grammar. Comrie (1981) has also provided some particular

examples of an 'integrative approach to language in context'. Adams (1980) and Bauer (1983) have also provided terminology and classifications in the study of word formation, from their works on English word formation, and also models of comparable studies.

1.5 The Hypothesis

As a vehicle of investigation, a hypothesis has been chosen which proposes a distinction between spontaneous and planned lexical development, based on a distinction between different types of lexeme formation. The hypothesis proposes that planned development has a tendency to make greater use of composite lexemes, particularly complex lexemes, than spontaneous lexical development.

By planned lexical development is meant the deliberate creation of new lexemes for general use. When this is done by an institution it will be termed formal planned lexical development, when it is necessary to distinguish it. Anything other than this will be considered spontaneous lexical development.

A composite lexeme may either be a derived or complex lexeme (one which consists of a root and one or more affixes) or a compound lexeme (one which consists of more than one root). The definition of lexemes is considered further in section 2.

1.6 Methodology

The specific methodology that has been used for testing the hypothesis is to develop in section 2 a principled framework of lexeme formation possibilities, which distinguishes the categories mentioned in the hypothesis. Then in section 3, using lexical data from a few modernising languages, this data is compared to and where possible

distributed to the framework. These analyses should confirm or disconfirm the hypothesis.

1.7 Sources of Data

Whilst it is not difficult to obtain lists of new planned lexemes, for those languages which are undergoing planned lexical development, one of the difficulties has been to obtain lists of spontaneous lexeme development. Consequently the number of languages that could be considered has been severely constrained.

The languages for which such material appeared to be available were Welsh and Swahili. A deliberate attempt was made to include a Third World language, as well as a language which was both European and Indo-European. However, in both section 2 and section 3 evidence from other languages has been considered when it is available and appropriate.

Welsh is a language which is receiving a great deal of attention at the moment from the point of view of modernisation. After a period earlier in this century when its usage for general institutional purposes was declining, it is now being used to a greater extent in education and the media. There is nevertheless still a problem here of obtaining lists of spontaneous development. However Watkins (1961) contains information on affixes that he states are no longer generally productive, based on word lists of the spoken language from a rural district of Wales. If it can be found that these affixes are still in use in planned lexical development then this should constitute evidence for the hypothesis.

Swahili provides fairly good documentation against which the hypothesis can be tested. Since colonial days a considerable amount of standardisation has taken place in Swahili (Whitely, 1969), and

since the independence of Tanganyika and Zanzibar (now Tanzania) an equal amount of effort has been expended on vocabulary modernisation. The Baraza la Kiswahili la Taifa (BAKITA - National Swahili Committee) regularly publishes lists of new lexemes in <u>Tafsiri Sanifu</u>. Ohly (1982) investigated the more informal process of lexical development which was taking place in a textile mill, and this report has been used as a source of information on spontaneous lexical development. Since it is technologically oriented it has been compared with a list of technological words from <u>Tafsiri Sanifu</u>.

2. CLASSIFICATION OF LEXEME FORMATION TENDENCIES

2.0   This section establishes a framework of lexeme formation possibilities, ranging from simple lexemes through to phrasal translations. Most classifications (e.g. Lyons 1977:512-50) use a threefold division of simple lexeme, derived or complex lexeme, and compound lexeme. However within these categories other distinctions have been made, and the framework established below will draw on these further distinctions. The framework is presented in Table 1.

Table 1. Framework of Lexeme Classification

	simple lexeme	derived lexeme	complex lexeme	compound lexeme			phrasal lexeme
				phon	syn	sem	
new form							
new meaning							

2.1 Lexemes

The term lexeme has been used by Matthews (1974:11f) and Lyons (1977:18f) to denote an underlying word-form. As an example find may be considered the underlying word-form or lexeme for find, finds, found and finding. As such the term lexeme has the important role of denoting 'the fundamental . . . unit of the lexicon of the language' (Matthews 1974:22), in the same way that phoneme and morpheme are used to denote the fundamental units of phonology and morphology.

In this sense it can be used to obviate the ambiguity of the term word. Both Matthews (1974) and Lyons (1977) use the term compound lexeme to denote a form that consists of more than one free root, but nevertheless needs to be considered as a unit from a lexical

point of view. Matthews uses the example of _ice-cream_ (1974:33) and Lyons the examples of _country_ _house_ and _washing_ _machine_ (1977:535,542).

2.2 Simple Lexemes

A simple lexeme is defined as one which only has one morpheme, that is to say it is monomorphemic, apart from inflections. There are a number of ways that a simple lexeme can become a new lexeme in a language.

> (i)   borrowing
> (ii)   extension of meaning
> (iii)   metaphor
> (iv)   creation of new lexeme

2.2.1 Borrowing

This is a very common way for a language to increase its stock of morphemes. It is widely used by Third World languages undergoing modernisation. It is usually readily identifiable, but can be complicated by a change in meaning compared to the source language, and by the existence of loan translations whereby a lexical construction or metaphor is borrowed but the forms are replaced by those of the borrowing language.

2.2.2 Extension of Meaning

Words and lexemes are constantly being reapplied to associated concepts in a language. Andrzejewski (1984:79-81) gives a number of examples of the reapplication of lexemes in this way in the Somali language - _magac_ for _noun_ (original meaning _name_); _taban_ for _negative_ (original meaning _missing_); and _neef_ for _gas_ (original meaning _breath_). It is often difficult to decide when the extension of meaning has gone far enough for a new meaning to be present. For instance the terms _jump_, _loop_, _stack_ and _heap_ are used as computer programming terms but it is not clear whether these should be

considered new lexemes. This is the traditional problem of polysemy. However at some point a distinctly new meaning can be formed - for instance it can be claimed that the word-form program has acquired a distinctly new meaning in the field of computers.

2.2.3 Metaphor

It is not always easy to distinguish metaphor from extension of meaning; however there are clearly cases where a word-form acquires a new meaning by means of metaphor. (Metaphor, as a general term, will be taken to include metonymy and synecdoche (cf Lyons 1977:548)). In computer terminology, a vanilla terminal configuration is a computer (software) configuration which is suitable for any make of terminal, without taking advantage of the particular features of any one terminal. The word-form vanilla therefore has acquired a new meaning by means of metaphor.

(Note that metaphor is not restricted to simple lexemes. In Somali dabadhalif (tail + sheep) has the metaphorical meaning of political stooge from its original meaning of tip of sheep's tail (Andrzejewski 1984:78). Similarly in English the compound live wire has acquired a metaphorical meaning.)

2.2.4 Creation of New Lexemes

This appears to be less common than the three categories previously mentioned. It may or may not involve the creation of a new word-form. For instance radar is a new lexeme and a new word-form. Bit, as a computer term, is a new lexeme but not a new word-form. Byte is a new lexeme and a new graphical word-form but not a new phonological word-form. New word-forms can be created in a number of ways - onomatopoeia, blending and acronyms being the most common.

2.3 New Form and New Meaning

As a simplified but principled categorisation we can divide new lexemes into those that are based on an existing lexeme and those that are not based on an existing lexeme in the language. Metaphor and extension of meaning are innovations based on existing lexemes and therefore will be considered new meanings. Borrowing and creation will be considered new forms. Clearly <u>bit</u> and <u>byte</u> quoted above could pose some problems to this categorisation. In this thesis, however, they would be considered new forms since their creation was not based on an existing form. Note that it is possible for lexemes other than simple lexemes to take on new meanings, as noted for metaphor in section 2.2.3.

2.4 Derived Lexemes

The approach adopted by both Matthews (1974) and Lyons (1977) is to divide non-simple lexemes into two classes: derived or complex lexemes, and compound lexemes. In this thesis a threefold division will be adopted: derived lexemes; complex lexemes; compound lexemes.

If Matthews' and Lyons' works are examined it becomes apparent that they make no distinction between affixes which basically have the effect of changing the class of the lexeme, e.g. -<u>ance</u> (<u>inherit</u>, <u>inheritance</u>), -<u>ness</u> (<u>coarse</u>, <u>coarseness</u>); and those that add a distinct semantic component, such as <u>pre</u>- (<u>natal</u>, <u>prenatal</u>), <u>super</u>- (<u>cooled</u>, <u>supercooled</u>) or <u>micro</u>- (<u>computer</u>, <u>microcomputer</u>).

In this thesis a distinction will be adopted such that class-changing affixes are considered to form derived lexemes; whilst affixes which have a distinct semantic component of their own will be considered to form complex lexemes. It is recognised that there may well be borderline cases (-<u>hood</u>, e.g. <u>sisterhood</u>, may be an example) where both aspects are involved; nevertheless the distinction has proved itself important in examining the hypothesis.

2.5 Complex Lexemes

As explained in the previous section, a complex lexeme is defined as one which contains an affix having a distinct semantic component of its own. Complex lexemes are therefore intermediate between derived lexemes and compound lexemes, and no doubt there are borderline cases which will be difficult to assign.

Table 2. Classification of Affixes in Webster's Third New International Dictionary and the Concise Oxford Dictionary

	Webster's	O.E.D.		Webster's	O.E.D.
un–	A	A	hyper–	A	A
non–	A	A	inter–	A	A
mis–	A	A	pan–	CF	CF
de–	A	A	ultra–	A	A
dis–	A	A	fore–	CF	A
			trans–	A	A
re–	A	A	anti–	A	A
co–	A	A	counter–	A	A
			contra–	A	A
uni–	A	CF			
bi–	A	A	micro–	CF	CF
di–	CF	CF	macro–	CF	CF
tri–	CF	CF	mini–	–	CF
multi–	CF	A	maxi–	–	CF
poly–	CF	A			
			eco–	CF	A
arch–	A	A	hydro–	CF	CF
vice–	A	A	techno–	CF	–
			bio–	CF	CF
pre–	A	A	geo–	CF	CF
post–	A	A	thermo–	CF	CF
sub–	A	A			
super–	A	A	auto–	A	A
under–	A+	Adj/Adv/Prep	self–	CF	A

A ≈ prefix	CF ≈ combining form	– ≈ not listed	
Adj ≈ adjective	Adv ≈ adverb	Prep ≈ preposition	
+ ≈ other designations also given			

The classification of affixes which form complex lexemes is generally considered problematic. As an example consider Table 2 which gives the classification of such affixes in English in two competent dictionaries, Webster's Third New International Dictionary (1981) and the Concise Oxford Dictionary (1982). (The fact that all these affixes are prefixes is considered arbitrary in this thesis, resulting from the nature of the Indo-European language family.

Compare Saussure (1959:188) where he points out, presumably referring to Indo-European languages, 'a prefix is more sharply delimited, for it is easier to separate from the word as a whole'.)

It is clear that an additional classification is used, that of combining form. The basis for this classification is not obvious; it is not clearly explained by either dictionary. Part of the problem of classifying these affixes is evidenced by the fact that the dictionaries do not agree between themselves as to which morphemes to classify as affixes and which to classify as combining forms.

Part of the difficulty is etymological, and Adams (1973:128-33) classifies some lexemes, using these affixes, as neo-classical compounds. Bauer (1983:213-6) also considers some of these types of affix, and uses the terms Final Combining Form and Initial Combining Form to describe them. However, in the interests of a cross-linguistic classification, in this thesis they will all be considered affixes capable of forming complex lexemes.

Nevertheless a distinction can be observed amongst these affixes. It is possible to sub-classify the affixes of Table 2 (cf Adams 1973:161). There are 'logical' affixes (called 'negative', 'privative' or 'reversative' by Adams) such as un-, non-, mis-, de-, dis-; there are 'numerical' affixes - uni-, bi-, di-, tri-, multi-, poly-; there are 'relational' affixes - pre-, post-, sub-, trans-, etc; there are 'intensifying' affixes - super-, hyper-; there are 'adjectival' affixes, normally a simple adjectival quality such as size - micro-, macro- etc; there are prefixes of 'eminence' - arch-, vice-; there are 'substantive' affixes - eco-, hydro- etc; finally there are 'reflexive' affixes - auto-, self-, which are in a sense substantive, but seem worthwhile classifying separately.

The importance of recognising distinctions amongst the affixes in this way, for lexical development in minority languages, can be seen by considering the folowing data from Welsh. In that many Third

Table 3. Percentage Use of Indigenous Morphemes in the Translation of the English Affixes of Table 2 into Welsh

logical	94%
intensifying	85%
numerical	76%
relational	75%
eminence[1]	
reflexive	55%
adjectival	31%
substantive	26%

[1]One of the Welsh affixes of eminence is <u>arch-</u> and it consequently was not possible to tell when this was a loan and when it was not.

World languages attempt to translate English or French terms, the fact that some of these types of affix may give greater difficulty in translation would seem to be a point worth noting.

2.6 Compound Lexemes

There is in general agreement as to the existence of compound lexemes (e.g. Matthews 1974:33-4,188-94, Lyons 1977:534-50, Adams 1973, Bauer 1983:201-13). They are lexemes that are made up of more than one root. However the ways in which they can be recognized are less easy to state (see in particular the discussion by Matthews 1974:188-94).

A threefold distinction will be adopted here. First of all there are those that can be recognized phonologically due to the fact that they have a phonological pattern characteristic of words in that

language (e.g. blackbird, desktop, push chair; or in Welsh oergell (refrigerator ≈ cold + cell), ansoddair (adjective ≈ quality + word)).

Secondly there are those that can be recognized by using syntactic criteria. For example, in Welsh the normal phrase order is head + modifier. However in compound words (geiriau cyfansawdd) the normal order has been modifier + head. These are termed by Watkins (1961:115) 'proper compound words' ('geiriau cyfansawdd rhywiog'). When this is not the case he terms them 'improper compound words' ('geiriau cyfansawdd afrywiog'). Where the order is different from the normal syntactic order this would appear to be evidence of lexeme formation.

Dumont and Mbodj (1983:452) find a similar situation in Wolof. The normal phrase order, they state, is head + connector + modifier. This, they state, is rigorously observed. Even phrases of foreign origin conform to this pattern. They give the following examples: jang bu raft (une jolie jeune fille ≈ jang jeune fille + bu + raft jolie), lam bu rafet (un joli bracelet ≈ lam bracelet + bu ≈ rafet jolie), nit ku nuul (un homme noir ≈ nit homme + ku + nuul noir), simis bu soon (chemise jaune ≈ simis chemise + bu + soon jaune). However when it comes to compound lexemes (1983:458) they divide them into 'syntactic compounds' ('les composes syntactiques') and 'semi-syntactique compounds' ('les composes semi-syntactiques'). In the case of semi-syntactic compounds the connector is, or can be, omitted. It would appear that the absence of a connector is again evidence of lexeme formation.

Thirdly there are those that can be recognized by semantic criteria. This is far less easy to state. The basic criterion will be recognised here that they represent a single concept, compare Matthews 1974:33 - 'Ice-cream is a distinct unit from the semantic or lexicographical viewpoint'.

It is not possible to recognize compound lexemes simply on the basis of missing semantic information (cf Lyons 1977:542-3) since many noun phrases depend for their meaning on extralinguistic information. For instance, the difference in the semantic relationship between the head and modifier in <u>plastic bottle</u> and <u>milk bottle</u> is based entirely on extralinguistic knowledge[1]. However institutionalisation of meaning can also be evidence of unity of meaning (e.g. <u>washing machine</u>).

The approach will be adopted here that if a compound lexeme can be recognised by means of phonological criteria then it will be considered a phonological compound, even though it may also be recognised by means of syntactic criteria and on semantic grounds. Similarly, if a compound can be recognised on syntactic grounds (but not phonological ones) then it will be considered a syntactic compound, even though it could also be recognised by means of semantic criteria. Only if it can not be recognised by either phonological or syntactic criteria will it be considered a semantic compound.

The boundary between semantic compounds and phrasal translations (see below) will generally not be easy to recognise.

2.7 Phrasal Translations

This category is introduced since it is needed in the analysis. The analysis of the two sets of Swahili data has been restricted to lexemes or phrases that correspond to a single English

---

[1] 'En la interpretacion de los compuestos - prescindiendo de las posibles fijaciones por la norma -, deben distinguirse precisamente tres fases racionalmente sucesivas y que corresponden a la aportation del sistema de la lengua, al conocimento general y al conocimento 'actual' de las cosas y de los estados de cosas extralinguisticas' (in the interpretation of compounds - disregarding the possibility of these being fixed by the norm - precisely three rationally successive stages must be distinguished, which correspond to the contribution of the language system, to general knowledge and to the contemporary knowledge of extralinguistic things and the state of extralinguistic things) (Coseriu 1981:203-4).

(graphical) word, in order to be sure of comparing like data with like. Consequently this category allows for a single word in English to be translated by a phrase in Swahili which is not considered to be a semantic compound. See the Appendix for examples.

2.8 Framework of Lexeme Formation Possibilities and the Hypothesis

The framework of Table 1 will be used in section 3 of this thesis to test the tendencies of planned and spontaneous lexeme formation. The hypothesis states that planned lexeme formation will make greater use than spontaneous lexical innovation of the central categories of this framework, particularly complex lexemes, as marked in Table 4.

Table 4. Hypothesised Tendencies of Planned Lexical Innovation

| | simple lexeme | derived lexeme | complex lexeme | compound lexeme | | | phrasal lexeme |
				phon	syn	sem	
new form		(X)	X	(X)	(X)	(X)	
new meaning							

3. TENDENCIES IN LEXEME FORMATION

3.0 This section will look for evidence in a number of languages for the hypothesis stated in section 1.5, using in general the framework developed in section 2. As explained in section 1.7, the two languages that will be looked at in some depth are Welsh and Swahili. In section 3.3, the evidence from articles on lexical development in a few other languages will be considered.

3.1 Welsh

The Welsh language belongs to the Indo-European family of languages. As a result it shares a number of features with other Indo-European languages. One of these features is a fairly extensive range of affixes which are used in lexeme formation.

3.1.1 The Data

The reference source for a description of the Welsh language that has been used in preparing this thesis is Watkins (1961). This reference manual contains one chapter devoted to prefixes ('Rhagddodiaid') and another devoted to compound lexemes ('Geiriau Cyfansawdd'). (As stated in section 2, the fact that most affixes that are used to form complex lexemes in Indo-European languages are prefixes is considered – as far as this thesis is concerned – to be incidental).

The chapter on prefixes contains a number of statements which are of relevance to this thesis. In the first place in this chapter, Watkins states a problem in classifying prefixes. He notes that the list of Welsh prefixes found in a standard dictionary, such as the Geiriadur Newydd, contains forms that can also be used independently.

In the second place he considers the productivity of affixes. In this connection he considers three types of affixes.

21

His first list is of two affixes that are no longer productive – a-/add- and cyfr-. Throughout this chapter he refers continually to a study performed by Fynes-Clinton (1913), The Welsh Vocabulary of the Bangor District. This study analysed the spoken language in the area of Bangor. With respect to a-/add- and cyfr- Watkins notes that 'in the spoken language very few of the above words in a-/add- and cyfr- have remained alive' (1961:109)[1].

His second list consists of affixes which he states 'have remained productive, at least in literature and lexicography' (1961:109)[2], or 'show striking activity in literature, especially to translate complex or compound technical words from English' (1961:111)[3]. This list consists of: ad- (again, re-), arch- (arch-), ar- (on, super-), all- (other, outside, ex-), cyd- (with), gwrth- (against), dad- (un-) and gor- (too much, over-). However with respect to ad- and ar- he states 'not one of the recent compositions in ad- and ar- have passed into the spoken language' (1961:110)[4].

His third list consists of morphemes which can be used independently but can also be used as affixes: ail (second), cyn (before), rhag (before), is (below), uwch (above), ol (after), croes (cross, trans-), hunan (own, self), un (one), cam (false, mis-), aml (many) and blaen (forward).

With respect to all of these prefixes he states: 'in the spoken language very few prefixing morphemes have remained productive'

[1] 'yn yr iaith lafar ychydig iawn o'r geiriau uchod yn a/add a cyfr sydd wedi aros yn fyw'.

[2] 'sydd wedi para yn gynhyrchiol, o leiaf yn llenyddol ac yn eiriadurol'.

[3] 'yn yr iaith lenyddol maent yn dangos gweithgarwch trawiadol, yn enwedig i gyfieithu geiriau technegol cymhleth neu gyfansawdd o'r Saesneg'.

[4] 'nid oes un o'r cyfansoddiadau diweddar yn ad ac ar wedi treiddio i'r iaith lafar'.

(1961:112)[5]. The exceptions he notes are: di- (un-), dad- (un-), cyd-
(with, co-), an- (un-, non-) and possibly ail (second).

These quotations from Watkins are in themselves important
supporting evidence for the hypothesis, at least as far as the Welsh
language is concerned, since they show that affixes used to form
complex lexemes are in use in planned lexical development, but show
little or no use in spontaneous lexical development. However it seems
worthwhile to consider some of these morphemes further and attempt to
look first hand at their actual productivity in a modern planned word
list.

As a method of doing this, three dictionaries were chosen for
comparison: the Collins-Spurell Welsh Dictionary, a 12,000 word pocket
dictionary published in 1960; Y Geiriadur Cymraeg Cyfoes (The
Dictionary of Modern Welsh), a 19,500 word dictionary published in
1981, and claiming to reflect many of the recent lexical developments
in Welsh; and Geiriadur Termau (Dictionary of Terms) published in
1973, a 16,500 word dictionary, which specifically gathers together
various planned word lists produced by organisations or journals, such
as the Faculty of Education of the University College of Wales, the
Board of Celtic Studies, Y Gwyddonydd (The Scientist), and the Welsh
Office. Of these Geiriadur Termau should clearly reflect the
morphological composition of planned lexemes to a greater extent.

The affixes of the second and third lists produced by Watkins
were looked up in these three dictionaries and their occurrence listed
in Table 5, and also adjusted proportionately to a dictionary of
15,000 entries in Table 6. There is no reason to believe that
Watkins' choice is biased, and therefore his list of affixes can be
taken as an arbitrary one.

5 'ar lafar ychydig iawn o forffemau rhagddodiadol sy'n dal yn
gynhyrchiol' (1961:112).

Table 5. No. of Entries per Prefix in Three Welsh Dictionaries

	CS	GCC	GT		CS	GCC	GT
ad-[1] (again)				rhag (forward)	46	66	54
arch-[2] (arch-)				is (below)	12	40	72
ar-[1] (on, super-				uwch (above)	2	20	29
all- (ex-)	7	47	56	ol (back)	3	19	39
cyd- (with,co-)	34	76	116	croes (cross, trans-)	9	21	26
gwrth-(against)	36	86	89	hunan (self)	6	14	31
dad- (un-,non-)	20	43	45	un (one)	22	49	25
gor-[1] (over, super-)			106	cam (false)	40	46	25
ail (second)	6	12	29	aml (many)	9	18	17
cyn[1] (with, co-)				blaen (forward)	10	25	29

[1] ad-, ar-, and cyn- are prefixes that have been in use for many centuries and the roots used with them often proved difficult to identify, resulting in the words as a whole not being able to be analysed. Consequently, reliable numerical data was not able to be extracted from any of the dictionaries. Gor- followed the same pattern in two of the dictionaries

[2] The Welsh prefix arch- has the same meaning as the English prefix arch-. Consequently it was not possible to tell when an entry was a genuine composition or a loan translation.

Table 6. No. of Entries per Prefix Adjusted to 15,000 Dictionary
Entries

	CS	GCC	GT		CS	GCC	GT
ad–				rhag	56	51	49
arch–				is	15	31	65
ar–				uwch	2	15	26
all–	9	36	51	ol	4	15	35
cyd–	42	59	105	croes	11	16	24
gwrth–	48	66	81	hunan	7	11	28
dad–	24	33	41	un	27	38	25
gor–			96	cam	50	35	23
ail	7	9	26	aml	11	14	15
cyn				blaen	12	19	26
				TOTALS	325	448	620

(excluding gor–)

3.1.2 Relationship of Data to Hypothesis

It is clear that the majority of these affixes are productive in planned lexical development. There seem to be two ways that this data specifically supports the hypothesis.

In the first place, Watkins states that very few of these affixes are productive in the spoken language. He bases this statement in part on the word–lists produced by Fynes–Clinton (1913), which reflect the spoken language of an area of Wales. Watkins' statements on productivity in the spoken language can be held to correspond to spontaneous lexical creativity. However in the data of Tables 5 and 6, if the frequency of use in the Geiriadur Termau reflects planned lexical creation, then it is clear that most of these affixes are still productive in planned lexical creation.

Secondly, the data from the three dictionaries having been compared, a tendency emerges. Many of the morphemes show a regression whereby the Geiriadur Termau shows greater use than the Geiriadur Cymraeg Cyfoes, which in turn shows greater use than the Collins-Spurrell dictionary. If, because of the nature of the dictionaries, the Geiriadur Termau is most likely to reflect planned lexical development and the Collins-Spurrell dictionary least likely to reflect planned lexical development then this regression is, by itself, evidence of the greater use of affixes in planned lexical development.

3.2 Swahili

Swahili is a Bantu language. As such it exhibits very few affixes used to form complex lexemes. As a result of these characteristics it could be considered a very inappropriate language to investigate as far as the hypothesis is concerned. However, if the hypothesis claims to be true of all languages, then a language such as Swahili is of value in seeing what tendencies of this nature do exist in it, and therefore to what extent the hypothesis can be claimed for such a language. In addition, the data for making a comparison between planned and spontaneous development was available in Swahili.

3.2.1 The Data

As already explained there is considerable difficulty in finding clear information on spontaneous lexical development. For this reason an article published in 1982 by the Polish scholar, Rajmund Ohly, on lexical use at a factory in Dar Es Salaam (The Friendship Textile Mill) was very valuable, and became the source data for spontaneous language use in Swahili. It will be referred to in future as LRFTM (Lexicographical Research at the Friendship Textile Mill).

Since the article by Ohly was concerned with technical language, by way of comparison a technological word list, produced by BAKITA (Baraza la Kiswahili la Taifa), published in <u>Tafsiri</u> <u>Sanifu</u> (1980:67-73), was taken as the comparison data for planned lexical development.

In order to make a like for like comparison, only those lexemes were used which corresponded to single graphical words in English. This resulted in 55 words from the LRFTM list being used. Assistance in analysing the words was also obtained from the Swahili instructor in the Linguistics Department at MSU. Since some of the words were specialised, only those words he was familiar with were used. This resulted in 144 words being used from the BAKITA list. The words, together with their classification and translation, are listed in the Appendix.

Table 7. Distribution of LRFTM Word List to Framework of Lexeme Classification (Spontaneous Development)

| | simple lexeme | derived lexeme | complex lexeme | compound lexeme | | | phrases |
				phon	syn	sem	
new form	20 (36%)	3 (5%)					22
new meaning	10 (18%)						(40%)

Table 8. Distribution of BAKITA Word List to Framework of Lexeme Classification (Planned Development)

| | simple lexeme | derived lexeme | complex lexeme | compound lexeme | | | phrases |
				phon	syn	sem	
new form	70 (49%)	15 (10%)	6 (4%)	4 (3%)	1 (1%)	4 (3%)	6
new meaning	38 (26%)						(4%)

Using the framework developed in section 2, the analyses of Table 7 and Table 8 were made of the LRFTM and BAKITA data. The percentage figures denote the percentage in each category of the total number of words being distributed in that table.

3.2.2 Relationship of Data to Hypothesis

The following points can be made concerning the data.

1. Complex Lexemes. The only affix that produced complex lexemes in the data is ki- (diminutive). As might be expected, therefore, only a few words had this affix. It is also a grammatical affix in Swahili since it is a class prefix. Only those words for which this affix had a diminutive function were considered to be complex lexemes. However, notably, words with this affix having this function were only discovered in the BAKITA data (4% of sample).

2. Derived Lexemes. Care has to be taken with this category since it is in general very productive in the language and a different percentage might be obtained if the sample listed, say, nouns and verbs, rather than just nouns. Nevertheless very few noun-verb pairs appear in the data. Consequently the fact that the BAKITA list has more words of this category (10% versus 5%) has some significance.

3. Compound Lexemes. Only the BAKITA list shows use of compound lexemes (7%).

4. The main feature of the data which is not reflected in the hypothesis is the very large number of simple lexemes in the BAKITA data (75% versus 54%). This is achieved by a much higher percentage of borrowed words, 49% as against 36% (see appendix), and the presence of metaphorical lexemes in the BAKITA list (8%) as against none in the

LRFTM list. The percentage of extended lexemes was almost the same in each set of data (19% versus 20%).

5. This is balanced out by the very much larger percentage of phrasal translations in the LRFTM list (40% versus 4%).

The data supports the hypothesis in that only the planned lexical data shows use of complex and compound lexemes, and also greater use of derived lexemes. The total percentage difference for these three categories is 16%. A t-test on this difference (which considered the five categories to lie evenly along a scale, symmetrically distributed around the central category of complex lexeme) showed it to be significant at the 1% level (t = 2.676).

One striking difference is the larger use of phrasal translations in the LRFTM list. This may be exaggerated by the fact that the LRTFM list included data from a section on The Language of Instruction, which may well not reflect spontaneous use. Table 9 shows the LRFTM data with this data excluded. The basic features of the data are not changed. The same t-test mentioned above gave a significance level of 3% (t = 2.227).

Table 9. Distribution of LRFTM Word List to Framework of Lexeme Classification Excluding Language of Instruction Phrases (Spontaneous Development)

| | simple lexeme | derived lexeme | complex lexeme | compound lexeme | | | phrases |
				phon	syn	sem	
new form	20 (45%)	3 (7%)					11
new meaning	10 (23%)						(25%)

3.3 Other Languages

3.3.1 Filipino

The article by Del Rosario (1968) in Vol 6 of Asian Studies has already been referred to in Section 1. However it also contains information on the attitudes of this scholar to preferred methods of lexical creation. Del Rosario makes the following general statement on preferred lexical creation, particularly for education.

'. . . a national language must have two qualities: modernity, to cope with the progress of science and technology, and lexical consistency, to be understandable to children just entering the school system' (1968:3).

Del Rosario's advocating of 'lexical consistency' involves the use of combinations of indigenous morphemes as far as possible. The following are examples he gives in the Filipino language (1968:12).

numeral	pamilang	pang- (instrument prefix) + bilang (number)
integer	buumbilang	buo- (whole) + bilang (number)
fraction	bahagimbilang	bahagi- (part) + bilang (number)
numerator	panakda	pang- (instrument prefix) + takda (schedule)
denominator	pamahagi	pang- (instrument prefix) + bahagi (part)

Del Rosario's approach to planned lexical creation clearly involves using complex and compound lexemes wherever possible and therefore is consistent with and supports with the hypothesis.

3.3.2 Bemba

Epstein (1959) discusses the use of Bemba in the Copperbelt of Northern Rhodesia (Zambia). He is particularly concerned with language innovation and its relationship to culture. However in the

process he provides interesting examples of spontaneous lexical development.

Most of the words and lexemes he considers are loans from English. This is natural since he is describing the contact of the Bemba speaking people with urban and Western culture. However in this list of words there is much evidence of metaphorical creativity rather than use of composed lexemes. The following are a representative set of examples. (The prefixes in the words are all class prefixes and therefore do not form complex lexemes.)

muchampioni	young girl who dresses in a modern fashion
kapenta	minnow – a simple-to-cook fish such as would be prepared by a married man's girl friend (cf painted lady)
spare wheel	mistress
diesel	a thick beer
malinso	informer – from the Bemba word for lizard – in African folklore the lizard was a young crocodile who failed to return to the water
bakapepala	casual workers – workers that had chits (paper) for food rations

The complete absence of complex lexemes in his article constitutes supporting evidence for the hypothesis.

3.3.3 Marshad on Swahili

Marshad's article on Swahili in Language and Development (1980) is a good example of the preference of a language scholar for composite lexemes. The following statements from his article on the task of modernising Swahili for scientific use can be quoted.

'. . . I find the Tagalog approach [analytical method rather than affix translating method] preferable in the light of the Swahili morphological structure' (1980:11).

'Consider, for example, the term anti-/counter-clockwise, rendered in Tafsiri Sanifu (No. 2/1976:27) as kinyume saa (opposite of clock). Instead of using this translation method, which often results in cumbersome phrases, it would be far better to employ a derivational approach involving the negative marker -/to/- (often used in certain tenses) and the adverbial marker -/ki/-, as follows: (a) saa "clock/time;" (b) ki-saa "in the manner of clock/clockwise"; and (c) ki-to-saa "counter-/anti-clockwise"' (1980:11).

'This approach can be extended to the derivation of compounds with interesting results' (1980:11).

A clear preference emerges for derived, complex or compound lexemes which are as concise as possible; although, he states, BAKITA have not always followed his preferences. The word 'analytical' is interesting in these quotations, and the possibility that this may characterise a planned approach to lexical development is considered further in section 4.

4. DISCUSSION

This study has presented a variety of data concerning lexical development, focused around a hypothesis which proposes that planned and spontaneous lexical innovation have different tendencies in the types of lexemes that they create.

The most direct data that has been presented is from Swahili. This data suffers from the drawback that very few complex lexemes, of the type defined in the thesis, exist in Swahili. The only complex lexemes in the data were those using the diminutive prefix ki-. However the hypothesis is supported in that only in the planned data were complex and compound lexemes found, and more derived lexemes were found in the planned data than in the spontaneous data. The difference between the two sets of data was found to be statistically significant at below the 5% level.

The other comprehensive data presented was from Welsh. Although no data was directly available on spontaneous innovation, two indirect comparisons with spontaneous innovation were possible. The data that was presented also strongly supported the hypothesis.

The articles by language scholars involved in lexical planning, discussed in section 3.3, also supported the hypothesis. Statements by the authors showed a clear preference for the use of derived, complex or compound lexemes (although in one case these were not always followed by the national committee involved in lexical creation).

The data presented suggests that the hypothesis could be restated in a more general fashion. It could be stated that planned lexical creation favours more analytical (or transparent) and compact lexemes than spontaneous lexical creation. The analytical aspect of

33

this is reflected directly by Marshad (see section 3.3.3) when he refers to 'an analytical method' of lexical creation. If this is true, then there are two opposing tendencies that lexical planners attempt to balance out. The analytical tendency pushes lexical planners towards multi-morphemic lexemes; the tendency towards compactness encourages them to reduce the number of morphemes to as few as possible. The result is a greater preference for composite lexemes, and in particular complex lexemes, then spontaneous lexical innovation. The same counter-balancing tendencies can be seen in many of the prescriptive statements on lexical planning by Tauli (1968).

An objection may be raised that the concepts that planned lexical development generally deals with, such as science terms, are by their nature more analytical or structural than those dealt with by spontaneous lexical development. This appears to be a realistic objection and may well account for some of the data. However it does not account for all of the data, since in the case of most of the lexemes quoted it is possible to think of more compact or more analytical lexemes. However these two aspects may go hand in hand - a more concise analytical approach to concepts which can be handled in this fashion. By contrast much spontaneous lexical creation attempts, not to be analytical, but to be descriptive or, as the data on Bemba shows (section 3.3.2), to be attention catching and vivid.

If the hypothesis is correct, it is also interesting to consider whether there are any psychological or sociolinguistic conditions which might contribute to the lexical preferences mentioned in the hypothesis. Two broad possibilities may be mentioned here. First, in that many lexemes generated by lexical planners are modelled on English or French originals, it is possible that lexical planners may be reflecting the morphological structure of

corresponding English or French terms (which in turn may well reflect the original morphological patterning of Greek and Latin). Second, in a more tentative tone, there may be a psychological preference for composite lexemes, perhaps as a result of an impressing of the values of conciseness and analyticity during education, and emphasis on 'le mot juste'.

What are the implications of the hypothesis? If it is true that planned lexical creation favours composite lexemes, and in particular complex lexemes with affixes, then it can be suggested that the hypothesis implies that the perpetuation, and possibly now the generation, of affixes may be in part conscious or deliberate. The Welsh data is particularly clear in this respect, where existing and new affixes are now being used in planned lexical development which are apparently not used at all in spontaneous lexical creation.

The hypothesis can also be used as a basis for suggesting some policies for lexical planners. It can be suggested that lexical planners should be careful of a bias towards derived, complex or compound lexemes. The reason for this becomes clear when acceptability is considered. As mentioned above, much spontaneous lexical creation is attention catching and vivid, and although some of it is ephemeral, the lexemes produced by it may be more acceptable to the generality of language users than analytical lexemes. In a similar fashion it can be suggested that lexical planners should pay closer attention to lexemes created spontaneously, and consider the possibility of these being a source of lexical innovation that could be officially sanctioned. This appears to have been Ohly's motivation in undertaking his study of language use at the Friendship Textile Mill.

5. SUMMARY

This thesis has considered the phenomenon of lexical development or creativity in a number of languages. The two languages that have been investigated in particular are Welsh and Swahili. It attempted to find evidence for a hypothesis which contrasts spontaneous and planned lexical development. The hypothesis states that planned lexical development would use more composite lexemes than spontaneous development (see section 1.5).

A framework of lexeme formation possibilities was derived in section 2 and used in classifying and analysing the data. The framework made a distinction between simple, derived, complex and compound lexemes and phrasal translations, and also between existing lexeme-forms with new meaning and new lexeme-forms. The distinction that it made between derived lexemes, that use a class-changing affix, and complex lexemes, that use other types of affix, was novel. Derived, complex and compound lexemes were all considered examples of composite lexemes.

The only direct evidence that could be obtained was from the Swahili language. Very few complex lexemes, of the type defined in the thesis, were found in Swahili. Nevertheless only in the planned data were complex and compound lexemes found, and the planned lexical data also showed greater use of derived lexemes. The results were found to be statistically significant. The Welsh data, using both indirect and secondhand evidence, also gave strong evidence for the hypothesis. Attitudes to lexical development for other languages were considered from articles, and found to be in agreement with the hypothesis.

It was suggested in the Discussion that lexical planners attempt to balance out two opposing tendencies – a tendency to create

36

transparent or analytical lexemes, and a tendency to be as compact as possible. This, therefore, results in a greater propensity to use complex lexemes, which balances out these two tendencies, than is the case with spontaneous lexical innovation. The possibility that planned lexical development may also tend to be concerned with more analytical concepts, such as in science, which are then capable of being represented by morphologically more analytical forms was considered, and accepted in part, but not felt to account for all of the data.

The thesis opens up an interesting sociolinguistic perspective on language change – that language change, particularly lexical change, is not entirely outside of human control or consciousness. In particular it suggests that the generation and perpetuation of affixes may, at least nowadays, be affected by deliberate language use. This was particularly clear from the Welsh data.

The hypothesis was also be used as a basis for suggesting that lexical planners should be careful of a bias towards the creation of composite lexemes, and should perhaps consider that other types of lexical innovation may often be more acceptable to society at large.

APPENDIX

LIST OF SWAHILI WORDS USED IN SECTION 3

LRFTM LIST

Extended Lexemes

belt	mkanda	needle	sindano
strapping	koa	knife	kisu
blade	kisu	hook	kisu
pulley	gurudumu	bobbin	gurudumu
cone	gurudumu	tappets	magurudumu

Borrowed Lexemes

pump	pampu	beam	bimu
copies	makopi	grease	girisi
hank	hanki	roller	rola
screen	skrini, skriini, sikirini		
warp	wapu	chromium	kromiamu
cirasol	siraso	enzyme	enziamu
fluorite	floritu	lap	lapu
machine	mashine, mashinu	rail	reli
sliver	slaiva	sodium	sodiumu
bobbin	bobini	cone	koni
yarn	yani		

Derived Lexemes

spinning	usokataji	scrappers	vikandamizio
bleach	ng'arisha		

Phrasal Translations[1]

dye	tia rangi	pick	pigia ngalawa
starch	tia wanga	bundling	kufunga mizigo
drafting	kukaza uzi	picking	kupigia ngalawa
sizing	kuwekea wanga	grading	kutia greidi
dying	utiaji rangi	electrolyte	tudi la chuma
loom	mashine ya kufumia		

[1] See footnote 2.

```
calendaring    mashine ya kupigia pasi
steaming       mtambo wa kukaza rungi
pre-shrinker   mashine ya kunyooshea kitambaa
platter        mashine ya kupangia nguo
pits           mashine ya kulowekea nguo
pitter         mashine ya kusambazia nguo
jiggers        mashine za kutilia rangi
repairing      kutengeneza sehemu mbaya
electroplating  kupaka chuma cha kromiamu kwenye chuma kingine kwa
                   kutumia umeme
radiators      mabomba yakuchuayo mvuke wa moto
singeing       kuchoma vinyuzi vilivyotikeza kwenye nguo
```

BAKITA WORD LIST

Extended Lexemes

accelerate	chapuza	adhesion	mng'ang'anio
allowance	mwanya	belt	mkanda
circulate	eneza	clamp	bana
draughtsman	mrasimu	elastic	a kunyumbaka
elasticity	mnyumbuko	frequency	marudio
friction	msuguano	gate	lango
gravity	uvutano	ignition	mwasho
instant	ghafla	lime	chokaa
loop	kitanzi	mould	finyanga
pair	jozi	parallelogram	msambamba
penetrate	penya	plan	ramani
polish	ng'arisha	property	tabia
project	chomoza	resistance	ukinzani
tension	mvuto		

Borrowed Lexemes

acetylene	asetilini	alloy	aloi
aluminium	aluminiamu	ampere	ampea
apron	aproni	arc	aki
atom	atomi	barometer	barometa
battery	betri	blue	bluu
body	bodi	brake	breki
brass	brasi	brush	brashi
caliper	kalipa	cam	kem
carbon	kaboni	cartridge	katriji
cassettes	kaseti	chimney	chemni
circuit	saketi	clutch	klachi
compass	dira (Arabic)	conserve	hifadhi (Arabic)
cylinder	silinda	dot	doti
dynamo	dainamo	electron	elektroni

element	elementi	engineer	mhandisi (Arabic)
exhaust	ekzosi	fuse	fyuzi
gear	gia	gearbox	giaboksi
generator	jenerata	geometry	jometri
glove	glavu	governor	gavana
indicator	indiketa	inertia	inesha
lintel	linta	machine	mashine
matter	maada	metal	metali
molecule	molekyuli	moment	momenti
momentum	momenta	motor	mota
neutron	nyutroni	nitrogen	naitrojeni
nozzle	nozeli	nut	nati
phase	fesi	plasticine	plastisini
plug	plagi	proton	protoni
pulley	roda (Portuguese?)	radiation	rediesheni
screw	skurubu	socket	soketi
switch	swichi	topic	mada (Arabic)
transformer	transfoma	tube	tyuba
tyre	tairi	valve	vali (?)
velosity	velositi	voltage	volteji
volt	volting	description	aridhio (Arabic)

Metaphors

axis	mhimili	conductor	kipitisho
nozzle	pua	relative	a uwiano
sag	legeo	section	mkato
universal	kienezi	viscocity	mnato
volume	mjao	current	mkondo (loan metaphor)
development	mkunjuo		

Derived Lexemes

amplification (process)	mkuzo		
amplification (result)	ukuzaji		
casting	kusubu	cohesion	mshikamano
indicator	kiashiro	interview	hojaji
protrude	tokeza	purify	halisisha
radiation	mnururisho	spool, reel	kibiringo
volume	ukubwa	continuity	mwendelezo
convex	mbonyeo	dampen	fifiza
lubrication	ulainishaji		

Complex Lexemes

lintel	kizumba	liquid	kimiminiko
orifice	kitundu	regulator	kirekebisho
roller	kiviringisho	scriber	kikwaruzo

Compound Lexemes - Phonetic

| horsepower | nguvufarasi | technician | fundisanifu |
| artisan | fundistadi | craftsman | fundistadi |

Compound Lexemes - Syntactic

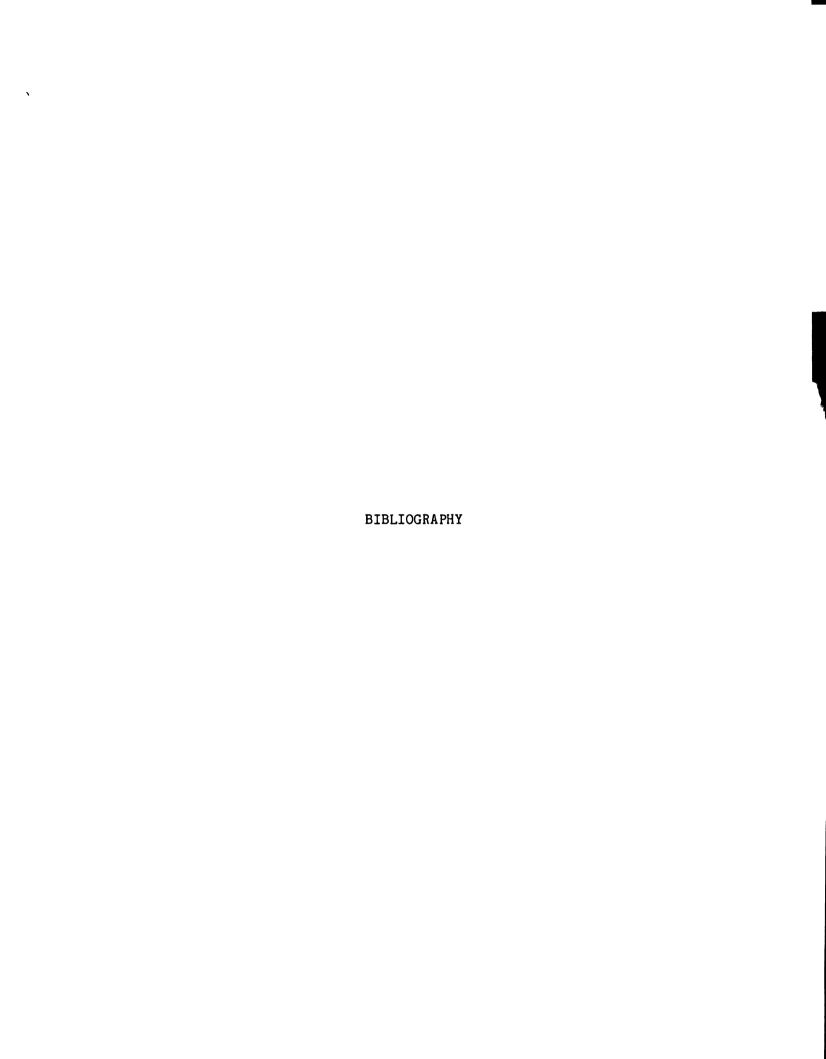atmosphere   angahewa

Compound Lexemes - Semantic

| liquid | maji maji | steel | chuma cha pua |
| brass | shaba nyeupe | plank | ubao wa jukwaa |

Phrasal Translations[2]

coal	makaa ya mawe	coke	makaa zimwe
envelope	mwako wa nje	maximum	kiwangi cha juu
minimum	kiwango cha chini	barometer	kipima hewa

[2] If these phrasal translations represent a conceptual unity then they should in fact be classified as semantic compound lexemes. They were left as phrasal translations either through uncertainty or because they were felt not to realise a conceptual unity.

BIBLIOGRAPHY

BIBLIOGRAPHY

Adams, V. 1973. An introduction to English word-formation. London: Longman.

Andrzejewski, B.W. 1984. Language reform in Somalia and the modernisation of the Somali vocabulary. Language reform, vol I, ed. by J. Fodor and C Hagege, 69-84. Hamburg: Buske Verlag.

Ansre, G. 1977. Four rationalisations for maintaining the European languages in education in Africa. Kiswahili 47.55-61.

Bauer, L. 1983. English word-formation. Cambridge: Cambridge University Press.

Chishimba, M. 1984. Language policy and education in Zambia. International education journal 1.151-180.

Comrie, B. 1981. Language universals and linguistic typology. Oxford: Basil Blackwell.

Coseriu, E. 1962. Teoria del lenguaje y linguistica general. Madrid: Editorial Gredos.

Coseriu, E. 1981. Principios de semantica estructural. Madrid: Editorial Gredos.

Coseriu, E and H. Geckeler. 1981. Trends in structural semantics. Tubingen: Gunter Narr Verlag.

Del Rosario, G. 1968. Modernisation-standardisation plan for the Austronesian-derived national languages of Southeast Asia. Asian studies 6.1-18.

Dik, S.C. 1981. Functional grammar. Dordrecht: Foris Publications.

Downing, P. 1977. On the creation and use of English compound nouns. Language 53.810-842.

Dumont, P and Ch. Mbodj. 1983. Le wolof, langue de developpement. Etudes des procedes d'enrichissement de la langue wolof. Language reform, vol I, ed. by J. Fodor and C. Hagege, 449-462. Hamburg: Buske Verlag.

Epstein, A.L. 1959. Linguistic innovation and culture on the Copperbelt, Northern Rhodesia. Southwestern journal of anthropology 1-16.

Evans, H.M. 1981. Y geiriadur Cymraeg cyfoes/The dictionary of modern Welsh. Llandybie: Hughes.

Fillmore, C.J. 1971. Types of lexical information. Semantics: An interdisciplinary reader in philosophy, linguistics and psychology, ed. by D.D. Steinberg and L.A. Jakobovits, 370-91. Cambridge: Cambridge University Press.

Fillmore, C.J. 1978. On the organisation of semantic information in the lexicon. Parasession on the lexicon, 148-73. Chicago Linguistic Society.

Fynes-Clinton, O.H. 1913. The Welsh vocabulary of the Bangor district. Oxford: Oxford University Press.

Hymes, D. 1972. Review of Lyons (1970). Language 48.416-427.

Ikwue, I. 1984. Effective educational language planning in Nigeria. International education journal 1.39-60.

Lakoff, G. and M. Johnson. 1980. Metaphors we live by. Chicago: University of Chicago Press.

Lewis, H. 1960. Collins-Spurrell Welsh dictionary. London: Collins.

Lyons, J. 1970. Noam Chomsky. New York: Viking Press.

Lyons, J. 1977. Semantics. 2 volumes. Cambridge: Cambridge University Press.

Marshad, H. 1980. On developing a scientific register: the case of Swahili in Kenya. Language and development. 2,11.

Matthews, P.H. 1974. Morphology. Cambridge: Cambridge University Press.

Ohly, R. 1982. Report of lexicographic research at the Friendship Textile Mill. Kiswahili 49.73-86.

Ohly, R. 1977. Pattern of new-coined abstracts terms (nominal forms) in modern Swahili. Swahili studies 47.1-14.

Opalka, H. and F. Passierbsky. 1977. Amalgamation processes in Swahili. Kiswahili 47.70-88.

Saussure, F. 1959. Course in general linguistics. New York: McGraw-Hill.

Sharma, P.G. 1968. Problems of Hindi terminology. Asian studies 6.383–394.

Tauli, V. 1968. Introduction to a theory of language planning. Uppsala: University of Uppsala.

Tumbo, Z. 1982. Towards a systematic terminology development in Kiswahili. Kiswahili 49.87–98.

Ullmann, S. 1962. Semantics. An introduction to the science of meaning. New York: Barnes & Noble.

Ullmann, S. 1957. The principles of semantics. New York: Basil Blackwell & Mott.

UNESCO. 1951. The use of vernacular languages in education. The report of the UNESCO meeting of specialists, 1951. Language problems in developing countries, ed. by J.A. Fishman, C.A. Ferguson and J. Das Gupta, 688–716. New York: Wiley.

Watkins, T.A. 1961. Ieithyddiaeth. Agweddau ar astudio iaith. Cardiff: Welsh University Press.

Weinreich, U. 1963. Lexicology. Current trends in linguistics, vol 1, ed. by T.A. Sebeok, 60–81. The Hague: Mouton.

Williams, J.L. 1973. Geiriadur termau/dictionary of terms. Cardiff: University of Wales Press.

Whiteley, W. 1969. Swahili. The rise of a national language. London: Methuen & Co.

General References

Ashton, E.O. 1959. Swahili grammar. London: Longmans.

Neustupny, J.V. 1968. Some general aspects of 'language' problems and 'language' policy in developing countries. Language problems of developing countries, ed. by J.A. Fishman, C.A. Ferguson and J. Das Gupta. New York: Wiley.

Ray, P.S. 1963. Language standardisation. The Hague: Mouton.

Schadeberg, T.C. 1984. A sketch of Swahili morphology. Dordrecht: Foris Publications.