

An investigation into Lemmatization in Southern Sotho
by **Kelebohile Hilda Makgabutlane** submitted in
fulfilment of the requirements for the degree of Master
of Arts in the subject of African Languages at the
University of South Africa.

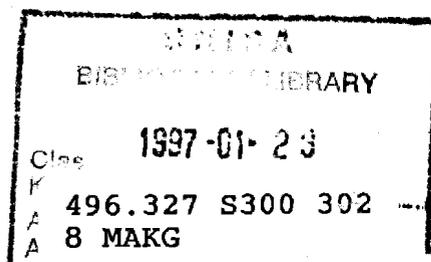
Supervisor:	I.M. Kosch
Joint Supervisor:	Mrs L.J. Kock
Date submitted:	January 1996

ACKNOWLEDGEMENTS

I would sincerely like to thank the following persons for helping me: Mrs Hleziphi Napaai, subject librarian African Languages UNISA; my supervisor and joint-supervisor Prof I.M. Kosch and Mrs L.J. Kock respectively;

Dr Mariëtta Alberts,
Prof D.J. Prinsloo, and
Prof E.B. Van Wyk.

Without their helpful comments, suggestions and various publications that they referred me to, this research work would not have been successful.



1660284

SUMMARY

Lemmatization refers to the process whereby a lexicographer assigns a specific place in a dictionary to a word which he regards as the most basic form amongst other related forms. The fact that in Bantu languages formative elements can be added to one another in an often seemingly interminable series till quite long words are produced, evokes curiosity as far as lemmatization is concerned. Being aware of the productive nature of Southern Sotho it is interesting to observe how lexicographers go about handling the question of morphological complexities they are normally faced with in the process of arranging lexical items.

This study has shown that some difficulties are encountered as far as adhering to the traditional method of alphabetization is concerned. It does not aim at proposing solutions but does point out some considerations which should be borne in mind in the process of lemmatization.

I declare that 'An Investigation into Lemmatization in Southern Sotho' is my own work and that all the sources that I used or quoted have been indicated and acknowledged by means of complete references.

TABLE OF CONTENTS

1.	INTRODUCTION	1
1.1	SCOPE OF INVESTIGATION	1
1.2	THE RELATIONSHIP BETWEEN THE LEXICOGRAPHER AND THE DICTIONARY-USER	2
2.	TYPES OF DICTIONARIES	8
2.1	INTRODUCTION	16
2.2	DICTIONARIES FOR THE SPEAKERS OF THE SOURCE LANGUAGE VERSUS DICTIONARIES FOR THE SPEAKERS OF THE TARGET LANGUAGE	16
2.3	DICTIONARIES OF THE LITERARY LANGUAGE VERSUS DICTIONARIES OF THE SPOKEN LANGUAGE	20
2.4	DICTIONARIES FOR PRODUCTION VERSUS DICTIONARIES FOR COMPREHENSION	20
2.5	HISTORICAL DICTIONARIES VERSUS DESCRIPTIVE DICTIONARIES	22
2.6	LEXICAL VERSUS ENCYCLOPEDIA DICTIONARIES	22
2.7	GENERAL DICTIONARIES VERSUS SPECIAL DICTIONARIES	24
3.	ENTRIES	27
3.1	INTRODUCTION	27
3.2	LEMMA	30
3.3	CONSTRUCTION OF ENTRIES	33
3.3.1	The character of an entry	34
3.3.1.1	Lexicographic definition	35
3.3.1.2	Synonyms	36
3.3.1.3	Types of entry clarification	38
3.4	ARRANGEMENTS OF ENTRIES	40
3.4.1	Root Arrangement	42
3.4.2	Morphological rewriting	43
3.5	MULTI-WORD LEXICAL UNITS	44
3.6	DICTIONARY-LABELS	47

4.	POLYSEMY AND HOMONYMY IN DICTIONARIES	51
4.1	INTRODUCTION	51
4.2	HOMONYMY AND POLYSEMY	52
4.2.1	Polysemy and its presentation	54
4.2.1.1	Meaning discrimination	55
4.2.1.2	Sequence of senses	59
4.2.2	Homonymy and its presentation	61
5.	MORPHEMIC ANALYSIS	65
5.1	INTRODUCTION	65
5.2	THE NATURE OF THE MORPHEME	66
5.3	LEXICOGRAPHICAL PROBLEMS OF WORD FORMATION (WITH SPECIAL REFERENCE TO SANDHI)	68
5.3.1	Sandhi rules	69
5.3.1.1	Palatalization	70
5.3.1.2	Assimilation	71
5.3.1.3	Plosivation	72
5.3.1.4	Alveolarization	74
5.3.1.5	Vowel elision	75
5.3.1.6	Consonant insertion	76
5.3.1.7	Vowel coalescence	77
6.	LEMMATIZATION OF DIFFERENT WORD CATEGORIES	79
6.1	INTRODUCTION	79
6.2	LEMMATIZATION IN THE NOMINAL CATEGORIES	79
6.2.1	The role of prefixes in the nominal category ...	80
6.2.1.1	Class prefix	81
6.2.1.2	Prefixes in qualificative constructions	81
6.2.1.3	The possessive	83
6.2.1.4	The absolute pronoun	84
6.2.1.5	The demonstrative pronoun	85

6.2.1.6	The quantitative pronoun	85
6.2.2	The role of suffixes in the nominal category ..	86
6.2.2.1	Diminutive suffix	86
6.2.2.2	Feminine suffix	87
6.2.2.3	The locative suffix	88
6.3	LEMMATIZATION IN THE VERBAL CATEGORY	88
6.3.1	The role of prefixes in the verbal category	88
6.3.1.1	The infinitive prefix <u>ho-</u>	89
6.3.1.2	The object concord of the first person singular <u>N-</u>	89
6.3.1.3	The object concord of the third person singular <u>mo-</u>	90
6.3.1.4	The reflexive prefix <u>i-</u>	90
6.3.2	The role of suffixes in the verbal category	95
6.3.2.1	The perfect morpheme	95
6.4	LEMMATIZATION OF CONJUNCTIONS, INTERJECTIONS AND IDEOPHONES	98
6.5	CURRENT STRATEGIES FOR THE LEMMATIZATION OF NOUNS AND VERBS AND POSSIBLE WAYS OF LEMMATIZING NOUNS AND VERBS	103
7.	DERIVATION AND INFLECTION	109
7.1	INTRODUCTION	109
7.2	DEVERBATIVES	110
7.3	VERBAL EXTENSIONS	113
7.4	IRREGULARITY AND VARIANTS	115
8.	LEXICOGRAPHY AND THE COMPUTER	118
8.1	INTRODUCTION	118
8.2	MAJOR STAGES IN DICTIONARY-MAKING AND THE CONTRIBUTION OF THE COMPUTER	118
8.3	THE MICROSTRUCTURE OF A DICTIONARY DATABASE	122
8.4	ADVANTAGES OF THE COMPUTER IN LEXICOGRAPHICAL WORK	123

8.5	A COMPARISON OF THE DIFFERENT CHARACTERISTICS OF A PAPER DICTIONARY AND A COMPUTERIZED DICTIONARY	124
8.6	THE POTENTIAL FUTURE OF LEXI-COMPUTING	125
9.	CONCLUSION	127
	BIBLIOGRAPHY	128

CHAPTER 1

INTRODUCTION

1.1 SCOPE OF INVESTIGATION

Southern Sotho, like other members of the Bantu language family, is a predominantly agglutinating language. A characteristic feature of such languages is the adhesion of various formatives to roots, resulting in complex or compound words, the constituent parts of which are generally still recognisable as separate entities.

In certain instances, however, fusional structures, characteristic of inflectional languages, are observed. Thus, although Southern Sotho is typified as an agglutinating language, it also displays certain features of inflectional languages.

Lemmatization in Southern Sotho is directly affected by the merging of formatives and as a result becomes a complex matter. This work seeks to investigate how lexicographers systemize lexical entries taking into account their semantic, syntactic, phonological and morphological levels, and thereby deciding on certain forms as lemmas.

The word 'lemma' indicates a headword, or 'catch word'. It is the head entry under which other forms are indexed which are morphologically derived from it. Thus the identification of lemmas is closely related to the morphological level of lexical items. In the following chapters an attempt will be made to point out difficulties which lexicographers come across, or considerations which they have to bear in mind in the

process of lemmatization. The intention to propose solutions, however is beyond the scope of this work.

Examples to illustrate how lexicographers tackle some of the important issues in lemmatization have predominantly been taken from the 8th edition of Mabile and Dieterlen's Southern Sotho-English Dictionary (1983).

1.2 THE RELATIONSHIP BETWEEN THE LEXICOGRAPHER AND THE DICTIONARY-USER

The task of the lexicographer is definitely not an easy one:

- providing each lexical item with sufficient information to describe its behaviour in the language
- expressing the relationship, sub-regularities and generalizations among lexical items of the language
- choosing the basic lexical item (lemma) in preference to other items which are related to it. (cf. Jackendoff, 1975:639)

Economizing on space, the lexicographer does all this, trying as much as he can to avoid superfluity.

Dictionaries are lists of lexical entries. Each entry is introduced by the lemma in its orthographic representation and alphabetized in terms of it. A dictionary assumes some knowledge of the morphological rules of the language. It is necessary that the linguistic principles on which a dictionary is based

should be applied consistently. This is not easy taking into account the complex derivational system which is characteristic of Bantu languages.

The first task of the lexicographer is to decide what kind of person will use his dictionary. The level of complexity with which he presents his work will be against the background of his anticipated readers. It is this user-friendliness approach component which this introductory chapter seeks to address.

The lexicographer's work is also creative. Because of the authority that his work has on the structure of the language, he not only concerns himself with the language as it exists, but finds new solutions to problems as yet unsolved, cf. Zgusta's (1971:20) advice to the lexicographer in this respect is that he should know the relevant properties of his lexical units as well as possible. Since a dictionary presupposes a grammar, he should also have a firm grasp of the morphology and syntax of the language or languages he is dealing with (cf. Al-Kasimi, 1977:49). He should have a clear picture of his future dictionary's disposition and of its purpose. The lexicographer should also know the history of his language. Thus although a dictionary presupposes a grammar, it should also function as an amendment to it (op cit:51).

The criteria for lemmatic status must be accessible to the dictionary-user. Sufficient efforts must be made especially if the target-user is a non-native speaker, for example, to give an indication as to how derived forms are treated, in as clear and systematic a procedure as possible. The following are helpful in this respect:

- (1) The front matter of the dictionary, which gives the phonological description of the language. It should explain the phonemes and allophones and their complementary distribution. Al-Kasimi (1977:50) rightly suggests that (in a bilingual dictionary) word-formation habits of the foreign language, such as compounding, suffixation, zero changes, shifting (a word from one class to another) and re-duplication, should be explained in the grammar in order to cut down the need for redundant information in the dictionary. The front matter should be organized in such a way as to facilitate the use of the dictionary and give adequate definition of the grammatical identifications which are given.
- (2) The typography-certain measures are taken to simplify the usage of the dictionary.

So word die lemma byvoorbeeld meerendeels vetgedruk sodat die lemma duideliker uitstaan; terwyl voorbeeldsinne en verbandsinne meesal gekursiveer word. (Alberts & Mtintsilana, 1988:11)

Diacritic symbols are also helpful in this regard, for example in indicating tone in Southern Sotho, which can be the only semantic differentiating factor in some lexical items.

The format of the dictionary should be eye-pleasing and attractive, pages well-printed, entries in bold-face etc. and typographical errors kept to the minimum.

- (3) Illustrative examples give the word a live

context, thereby enhancing the user's understanding of the semantic rules governing its usage.

(4) Labels, also,

verhoog die gebruiksvriendelikheid van die woordeboek en vergemaklik die kitsherwinning van inligting. (Alberts & Mtintsilana, 1988:10)

(For a detailed discussion of illustrative examples and (dictionary) labels cf. Chapter 3).

In the discussion it is clear that although the lexicographer's task and the user's needs are very diverse, they are, on the other hand, quite closely related. Lexicographers provide the lexicographical information that users expect to find. Thus lexicographers are not only concerned with providing this information, but they are also concerned with, amongst others, operations, activities and skills users rely on during the look-up operations.

It is the task of the lexicographer to present all this information in an organised manner, which is comprehensible to the dictionary-user. The dictionary must present its matter in a way that is convincing to its user. James (1989:19) points out that if the user agrees with the dictionary in cases where he is sure of his judgement, then he is more likely to rely on the dictionary in cases where he is not.

In this way, the dictionary not only provides the user with the knowledge he thinks he needs, but

much more with the knowledge he does not know that he needs until he studies the article in detail. In this respect language teachers should provide their students with lexicographical education.

The neglect of this important aspect of linguistic education results not only in the student's inefficient use of the dictionaries but in many misconceptions about their nature and function. (Al-Kasimi, 1977:107)

Since dictionaries are capable of more uses than students suspect, this lexicographical education will enable them to get the most out of using their dictionaries.

By way of concluding this section, reference is made to axioms of lexicography by E B van Wyk (1989). He mentions, amongst other things, that:

- (1) a dictionary primarily provides lexical information;
- (2) a dictionary presupposes linguistic knowledge on the part of the user. The productive aspects of its morphology are normally assumed to be known;
- (3) the linguistic principles on which a dictionary is based should be applied consistently;
- (4) a dictionary should be user-friendly. It should be less technical and should provide the least troublesome way of looking up a

desired lexical item;

- (5) a dictionary should be aimed as a specific target group.

CHAPTER 2

TYPES OF DICTIONARIES

2.1 INTRODUCTION

A dictionary has been defined by Al-Kasimi (1977:66) as a book containing a selection of words, usually arranged alphabetically, with explanations of their meanings and other information concerning them, expressed in the same or another language.

Zgusta (1971:197) adopts a definition by C.C. Berg:

a systematically arranged list of socialized linguistic forms compiled from the speech habits of a given speech-community and commented on by the author in such a way that the qualified reader understands the meaning of each separate form, and is informed of the relevant facts concerning the function of that form, in its community.

These definitions reveal that a dictionary is a bi-sectional system. These 'sections' are called a 'macro-structure', i.e. an ordered series of lexical items, entries, or lemmata; and a 'microstructure' i.e. linguistic information contained inside the entry (Hartmann 1983:77).

Dictionaries play an important role in the standardization and codification of languages. The word 'dictionary' in itself suggests supreme scholarship and precision. It also carries in its weight authority as regards comments that the author makes about language. Talking about dictionaries,

Hartmann (1983:166) points out that

Their listing of certain words or phrases as having a defined sense, however tenuous or dubious that definition may turn out to be, seems to confer an uncontested authority upon those listed items.

The evidence can be seen from various purposes users want to achieve in consulting a dictionary. These are summarised in tabular form in the following manner.

INFORMATION	OPERATIONS
meanings/synonyms	finding meanings
pronunciation/syntax	finding words
spelling/etymology	translating, etc.
Situations of dictionary use	
USERS	PURPOSES
child	extending knowledge of the mother-tongue
pupil/trainee	learning a foreign language
teacher	composing a report
scientist/secretary, etc.	reading/decoding FL texts, etc.

Bearing these considerations in mind one tends to agree with Ilson (1985:1) that the dictionary is the most successful and significant book about language. Dictionaries can be classified according to many criteria. In his book Linguistic and Bilingual Dictionaries, (Al-Kasimi (1977) discusses different typologies of various linguists.

Malkiel's typology is based on three criteria:

1. Classification by range. In this criterion the following factors are considered:
 - (a) density: how much of the total lexicon of the language is covered, and the depth of coverage, i.e. how many meanings are listed under the entry;
 - (b) number of languages involved; accordingly there are
 - monolingual dictionaries
 - bilingual dictionaries
 - trilingual dictionaries
 - quadrilingual dictionaries and
 - multilingual dictionaries;
 - (c) extent of concentration of lexical data, i.e. how encyclopedic is the work.
2. Classification by perspective; the following factors are considered:
 - (a) The fundamental dimension, i.e. whether it is synchronic (descriptive) or diachronic (historical);
 - (b) Pattern of organization, i.e. whether the compiler used an alphabetic, semantic or nonsystematic pattern;
 - (c) Level of tone, i.e. whether it reports facts objectively, didactically or facetiously.
3. Classification by representation: here

dictionaries are classified in the light of the following aspects:

- (a) definitions
- (b) exemplification
- (c) graphic illustrations
- (d) special features such as marking of pronunciation and usage information.

Al-Kasimi rejects Malkiel's typology on the grounds that it does not result in discrete, mutually opposed dictionary types. Almost all features can co-occur freely in all dictionaries. Another objection pertains to its terminal objectives. He asserts that this kind of typology is

... more valuable in the library cataloguing room than in the office of the lexicographer. (Al-Kasimi, 1977:16)

Al-Kasimi goes on to present a neat argumentation of the typologies of other linguists, such as those worked out by William S. Cornyn and I.J. Gelb. They classify dictionaries on the basis of the amount and type of information contained in the dictionary.

Cornyn distinguishes three kinds of dictionaries:

- (a) encyclopedic, i.e. treating all aspects of cultural contexts;
- (b) usage, i.e. attempting to portray the morphological and syntactic features; and
- (c) glossary, i.e. a list of forms with a minimum information.

The objection against this classification is based on the fact that it fails to take into account the fact that most dictionaries are not limited to one type of information, but that they are rather a mixture of two or all of the types mentioned.

Gelb's classification is based on the point of view of the compilers. Thus we can distinguish the 'philological', linguistic and anthropological and the general purpose dictionaries by the philologist, anthropologist and the one produced by commercial undertakings, respectively. Al-Kasimi (op. cit.) cites two difficulties set by this typology:

- (1) What is meant by the terms themselves, i.e. philologist, linguist etc.
- (2) the failure to point out differences among the suggested types of dictionaries.

Before looking at the typology proposed by Al-Kasimi, a brief survey will first be made of some factors which according to Landau (1989:7 et sqq) in his book entitled **Dictionaries: the art and craft of lexicography**, are important in classifying dictionaries.

- (1) Number of languages

Dictionaries differ in the number of languages they contain. A monolingual dictionary is written entirely in one language. The chief purpose of the monolingual dictionary is to explain, in words likely to be understood by the native speaker, what other words mean. A bilingual dictionary consists of a list of words in one language (the source language) for which

equivalents are given in another language (the target language). A bilingual dictionary may be unidirectional or bidirectional. In a unidirectional dictionary the lemma is indicated in one language and translated into another. In a bidirectional dictionary there are in actual fact two dictionaries, e.g. South Sotho-English and English-South Sotho. There are two main purposes for using a bidirectional dictionary:

- (a) comprehension (as in reading), and
- (b) as an aid in expressing oneself, as in writing, in the target language.

(2) The Manner of Financing

Dictionaries differ in how the projects are being financed and the profit expected.

- scholarly dictionaries; often financed by government addition to higher institute of learning, and are not designed principally to make a profit
- commercial dictionaries; supported by private investors who are looking forward to some profit through their sale.

(3) The age of users

Some dictionaries are aimed at children; others at adults.

(4) The size of dictionaries

Dictionaries differ in exhaustiveness, i.e. in how fully they cover the lexicon.

(5) The scope of coverage by subject

Dictionaries differ in scope with respect to the subjects they cover. There are those that are confined to a special subject, such as medicine, and those limited to an aspect of language such as a pronunciation dictionary. Dictionaries limited to one aspect of language are called "special-purpose dictionaries".

(6) The lexical unit

Dictionaries differ in the size of the unit catalogued. Some confine themselves to words or lexical units; others to phrases, set expressions or larger quotations.

(7) The primary language of the market

Dictionaries differ in the primary language of their intended users. Some are designed for those who are learning the language as a second language, as distinguished from native speakers. Dictionaries intended for second language speakers combine many features of foreign language instruction such as providing information on pronunciation, verb patterns, with some characteristics of children dictionaries: definitions are expressed in a simplified language and sometimes in a controlled vocabulary. In this way they have an essentially pedagogical purpose.

(8) The period of time covered

Dictionaries differ in the period covered by their lexicons, i.e. whether they are synchronic

or diachronic. Diachronic dictionaries deal with an extended period of time, with the chief purpose of tracing the development of forms and meaning of each lemma over a period of time, while synchronic dictionaries focus on the language at its present state.

(9) The linguistic approach chosen

Dictionaries differ in their attitude toward usage, i.e. whether the editor's approach is prescriptive or descriptive.

(10) The means of access

Dictionaries differ in the manner in which access to their information is provided, specifically as to whether their word lists are arranged alphabetically, conceptually, or in some other way.

The typology proposed by Al-Kasimi (1977:19) presents differences in:

- (1) source
- (2) scope
- (3) purpose

He has the following points (amongst others) to mention about this proposal:

- it is meant to be a guide to the dictionary-user in selecting the appropriate dictionary, i.e. the one that can help him the most;
- it is limited to bilingual dictionaries only, the primary intention of which is to assist the

lexicographer to produce better bilingual dictionaries.

Here are some of the different contrasts which this typology sets up:

- (1) Dictionaries for the speakers of the source language versus dictionaries for the speakers of the target language.
- (2) Dictionaries of the literary language versus dictionaries of the spoken language.
- (3) Dictionaries for production versus dictionaries for comprehension.
- (4) Historical dictionaries versus descriptive dictionaries.
- (5) Lexical dictionaries versus encyclopedic dictionaries.
- (6) General dictionaries versus special dictionaries.

In the following section the distinguishing features of different types of dictionaries will be discussed, based on the above typology.

2.2 DICTIONARIES FOR THE SPEAKERS OF THE SOURCE LANGUAGE VERSUS DICTIONARIES FOR THE SPEAKERS OF THE TARGET LANGUAGE

By the source language is meant the language of the entries. The target language is the language of the equivalents.

It is not possible to satisfy the needs of speakers of

both (the source and target) languages equally. The following facts show how monolingual and bilingual dictionaries differ in this respect:

- (a) Although both help in understanding a foreign language, with a monolingual dictionary, the student is forced to use the foreign language in order to understand it, and there is no guarantee that the definitions, examples or metalanguage notes (since all the vital information is in a foreign language) are comprehensible. In the case of a bilingual dictionary however, target language equivalents are given for headwords, derived forms and examples, and the metalanguage is L1: the student thus uses L1 in order to understand L2.

- (b) When students are trying to express themselves in the target language L2, for example writing an essay, the monolingual dictionary offers little immediate help and demands much more from its users than does the bilingual dictionary. Users of a monolingual L2 dictionary can access the material in it only by means of a foreign language headword. It might be just that word that they do not know. If that is the reason for the difficulty, the situation becomes circular, and there is no way out. The user of a bilingual dictionary, on the other hand, gradually learns to operate in L2 without the L1 barrier constituting a hindrance to progress.

Some semantic problems can however be evidenced in the case where the user (of a bilingual dictionary) has to operate in L2.

- (c) The choice of equivalents, i.e. to find

appropriate equivalents in the target language to the units of the source language. Two factors which particularly contribute to this difficulty are:

- culture-bound words which denote objects peculiar to the culture of the source language, as can be seen in the following example:

Telejane¹ n., ho botsa - to ask for advice before getting married;

- the scientific and technological terminology which does not exist in other languages.

In his attempt to create the closest equivalence in translation, the lexicographer has to consider the following:

- the translation must represent the customary usage of the receptor language
- the translation must make sense, and
- the translation must conform to the meaning of the original, according to Al-Kasimi (1977:62).

- (d) Whereas it is nearly always possible to establish translation equivalence between sentences, it is often difficult to do so between lexical items. The following facts indicate why differences between two related items in two languages may

¹ according to the Southern Sotho culture a girl and/or her parents would secretly inquire the young man's manners, parents, etc., before consenting to marriage. The point here is that since **Telejane** is peculiar to the English culture, it would be difficult to choose an equivalent.

exist.

- Languages differ in the (semantic) implication of their grammatical categories, for example, the tense aspect in South Sotho can also imply respect as in Ke ne ke kopa ho bona mosuwe-hlooho 'May I see the principal?'
- Languages differ in their parts of speech, for example English prepositions may be expressed by means of nouns which are used as adverbs in South Sotho, as in:

O behile buka hodima tafole. 'She placed the book on the table.'

- A lexical unit in one language may not even have a corresponding lexical unit in another language. The meaning of that lexical unit might be expressed by a syntactical device in the other language, e.g.

hetla (to)²'look behind oneself'
motswetse woman just confined.

- Two different languages may have two grammatical patterns to determine certain aspects of experience, e.g.

ke lebohile 'thank you' (lit. I have thanked).

- A lexical unit in one language may have two

²

The translations will include to (as in the dictionary by Mabile and Dieterlen) although it should strictly speaking be left out since the infinitive prefix (ho-) does not appear in the Southern Sotho equivalent.

components; but its equivalent in another language may have one component only, as in Fonaneng 'Good-night'; matsale 'mother-in-law'.

- Two related items in two different languages may not cover the same semantic range, e.g. ngwetsi 'daughter-in-law'; 'newly-married woman'. Uncle can either mean malome 'maternal uncle' or rangwane 'paternal uncle'.

Bohloko can mean grief; sadness; suffering; torment; affliction; pain.

2.3 **DICTIONARIES OF THE LITERARY LANGUAGE VERSUS DICTIONARIES OF THE SPOKEN LANGUAGE**

Speech is the fundamental form of language activity, and writing is just a representation of speech. Where the literary and the spoken languages differ functionally and structurally it is necessary that the two should be dealt with in separate dictionaries. As far as Southern Sotho is concerned there is no marked difference with regard to this aspect. Moreover, although there may be a number of variations in the spoken language, very few alternative forms are allowed in the written language. Consequently, Southern Sotho is one of the most uniform written Bantu languages of South Africa.

2.4 **DICTIONARIES FOR PRODUCTION VERSUS DICTIONARIES FOR COMPREHENSION**

The distinction between these two types of dictionaries lies in the choice of the source and target languages. For example, if we intend to make a dictionary which is a Southern Sotho-English dictionary for Southern Sotho speaking people, we will

choose Southern Sotho as the source language if we intend this dictionary as one for production, and English if we intend it as one for comprehension. A dictionary which is intended to help speakers of Southern Sotho in both comprehension and production of English sentences should be bidirectional. Alberts & Mtintsilana's (1988:15) categorisation of bilingual dictionaries, based on Afrikaans and Xhosa, for example, may be referred to to exemplify this notion of the choice of the source and the target languages.

They distinguish four categories:

- (1) Xhosa-Afrikaans: To help the Xhosa speaker to express himself in Afrikaans. (Dictionary for production where Xhosa is the source language and Afrikaans the target language).
- (2) Xhosa - Afrikaans: To help the non-speaker of Xhosa to understand Xhosa: This would be a dictionary for comprehension where Xhosa is the target language and Afrikaans is the source language.
- (3) Afrikaans-Xhosa: To help the Xhosa non-speaker to understand Xhosa: This would be a dictionary for comprehension where Afrikaans is the source language and Xhosa is the target language.
- (4) Afrikaans-Xhosa: To help the Xhosa speaker to understand Afrikaans. This would be a dictionary for comprehension where Xhosa is the source language and Afrikaans is the target language.

Another difference between the dictionary for production and the one for comprehension lies in the content structure of the entries. The entry words in

a dictionary for production provides the user with more information than the dictionary for comprehension, about the morphological and syntactic behaviour of those words. Such a dictionary will help the user to express him - or herself appropriately in the foreign language.

2.5 HISTORICAL DICTIONARIES VERSUS DESCRIPTIVE DICTIONARIES

Descriptive dictionaries describe the lexicon of a language as it is in its present state. Historical dictionaries describe a language as it was in one or more of its previous periods. It arranges the senses of its entries in such a way as to show how meanings have developed from one another. In this way, the semantics is historically orientated. Hartmann (1983) points out that the historical dictionary is the most comprehensive type of an academic dictionary, and succinctly summarizes its aim as:

to trace the development of each word from its first known appearance by means of a series of quotations and references selected from literary and non-literary sources and arranged in chronological order as to give evidence of:

- (1) Each word's existence in different periods, places and genres.
- (2) The changes in the word's meaning, use, form and spelling.
- (3) Its idiomatic phrases and habitual collocations.
- (4) Its external, or comparative, etymology as well as its internal derivation.
- (5) Its morphological and syntactical characteristics. (Hartmann, 1983:123)

2.6 LEXICAL VERSUS ENCYCLOPEDIAIC DICTIONARIES

Lexical dictionaries deal with the lexical units of the language, i.e. all their linguistic properties, hence the other name for these type of dictionaries, linguistic dictionaries. Lexical dictionaries can be divided into synchronic and diachronic types. The purpose of the former is to deal with the lexical stock of a language at one stage of its development, whilst the latter deals with the development of a language, both with respect to form and meaning (Zgusta, 1971:198). Examples of diachronic dictionaries are the historical and etymological dictionaries. Etymological dictionaries focus their work on the origin of words hence the 'pre-historic' sense attached to their manner of approach in dealing with words.

In contradiction to lexical dictionaries, encyclopedic dictionaries give information about the extralinguistic world.

The three defining characteristics of encyclopedic information are:

- (1) the inclusion of names of persons, places and literary works;
- (2) coverage of all branches of knowledge; and
- (3) extensive treatment of facts.

Al-Kasimi (1977:31) point out that encyclopedic information related to the foreign language is more often looked for in bilingual dictionaries.

He (the foreign student) needs information about the famous persons and places, both real and literary, that native speakers respond to as part of their cultural heritage and education ... If he comes across a reference to Old Glory, Bucephalus, Mrs Maluprop, Madison Avenue, or even Tom, Dick and Harry, he cannot be expected to rush to the library and look it up (in encyclopedias, gazettes, and almanacs). He needs the reference immediately. He needs the reference in a dictionary right here on his desk.

2.7 **GENERAL DICTIONARIES VERSUS SPECIAL DICTIONARIES**

General dictionaries include learners', scholars' and translators' dictionaries.

Burchfield (1987: 77 & 78) cites the following as the main differences between learners' dictionaries and the conventional monolingual dictionary:

- (1) the removal of historical and etymological information
- (2) indication of pronunciation based on a pedagogically sound and widespread system
- (3) the International Phonetic Alphabet, instead of the use of 'figured pronunciation'
- (4) the incorporation of information of particular importance to learners but of less concern to native speakers, e.g. irregular plurals, special grammatical features of the headword, the use of controlled and simplified language in framing definitions.

- (5) copious use of citations and examples to support the bare definition
- (6) the incorporation of a detailed Introduction which is intended to help the learner to have access to the great range of lexical semantic, grammatical, varietal, and cultural information residing within entries.

An example of a special dictionary that will be briefly discussed is the "technical" dictionary. The technical dictionary is compiled on the basis of criteria provided by particular target groups and their professional or special-interest needs.

The lexicographer of a technical dictionary has to consider the level of linguistic proficiency that he expects the users to have, because, unlike the lexicographic approach in learner's dictionaries, his is basically non-linguistic. The user of his dictionary is generally not too interested in language. He is, rather, preoccupied with particular technical issues.

While having a certain command of one or more languages, a technical expert is usually quite aware of his limitations in this area and feels more at ease with a linguistic approach rather below his real level of proficiency. Adding to this the conditions of stress that often accompany the consultation of technical dictionaries, we can readily appreciate the general requirement of technical lexicography that only the most basic language handling skills may be assumed on the part of the user. The first and foremost of the required skills is the ability to find and identify lexical items in a total word-list (Hartmann, 1983:172).

Consequently, user convenience received high priority in technical dictionaries.

The above discussion shows that the dictionary is the most important record of language. Since linguists classify dictionaries according to a variety of different contradictions, dictionary-classification typologies are, as a result, also diverse. It is doubtful whether there are characteristics that will outrightly distinguish one type from another. Also, considering the scope, it is impossible to determine or maintain what any other type of dictionary should do, because, as already pointed out, most dictionaries are not limited to one type of information, but rather a mixture of two or all the types mentioned. In any way, any dictionary must be accurate in what it includes, and comprehensive within the limits of its original design.

Thus the purpose of a dictionary plays a significant role in determining its format.

Looking at the present situation regarding Southern Sotho dictionaries one realises that there are only bilingual dictionaries (i.e. no monolingual dictionaries) at present. Alberts & Mtintsilana (1988: 29-30) give a list of these (as well as word-lists). These bilingual dictionaries either have Afrikaans or English, as target languages, and Southern Sotho as the source language. Thus there are no lexicographic resources that are particularly provided for the speakers of the Southern Sotho language. The language probably has to undergo more research before a standardized monolingual dictionary can be written.

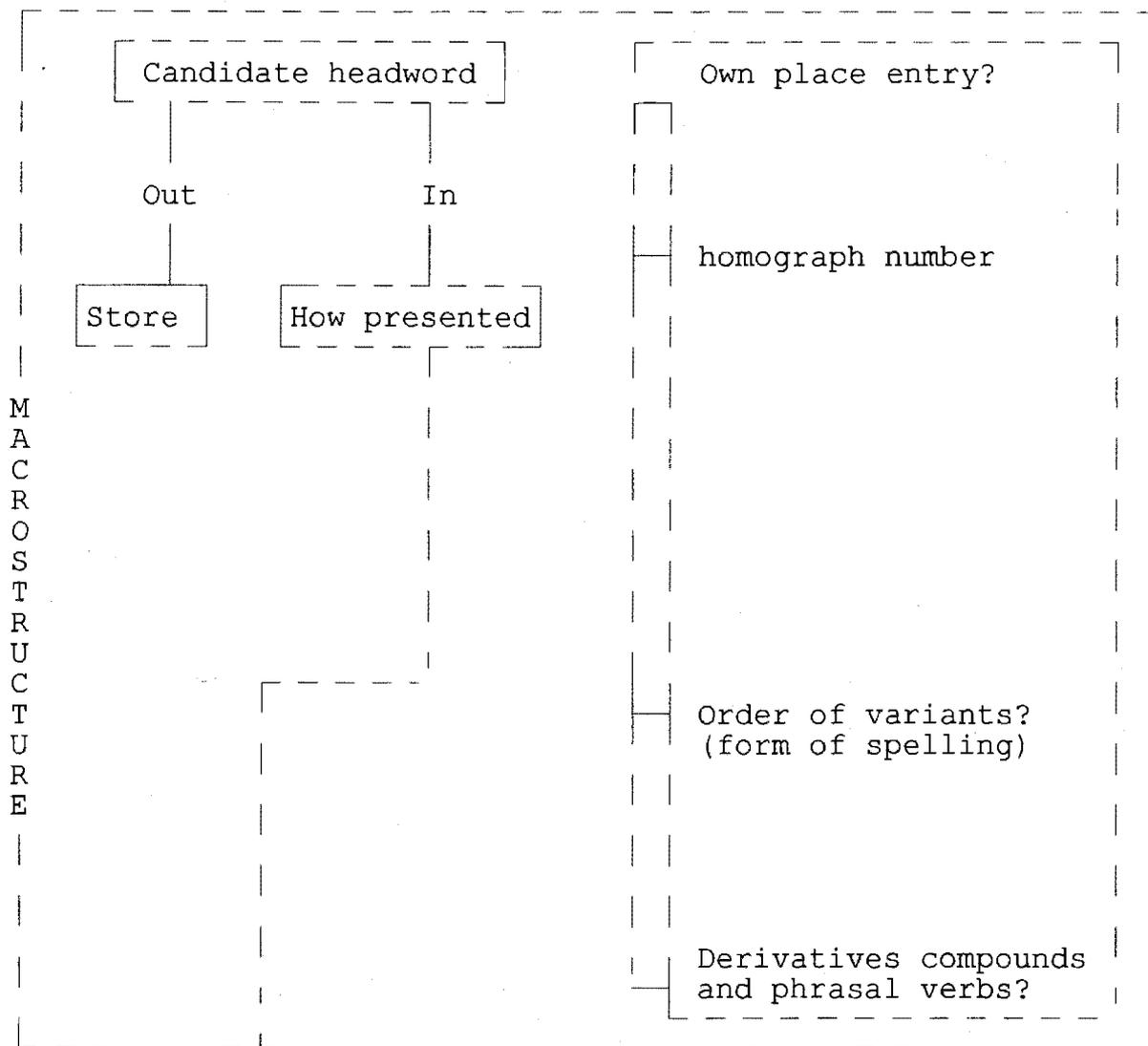
CHAPTER 3

ENTRIES

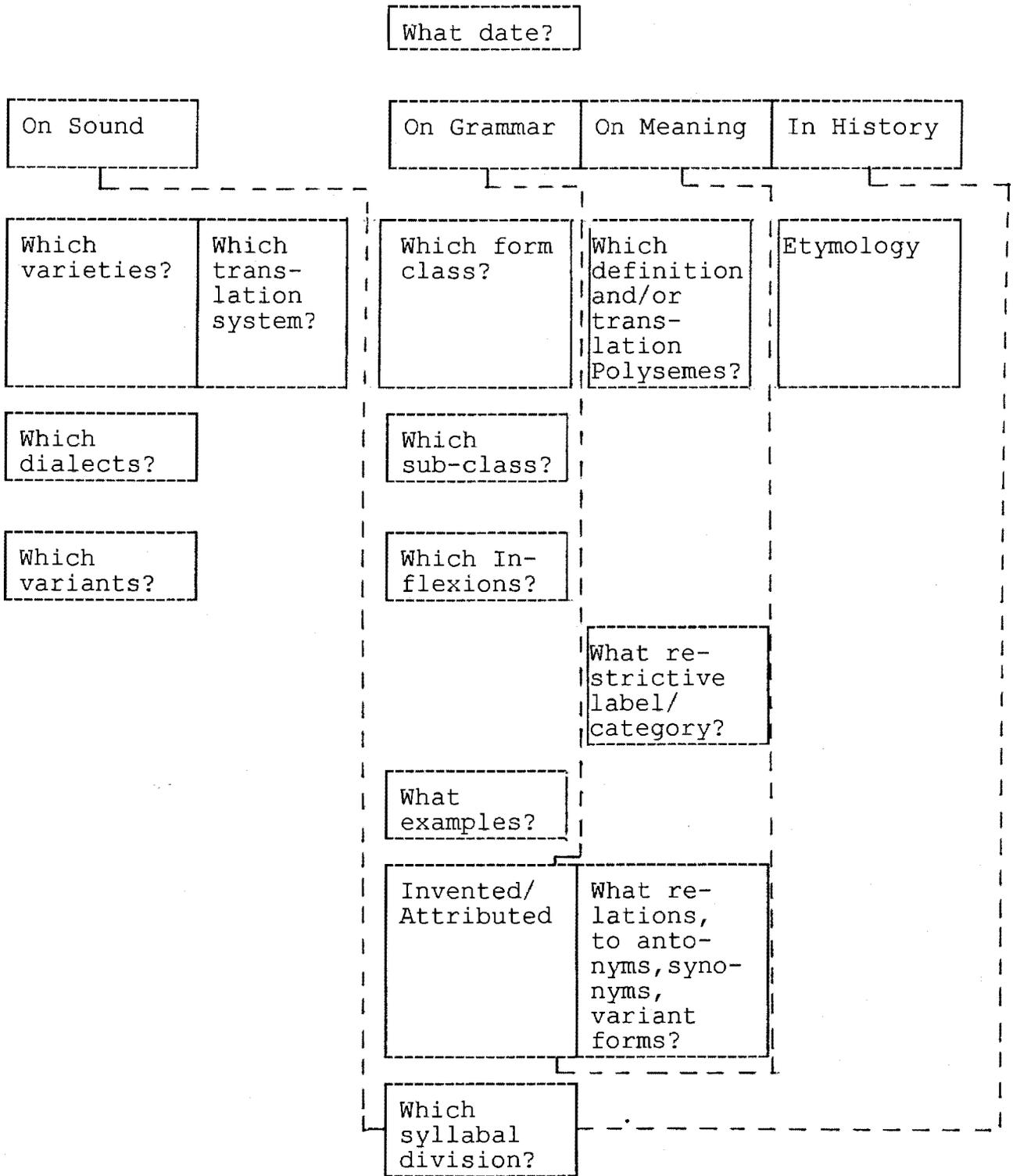
3.1 INTRODUCTION

As already mentioned earlier on, dictionaries have a macrostructure and a microstructure. The macrostructure involves the outer form and size of the lexicon. Thus the choice of lemmas determines the macrostructure of a dictionary. The microstructure entails the grammatical information that is found in the entries of the dictionary. Alberts & Mtintsilana (1988:8) further distinguish two types of information normally found in the microstructure, i.e. the implicit and explicit grammatical information. Implicit information includes word-type indications, plurals, diminutive and past form indications. Explicit information includes illustrative examples and information on collocations. The following diagrams illustrate further what is meant by the macrostructure and the microstructure of a dictionary.

Sources



The Macrostructure of a dictionary, Alberts & Mtintsilana (1988:7), as quoted from McGregor (1985:127).



The Microstructure of a dictionary Alberts & Mtintsilana (1088:8) as quoted from McGregor (1985:128)

3.2 LEMMA

A lemma is the unit at the beginning of an entry, which is usually semi-bold. It is the principle part of an entry because it introduces the entry, hence the other names 'headword' or 'catch-word'.

Criteria for lemmatic status differ from dictionary to dictionary. Inflected forms of the lemma can also occur as sub-lemmas. For example, we may find the noun lehleke 'disorderly things' as a sub-lemma of the ideophone hleke (hore hleke 'to be in disorder'). These inflected forms are semi-bold and they do not begin the entry. Indicating inflected forms of a lemma is particularly important when the inflections differ markedly from the canonical forms, (the canonical form of a word being the form chosen to represent a paradigm) like for instance, tjeo 'eating, food' which is a sub-lemma for the lemma-ja 'to eat'.

The purpose of the lemma is to identify the lexical unit, to locate it in the morphological system. In this way, lexicographically, the lemma has an administrative function, as it is the basis for the alphabetical ordering of the entries.

Apart from this function the lemma also serves a number of different purposes:

- (1) it indicates the preferred spelling
- (2) it indicates the usual printed form of the lexical unit, i.e. whether capitalized or not
- (3) it supplies phonetic information, which consists of the following categories:

- (a) prosodic features, which inform us about stress
- (b) segmental features, which indicate the pronunciation using a phonetic notation like the IPA (The International Phonetic Alphabet)

There are other information types, i.e. apart from phonetic information, that can be found in the entry about the lemma. This information corresponds to the various linguistic disciplines:

(1) Etymological information

Etymological information gives information about the historical origin of the word or language from which it was borrowed.

(2) Graphical information

Graphical information tells us how words appear in writing, for example the orthography of a word.

(3) Grammatical information which includes the morphological and syntactic categories.

In the light of this category we can distinguish between the following:

- Morphological sub-category, which specifies how the word in question is formed or can be used to form other words and compound words, as well as the kind of words that the lemma can enter into compounds with. Similarly, with derived words, information can be given on the morphemes they

can be combined with to form derived words. This sub-category also looks into inflectional forms, for example,

-kojwa 'be bent' is an inflectional form of which koba 'bend' is a lemma.

- Syntactic sub-category; which includes sub-categories, valency and syntactic function. Valency refers to the combinability of a certain word with certain other words, for example a transitive verb. Syntactic function also includes mentioning syntactic restrictions, if they exist.

Syntactic function indicates that a word can perform a certain function in a sentence, e.g. ka, 'at, in', used to recognise an adverb. This sub-category also involves stating that a certain lemma is a verb, noun etc.

Syntactic restriction entails a limitation regarding the syntactic function of a word, e.g. an intransitive verb such as ho shwa 'to die' cannot take an object.

(4) Pragmatic information

Pragmatic information refers to information categories that indicate the use of the word in the broadest sense. It includes examples of use as well as frequency of its use. As regards usage three dimensions are distinguished.

- the temporal dimension; e.g. word used in former times

- the spatial dimension; e.g. word used in a particular geographical area
- the social dimension; e.g. word used in a special social circle or in a specific context. This dimension will be discussed in detail in a following section, (cf.par.3.6).

(5) Semantic information

Semantic information serves to describe the content of the lemma - it provides its meaning in the broadest sense, for example subject classification is a sub-category of semantic information, which provides information about specialized fields or subjects.

3.3 CONSTRUCTION OF ENTRIES

As already noted, an entry-form consists of a single word. Multi-word lexical units are also considered as entries in the process of selection because they function as single words. The question of multi-word lexical units will be addressed fully later on (cf.par. 3.5). At this stage it is necessary to point out, that they are selected as entries. Zgusta (1971:241) points out that in some cases even a morpheme can enjoy lemmatic status, especially in languages with a complex compositional structure.

In those languages where the boundary of the word is not sufficiently clear the lexicographer will meet morphemes about which he cannot easily decide whether they are words of their own or not ... he will be well advised to allow them their own entries as if they were independent words,

eventually with some further special indications and specifications.

For example the object concord of the first person singular, N- is treated as a lemma, whereafter the syntactic function is given and exemplified.

Therefore, as these examples illustrate, it is fully legitimate to consider single morphemes as lemmas, describe their meaning and add some of the more stabilized words in which they occur.

The construction of entries, as can be seen from the foregoing discussion, is one of the basic tasks that have to be dealt with in the process of planning a dictionary. Each entry must contain all the relevant information concerning the lexical unit for the purpose of the dictionary. The entries should be constructed in a uniform manner. Although cross-reference is considered a necessary procedure, each entry should nevertheless be treated as a compartment of its own, i.e. it should concentrate upon the individual properties of the lexical unit in question.

3.3.1 The character of an entry

Ilson (1995:16) explains that

a dictionary entry will consist of some or all of the following components:

- (1) the lemma
- (2) an indication of pronunciation
- (3) detail of the word classes
- (4) morphology: inflections of the lemma
- (5) syntax: syntactic potential as well as syntactic restrictions

- (6) various senses of the lemma
- (7) exemplification of usage
- (8) derived forms of the lemma
- (9) cross-reference
- (10) classification of the semantic content
- (11) stylistic indication (where relevant)
- (12) usage material for the purpose of further classification
- (13) etymological clarification

Certain important issues regarding the character of an entry will be discussed below.

3.3.1.1 **Lexicographic definition**

In the entry, the core definition of the lemma is normally given. The editor of a dictionary must decide on how definitions as well as their level of complexity will be presented in his dictionary. Definitions presented in a dictionary for an elementary school child, for instance, will differ markedly from those prepared for a scholarly specialist. In presenting the definition the lexicographer will therefore have to bear in mind the background of the target group which will be consulting the dictionary.

There are different types of definitions used in dictionaries.

TYPE 1: FUNCTIONAL, where the object is defined by its use or by the function it could perform
e.g. molayi ' one who reprimands'.

TYPE 2: CONCRETE, where the definition refers only to the object's descriptive or

perceptual attributes

e.g. seqhahabola 'thin porridge or drink made of maize'.

TYPE 3: RELATIONAL, where the definition is offered by use of an analogy; a large number of such definitions is by synonym. This type also relies on relations of functions

e.g. hoballa 'to walk conspicuously, as a tall person'.

TYPE 4: COMBINATORIAL, where two or more of the categories mentioned above are combined
e.g. mokwetla 'white clay used for whitewashing'.

Looking closely at the examples outlined in the above discussion of definitions, one cannot fail to realize that a lexicographic definition enumerates only the most important semantic features of the defined lexical item, which is just enough to distinguish it from other items. The definition should consist exclusively of words which are explained in the dictionary.

3.3.1.2 **Synonyms**

Synonyms are phonetically different forms which express the same meaning. From the phonetic end, synonymy is neutralization of semantic contrast. From the semantic end, it can be seen as phonetic diversification according to Ohlander (1976:50).

The narrower sense of synonyms is limited to the identity of meaning only between words. Terminologically however, synonymity relations may be

assumed to characterize the words of larger linguistic units.

Hence there is a distinction between synonymy and partial synonymy, but practically they express the same idea.

Absolute synonyms would be lexemes sharing all of the same syntactic functions as well as of the same connections to the semantic level. It happens to be quite correct that such units do not exist, but this argument does not require absolute synonymy. In other words, synonymy is always partial, never complete. The lexemes *big* and *large* are synonymous because they both connect to a single sememe ... But they are not absolutely synonymous, cf. Hill (1969:47)

The argumentation accepts the fact that synonymy be somehow measured in terms of degree of distributional similarity. Hence not only words but morphs, phrases, clauses, sentences and sequences of sentences may be cited as examples of synonymous examples, cf. Harris (1973:1).

Synonyms are usually indicated as an addition to the definition, e.g. -qwela 'to sink in water; to dive'.

Synonyms help to clarify the definition of the lexical unit in question, as in -qhala 'to disperse, to rout, to scatter, to waste'.

The lexical unit dealt with is located in a group of semantically most related words. The resulting contradistinctions and comparisons clarify the whole semantic description; disambiguates one near-synonym

from another so as to convey all the critical features of the lexical unit. Consequently, this may help the user to find an expression more suitable for his purpose than the one he originally had in mind and checked up in the dictionary cf. Zgusta (1971:260).

3.3.1.3 **Types of entry clarification**

- (1) There are different ways in which a lexicographer can help clarify the meaning of a lexicographic unit to the dictionary user. Illustrative examples are phrases or sentences that illustrate the use of the defined item. Citations and quotations also serve as illustrative examples, although they are limited to those illustrative examples which are found in actual speech or writing.

The use of illustrative examples in dictionaries serves a number of purposes:

- (a) to prove that a word or particular sense of a word exists in the language;
- (b) they serve to illustrate the meaning of the word defined;
- (c) to illustrate the grammatical (phonological, morphological and syntactic) behaviour of the lexical unit;
- (d) to indicate the stylistic value of the entry;
- (e) they are a pedagogical device, serving ...

to contribute to the user's interest by showing the word in a live context, and to

enhance his understanding of the grammatical and semantic rules governing the usage of the word and by showing these rules in action (Al-Kasimi, 1977:9).

Illustrative examples are either taken from the texts of the language or constructed by the lexicographer with the help of his informants.

(2) Constructed examples

The advantage of constructed examples is that since they are constructed by the lexicographer himself they can be of great use in illustrating difficult points, thus they can be constructed in such a way that they invite analogical application by the dictionary-user. Thus they often have great generative power.

Secondly, constructed examples are usually very short, thus economising in space, e.g. -sala 'to remain, stay', ho sala morao 'to stay behind'.

(3) Quoted examples

Quoted examples help clarify the definition by reflecting the sole lexicographic context of the entry word. Dictionaries quote examples which are grammatically complete. The context in which these examples are quoted helps the user to perceive in which signification the word in question is applied, and to perceive at least some of its semantic features and grammatical properties; cf. Zgusta (1971:227).

(4) Set expressions

Sometimes it becomes imperative that the meaning of a set expression be presented in the dictionary, especially if the common reader cannot deduce its meaning from the constituent parts and if the expression is frequently used, e.g. nonyana 'bird' may be find nonyana e dula motho meaning 'emotion is high'.

A question that lexicographers have to answer concerning set expressions is where to list them, for example having to decide to which of the senses of a polysemous word the respective set expression belongs. Zgusta (1971:267) suggests in small dictionaries these should be presented at the end of the whole entry. In large dictionaries they should be presented at the end of the examples of free combination of words which illustrate the definitions of the single senses.

For example if a lexicographer considers boroko 'sleep' as a polyseme he would present the definition as follows to illustrate the definition as single senses

- 'to sleep' ho robala boroko bo monate 'to sleep well'
- 'to accommodate for sleep' kamore e robala batho ba leshome
- 'the room has sleeping room for ten people'.

3.4 ARRANGEMENT OF ENTRIES

With regard to the arrangement of entries each dictionary seems to follow its own system, which is considered most convenient to its user. The most

common method of arranging entries is by the alphabetic sequence of lemmas. Other possible arrangements are by, semantic connection and derivation of words. Most important in the discussion of arrangement of entries is the notion of nests. A nest is a group of entries which is conflated into one, in which case the second and other entries are explained by reference to the first member of the nest. Zgusta (1971:284) points out that the more frequent situation is that the single members of the nest are presented separately as any other entries, with their own eventual polysemy and sometimes even with their own examples.

e.g. tswa 'to come out; to emit'
tswa tlung 'to leave the house'
ho tswa madi 'to bleed'
tswa, 'deficient verb to form an immediate past tense'; ke tswa kena 'I have just entered';
motswa 'one who comes from' motswa disa o ja
motoho sedula hae se ja masepa, proverb: 'it is right that one who works gets the advantages';
matswathaka, (mosadi wa matswathaka 'first wife').

In languages where regular derivations are numerous, such as Southern Sotho, it is possible to construct extensive nests. The morphemic status of such derivations can disagree with the alphabetical order of such nests.

Dictionaries deal with problems arising from this situation in different ways. As far as this matter goes, some dictionaries expect a certain degree of analytic ability on the part of its user. In some dictionaries the nest is presented as a whole with dubious cases, but the morphemically transparent words

are entered once more in their alphabetically proper places among the other entries, frequently rather far away from the nest itself, and have a cross-reference to the nest, e.g. hautshwane 'diminutive of haufi near' -entse perfect of -etsa 'to do'.

Root arrangement and morphological rewriting are some of several alternatives which present themselves in the search for a technique which might enable us to indicate relationships among words belonging to the same family in the dictionary.

3.4.1 Root arrangement

According to this arrangement vocabulary is presented in the word families. Some linguists consider root arrangement as the only type of arrangement which can clarify derivational affinities within each root, semantic evolution and etymological relationship cf. Al-Kasimi (1977:76).

This type of arrangement of entries is very popular in the dictionaries of derivational language.

However, the root arrangement has some obvious disadvantages. First, it requires a considerable grammatical sophistication on the part of the dictionary-user, for instance expecting him to guess that tjeo 'eating' is listed under -ja 'to eat' or mahlo 'eyes' under the lemma -ihlo. Secondly, even if the dictionary-user was able to deduce the required root, he would still have to read right through the article to find the word he was looking up.

The term article in this regard refers to all the information contained in the entry.

These inconveniences would not make a good dictionary which, according to Al-Kasimi (1977-77) is the one in which you can find what you are looking for - preferably on the first try.

3.4.2 Morphological rewriting

The lemma should include morphological rewriting, which is achieved by using hyphens, i.e. in the case of a dictionary which opts for this alternative of presenting entries. These hyphens are used to mark morpheme boundaries. Thus the lemma -latela 'to follow' is followed by its morphological rewriting - lat - ela.

This proposal also, has some problems associated with it. Firstly, its application to lexicography presupposes clear-cut boundaries between morphemes. Unfortunately morphologists have not yet reached overall agreement on morpheme boundaries for example the inflection of mono-syllabic verb stems in Southern Sotho. Secondly, this method might work when applied to derived words whose constituent morphemes are distinct, e.g. -latela, but not to words whose constituent morphemes are not so obvious, e.g. tjeo, and more specifically, words which show sandhi forms.

Some dictionaries consider 'nests by the stem' as a good practical solution to problems caused by word-formation. Considering the stem for lemmatization and at the same time bearing in mind the productive capacity of affixational morphology the lexicographer must decide which lexical forms to list in the dictionary and in what manner to list them.

For example, taking tsamaya, 'to walk' as the primary verb form, should derivatives and extended forms

(verbal extensions) be listed as independent entries or as run-ons under the lemma tsamaya 'walk'? Bearing in mind the question of consistency this is not an easy decision to make. The argumentation concerning arrangement of lexical items will be looked at under the discussion on Derivation and Inflection (cf. Chapter 7), but as regards the bearing that this matter has on the question of arrangement of entries.

the decision is whether derivational suffixes are to be treated simply as grammatical formatives analogous to the structural prefixes, or as constituting some lexical property which might necessitate lexicographic treatment (James, 1989:7).

3.5 MULTI-WORD LEXICAL UNITS

Before a sequence of phonemes can be lexicalized or lemmatized it must be coded. Codedness is the word at the centre of lemmatization of multi-word lexical units or syntactically rigid sequences, in fact it is an important notion in the whole field of lexicography, because the word list of a dictionary can only be made up of coded units. James (1989:2) points out that the two criteria which determine codedness are frequency (in a corpus being evidence of the fact that it is used by the community at large) and syntagmatic rigidity (i.e. stored as a unit whereby no or very little syntactic or lexical variation is allowed.)

Zgusta (1971:145-146) discusses some principles that can help us to detect a multi-word lexical unit.

(1) Substitution

This is the criterion that he deems more important than all the others, in deciding whether a sequence is a multi-word lexical unit. Substitution is not possible in a multi-word lexical unit, e.g. it is not possible to say, i.e. in the case of seqatamajwana, seqataditlhaka or substitute majwana 'little stones' by another word. The meaning of seqatamajwana 'a kind of game played by boys by throwing stones' can only be grasped by both units used as one multi-word lexical unit.

- (2) Sometimes it is impossible to add something to the unit, since it already conveys the complete idea on its own, e.g. Setsweng ke hole meaning 'that is impossible' (lit: the elbow is far), i.e. no one can kiss his/her elbow.
- (3) Different semantic phenomena are internal or at least indicative, so the meaning of the unit is not derivable from that of single parts as in the case with setsweng ke hole. The meaning is less derivable than, in the case of, sekolong ke hole 'school is far', therefore the common procedure will be to accept setsweng ke hole rather than sekolong ke hole as a syntactically rigid sequence.
- (4) A constituent part of a set combination of words may be severely or exclusively restricted to it, e.g. mahlo ke diala 'eyes see anywhere' mohla monene 'that day' and letsoha-phora 'late-ariser' diala, monene and phora seem to be restricted to these combinations respectively.

- (5) The multi-word lexical unit may have a synonym or near -synonym which consists of one word only. For example lenyeka-thipa or mofutsana 'a poor person'.

Coming to lexicalization of multi-word lexical units and set expressions, the point of concern is where to list them. For example' the meaning of multi-word lexical units and multi-morphemic words does not generally correspond to the sum total of the meaning of their constituent elements, as has been found in the discussion of the above examples. As a result they are oftentimes treated as entries of their own. Zgusta (1971:290) points out that it is easier to deal with them as sub-entries, i.e. where they are inserted into the main bulk of the entry. The method of sub-entries can be of great value to both the lexicographer and dictionary-user if the sub-entry word and lemma are typographically well expressed, and the choice is done with sufficient care. Just as the entry is a lexical abstraction, it follows that a sub-entry is also a lexical abstraction, but of a lower level.

Coming now to the question of where to list the set combination of multi-word lexical units a number of factors are considered for instance, to list mahlo ke diala 'eyes see anywhere' under mahlo 'eyes' or diala (from -ala meaning (to) spread out) or modiha-nyewe 'one who wins a court case' under -diha '(to) cause to fall' and nyewe 'case'. There is no single unequivocal case reply but based on Zgusta's (1971:268) opinion the following points are decisive:

- (1) Elements which are of minor significance are generally not suitable: for instance nobody will probably look for mahlo ke diala under ke.

- (2) In the case of a set combination like seqatamijwana 'game of throwing stones at each other' (played by boys) the user will have no trouble with the word majwana (diminutive of 'stones'). The possibility is that he will have trouble with -qata, so he will look up the meaning of seqatamajwana under : -qata.³
- (3) In certain cases the meaning can be inserted under more than one entry in the case of the multi-word or set expressions. For example, letsohaphora is inserted under both -tsoha '(to) waken' and phora 'late'.

If it is not possible for reasons of space, then these expressions could be inserted under items which indicate the core, i.e. not under those elements which attribute, e.g. leihlolakgomo 'golden autumn star' would be inserted under the lemma -ihlo (leihlo 'eye') and not under kgomo 'cow'. The best solution to combinations which offer little or no arguments for the preference of one of their elements is to insert the sub-entry into both the entries, as is the case in letsohaphora above. In the dictionary by Mabile and Dieterlen for example the item is listed under both -tsoha '(to) waken' and phora 'late'.

3.6 DICTIONARY-LABELS

A dictionary should inform the user about the attitude of the society or certain sectors of it towards

³ following context this word probably means '(to) throw', but this sense is not provided in the dictionary, hence seqatamajwana is listed as a lemma and not a sub-entry of which -qata is lemma. One would assume that the sense in which some of such words are used has become obsolete.

particular linguistic forms, i.e. whether a particular linguistic form is formal or informal, and on what occasions it can be used. The user must be helped to understand the social implications of its use. Knowing the appropriateness of its use can save a lot of embarrassment and confusion.

The lexicographer provides this information by means of labels.

Die term etiket verwys na daardie inskrywing(s) in 'n woordeboek wat as merker(s) optree om die lemma of 'n ander inskrywing in die woordeboek-artikel se beperkinge ten opsigte van byvoorbeeld stylaard, gebruik sfeer en geografiese voorkoms aan te dui (Alberts & Mtintsilana, 1988:8).

The use of the dictionary-labels is influenced by the type of the dictionary. Descriptive dictionaries have a semantically neutral inclination whereas a prescriptive dictionary has an analytical nature.

There is an important connection between a word and its use, and as pointed out earlier on, to know a word is to know its use in the language. Lexicographical labelling is an invaluable tool in achieving this goal.

Leksikografiese etikettering is die enkele inskrywingstipe waarin taalkundige sowel as pragmatiese inligting dus oorgedra kan word en waarin die algemene taalvermoë van die woordeboekgebruiker verhoog kan word (Alberts & Mtintsilana, 1988:9).

Labelling is done by means of symbols or devices of some kind, such as abbreviations. These are explained in the Introduction of the dictionary. The signification of labels employed is also precisely defined in the front matter.

There are different types of labels:

- (1) Temporal labels indicate the chronological marking on a lexical item. For example a word may be labelled archaic if there is little evidence of its recent use.
- (2) Another set of labels are regional or geographical indicating the use of the word in a specific area.
- (3) Stylistic labels indicate the sociolinguistic value of the lemma as well as its use, in a typical conversation situation. In this set of labels we have the following as examples: vulgar, obscene, figurative, and colloquial.
- (4) Slang is a type of label which is used for words which are restricted to relatively small groups and have a flair to them.
- (5) The status label indicates social level. One example is the label non-standard, characteristic of those of little education.

Although dictionary-labels are useful, there are some problems associated with their use, the most significant being the following three. Firstly, labels mean different things to different lexicographers and different users. They do not always agree for instance on what 'slang' or 'archaic' mean.

Secondly, as a result of this disagreement, there is no consistency in labelling words, cf. Al-Kasimi (1977-88). Hence the same word used the same way might be labelled differently in different dictionaries.

Lastly, no set of usage labels will be valid for all time. At some stage a word can be regarded standard or formal when at some other time it was labelled informal or even vulgar. Labels and their implications change as practices and opinions change, cf. Congleton (1979:34).

We therefore conclude, that the best way to get information about usage is by carefully observing what kind of people use the linguistic form and to what kind of people they use it, and lastly, by examining the status of the linguistic form which is constantly changing.

CHAPTER 4

POLYSEMY AND HOMONYMY IN DICTIONARIES

4.1 INTRODUCTION

Meaning is an important factor that lexicographers have to deal with in the process of constructing lexical entries. The concept 'meaning' will be briefly discussed before going into further details on its presentation in a dictionary.

After indicating the lemma the lexicographer must also indicate the meaning of the lexical unit in all its aspects. The meaning which he assigns to the entry is what has been referred to as lexical meaning. Lexical meaning is not independent of situation. A word is its use in the language, therefore to know the meaning of a word (any piece of language) is only meaningful when placed in situations. Thus the meaning of a word cannot be provided say, by just saying in words what the word means, but by carefully categorized instances of its use.

Any strict separation of linguistic-, lexical- and extralinguistic-factual information is difficult, if not impossible. Lexemes ... can be described and explained only by reference to the contents in which they are used (Hartmann, 1983:77).

We can thus see from this explanation that meaning, although it is an abstract concept, is observable in use. The lexicographer must take certain steps to try and make it accessible to the dictionary-user. Just as performance is the only evidence for competence, so

actual meaning is the only evidence for lexical meaning, cf. Van Rensburg (1982:117).

Problems presented by polysemy and homonymy come as a result of lexical ambiguity, which is due to alternative meanings of an individual lexical item. The lexicographer must do all he can to enable the dictionary-user to get the appropriate meaning for his particular purpose for consulting the dictionary.

4.2 HOMONYMY AND POLYSEMY

Every lexicographer has to consider the problem of the lexicographical treatment of homonymous and polysemous lexical items when planning a dictionary.

Let us examine the following pairs of items:

- | | | |
|-----|-------------------|--|
| (a) | <u>kgopo</u> | (i) crooked
(ii) unrighteous |
| (b) | <u>-hlokahala</u> | (i) to be rare
(ii) to die |
| (c) | <u>bohloko</u> | (i) pain
(ii) small white bowels in
an animal |

Words have two related aspects:

- (1) a phonetic or graphic image, and
- (2) a semantic content

The problem with examples such as those above is that the relation between meaning and form is not one to one. This results in various senses being conveyed by the same word form. Polysemy and homonymy are the two

cardinal types of semantic ambiguity.

The boundary between homonymy and polysemy is not always clear-cut. Linguists and lexicographers make use of different criteria to decide whether a certain lexical unit is homonymous or polysemous, such as formal grammatical differences, etymology, semantic distinctiveness and collocational sets.

When we look at the examples (a) kgopo
 (i) crooked
 (ii) unrighteous

 (b) -hlokahala
 (i) to be rare
 (ii) to die,

we can conclude from the description of these words that the second meanings, i.e. 'unrighteous' and 'to die' have been derived from 'crooked' and 'to be rare' respectively. Thus, 'crooked' and 'to be rare' can be referred to as parent lexemes, of which 'unrighteous' and 'to die' are derivative forms. This notion outlines semantic relations that lexicographers consider in their dealing with polysemous and homonymous items. In this case, i.e. 'kgopo' as well as '-hlokahala' we can conclude that these items are polysemous, because other senses can be traced back to one original sense of the word.

Mtintsilana (1990:69) notes that this procedure is commonly employed in etymological dictionaries.

Otherwise, if the origin of the words cannot be traced, the words are homonymous. But this notion of assuming polysemy or homonymy on the grounds of relatedness of meaning is also criticized on the

grounds that it is a matter of degree, so that even if some native speakers may see a connection between words, others may not. This view comes down to the fact that: whilst speakers see a polysemous relation between a pair of lexical items, others see it as homonymous. Zgusta (1971:17) justifies this reliance on the subjective interpretation of speakers in determining whether a pair of lexical items is polysemous or homonymous in the following way:

After all, a language exists to be spoken and understood, and it exists by being spoken and understood, so that the intersubjective understanding of the speakers can be considered a criterion.

4.2.1 Polysemy and its presentation

As a general term, polysemy designates the fact that a single word can have different senses. The expression 'polysemy of a word' thus means all the senses that a word has.

The presentation of polysemy is one of the difficult tasks that a lexicographer has to deal with. This process is made more complicated by the role subjective interpretations of speakers play in its description.

The lexicographer has his data in the form of citation-slips, which he groups according to his perception from the context of the various senses of the word which they represent.

Although this is a very simple procedure in principle, in practice the difficulty is that for most words various alternative groupings are conceivable, and the

lexicographer has to rely on his expert intuition to decide which is best. Van Rensburg (1982) further warns that even an expert's intuition is not without error, so his decisions cannot be consistently good.

... and to make matters worse, it usually takes another expert to detect the bad ones (Van Rensburg, 1982:113).

Since intuition plays an active role in the treatment of polysemous lexical units, any inconsistencies are justifiable on the grounds of the fact that speaker's intuitions are subjective and unreliable.

4.2.1.1 **Meaning discrimination**

At the centre of the process of presenting polysemous lexical items lies the notion of meaning discrimination.

If an entry-word is polysemous, the dictionary-user may experience some difficulty in selecting the proper sense of an equivalent. In such cases the dictionary must provide meaning discrimination. For example let us say an equivalent of tao is being looked for in a Southern Sotho-English dictionary. The dictionary-user will find a number of words in the entry:

tao n., bed, place, situation, position

He will have some difficulty in selecting the equivalent which fits his context unless he has a good command of Sotho, in which case it is unlikely that he would have needed the dictionary, cf. Hartmann (1983:160).

Considering this difficulty that may be experienced by the dictionary-user as can be seen in the above illustration, it is imperative that meaning discrimination be provided in a systematic fashion. Otherwise the bilingual dictionary cannot be a dependable guide to proper equivalents and translation efficiency.

To solve this problem the lexicographer has to follow a particular pattern in order to provide meaning discrimination.

The following procedure is suggested by Hartmann (1983:161). Firstly, the lexicographer must answer the following questions.

- (1) When does a polysemous source language or target language word require meaning discrimination?
- (2) Should meaning discrimination be provided in the source language or target language?
- (3) How should meaning discrimination be best presented in the entries of the dictionary?

Considering firstly questions (1) and (2), we realise that the need for meaning discrimination depends on the purpose of the dictionary, i.e. whether it is meant as an aid to comprehension or production, and whether it is intended for speakers of the source language or the target language. If the dictionary is designed as an aid to production, no meaning discrimination is necessary in the following two cases:

- (a) if the source language word has one meaning for which the target language has one meaning, e.g.

seledu 'chin' and

- (b) if the source language has one meaning for which the target language has a polysemous word.

Meaning discrimination is necessary in the following two cases:

- (a) the source language word is polysemous and for each of its meanings the target language has a separate word meaning.
- (b) the source language word is polysemous and for each of its meanings the target language has two or more polysemous words.

If the dictionary is intended for comprehension only (i.e. for target language speakers), no meaning discrimination is needed in the following two cases:

- (a) the source language word has one meaning for which the target language has one word or only one meaning, and
- (b) the source language word is polysemous and for each of its senses the target language has one word or only one meaning.

Meaning discrimination is required in the following case:

the source language word has one meaning and for each of its senses the target language has two or more polysemous words.

The language in which meaning discrimination is supplied should be the language of the speakers for

whom the dictionary is intended.

Considering question (3) i.e. How should meaning discrimination be best presented in the entries of the dictionary? the lexicographer has several devices which are available to him. The following are some of them:

- (a) Punctuation: separation of synonyms or near synonyms by commas, and different meanings by semicolons, e.g. phala 'trumpet, bugle, whistle; headman, tube, pipe; stem on which a glower blossoms'.
- (b) Definitions: i.e. to show different meanings of a polysemous word, e.g. -lapisa 'to cause to become hungry; to excite the appetite'.
- (c) Synonyms: they can provide briefer meaning discrimination than definitions

tao 'bed; place, situation, position';

- (d) Illustrative examples: illustrative phrases or sentences may serve the purpose of meaning discrimination well, but they use a lot of space.

-nkgisetsana, 'to cause one another to smell' ho nkgisetsana mahafi, 'to struggle to wrestle'.

- (e) Part of speech labels can sometimes provide meaning discrimination in a polysemous word e.g. (1) hore conj., 'that, in order that' (ii) hore adv. 'in a certain manner' hore, horeng n. 'a certain thing'.

- (f) Context words or phrases which give only just enough of the context in which a polysemous word is used, e.g. morebele, 'quick running, swiftness'; ho bea motho morebele, 'to follow, to pursue, to worry someone'.

An important and basic decision that a lexicographer has to make in connection with the presentation of polysemy is the depth and detail with which the polysemy is to be presented.

Since the notion polysemy implies more than one sense as far as lexical meaning is concerned, the lexicographer must decide on a selection of the senses and how he will co-ordinate these with the selection of entries. This is no easy task because the selection must be presented systematically throughout the dictionary, i.e. the selection of the lexical entries and the selection of the single senses of the polysemous entries must be governed by roughly the same principles (Zgusta, 1971:273).

4.2.1.2 **Sequence of senses**

As already pointed out, the use of a polysemous word is occasioned by diverse kinds of situations. The sequence in which the single senses of a polysemous word are to be presented is determined by the lexicographer.

Various arrangements are possible depending on factors such as:

- (1) for whom the dictionary is intended and
- (2) for what use it is intended.

For instance a historical dictionary will arrange according to historical developments, an encyclopedic dictionary will stress the technical senses.

Both the grammatical as well as the semantic considerations are regarded in ordering the sequence of senses. (Zgusta (1971:276-277) outlines the following suggestions as far as ordering is concerned.

- (1) The dominant sense is indicated first.
- (2) If the polysemous word does not have any dominant sense, the sense which has the broadest application and no marked connotation should be indicated first.
- (3) All other senses which are more specialized or which belong to the restricted language or styles, or which are obsolete, regional or vulgar should then follow in the order which seems to be optimal to the lexicographer,

e.g. (a) **-taba** 'to mix snuff with other ingredients; to become grey, (pepper and salt), to become aged; to mix in order to season'.

(b) **-seha** 'to cut with a knife or scissors; to hurt; to move one end so as to turn; intransitive verb, to run'.

It is difficult to maintain a consistent way of dealing with the semantic relationships of a polysemous nature, because the methods used for analysing polysemous lexical items depends almost entirely on the analyst's intuition.

If the contribution of intuition could be reduced by means of a formal methodology, the resulting analysis should be better (Van Rensburg, 1982:113).

4.2.2 Homonymy and its presentation

Homonymy is encountered where two or more different words have the same form. It begins at the point when the speakers of a language are unable to conceive different senses as connected, cf. Zgusta (1971:74).

There are two types of formal homonyms:

- (1) homophones: homonyms identical in sound and not in spelling. There are no homophones in Southern Sotho because the spelling system does not allow such a situation.
- (2) homographs: homonyms identical in spelling but not in sound.

- | | | |
|-----|----------------|---|
| (a) | <u>lekgolo</u> | [lekxholo] (i) 'hundred' |
| | | [lekxholo] (ii) 'black or brown animal with a yellowish muzzle' |
| (b) | <u>-lema</u> | [lema] (i) 'to plough' |
| | | [lema] (ii) 'to let someone get into a habit' |
| (c) | <u>-otla</u> | [otla] (i) 'to feed' |
| | | [ɔtla] (ii) 'to strike, to beat' |

Homonyms in the strict sense are words both spelt and pronounced alike. Hence to assume (true) homonymy two factors must be satisfied:

(1) the written and the spoken forms must be identical

(2) the pronunciation must be similar

'to spread out'

An example like nama -----

'flesh, meat'

qualifies as a case of homonymy, having two completely distinct meanings. Ohlander (1976:49) observes in a case such as the one presented above that the neutralization of the 'expected' phonetic contrast is completely accidental. There is, as it were, no logical reason why these two unrelated meanings should be expressed by the same phonetic form.

Given the fact that nama can mean 'flesh' there is no way of predicting that the same form could also mean 'to spread out'.

Therefore, owing to its totally idiosyncratic nature, the kind of phonetic neutralization displayed in homonymy cannot be captured in a phonological rule system, it is an exclusively lexical concern ... the phonetic neutralization is completely accidental, there being no phonological or other rules to account for it (Ohlander, 1976:49-50).

In the Bantu languages we distinguish homotonal and heterotonal homonyms,

determining structural and quantitative characteristics of homonymic groups, can lay the foundation for elaborating the principles governing the rational lexicographical description of homonyms, cf. Burchfield (1987:36).

What follows are some examples to show how homonymy is presented in existing works.

- (1) -tjeka¹ 'to dance, to turn' (--)
tjeka² hlonopo, substitute for ntlo, 'house'
 (-)
- (2) hoba¹ 'conjunction ba (--)
-hoba² 'to smear, to besmear, to anoint'
 (--)
-hoba³ 'to fail (of marriage)' (--)
-hoba⁴ 'to turn; to dance' (--)
-hoba⁵ 'to beat with a stone' (--)

Polysemy is not a defect but an essential feature of language; the alternative would be to have separate words for any possible subject we may wish to talk about. Homonymy, on the other hand, is in no way essential, except to the punster: a language without homonyms is perfectly conceivable, and would in fact be a more efficient vehicle of communication (Minnis, 1971:81).

CHAPTER 5

MORPHEMIC ANALYSIS

5.1 INTRODUCTION

Word formation is a serious aspect that lexicographers have to deal with in compiling a dictionary. Criteria for recognising a unit 'word' differ in detail from language to language. These criteria fall into three sets:

- (1) Phonological criteria, for example, stress
 - there will be a certain stress pattern applying over the unit word.

- (2) Grammatical criteria, for example, cohesiveness
 - all the parts of a word must occur together and cannot be scattered through the sentence.

- (3) General criteria, for example
 - a word is generally a 'minimal free form', the smallest unit that can be pronounced as a complete utterance;

 - a word will generally have psychological reality for speakers. They will talk of meanings of words and not, for example, morphemes;

 - breaths (or pauses) and interpolation of expressions like swear words are more likely to occur between words than in the middle of a word.

When dealing with word formation processes it is important that the phonological word and the grammatical word are clearly distinguished. The branch of the science of language, which studies the patterns on which a language forms words is known as word-formation. The next discussion will look into the processes of formation and the results or products of this formation.

The choice of items to be included in a particular dictionary and their presentation within it constitute problems for the practical lexicographer. Ilson (1985:38) points out that in dictionaries for the foreign learner there is generally a stronger emphasis on the encoding language needs of the learner. This in practice means, for instance, that a learner who knows the meaning of tshweu 'white' and studies the article -mothofatsa, 'personify' should be made to realize the existence of an item like -tshweufatsa 'to make white' in the lexical stock. He should be able to establish appropriate links between related items. The choice of items to be included in the dictionary presupposes a description of formative elements.

5.2 THE NATURE OF THE MORPHEME

The morpheme is the minimal unit of grammatical analysis. The word 'morpheme' is derived from the Greek word 'morphe' meaning 'form'. Hence a morpheme signifies the elemental unit of grammatical form.

Morphemes do not always have an independent form. They are combined in various ways to form complete words. Fromkin and Rodman (1983:114) mention that there is an arbitrary union of a sound and a meaning. If one looks at the following example, for instance:

1st speaker: Ke batla di ...

2nd speaker: Di eng?

it is evident that it is the class prefix that enabled the second speaker to formulate the question appropriately, i.e. he would not say Mo eng? or Se eng?

In this way morphemes have a certain degree of semantic significance. they are thus

" ... the smallest individually meaningful elements in the utterances of a language" (Hockett, 1969:123).

Thus, although the word is the principal unit in lexicography (cf. Lyons, 1979:194), the acceptance of a word as a lexical item presupposes the acceptance of an analysis of its constituent morphemes, or their associated morphs. For example, the lexicalization of -hlotjhwa 'be bothered by' under the lemma -hlopha 'bother', presupposes the knowledge of the change ph > tjhw in the passive form.

There are two kinds of morphemes namely roots and affixes. Roots are basic word constituents from which words may be formed. They cannot be reduced any further if the semantic core is to be maintained. A root itself can be a word, e.g. ideophones, such as tseke! (i.e. the flashing of lightning) or interjections such as kgele! (surprise), towe! 'you' (exclamation), which are complete words. These are termed free morphemes, they occur 'unattached'. These single morpheme words are simple lexemes, i.e. they do not have any morphological variation and are thus

lemmas as far as classification of lexical entries is concerned.

Affixes are additions to the root. Hence they are essentially bound morphemes. They may be prefixed, i.e. placed in front of the root, or suffixed, i.e. placed after the root. Affixes have important grammatical functions and may also have derivational properties.

(1) Grammatical Functions:

e.g. To indicate class gender and number, e.g.

mosadi 'woman': class 1, singular.

basadi 'women': class 2, plural.

(2) Derivational Properties:

To effect a change from one part of speech to another, e.g. mo- -tsamay-. i (>motsamai) 'one who walks'. In this example the prefix mo- and the suffix -i play a role in deriving a noun from a verb.

Affixation is the most important morphological process for the Bantu languages, since most of their categories are formed by it.

5.3 LEXICOGRAPHICAL PROBLEMS OF WORD FORMATION (WITH SPECIAL REFERENCE TO SANDHI)

This section deals with how sandhi may complicate the identification of lemmas in Southern Sotho. An investigation is made as to how instances of sandhi are accounted for in the lemmatization process.

Sandhi is a term that has been borrowed from Sanskrit,

an ancient language of India, and means joining. Basic forms may undergo a change or modification as a way of adapting themselves to other forms which surround them. Sandhi describes the process by which these adaptations occur. A sandhi form thus indicates a new form which a basic form adopts in an included position. Morphemes often undergo a phonological change. Both roots and affixes may display sandhi forms, e.g. the morph of the root morpheme in the verb stem -soma '(to) mock' is /-song-/ in the passive, and the morph of the class prefix (mo-) is /ngw-/ as in ngwana 'child' and ngwetsi 'daughter-in-law'.

Thus sandhi processes may result in different morphs of the basic morpheme, whether they be roots or affixes.

In some cases these processes may result in fused morphs consisting of more than one morpheme in which case the sandhi form is not readily segmentable, as can be seen in the following examples:

<u>-nany-</u>	+	<u>-ile</u>	>	<u>-nantse</u>	'crept to'
<u>-hony-</u>	+	<u>-ile</u>	>	<u>-hontse</u>	'plucked out'

Different sandhi rules will subsequently be examined and it will be indicated how these may affect the placing of lemmas in the dictionary.

5.3.1 Sandhi rules

Sandhi rules are those rules which underlie the change or modification that basic forms undergo. These rules are broadly categorized into assimilation and dissimilation. In assimilation forms which undergo modification acquire characteristics similar to those of the forms surrounding them, e.g. the prefixation of

the object concord of the first person singular to verb stems beginning with a bilabial consonant results in assimilation in that this prefix also becomes bilabial, as in the following examples:

<u>n-</u>	+	<u>-makatsa</u>	>	<u>-mmakatsa</u>	'surprise me'
<u>n-</u>	+	<u>-binela</u>	>	<u>-mpinela</u>	'sing for me'
<u>n-</u>	+	<u>-phehisa</u>	>	<u>-mphehisa</u>	'cause me to cook'

Dissimilation occurs when two similar sounds in juxtaposition acquire dissimilar characteristics e.g. when the object concord n- is attached to verb stems commencing with voiced consonants. In certain cases the initial consonant of the verb stem becomes voiceless to contrast the n- e.g. -nlata > -ntata 'fetch me'.

Two types of sandhi are distinguished. Internal sandhi explains a sandhi rule that applies within the phonological word, whilst external sandhi applies outside the phonological word. Since lemmatization is the main issue in this case the former will be discussed in more detail and it will be indicated thereby how sandhi may affect lemmatization.

5.3.1.1 Palatalization

Palatalization involves the substitution of a non-palatal consonant. It usually occurs with the passives of verbs, diminutives of nouns and colour adjectives.

Diminutives

Noun:	<u>lerole</u>	'dust'
Root morpheme:	<u>-role</u>	
Morph:	/-roj-/	

Diminutive noun: lerojana 'a little dust'.

Nominal diminutives are indexed under the nouns from which they derive. Lerojana 'a little dust' is thus placed under the lemma -role. This way of placing items presupposes a thorough knowledge of changes that may take place in the phonological word as a result of these processes. As a matter of convenience these forms are also lemmatized in their alphabetical place but not in bold-face.

5.3.1.2 Assimilation

Assimilation is a process in which two sounds in close proximity tend to become similar or acquire similar properties. It is mostly accompanied by vowel elision in cases which involve consonants. In cases which involve vowels it takes place in the form of vowel raising (or vowel coalescence).

In the following example:

mmuso 'government' < mo- + -bus- -o

b is assimilated to m after the elision of the vowel of the prefixal element. Since b has completely assimilated to the preceding element it is called complete progressive assimilation.

An example of partial assimilation can be illustrated by means of the affixation of the object concord of the first person singular to verb stems with a bilabial consonant in the initial position,

e.g. n- + -bon- + -a > /m-/ + /-pon-/ + /-a/
'see me'. Here the morph of the principal element is /m-/. Thus the morpheme N assimilates partially to the initial consonant of the root morpheme by becoming

bilabial as well. This (partial) assimilation is regressive, in that the feature 'bilabial' of the root morpheme has an effect backwards (i.e. on the preceding phoneme). Hence this process is called partial regressive assimilation.

Coming to identify lemmas in words where assimilation has taken place, it is necessary that the dictionary-user be conversant with the changes that may take place as a result of assimilation.

For example to be able to find the meaning of mmuso 'government', he has to look it up under the lemma -busa 'govern' which presupposes a thorough knowledge of assimilation on his part.

5.3.1.3 Plosivation

This process is activated by the following prefixes:

- (1) The prefix of class 9 N-
- (2) The object concord of the first person singular N- and
- (3) The reflexive prefix i-

The prefix of class 9 N-

In this case plosivation occurs in the formation of nouns and adjectives.

Nouns: e.g. mpho 'gift' which is derived from the verb stem -fa. This shows that because n and f cannot occur in juxtaposition, f changes to ph under the influence of the nasal and n changes to m, because in a word a bilabial cannot be preceded by an alveolar consonant.

As regards lemmatization, mp_ho is classified under the lemma -fa 'give'. This would imply a lot of cross-reference unless the dictionary-user understands the morphological aspect of the change that occurs.

Adjectives: e.g. mpe 'ugly' is an item that is classified under the lemma -be. So in the same way as in the case with mp_ho above, the change b > p occurred as a result of the noun class prefix before the adjectival root, nbe, hence the morph /-pe/ of the root morpheme (-be).

The object concord of the first person singular (N-)

e.g. verb stem -bona 'see' root morpheme -bon- morph /-pon-/ as in -mpona 'see me'.

To be able to identify the lemma of a word that has undergone a phonological change as a result of prefixation of this element, the dictionary-user must have a thorough understanding of the effect that this element (N-) can have, as words can be morphological complex, as for example, in mponeng 'a conceited person' which also derives from the lemma -bona '(to) see'.

The above examples show how plosivation can complicate the process of identification of lemmas. To alleviate this problem a thorough explanation of phonological changes is usually given in the introductory matter. For example, in their introduction to the 8th edition, Mabile and Dieterlen (1983:XIV) indicate that certain phonological changes may occur as a result of the prefix (n-), furthermore, footnotes are provided as a help to users who are not familiar with the Sotho system of prefixes.

The difficulty that is particularly posed by this prefixation process is that words ultimately look far different from their basic underived forms as is the case, for example, with mono-syllabic verb stems e.g. -ntja 'worry me' (lit. eat me), which is derived from -ja '(to) worry' (or lit. eat).

The reflexive prefix i-

e.g. verb stem -bolaya '(to) kill'
 root morpheme -bolay-
 morph /-polay-/
 as in -ipolaya '(to) kill oneself'
 (i.e. commit suicide).

The meaning of -ipolaya 'kill oneself' appears under the lemma -bolaya. in the same way -ithuta 'teach oneself' (i.e. learn) appears under -ruta '(to) teach' etc.

5.3.1.4 Alveolarization

Alveolarization occurs when an alveolar sound substitutes a sound of some other organic position.

e.g. Nominal diminutives

Noun morifi 'saucer'
 Root morpheme -rifi
 Morph /-ritshw-/
 Diminutive: moritshwana 'little saucer'

Nominal diminutives are indexed under the noun from which they derive. Hence the meaning of moritshwana 'little saucer' can be looked up under the lemma -rifi.

Alveolar transference

The consonant d can only be followed by i or u. Because the diminutive suffix commences with an a, attachment of the morpheme to nouns ending with a d for instance, causes d to be transferred to another alveolar sound which will be in compliance with the morpheme structure condition, i.e. ts, e.g.

Noun: naledi 'star'

Root morpheme naledi

Morph /-nalets-/

Nominal diminutive: naletsana 'little star'.

This can also be seen in diminutives of colour adjectives,

e.g. Adjective kgwadi 'black with white spots'

Root morpheme kgwadi

Morph /-kgwats-/

Diminutive kgwatsana 'a little black and white'

5.3.1.5 Vowel elision

This process explains certain instance where vowels are elided, as in the formation of the plural form of certain nouns, e.g. leihlo > mahlo 'eyes', whereby i is elided. So one may wonder where to look up the meaning of such words in this case mahlo 'eyes'. The meaning is classified under the lemma -ihlo from which leihlo 'eye' is derived. In the same way heso 'at my native place' is classified under the lemma -eso 'of our family' (i.e. belonging to me.) (Heso is a word that has undergone vowel elision since the long form is haeso).

5.3.1.6 Consonant insertion

The occurrence of a morphological change may necessitate the insertion of a consonant in a word, e.g.

(1) the insertion of k

This consonant is inserted between (a) the reflexive prefix and the verb stem whenever the stem begins with a vowel or semi-vowel; (b) the object concord of the first person (singular) and the verb stem whenever the stem begins with a vowel or semi-vowel; (c) between a zero morpheme and the verb stem when forming deverbatives in verb stems beginning with a vowel or semi-vowel. The three instances are illustrated, respectively, as follows:

-oka '(to) nurse' -ikoka 'nurse oneself'
-utlwa 'hear' nkutlwa 'hear me'
-otla 'punish' kotlo 'punishment'

This process must be explained to the dictionary-user to help him in identifying lemmas, i.e. he must be in a position to realise that, for example the meanings for -ikoka, -nkutlwa and kotlo may be found under the lemma oka, -utlwa and -otla, respectively.

(2) the insertion of n

A rare case exists whereby the consonant n is inserted within a word, e.g. the perfect form of the verb stem -etsa 'do' is -entse 'done'. This form is found in its correct alphabetical place, i.e. separate from the lemma -etsa because of its morphological irregularity, but the perfect form

is also indicated under the lemma -etsa.

In the verbs which have ny as the terminal consonant of the root, homorganic nasal realisation takes place whereby the existing palatal ny adopts an alveolar articulation,

e.g. -thinya '(to) turn'

Perfect: -thintse 'has turned'

Causative: -thintsha 'cause to turn'.

It is important, that the user be able to recognise the changes that have taken place so as to be able to identify the lemmas in these cases.

5.3.1.7 Vowel Coalescence

This is a process whereby the different vowels fuse to form another vowel of a different form. Examples given below illustrate the fusion:

a + i > e

(1) The perfect form of monosyllabic verb stems,

e.g. -shw- + -a- + -ile > -shwele 'died'

-j- + -a- + -ile > -jele 'eaten'.

The perfect form where such changes occur is usually indicated after the lemma. This involves a method of trial-and-error on the part of the dictionary-user. For example if a person reads the sentence Dineo o jele dijo 'Dineo has eaten (the) food', he can either look for * jel.a or * -ja as the lemma so as to get -jele as a sub-entry or a run-on derivative. Thus a thorough knowledge of the behaviour of monosyllabic

verb stems in the perfect is necessary.

- (2) The plural form of certain nouns

e.g. ma- + -ino > meno 'teeth'.

The meaning of meno is indexed under the lemma -ino 'tooth' (nominal stem). Here also it is a matter of a lucky user finding the meaning of meno on the first try.

- (3) The causative verbal extension in monosyllabic verb stems

e.g. -kgw- + -a- + -is- + -a > -kgwesa
'wean'.

In this case the final vowel of the verb stem and the vowel of the verbal extension element coalesce thus -kgw- a- + -is- + -a. It may not be easy to realise that -kgwesa derives from -kgwa 'to become weaned'. As a result -kgwesa is also lexicalized in the proper alphabetical place with a cross-reference to -kgwa.

This section shows that sandhi processes can complicate the process of identification of lemmas. Notes of the introductory matter cannot cover all aspects of morphology that are involved. As a result much is still expected of the user, i.e. his critical and analytic thinking. Consequently he basically employs a trial-and-error method in finding certain lexical items.

CHAPTER SIX

LEMMATIZATION OF DIFFERENT WORD CATEGORIES

6.1 INTRODUCTION

The classification of words in terms of morphological structure basically yields a binary classification namely a nominal and a verbal category. These categories are morphologically complex. Different morphological processes affect the structure of constituent morphemes and finally the placing of the word in a dictionary.

This part of the discussion concentrates on how lexicographers account for the morphological component of word categories since the form that words finally take in the dictionary is determined by morphological processes. The situation can be made even more complex, as already indicated earlier, by the productive capacity of these morphological processes.

6.2 LEMMATIZATION IN THE NOMINAL CATEGORY

The noun is the head word category for nominals. Other word classes which fall under this category are the pronoun and the possessive, the adjective and the nominal relative which have been termed qualificatives since they qualify nouns.

The general morphological structure of the nominals is a class prefix and a stem. There are, however, classes where the class prefix has disappeared or where it has been reduced to certain elements as a result of certain processes. These, however, reappear in the

concord system of the language.

Affixes play a significant role as far as identification of lemmas is concerned. The attachment of some of them lead to morphological changes which may even lead to semantic shifts. Since the lemma may appear in the form of a stem or root it is necessary to note how affixes may affect their basic structure.

6.2.1 The role of prefixes in the nominal category

The nominal stem only acquires a specific meaning when conjoined to its prefix. The prefixes in the following words are the only semantic differentiating factor:

sesadi 'pertaining to a woman'

bosadi 'what pertains to the feminine sex'

mosadi 'woman'

(n) tshadi 'womanfolk'

Prefixes thus govern the nominal category and thus have an important role in identifying lemmas. Mabile and Dieterlen (1983) have preferred to classify words which display prefixes according to the stems, or as near to them as could be ascertained. Thus the meaning of the word mosadi 'woman', for instance is found under -sadi. By the same token motswadi 'parent' is classified under the verb stem -tswala 'give birth to'. The latter case involves derivation and hence the lemmatization of such a word presupposes a sound knowledge of sound changes of the language on the part of the dictionary-user.

6.2.1.1 Class prefix

The class prefix indicates class gender and number. For example the prefix mo- in mosadi 'woman' indicates that it is a noun of class 1 (class gender) and is in the singular form (number).

Nouns are classified according to their singular form as far as possible, i.e. to the extent to which the singular form can be established. As a result Mabile and Dieterlen (1983) provide a list of the class prefixes in their introduction so as to aid a dictionary-user looking up the meaning of a noun in plural form. The lemma, is given in stem form, hence, the meaning of mosadi 'woman', for example, will be found alphabetically under the lemma -said, indicated in singular form as pointed out earlier.

6.2.1.2 Prefixes in qualificative constructions

A qualificative consists of a concord and a stem. It is always in concordial agreement with the noun it qualifies. The following are examples of adjectival stems:

- be 'ugly' as in bosiu bo bobe
 '(an) ugly night'
- tjha 'new' as in seeta se setjha
 '(a) new shoe'
- holo 'big' as in leeto le leholo
 '(a) great journey'

Apart from these adjectives which are called proper or common adjectives, we also have adjectives which indicate number and those which indicate colour.

e.g. number: -raro 'three' as in

bana ba bararo 'three children'

-hlano 'five' as in

mekotla e mahlano

'five bags'

colour: -fubedu 'red' as in

moriri o mofubedu

'red hair'

-tsho 'black' as in

majwe a matsho 'black stones'

The adjectival concord is composed of the demonstrative element and the noun class prefix, (according to Doke and Mofokeng, 1985:119). That noun class prefix often undergoes changes (including deletion) as a result of the influence of the adjectival stem, e.g.

n- + -be > mpe 'ugly' as in ntja e mpe an 'ugly dog'

di- + -holo > kgolo 'big' as in dilepe tse kgolo 'big axes'

As can be seen in the above examples the morphological changes that may occur on the adjectival stems can affect their identification in the dictionary.

For example, a point of concern arises as to whether mpe and mobe both from -be 'ugly' should appear as separate lemmas or whether one should consider them as derived forms and -be as the lemma. In the earlier edition of Mabile and Dieterlen's work (1979) for example, setle and motle both from -tle 'beautiful' appear as separate lemmas. In a later edition (1983), however, preference is given to the approach whereby

stems are given as lemmas. The explanation offered in support of this method is that, although prefixes do govern adjectives, nouns, etc., the prefix is a mobile and exchangeable element.

The reason for resorting to stem form is because

'... this has a great advantage of bringing words together which are similar in origin and related in meaning and of showing better the relative place of a given word in the language.' (Mabille & Dieterlen, 1983:xiii)

This assists the user in identifying lemmas. Hence mpe, which is derived from -be, and all adjectives are likewise classified in stemform.

6.2.1.3 The possessive

The possessive consists of the possessive concord and stem. The possessive concord refers to the possession and the stem to the nominal possessor, e.g. in kgomo ya ntate 'father's cow', kgomo 'cow' is the thing possessed and ya the possessive concord referring to kgomo 'cow', and ntate 'father' is the nominal possessor of kgomo.

Coming to lemmatization, we realise that the possessive is written as two 'parts' in the practised orthography, e.g.

tsa rona 'ours'

sa hao 'yours'

These parts are lemmatized disjunctly. Thus a person looking up the meaning of tsa rona 'ours' for example,

will have to look up the meaning of t_sa and r_ona which are given as

possessive concord 4 and 5 plural
(Mabille & Dieterlen 1983:558)

and

abs. pronoun 1st person plural,
'we', 'us' (op cit:437) respectively.

Contractions that may occur in possessives e.g. m_mae < m_me wa hae 'his/her mother' can also cause some difficulty in identifying lemmas.

6.2.1.4 The absolute pronoun

The basic structure of the absolute pronoun can be established by means of reconstruction. The structure is as follows: prefix (derived from the class prefix of the noun) + root + suffix -na. Noun classes 2-18 have -o- (phonetically [ɔ]) as root, e.g. bona 'they/them'. The second person singular and class 1 use -e- as root: wena 'you' (singular) and yena 'him/her', The 1st and 2nd persons plural have -o- (phoneticall [O]) as root; e.g. r_ona 'we', l_ona 'you' (plural). The suffix is -na for all classes.

The absolute pronoun is lemmatized as it appears in the practical orthography and there are practically no complexities as regards identification of lemmas, e.g. wena absolute pronoun 2nd person singular, 'thou'; (Mabille & Dieterlen 1983:596).

6.2.1.5 The demonstrative pronoun

The structure of the demonstrative is as follows: prefix (derived from the class prefix of the noun) + suffix. Diachronic analysis will reveal there is also a root. There are three positions for the demonstrative as well as two sets of demonstratives for each of the three positions, e.g.

Class le-

I	II	III
lena lee	leno leo	lane lela

As regards lemmatization of demonstratives the suffix forms the lemma, for example -na is an example of a lemma of the demonstrative of the first position, and -no and -la the second and third positions respectively.

6.2.1.6 The quantitative pronoun

Southern Sotho has one quantitative root: -ohle, meaning 'all'. The structure of the quantitative is as follows: concordial morpheme resembling the subject concord + root -ohle, e.g. ba- + -ohle > bohle.

In class 2 the vowel of the subject concord is deleted before the root. With classes 8 and 10 the sequence di- of the subject concord changes to the affricate ts, e.g. di- + -ohle > tsohle.

Consonantalization occurs with classes of which the subject concord is e or o, e.g. yohle and wohle (excluding the second person singular which does not have a quantitative).

As regards lemmatization all quantitatives appear under the lemma -ohle (i.e. the quantitative root, meaning 'all') as well as in their normal alphabetic position for the convenience of the dictionary-user. In the dictionary -ohle is defined thus:

-ohle, adj., all; joined to the rel. conc., forms a set of indefinite adj., as follows:

cl.	1	2	3	4	5	6
sing	-	ohle	lohle	sohle	yohle	bohle johle
plur.	bohle	yohle	ohle	tsohle	tsohle	ohle
		7				
sing.		hohle.				
plur.		-				

In addition to this definition, each quantitative is given in its alphabetic position, with a cross-reference. Hence bohle for instance is defined thus:

bohle, adj., all; ref. -ohle.

6.2.2 The role of suffixes in the nominal category

Certain changes occur when the suffix is affixed to the root or word, which affect the process of lemmatization. A dictionary-user has to have some knowledge of the effects of those nominal suffixes in order to know how a particular lemma can be identified.

6.2.2.1 Diminutive suffix

When the diminutive suffix -ana is attached to a nominal stem, the final consonant of the stem may

undergo a change, e.g. se- + -lepe '(an) axe' > se- + -letsw- + -ana '(a) little axe'. Incidentally, seletswana is also the diminutive of seledu 'chin', which may imply a lot of cross-reference as the attachment of the diminutive suffix may result in some modification in the stem. Furthermore, nominals or adjectives ending with -ana, may express derogation e.g. monnana (from monna) 'despicable man', or appraisal/endearment, e.g. e motlenyana 'particularly beautiful' (from -tle 'beautiful').

As regards lemmatization, an explanation is given in the introduction of the dictionary (Mabille & Dieterlen) that diminutives have been entered under the basic stem with a cross-reference given only when the stem undergoes some modification.

6.2.2.2 Feminine suffix

Nouns and adjectives may take the suffix -ana and -hadi and thereby express femininity, e.g. kgunong (used to refer to a female animal) 'brownish'.

Here again the identification of lemmas is made difficult by changes which may occur in the stem, e.g. kgwadi > kgwatsana (or kgwadipana) 'black with white spots'.

The nominal suffix -hadi (which may also express augmentation) does not cause any phonological changes in the word to which it is attached.

Just like in the case of diminutives, nouns which take the feminine and augmentative suffixes appear under the basic stem except where affixation results in modification to the stem, hence kgwadipana for example, is classified under the lemma kgwadi.

6.2.2.3 The locative suffix

The locative suffix -ng may be attached to nominal stems in locative formation. In practical orthography, no changes occur except when the nominal stem ends with the vowel a, in which case the resulting locative ends with eng, e.g.

<u>tela</u> 'path'	-	<u>tseleng</u>
<u>thaba</u> 'mountain'	-	<u>thabeng</u>
<u>thota</u> 'veld'	-	<u>thoteng</u>

Since locatives, i.e. tseleng, thabeng and thoteng are not lexical items it is necessary that the dictionary-user be made aware of the fact that to get the meaning of these words he must look for the lemmas from which they derive. Therefore the meaning of these locatives will be classified under the lemmas tela 'path', thaba 'mountain' and thota 'veld' respectively.

6.3 LEMMATIZATION IN THE VERBAL CATEGORY

This category includes all classes which can act as predicates of sentences. Conjugation in this category includes the diverse inflectional and derivational properties of Bantu languages.

6.3.1 The role of prefixes in the verbal category

Lemmatization may be affected by certain prefixes whose affixation to verb stems may bring about phonological changes.

6.3.1.1 The infinitive prefix (ho-)

This prefix is also the prefix of class 15. It can therefore be interpreted with reference to both the nominal category and the verbal category. When interpreted with reference to the nominal category it is a prefix to nouns which constitute class 15, and non-indicative of number. With reference to the verbal category it is an infinitive prefix as can be seen in negative and derivative forms

e.g. ho se bue 'not to talk'
(negative)

ho tshedisana 'to comfort one another'
(derivative)

In the verbal category this prefix is used in an indefinite construction. When interpreted in the nominal category it may be used in di-transitive implication, e.g. ba rata ho nwa jwala 'they like drinking beer'.

The Southern Sotho orthography writes the infinitive prefix apart from the stem. This implies that the attachment of the prefix to verb stems does not bring about a morphological change. Thus it does not present any complications for lemmatization purposes, e.g. ho bua 'to speak' would have to be looked up under -bua '(to) speak'.

6.3.1.2 The object concord of the first person singular (N-) ⁴

In the practical orthography this prefix appears as m

⁴ n in practical orthography may represent alveolar [n], prepalatal [ɲ] or velar [ŋ].

if the first letter of the verb stem is a bilabial, or n in all other cases, in which case it becomes partially assimilated to the first consonant of the verb stem. Consequently this concord can manifest itself in different ways which may be a bit confusing to the dictionary-user. This prefix also causes changes in the phonological structure of the verb stem, which may further complicate lemmatization,

e.g. n- + -bona > -mpona '(to) see me'
n- + -lemosa > -ntemosa '(to) notify me'.

One notices that in such cases, a reader coming across -mpona '(to) see me' for example, is expected to know that the meaning can be found under the lemma -bona '(to) see'.

6.3.1.3 The object concord of the third person singular (mo-)

This prefix is conjoined to verb stems starting with the consonant b

e.g. mo- + -bitsa > -mmitsa '(to) call him/her'
mo- + bona > -mmona '(to) see him/her'

The sound changes can perplex the dictionary-user, who has to realize that to get the meaning of -mmitsa 'call him/her' one should consult the entry -bitsa '(to) call'.

6.3.1.4 The reflexive prefix (i-)

The reflexive prefix can also bring about morphophonological changes in the verb stem to which it is attached. The changes are similar to those that are caused by the prefix N-, e.g. i- + -bolaya >

-ipolaya 'commit suicide' (lit: kill oneself).

An added difficulty can be seen in cases where the initial sound of the verb stem is a vowel or a semivowel, in which case the insertion of k is observed,

e.g. i- + -otisa > -ikotisa '(to) cause oneself to lose weight'

i- + -utlwa > -ikutlwa '(to) hear oneself'

Morphological implications can therefore bring about some difficulty in identifying such words, for example, in the following words the k is not inserted as a result of the attachment of the reflexive prefix

e.g. -ikvetlisa 'exercise oneself'

-ikudisa 'fake illness' or 'cause oneself to become ill'

Consequently, the meaning of -ikotisa can be found under the lemma -ota 'lose weight', -ikudisa under -kula '(to) become ill' and not *-ula.

The insertion of k in the case of the attachment of the object concord of the first person singular to verb stems starting with a vowel affects the identification of lemmas in a similar fashion.

Prinsloo (1992b:1) sub-groups the various occurrences of reflexives into four categories:

(1) no sound changes nor any semantic shift

ho ithusa 'to help oneself'

- (2) no sound changes but semantic shift

ho ingodisa 'to register'

(i- = 'self', -ngodisa = '(to) cause to write')

- (3) sound changes but no semantic shift

ho ithata 'to love oneself'

(i- = 'self', -rata = '(to) love')

and

- (4) sound changes as well as semantic shift

ho ikemela 'to stand alone', 'to be independent'

(i- = 'self', -emela = '(to) stand (for) by, wait for').

The effects of sound changes have been dealt with in detail in the chapter on Sandhi (Chapter 5). Considering what the reflexive prefix results in apart from these sound changes, i.e. semantic shifts, as illustrated in the above examples, we may expect some difficulty in lemmatizing verb stems in a consistent manner. Therefore great care in lexicographical planning should be exercised in lemmatizing reflexives. Prinsloo mentions that as a result, frequency of use should be the most important consideration in the compilation of say, a pocket dictionary. Inconsistency can result in cases where highly used reflexives have been entered but hidden in articles of head words where they are not likely to be looked for. As an example he points out an instance in Pukuntsu (1983 and previous editions), where -ikemišeditse 'intend' is 'hidden' in the lemma -ikemela 'stand for oneself' which he asserts, from a

lexicographical point of view, is incorrect.

Prinsloo warns that ignoring the phenomenon of semantic shift, lexicographers actually create the impression that a reflexive semantically always consists of 'self' + the meaning of the verb stem. For example: -ikudisa as shown earlier, can mean 'fake illness' and not necessarily 'cause oneself to become ill'.

Normally, the tradition in lexicographical works as far as lemmatization of reflexives is concerned, is based on the rule: simply remove the i- and look the word up under the following letter, but as Prinsloo points out, certain changes might occur as a result of the attachment of the reflexive prefix, consequently presenting a situation whereby an unfair amount of knowledge is expected from the target user (cf. how he outlines the attempts of a dictionary-user who tries to find the meaning of -ikemišeditše and concludes that the chances of his having missed -ikemišeditše 'intend' in the article is 90% (Prinsloo, 1992b:6). The prevailing editorial policy expects the learner to know that -ikemišeditše consists of

a reflexive vowel i-

a consonant k (cf. k- insertion rule)
(when the reflexive i- is used before a vowel
commencing verb stem)

the verbal root -em-

the causative extension -is-

the applicative extension -el-

the perfect extension -il-

(Furthermore he has to realise that -el- and -il- have undergone sound changes giving rise to -dits-) and finally the verb ending -e.

Prinsloo cites the following reasons why this policy is disadvantageous.

Firstly, it is unlikely that the target user of this or such a dictionary will be able to analyse this word by reversing the sound changes that occur when the different suffixes are combined, as has been done above.

Secondly, even if he is able to do this he then attempts to establish the meaning of the word and then concludes that:

i- 'self' + -em- 'stand' + -is- 'cause' + -el- 'for' + -il- 'past' means 'caused yourself to stand for', compared to the actual meaning of 'intended'.

Prinsloo thus rejects the approach of lemmatizing reflexives randomly. He emphasizes rather, that reflexives should be lemmatized in such a user friendly way that they can be directly look up under i- without prior reference to the mini-grammar or knowledge of difficult language rules (cf. Prinsloo, 1992b:14-15).

So as to be able to lemmatize those reflexives most likely to be looked up by the target user, he emphasizes that the lexicographer (in addition to relying somehow on his intuition) should employ word frequency studies in order to ascertain the 'correct'

corpus of reflexives for the specific type of dictionary to be compiled, e.g. pocket size or medium size. It is not just the advantage but the purpose of frequency studies, to

- (1) Select just the right corpus of words (reflexives) for a specific dictionary and
- (2) Prevent the omission of essential words (reflexives)

6.3.2 The role of suffixes in the verbal category

Verbal roots display different phonological shapes in the process of suffixation, which can complicate the lexicographer's endeavour to lemmatize verbs consistently. For example, Prinsloo (1994:5) mentions five standard modifications to the verb namely:

- (i) plus passive,
- (ii) plus perfect,
- (iii) plus passive and perfect,
- (iv) plus reflexive (discussed in the preceding paragraph)
- (v) plus reflexive and perfect.

Then there are deverbatives, as well as deverbatives on all five standard modifications in any class. Lastly, the lexicographer also has to consider, in the case of lemmatization of verbs, derivations for each single verb which is to be provided for in the dictionary.

6.3.2.1 The perfect morpheme

The perfect morpheme -ile, may display different forms depending on the morphological structure of the root.

For example, in monosyllabic verb stems, the perfect form consists of the consonant or consonant cluster plus -ele,

e.g. -shw- + -ile > -shwele 'dead'
-w- + -ile > -wele 'fallen'
-j- + -ile > -jele 'eaten'

Apart from the ending -ele, the perfect form of the verb may have the ending -ile, -itse, -ne, -me, (cf. Mabile & Dieterlen, 1983: XV).

But one realizes that for the dictionary-user who is not familiar with the intricacies of verbal inflection in Southern Sotho, it may not be easy to analyse (i.e. morphologically) any word in perfect form so as to arrive at its meaning in the dictionary. For example, the lemma for a stem like -robetse 'asleep' (stative form, active voice) can be difficult, worse so, if it is in the passive form, i.e. -robetswe 'asleep'.

If we compare -robetse 'asleep' with -kgathala 'tired' (perfect form -kgathetse), we find that according to Mabile & Dietelen, it is not analysable, i.e. it is lemmatized as -robala, unlike -kgathala which is lemmatized as -kgathala. One would expect that to alleviate this difficulty the perfect form -etse could be indicated in the introductory notes, nevertheless, one notes that the authors have chosen the option of treating all such forms (i.e. those taking -etse in the perfect form) as cases of morphological irregularity, hence writing them in their normal alphabetical position (not in boldface), thereby enabling the user to get the word's original underived form.

Perfect forms of verb stems involving derivatives may

cause some difficulty as a result of morphological irregularities that may occur. For example if we consider the causative, we realise that verbs having -ny- in the final position of the root take the form -ntse as their ending in the perfect form.

e.g. -kopanya > -kopantse '(to) mix'

-fapanya > -fapantse '(to) place across one another'

-minya > -mintse '(to) drain'

Assimilation can also influence the morphological structure of the perfect form of verbs, and consequently the identification of lemmas as can be seen in the following examples:

-sunne from -suna '(to) kiss'

-hanne from -hana '(to) refuse'

-omme from -oma '(to) dry'

-kgamme from -kgama '(to) strangle'.

Coming to lemmatization of perfect forms, we realise that the verb conjugation that my result from the affixation of the perfect morpheme is explained in the introductory matter of the dictionary. From here these perfect forms are explained in their normal alphabetical place under the lemmas from which they derive.

6.4 LEMMATIZATION OF CONJUNCTIONS, INTERJECTIONS AND IDEOPHONES

Languages have a class of words that are identified on formal grounds, and which stand apart because of some phonological, morphological or syntactic feature. Doke & Mofokeng (1985:335) define an ideophone as

a word, often onomatopoeic, which describes a predicative, qualificative or adverb in respect to manner, colour, sound, smell, action, state, intensity, etc.

Samarin (1971:133) criticises this definition on the grounds that

we are ... too sophisticated linguistically these days to let ourselves believe that a formal category in language is required because its terms 'describe' a 'predicate' with respect, for example to 'smell'. Knowing something about African languages, I understand what is meant, but this is not sufficient to justify a grammar. I would therefore suggest that Bantu languages first be described in their own terms and then compared to other languages.

The word ideophone is self-explanatory:

idea + phone: 'sounding by itself'
(op cit: 153)

A summary of the characteristics of the ideophone has been given by Samarin (1971:154) as follows:

Ideophones

- represent or express actions or ideas
- echo, express, mirror sense, impressions or perceptions
- express or reflect emotions and feelings
- complete thought or create images
- vivify speech

Interjections, also, are used to express emotions and for the purpose of calling attention or giving command. (cf. Doke & Mofokeng 1985:365)

While ideophones virtually describe a predicate, qualificative or adverb, interjectives may of themselves, constitute complete 'sentences'.

Ideophones and interjections are both exclamatory in nature. They are gestures of speech.

With them one is in a special realm of spoken art. There is a roundness, a complete shape, not so vividly conveyed by more complex constructions, more formal expressions. They attempt to be vivid re-presentations or recreation of an event in sound ... One notices that ... in a discussion about their nature ... they bring a sense of ease and mirth ... Always they try to capture the freshness of an event and express it of themselves with nothing to dull or cloud the evocation'. (Fortune, 1962:6 in Samarin, 1971:153)

Thus ideophones and interjections give a picture of the situation.

Coming to consider lemmatization of this category. Madan (1911:40) in Samarin (1971:156) observes that ideophones get scanty treatment in dictionaries and that people have confessed difficulties in defining them. This is caused by the fact that the semantics of the word category cannot be defined in a more specific sense like the other word categories. The sounds convey a meaning quite sufficiently definite for the speaker's purpose, being defined by aspects which cannot be outrightly covered by the field of semantics, such as:

(1) his (the speakers')

- tone
- gesture
- look

(2) the context of the communication

(3) the hearer's sympathetic attention

(4) the circumstances generally with spoken language.

Hence the vagueness of meaning that has been attributed to ideophones. Their nature precludes any hard and fast meanings being attached to them (op cit:157).

Ideophones, interjections and conjunctions also enjoy lemmatic status. For example an interjection is given as follows:

jo, inter. expressing sorrow: alas!

To a great extent ideophones are radicals. Ideophones are neither transitive or intransitive. Furthermore,

derivational tendencies of these categories raise lexicographical concern as far as lemmatization is concerned.

Nouns and verbs may be derived from ideophones:

NOUNS:

tseke > letseke -tseke 'glittering'

tlwepe > letlwepe 'swollen neck of the rinkhals snake'

VERBS:

As far as de-ideophonic verbs are concerned the general rule is to suffix -ha for the intransitive, -la or -fala (in ideophones which indicate colour) for the transitive and -tsa for the causative (cf. Doke and Mofokeng, 1985:338)

e.g. tlere > tlerefala 'to become red'

thakgo > thakgoha 'to get up quickly'

bure > buretsa 'to crush instantly as with a blow'

There are certain methods which can be evidenced with some consistence, e.g. the use of -la as in -bohlela 'to diminish'; honye > honyela 'to shrink.

Doke & Mofokeng (1985:338-339) illustrate certain cases which show how ideophones, on their part, can be developed from other word categories, e.g. verbs, ... etc.

the use of -i

e.g. kgathadi 'of tiredness' from -kgathala 'be tired'

Furthermore, a type of reduplication is common with the use of ideophones,

e.g. qoqomuqomu 'of throwing stones into water'.

When looking at the lemmatization of ideophones one cannot fail to recognise some inconsistencies as far as their derivational tendencies are concerned. For example in Mabile & Dieterlen (1983) some deideophonic verbs are lemmatic

e.g. -buretsa, some are not, e.g. -bohletsa.

Conjunctions link sentences to each other. They are usually expressed by means of any part of speech that functions as such. Thus they are not bound up with form. There are extremely rare cases, called primitive conjunctions (cf. Doke & Mofokeng, 1985:351) which are radical e.g. -ha 'if', athe 'and yet' etc.

Conjunctions are lemmatized in different ways according to their types. For example, one notices that primitive conjunctions are lemmas (printed in bold-type)

e.g. athe, conj. whereas; athebare, athebo, athenyane, corr. athe (The abbreviation 'corr'. refers to 'correct to'. This means that athe is the form considered to be more correct than athenyane.)

One notices, however, that certain problems can be encountered as far as the semantics is concerned. For example one finds hore 'that' under the lemma -re 'to

think' and hobane 'because' as a sub-entry under the lemma (mo hobahoba) with a cross-reference to -ba 'to be'.

Interjections have been defined as

An isolated word or group of words of an exclamatory nature, used to express various emotions and for the purpose of calling attention, giving command, or conveying assent or dissent (Doke & Mofokeng, 1985:365)

e.g. ashe, interjection 'pshow' (expressing impatience)

hm, interjection 'expressing doubt'.

Interjections have no grammatical or concordial bearing on the rest of the sentence to which they may be attached. Thus their lemmatization does not create many lexicographical problems like the other word categories.

6.5 CURRENT STRATEGIES FOR THE LEMMATIZATION OF NOUNS AND VERBS AND POSSIBLE WAYS OF LEMMATIZING NOUNS AND VERBS

Lexicographical works that have been done so far (i.e. in Southern Sotho) involve bilingual and trilingual dictionaries. There are no monolingual dictionaries. Some of these works are basically word lists with translation equivalents into English and/or Afrikaans, with no profound treatment of derivational issues. Mabile and Dieterlen's work show a great deal of effort in trying to present derivational issues. The work by Hamel (1965) needs some updating. Lexical items are not sufficiently explained taking into

account the fact that the work is divided into six volumes:

- the first volume treats items beginning with A,B and C
- the second volume treats items beginning with D, E and F
- the third volume treats items beginning with G, H,I,J,K and L
- the fourth volume treats items beginning with M,N,O and P
- the fifth volume treats items beginning with Q,R and S.
- the sixth volume treats items beginning with T,U,V,W,X,Y and Z.

Learner's dictionaries include the work by Du Plessis et. al. which also contain a list of idiomatic expressions to help the user, and the **Learner's English - S/Sotho Dictionary** which is designed for Primary pupils, and contains items illustrated in colour. These efforts help the learner to understand better. These dictionaries lemmatize verbs according to the first letter and arrange them alphabetically. Derivational issues are not fully covered. In the case of nouns the prefix is left out in the process of lemmatization and the noun class system of Southern Sotho is explained in the introductory notes.

Derivation is a very important morphological process in Southern Sotho. Because of overlooking problems that may arise as regards ignoring derivational tendencies in the nominal and verbal categories, the dictionary-user may be confused. In some cases for example, the lexicographer may choose to deviate from a certain pattern of lemmatization without giving

lexicographically justifiable reasons. For example in the introduction of the Buka-Ntswa ya maleme-pedi Sesotho-Seafrikanse, it is mentioned that in the case of the passive and the reflexive the derivations are given in 'exceptional cases'. It is not clear which cases are being referred to.

Here are some of the problems that can be seen in some of the existing works:

(1). Inconsistency

Strategies are not employed consistently.

For example, one expects to find mosehli 'one who cuts' the sub-lemma of which -sehla '(to) cut' is the lemma because just on the previous page mosebi 'one who backbites' has been given as a sub-lemma of -seba '(to) backbite', in Mabile and Dieterlen (1974). Thus the strategy to lemmatize personal nouns (i.e. in this case) is not consistently applied. Inconsistencies can also be evidenced in many cases which involve both the nominal and the verbal derivation - the typographical presentation of these items is also not consistent.

(2). Unsystematic presentation

Strategies are not applied systematically.

If, for example in the case of multi-word lexical units, the meaning of the unit is given, it would only be systematic to also provide the meaning of the constituent units in their normal alphabetic positions, (unless if they are not elements of the lexical stock?). We may also have the meaning of the multi-word lexical unit provided for under

each constituent unit. If it is polysemous (or homonymous) the meaning of the multi-word should then appear under the sense closest to the sense applicable in the multi-word lexical unit.

(3) Notes

Notes can either be neglected or not understood by the dictionary-user, or he may feel he does not have time to go through the notes to study the case involving the specific lemma he is faced with at that time (i.e. if he is fortunate enough to realise that the case he is faced with involves a class prefix, a reflexive prefix or whatever the case may be). Although notes can be helpful, they should not be over-used to the extent of rendering a user-unfriendly product.

Although by definition the dictionary is meant for the 'qualified' reader, his knowledge should not be overestimated. As already mentioned, lexicographers provide the information that users need. It would be a sad case to provide the information that users do not know how or where to find.

(4) Repetition

Because certain items become hidden in long articles the lexicographer lemmatizes a word which has been fully explained elsewhere, again in its alphabetical place, with a cross-reference. In certain cases even notes are repeated. For example, this information

words beginning with i- which do not appear here will be found by consulting the letter

follow i-...

appears four times (i.e. on all pages containing words beginning with i-) in Mabelle and Dieterlen's work (1974). This strategy costs heavily on the volume of the work.

The techniques of cross-reference and notes have been so extensively used but one feels they need a more rigorous in-depth research that will result in a more efficient lexicographical strategy.

(5) Words that are not commonly used

Existing works contain a number of words that are not commonly used.

In some cases the sense of the word has changed. Certain categories and signs in Mabelle and Dieterlen's work, also leave one with the feeling that they need some changing e.g. the category 'correct to' as in -phadima corr. phatsima '(to) shine'.

Why not lemmatize -phatsima and leave -phadima? Using the sign '?' also, (meaning uncertainty of derivation) is not lexicographically accountable.

In order that such problems be avoided lexicographers should strategize lemmatization basing their ideas on the nature of Bantu languages themselves and not take a methodology that works for certain languages and then implement it in lemmatizing word categories in Bantu languages. The nature of the language should lead to the strategy, and not vice versa.

Lemmatizing words according to the first letter could solve a lot of problems. The lemma should then be followed by all the necessary information that should be supplied according to the type of the dictionary. User-friendliness should also receive high priority. Words should also be lemmatized according to practical needs. This means that the dictionary should contain words most likely to be looked for.

Words should also be lemmatized taking into account their frequency of use. Frequency of use is helpful in dealing with more productive combinations. The problem of words that are not commonly used, for example, could easily be solved by using this strategy. It means the selection is based on actual word usage. In this way, in designing say a pocket dictionary, essential words will be included, and non-essential words will be left out. Prinsloo (1994) is of the idea that the category of selection has proved to be the major pitfall concerning works done on Northern Sotho. This can also be said in the case of Southern Sotho.

Current strategies therefore, do not lemmatize verbs and nouns satisfactorily. These works underestimate the complexity of lemmatizing these categories. At the core of the matter, the lexicographer finds himself faced with the task of reducing the number of derivations, considering limitations on the volume of his dictionary. Prinsloo (1994:96) explains that there can exist 2000 potential entries for a single verb. The lexicographer should therefore apply some strategy for reduction. Making provision for these items should be based on a carefully worked out lexicographical plan.

CHAPTER SEVEN

DERIVATION AND INFLECTION

7.1 INTRODUCTION

Derivation and inflection are two types of morphological operations which may result in a change in word form. Inflection is the type of morphological process which results in a change within one and the same word class or lexical category but leaves the word's lexical meaning unchanged. In derivation a change in the form of a word implies a change in the lexical meaning itself.

Dictionaries do not list all words which are formed as a result of grammatical inflection, because of their formal regularity. Only cases which are not so uniform are listed. As far as derivation is concerned, the lexicographer studies these cases with even greater care, since he is concerned primarily with lexical meaning. Thus word variation is of two kinds: either the lexical meaning remains generally unchanged (grammatical inflection) or it is changed in a more or less regular pattern (derivation, word-formation), cf. Zgusta (1971:131).

In practice it can be observed that on the whole nests⁵ containing derivations tend to be more conflated than those which deal with different composed words. The fact that members of a nest really have the same lexical meaning at the level chosen by the dictionary, may require a lot of systematization on the part of

⁵ A nest is a group of entries which is conflated into one (Zgusta, 1971:284).

the lexicographer. He is also faced with the question of consistency in the treatment of lexical entries.

Lexicography in Bantu languages is faced with a lot of complications in the treatment of lexical entries as a result of inflectional and derivational affixes.

The central problem is particularly the method of the nominal and verbal items of the language, emanating from the complex morphological structure common to Bantu languages, of a nominal classification system categorising nouns by means of prefixes and a verbal derivational system forming new verbs by means of derivational affixes (James, 1989:5)

Because of the complex morphological structure of lexical items, the lexicographer has to decide which lexical forms should be listed in the dictionary and in what manner. The difficulty concerns the derivational morphology in which a very productive system is envisaged, as well as the selection and arrangement of these derived items.

7.2 DEVERBATIVES

We shall briefly look at nouns of a verbal origin, namely deverbatives. They are regularly derived words produced by the same word -formational means and are regular both as far as their form and as far as their meaning go. They may be formed from simple or derivative verb stems. They are formed by prefixing the class prefix to the stem, and changing the final vowel of the verb stem to -i in the case of personal nouns. Deverbatives can be derived from both simple verb stems as well as derivative verb stems, as indicated in the following sections:

Simple verb stems

e.g. moruti 'teacher; minister' is a personal deverbative formed from the verb stem -ruta 'teach'. The noun thuto 'teaching', which is an impersonal deverbative, is also classified under the same stem.

Derivative verb stems

Derivative verb stems are those which have been extended by means of a particular verbal extension. From the verb stem -ghanolla '(to) off-saddle' for example, the noun moqhanolli 'one who off-saddles' can be formed.

Deverbatives are to be classified under the verb stems from which they derive. Furthermore, nouns which have been formed from derivative verb stems are classified under the derived (extended) verb stems from which they derive. Hence moqhanolli 'one who off-saddles' derived from -qhana '(to) saddle' is classified under -ghanolla '(to) off-saddle', moqhatsetsi 'one who pours for' derived from -qhatsa 'to pour' is classified under -qhatsetsa '(to) pour for', and molwanedi 'one who fights for' derived from -lwana '(to) fight' is classified under -lwanela '(to) fight for'.

Deverbatives can, as a result of morphological affixation, attain great morphological complexity. An added difficulty is the variety of phonological changes which the base forms may undergo.

Deverbatives may also include an object concord or the reflexive prefix. In the case of an object concord it is usually the object concord of the first person singular that is used.

Object concord

e.g. Moreshwedi 'He who died for us'
Montsamaisi 'He who leads me'

These nouns are classified under the verb forms from which they derive, i.e. -shwela 'die for' and -tsamaisa 'lead (lit. cause to walk)' respectively.

Reflexive prefix

e.g. moipolai 'one who commits suicide'
seithati 'egotist'
moitwanedi 'one who fights for himself'

These nouns are classified under -bolaya '(to) kill', -rata '(to) love' and -lwanela 'fight for' respectively.

These morphological changes require careful lexicographical planning, as can be deduced from the foregoing illustrative examples. Otherwise it can be only a dictionary-user with a superb and sophisticated knowledge who will be able to find and understand the meaning of deverbatives in a dictionary. Zgusta (1971:273) remarks that this procedure (i.e. classifying derived forms under basic forms) entails such a quantity of cross-reference that some lexicographers prefer to treat the basic and the derived words separately, even if a great amount of repetition is involved. Another important point to be made in this case is that the lexicographer must check the regularity of these words together with their semantic identity. (This aspect will be looked into in the next section when semantic shift will be considered).

Zgusta (1971:273) is of the opinion that in large dictionaries separate treatment of all entry-words should be preferred.

Since Bantu languages are rich in regular morphology in general and word-formation in particular, the reduction of entries by means of classifying derived forms under basic forms can be beneficial in connection with space. As a matter of fact, in small dictionaries, as Zgusta (1971:272) points out, regularly derived words can be left out if the dictionary contains an appendix on derivation and word-formation. (Prinsloo, 1994:15) is of the opinion that frequency tests can be helpful in this regard.

7.3 VERBAL EXTENSIONS

In this section we shall look at how verbal extensions are considered in lexicographical planning.

Prinsloo (1994:14) outlines some Modules for Lemmatization of verbs, which show how verbal extensions can affect lemmatization. In Module I he outlines 90 possibilities of lemmatizing verbs. In Module II, 90 possibilities as well. In Module III to XX there are about 1 800 possibilities of root and suffixes for one single verb.

Considering these possibilities one realises the immense task lexicographers are faced with in lemmatizing verbs. For all possibilities they must be able to choose the most suitable word form as the lemma, having full details to account for the choice. This requires a systematic presentation of derivational affixes.

James (1989:15) urges that

any attempt at a description of affixational morphology should consider derivational affixes as central to the process of forming new words.

Considering these verbal extensions (op. cit:16,, James suggests that it may be fruitful under each verb to list all those suffixes which can be combined approximately with it to form new (deverbal) verbs, for example -is-; -el-; -an- etc. This would instantly indicate a possible lexical derivative and a user would thus be offered the opportunity to understand the combinational characteristics of such suffixes for a given base form. In cases where the lexical derivative form can act as the base for further derivation he, (i.e. James), proposes that lexical derivatives be entered separately and the suffix which functions as the head be clearly indicated typographically.

Alternative forms may exist, e.g. in the passive: - nyorilwe / -nyoruwe 'become thirsty'; (mo) holehwa / (mo) holehuwa 'one who is taken captive'.

The redundancy of such alternative forms, according to Fromkin and Rodman (1983:118) explains accidental gaps in the lexicon. Thus it is clear that the language constitutes only a subset of the words.

The following examples illustrate this fact:

apo! < -apoha 'to come apart' (of clouds).

In this case the ideophone apo! comes from the stem -apoha '(to) come apart'. But in an example like * nyolo > -nyoloha 'go up', although -nyoloha 'go up'

is a lexical item nyolo is not. The reversive verbal extension can also illustrate this very well. The verbal extension (-oll-) signifies reversive acts. The following examples show the possibility of -oll- having fossilised to a root, thereby losing its reversive signification.

-nyorolla 'quench' < * -nyora
-lokolla 'unite' < * -loka
-apolla 'to cut open' < * -apa

Some run-on derivatives may never have been used. Words that could exist but for which no record exists to prove that they have ever been used are called latent words (cf. Landau, 1989:78). This fact is clearly evidenced in the area of verbal extensions.

7.4 IRREGULARITY AND VARIANTS

Derivations can result in very rich nests. Zgusta (1971:288) mentions two difficulties in connection with these nests. Firstly, such nests can disagree rather considerably with the alphabetical order. Secondly, difficulty is caused by words of which the morphemic status is unclear.

Dictionaries rely on different methods in order to solve these problems. Some expect a certain degree of analytic ability on the part of the user and thus do not expand on the irregular item. Some go on with the nest up to the point where another word which does not belong to it, must be entered, according to the alphabetical order, and then combine with the nest, (op cit:288). In the case where a word appears in the nest to which it does not belong (i.e. morphologically), that word will be placed again in

its alphabetically correct place, with a cross-reference to the nest (e.g. -entse 'done').

Thus, the way in which morphological irregularity may be handled is by entering into the dictionary as separate lemmas all the irregular forms with a cross-reference to the basic form (which is -etsa '(to) do' in the case of the given example.

Irregularity can also be caused by semantic variation which occurs during affixation. If we take the word motshedisi 'comforter' for example, we realise that if we consider it a word derived from tshela '(to) cross', we can reason it out as a polyseme, i.e. also meaning 'one who causes to cross' because, after all, a comforter is someone who causes another to 'cross over' from an undesirable state to a more desirable one. But according to Mabile & Dieterlen (1983) motshedisi is not considered a polyseme, i.e. derived from tshela. It is considered, rather, a derivative of tshedisa 'cause to cross', and hence lemmatized as such.

The fact that morphemes are sensitive to environmental factors, also leads to variations in the morpheme form. In the process of morphemic analysis (the operation by which the analyst isolates minimum meaningful elements in the utterances of a language and decides which occurrences of which elements shall be regarded as occurrences of the same element, cf. Hockett (1947:229) in Ohlander (1976:93)) the same morpheme can be realised as different forms in different environments. Ohlander (1976:98) remarks that it is this alternation between phonetically different forms of the same morpheme which makes phonology

"rather a complex affair".

In Mabilille & Dieterlen's work (1983) one notices, as a result (i.e. certain forms being realised in different phonological forms) that certain forms, such as perfect forms of most verbs and diminutives of some nouns have been lemmatized in their respective alphabetic places far from the nest, with a cross-reference.

Derivation and inflection therefore have a lot of implications for both the compilers of the dictionary and its users.

Recognition of affixational morphology as a property of the lexicon and its improved presentation in the dictionary could help a (user-) learner enhance not only his/her retrieval techniques for particular portions of the vocabulary of the language but also the ability to produce well-formed lexical items appropriately (James, 1989:16).

CHAPTER EIGHT

LEXICOGRAPHY AND THE COMPUTER

8.1 INTRODUCTION

In the past few years computers have been used in the various phases of lexicographical work. They simplify the recording, description and presentation of data for the purpose of dictionary production. The field of lexicography continues to grow as technical development advances by the day.

For example, up to now lexicographical works have been presented in the form of printed books. The written word has predominantly been the medium of expression, with occasional usage of pictures incorporated, as in pictorial dictionaries. But in the future, a combination of writing, pictures and sound is conceivable, whereby one will listen to the pronunciation of the word enquired about.

8.2 MAJOR STAGES IN DICTIONARY-MAKING AND THE CONTRIBUTION OF THE COMPUTER

Dictionary-making involves five basic steps:

- gathering of data
- parsing and excerpting of entries
- filing of entries according to a certain arrangement
- writing of articles
- publication of the final product

This involves a lot of work and time on the part of

the lexicographical team. Computers simplify the task in a considerable manner as will be seen in the following discussion. The phases in which computers have been found to be of considerable use can be outlined as follows:

- Collecting and transferring of data from its original form to another, including citation retrieval (RECORDING)
- Sorting and alphabetizing, including lemmatization and homograph separation (DESCRIPTION)
- Educational and analysis procedures including definition, writing and the composition of entries (PRESENTATION)

Human intervention is still required to some extent, for instance in lemmatization and homograph separation.

We shall now look into the ways in which the computer can play a role in the major stages of dictionary-making. As can be gathered from the above explanation, three major functions of the computer are distinguished:

- collection of data on which the dictionary will be based:
- preparation of the entries, including the choice of canonical forms, the writing of pronunciations, usage notes, definitions and selection of illustrative examples;
- production of the finished work.

In order to highlight the computer, the discussion will be presented in the form of a comparison.

(1) Establishing a file of usage

Collecting citations for a dictionary has traditionally been carried out by trained readers, a group of house readers or volunteers. Because it is not easy to get a group of willing readers, the possibility of using a computer to take over the laborious work of establishing a file of usage comes to mind whenever new dictionary projects have been initiated. Excerpting tests as well as parsing can be very tiring for a human being as it is a tedious and repetitive job, but it is the sort of thing that a computer does well. Human intervention cannot, however, be completely done away with. The task of decision-making is still left to human judgement. For example, special usages is one of the lexicographer's main points of concern, and yet, as Scholler & Reidy (1973:41) rightly point out, the computer cannot approach the skill of even the most witless volunteer in collecting them unless some sort of oblique strategy is employed in searching for them.

Because of the high standards of programming that research has yielded, the following can be provided on a computer slip:

- context for the selected word
- full bibliographical information
- a note on the key-punching conventions for

special characters.

Scholler & Reidy (1973:44) foresee a great and important advantage of such a computer system:

More extensive use of this computer system will help build a file that will yield valuable information for the editing process without making the collection too enormous for its intended use.

(2) Easing the editor's burden

It is evident that the most interesting area in computational linguistics focuses on the interaction between man and machine. The computer is made to perform the task which it can do best. After the texts have been recorded in machine readable form, the computer can very well arrange lexicographical raw data, making it tidy and hence easier to work with.

(3) Producing the finished copy

After the editorial process the computer has a role in the production of the finished dictionary through the operation of computer-driven typesetting machines. Corrections and adjustments can thus easily be dealt with since the computer at this stage, offers the possibilities of preparing the entry material for storage in and retrieval from a database.

8.3 THE MICROSTRUCTURE OF A DICTIONARY DATABASE

The microstructure of a dictionary database is as follows:

(NB. For a detailed discussion on the structure of an entry cf. Chapter 3).

- K1 Spelling of the lexeme
- K2 Phonology, phonetic transcription, information on syllabification and stress
- K3 Lemma: information on the basic form found in each form on an inflectional paradigm
- K4 Word class: classification of word types defined in syntactic, morphological or semantic terms
- K5 Information on the morphology of inflections: details on the declension of substantive elements
- K6 Derivation data: details on the word-formation process using derivational morphemes
- K7 Syntactic environment: surface structure
- K8 Deep case: the required deep case frame
- K9 Semantic information: explanation of meaning (i.e. semantic definition) context, examples of use, marking of specialized fields (cf. DANLEX - Group 1987:7).

8.4 **ADVANTAGES OF THE COMPUTER IN LEXICOGRAPHICAL WORK**

(cf. James, 1989:87 & DANLEX - Group 1987:2)

- economy is an important benefit to be anticipated from computer production
- a unique description of the contents of the dictionary that can form a basis for comparative work
- the compilation of dictionary-entries with as much consistency and as little wasted time as possible
- the possibility of wholly, or partially automatic changes of lexical data among different dictionaries
- textual accuracy can be raised considerably by designing routines to check the editor's finished product. An added advantage if the final stages of the dictionary are mechanised is that it would not be difficult to produce abridgements of various sorts to fulfil student demands. The dictionary which is computer encoded also allows regular revision and alteration as new evidence comes to light
- the same dictionary can be presented in various ways in printed or other media as by-products of lexicography, such as the tape or disc containing the dictionary
- the potential to produce dictionaries with different types of content on the basis of the

same dictionary data.

8.5 A COMPARISON OF THE DIFFERENT CHARACTERISTICS OF A PAPER DICTIONARY AND A COMPUTERIZED DICTIONARY

A paper dictionary is static, reflecting a state of linguistic affairs of the period preceding its publication. A dictionary held in dynamic form (computer) is kept abreast of current language. Alterations can be done to any database. In this way there is no need to wait for accumulated corrections, additions and other changes to be sufficient to justify a new edition.

The cost of a printed dictionary has to be met immediately on acquisition and is fixed regardless of the level of use. The production involves teams of lexicographers in years of work, which leads to high prices. New editions imply new efforts by a large team and since they occur at irregular intervals it may be impossible for the original group to work on them. On the other hand, once the creation phase of a computer dictionary is over, the maintenance effort is small and steady.

While a computer dictionary points new routes to the data stored in it, a printed dictionary's potential routes are limited to the frontiers of what is contained in the dictionary combined with possible manipulations or instructions of these items of data (James, 1989:88). In a dynamic dictionary it is possible to gain access to any entry by means of the pieces of information composing it. A printed dictionary has a single method of gaining access, usually an alphabetic route.

Since, in the case of a computer the information is presented in a far more variable way than in a book, it can be of great value in languages with considerable morphological marking and numerous irregular forms. Complexities which are brought about by derivational affixation in a normal printed dictionary can be effectively dealt with in the dynamic form, through including varying forms at least as cross-references to the normal lemma (James, 1989:90)

A computer dictionary can, unlike the printed dictionary, effectively serve the needs of any individual user.

In view of the much greater wealth of information stored in a machine dictionary, it becomes possible to think in terms of a personal dictionary. A printed work must try to be all things to all users, and must therefore fail to be perfect for any one user. Since a machine dictionary will be in a database on a large mainframe computer, it will be accessible to authorised users by means of the quotation of some sort of identifier. This password can then automatically be used to ensure that each user's preferences are observed precisely (James, 1989:92).

8.6 THE POTENTIAL FUTURE OF LEXI-COMPUTING

The printed dictionary will not disappear overnight, nor perhaps ever, but the advantages of the machine dictionary, particularly an autonomous database, will probably be more favourable than words fixed forever on paper.

Burchfield (1987:55) points out that although computers are now indispensable in dictionary-making, they cannot replace the ultimate judgement and decisions of the (human) lexicographer.

The role of the computer, as the above discussion shows, is an important one in lexicography. The potential evolution of dictionaries in the future is a great one, as the effects of the computer revolution which has changed so many features of human society become more and more directly evident in the field of lexicography. Recent work reveals that effective techniques can be employed to make consultation easier.

In the near future, as research progresses, words could be clarified by the presence of sound.

Computers already have peripherals which can play tunes and make noises in a satisfactory way, and in the last few years digitised sound storage and reproduction have moved forward by leaps and bounds, so that personal microcomputers and even pocket calculators are now regularly able to realise the text which appears on their screens or is input through their keyboards.

This aspect, too, is certain in the near future to be pushed much further than the point it has currently reached, suggesting that it will soon be possible to demonstrate pronunciation directly, rather than relying wholly upon a phonetic transcription. Some entries, particularly onomatopoeic words, moreover, could well be classified by the presence of "sound effects". (James, 1989:91).

CHAPTER NINE

CONCLUSION

The understanding of the morphological structure of a word is important in identifying the lemma in a language such as Southern Sotho, since words achieve great morphological complexity because of derivational and inflectional affixes. Bennet (1986:4) points out that the option of arranging the lexicon by stem or root usefully groups related items and saves on cross-referencing but, in such a system the user must be able to identify the stem, which, given the sometimes complex morphophonemics of Bantu languages, may not be easy.

In general it may be asserted that words which can be derived by means of a formula may be removed from the dictionary, but because of the interrelationship between words as well as their close relatedness it may be difficult to decide at what point the degree of relatedness between the source item and the derived item is too remote as to enter them under the same lemma.

One also expects a certain level of uniformity regarding how morphological changes are accounted for in lemmatization, i.e. if montsamaisi 'one who leads me' appears under -tsamaisa 'lead (lit. cause to walk)' one would also expect to find the meaning of momphephetsi 'my challenger' under -phephetsa 'challenge'. Otherwise it is most likely that one will think the latter is not a lexical item in the language if it does not appear under the appropriate lemma.

We thus conclude that the question as to the proper arrangement of the Bantu lexicon is not yet settled. The complex and productive derivational system which is characteristic of Bantu languages, poses difficulties when coupled with the Western European habit of strict alphabetic arrangement (cf. Bennett, 1986:3). Much still has to be done in the field of lemmatization.

BIBLIOGRAPHY

- Al-Kasimi, A.M. 1977. Linguistic and Bilingual Dictionaries, Leiden: E.J. Brill.
- Alberts, Mariëtta & Mtintsilana, Priscilla 1988. Leksikografiese ontleding van Afrikaans en Afrikatale leksikons. Verslag LEXI 1-2. Pretoria: R G N
- Bennet, P.R. 1986 'Grammar in the lexocon: two Bantu cases': Journal of African Languages and Linguistics, 8:1-30.
- Bolinger, Dwight Le Merton 1965. Forms of English: accent morpheme, order. Cambridge: Harvard University Press.
- Burchfield, R.W. (ed) 1987. Studies in lexicography. Oxford: Clarendon Press.
- Burton, T.L. & Burton, J. (eds) 1988. Lexicographical and Linguistic studies.
- Capell, A. 1951. 'Bantu and North Australian: A study in agglutination', African Studies. Vol. 10. No. 2.
- Cassidy, Fred. 1982. Lemmatization: the case of Catalpa. In Language form and linguistic variation: papers dedicated to Angus McEntosh edited by John Anderson. Amsterdam: Benjamins.
- Congleton, J.E. (ed) 1979. Papers on lexicography in honour of Warren N. Cordell. Terre Haute: Indiana State University.
- Danlex Group 1987. Descriptive Tools for Electronic Processing of Dictionary Data, Lexicographica Series Maior 20. Tübingen Max Niemeyer Verlag.

- Doke, C.M. 1950. 'Bantu Languages: Inflexional - a tendency towards agglutination', African Studies Vol 9 No. 1.
- Doke, C.M. & Mofokeng, S.M. 1985. Textbook of Southern Sotho Grammar. Cape Town: Longman, Green & Co.
- Du Plessis, J.A. et,al. Bukantswe ya Maleme-Pedi Sesotho - Seafrikanse. Cape Town: Via Afrika.
- Fromkin, V. & Rodman, R. 1983. An Introduction to Language. New York: Holt, Rinehart & Winston.
- Gelb, I.Z. Lexicography, lexicology and the Accadian dictionary in Miscelenea; Homenaje a Andre Montinef E structuralisma E Histona p.4 and Cornyn WS comments in householder and saporta, p.274. In: Al-Kasimi, A M 1977: 17-18.
- Hamel, L.O.M.I. 1965. English Southern Sesotho Dictionary Catholic Centre, Mazenod-Lesotho.
- Harris, R. 1973. Synonymy and Linguistic Analysis. Oxford: Basil Blackwell.
- Hartmann, R.R.K. (ed) 1983. Lexicography: Principles and Practice London: Academic Press.
- Hill, A.A. (ed) 1969. Lexicology and Semantics (Sydney Lamb). In Linguistic Voice of America Forum Lectures.
- Hockett, C.F. 1969. A course in Modern Linguistics. USA: MacMillan Company.
- Hulbert, J.R. 1955. Dictionaries: British and American. London: Andre Deutsch Limited.

- Ilson, Robert (ed) 1985. Dictionaries, Lexicography and Language learning. Oxford: Pergamon Press.
- Jackendoff, R.S. 1975. 'Morphological and Semantic Regularities in the lexicon' Languages Vol 51 pg 639-671.
- James, G. (ed) 1989. Lexicographers and their works. University of Exeter.
- Kock, L.J. et al. 1984. Study Guide 3 for SSA 100-B Pretoria: University of South Africa.
- Kosch, I.M. & Louw, J.A. 1987. The only Study Guide for MORPOO - G Pretoria: University of South Africa.
- Landau, S.I. 1989. Dictionaries, the art and craft of lexicography. Cambridge: Cambridge University Press.
- Learner's English-S/Sotho Dictionary 1985. Alberton (RSA): Librarius.
- Louw, J.A. 1984. 'Word categories in Southern Bantu', African Studies Vol. 43, No. 2: 231-239.
- Lyons, J. 1979. Introduction to Theoretical Linguistics. Cambridge University Press.
- Lyons, J. 1983. Semantics 2. Cambridge University Press.
- Mabille, A. & Dieterlen, A. 1974. Southern Sotho - English Dictionary. Morija Sesuto Book Depot, Morija-Lesotho.
- Mabille, A. & Dieterlen, A. 1979. Southern Sotho English Dictionary. Morija Sesuto Book Depot, Morija-Lesotho.
- Mabille, A. & Dieterlen, A. 1983. Southern Sotho-English Dictionary. Morija Sesuto Book Depot, Morija-Lesotho.

- McGregor, C. 1985. From first idea to finished artefact: The general editor as chief engineer. In: Ilson, R. 1985: 123-132.
- Minnis, N. (ed). 1971. Linguistics At Large. (The Fourteen Linguistic lectures presented by the Institute of Contemporary Arts London 1969-70). London: Victor Gollanez Ltd.
- Moeketsi, R.H. 1991. The Concise Trilingual Pocket Dictionary English Southern-Sotho Afrikaans. Parklands: Ad Donker Publisher.
- Mtintsilana, P.N. 1990. 'Polysemy, homonymy and hyponymy in Xhosa bilingual dictionaries', South African Journal of African Languages, Vol. 10 No 2.
- Ohlander, S. 1976. Phonology, meaning, morphology: on the role of semantic and morphological criteria in phonological analysis. Goteborg: Acta Universitatis Gothoburgensis.
- Prinsloo, D.J. 1991. 'Towards computer-assisted word frequency in Northern Sotho', South African Journal of African Languages, Vol 11 no. 2.
- Prinsloo, D.J. 1992. a. Prescriptiveness versus descriptiveness - an approach for Northern Sotho lexicography in South African Journal of Languages, Vol 12 No.1.
- Prinsloo, D.J. 1992. b. Lemmatization of reflexives in Northern Sotho Lexikos 2. Afrilex-reeks, Buro van die woordeboek van die Afrikaanse Taal, Stellenbosch.

- Prinsloo, D.J. 1994. 'Lemmatization of verbs in Northern Sotho', South African Journal of African Languages, Vol 14 No. 2.
- Robins, R.H. 1980 (3rd ed) General Linguistics: An Introductory Survey. London and New York: Longman.
- Samarin, W.J. 1971. Survey of Bantu Ideophones. African Language Studies XII School of Oriental and African Studies University of London (Sole Agent Luzag and Company Ltd.)
- Scholler, H. & Reidy, J. 1973. Lexicography and Dialect Geography. Wiesbaden, Satz and Druck: F. Wolf, Heppenheim.
- Tomic, O.M. & Shuy, R.W. (eds). 1987. The Relation of Theoretical and applied Linguistics. New York and London: Plenum Press.
- Tomaszczyk, J. and Lewandowska - Tomaszczyk, B. (eds) 1990. Meaning and Lexicography. Amsterdam Philadelphia: John Benjamins Publishing Company.
- Van Rensburg, M.C.J. (ed) 1982. Linguistic Society of Southern Africa: Conference Paper. 18th Annual Congress.
- Van Wyk, E.B. 1989. Linguistic Principles in the lexicography of the South African Languages, unpublished paper read at the 5th ALASA Conference at the University of Pretoria.
- Wierzbicka, A. 1985. Lexicography and Conceptual Analysis. USA: Karoma Publishers, Inc.
- Zgusta, L. 1971. Manual of lexicography. Praha: Academica.

Ziervogel, D. 1969. A Handbook of the Northern Sotho Language.
Pretoria: J.L. van Schaik Ltd.

Ziervogel, D.Z. 1967. Handbook of the speech sounds and
sound changes in the Bantu languages of South Africa.
University of South Africa.

PAN SOUTH AFRICAN
LANGUAGE BOARD

Lexicography Training Course

Rufus Gouws
Department of Afrikaans and Dutch
University of Stellenbosch

Lexicography Training Course

Session 1
Lexicography: past, present and
future

The development of the
lexicographic practice

From clay tablet to computer disk and
the internet

Functional containers of information

The first dictionaries were developed as practical instruments.

- Assyrians in Babylon had problems with the Sumerian signs
- Early Arabic dictionaries gave access to the holy scriptures
- Early western dictionaries aimed at pedagogical purposes.

Rufus Gross

4

Looking for the *omne scibile*

- Early dictionaries tried to capture the *omne scibile* (=everything worth knowing).
- Hugues de Saint Victor, a 12 th century Scholastic: "Learn everything; later you will see that nothing is superfluous."

Rufus Gross

5

Thematic ordering

- Vocabularia: organised in themes or topics.
 - 1 God, heaven, angles, sun, moon, ...
 - 2 man, woman, parts of the body
 - 3 kinship ...
 - 4 diseases
 - 5 abstract terms
 - 8 birds

Rufus Gross

6

From theme to alphabet

- Glossa/glotta: Annotated word lists
- Partially alphabetical
- Glossae Collectae
- Gutenberg: Printing press.

Rufus Gowers

7

Early alphabetical dictionaries

- 1604: Robert Cawdrey *Table Alphabeticall*
- 1615: William Bathe *Ianua Linguarum* (the Gate of Tongues)
- 1631: Amos Komensky (Comenius) *Ianua Linguarum Reserata* (the Gate of Tongues Unlocked)

Rufus Gowers

8

Later products

- 1755: Samuel Johnson *A Dictionary of the English Language*.
- 1961: Philip Gove *Webster's Third New Dictionary of the English Language*.
- 1987: John Sinclair Collins *COBUILD English Language Dictionary*.
- Various dictionaries on CD ROM and on the Internet.

Rufus Gowers

9

The electronic era

- A new approach to the packaging of data.
- Added value of electronic dictionaries.
- Lexicographic intertextuality.
- The user as lexicographer.

Rufus Gouws

10

Prescriptivity and descriptivity

- Johnson wanted to fix the language by following a prescriptive approach.
- Gove tried to describe actual language usage.
- Users regard a dictionary as an authoritative source of information.
- Status of the dictionary.

Rufus Gouws

11

Dictionaries and language usage

- Dictionaries should reflect real language usage.
- Lexicographers have to apply objective criteria in their choice and treatment of material.
- The compelling need for a representative corpus.
- Lexicography is not a one-person job.

Rufus Gouws

12

The development of theoretical lexicography

The emergence of a new discipline

Identifying the field

- Theory of Lexicography
- Metalexigraphy
- Dictionary Research.

Rufus Gouws

14

Bridging the gap

- Lexicographers regarded linguistics as too theoretical.
- Linguists regarded lexicography as atheoretical
- Zgusta: *Manual of Lexicography* (1971).
- Lexicographic associations.
- Establishment of a new discipline.

Rufus Gouws

15

Metalexigraphy

- The endeavour to establish a general theory of lexicography.
- The contribution of Herbert Ernst Wiegand.
- A textual approach to dictionaries.
- Providing the lexicographic practice with a sound model.

Rufus Gouws

16

The present

- Lexicography: an art, a science, a craft ?
- Lexicography is a scientific practice, aimed at the production of dictionaries. This activates a further practice, i.e. the cultural practice of dictionary use.
- Successful dictionary consultation procedures presuppose well-structured dictionaries with a sound theoretical basis.

Rufus Gouws

17

Lexicography Training Course

Session 6 The Treatment of Meaning in Dictionaries

Meaning and the dictionary user

- Most dominant data category in dictionaries.
- Often regarded as the genuine purpose of a dictionary.
- Semantic bias which is detrimental to the overall assignment of linguistic dictionaries.

Rufus Gouws

2

Different aspects relevant to the treatment of meaning

- Homonymy and polysemy.
- Lexicographic definitions (the paraphrase of meaning).
- Translation equivalents.
- Lexicographer has to rely on insights from the field of lexical semantics.

Rufus Gouws

3

Homonymy and polysemy

- Homonyms
 - Lexical items with the same form and pronunciation but unrelated meanings.
- Polysemy
 - Different related senses of one lexical item.

Rufus Gouws

4

Homonyms

mole¹ n. 1 small burrowing mammal with dark velvety fur and very small eyes. 2 slang spy established in a position of trust in an organization. [Low German or Dutch]

mole² n. small permanent dark spot on the skin. [Old English]

mole³ n. 1 massive structure serving as a pier, breakwater, or causeway. 2 artificial harbour. [Latin moles mass]

The Pocket Oxford Dictionary

Rufus Gouws

5

Polysemy

mouse —n. (pl. mice) 1 small rodent, esp. of a kind infesting houses. 2 timid or feeble person. 3 (pl. -s) Computing small hand-held device controlling the cursor on a VDU screen. —v. (-sing) (of a cat, owl, etc.) hunt mice.  mouser n. [Old English]

The Pocket Oxford Dictionary

Rufus Gouws

6

Homonymy and polysemy in dictionaries (2)

- Homonyms are entered as separate lemmata to be the guiding elements of unrelated articles. Homonymy holds between different lexical items.
- The polysemous senses of a lexical item are accommodated in one article. Polysemy holds between the senses of one article.

Rufus Gouws

7

Ordering of senses

- The ordering of the senses of a polysemous lexical item may not be done in an arbitrary way.
- Ordering should be done according to fixed criteria
 - historical ordering (typically in a dictionary based on historical principles)
 - ordering according to usage frequency.

Rufus Gouws

8

Senses and uses (1)

- A polysemous sense is part of the meaning of a lexical item.
- Different senses can easily be identified.
- The context in which a word is used activates one sense and neutralises all other.
- The context does not give a word a new sense.
- The sense is not derived from the context.

Rufus Gouws

9

Senses and uses (2)

pipe —n. 1 tube of metal, plastic, etc., used to convey water, gas, etc. 2 a narrow tube with a bowl at one end containing tobacco for smoking. b quantity of tobacco held by this. 3 a wind instrument of a single tube. b any of the tubes by which sound is produced in an organ. c (in pl.) = *bagpipes. 4 tubular organ, vessel, etc. in an animal's body. 5 high note or song, esp. of a bird. 6 a boatswain's whistle. b sounding of this.

The Pocket Oxford Dictionary

Rufus Grewes

10

Senses and uses (3)

- do**
1. To perform, to carry out (*do a job*)
 2. To produce ..
 3. Deal with, put in order

She usually does my hair.

She paints a lot of things but she prefers to do flowers. Do = paint ??

The context gives *do* a specific meaning.

Rufus Grewes

11

Senses and uses (4)

- *A short pencil.*
- *A short lecture.*
- *A short man.*
- *A short throw*

How many senses?

Rufus Grewes

12

Senses and uses (5)

short (ort) *adj.* [[ME < OE *scort*, akin to ON *skort*, short piece of clothing, OHG *scurz*, short < IE *(s)kerd- < base *(s)ker-, to cut, SHEAR < CURT]] 1 not extending far from end to end; not long or not long enough 2 not great in span, range, or scope /a *short* distance, journey, throw, view, etc./ 3 low or relatively low in height; not tall 4 a) lasting only a little time; brief b) passing quickly /a few *short* weeks/

– The Complete Reference Collection. Copyright © 1994, 1995, 1996, 1997
The Learning Company, Inc. All Rights Reserved.

Rufus Gouws

13

The lexicographic definition (1) (paraphrase of meaning)

- Definition: Definiendum + definiens
- Different types of definitions, e.g.
 - Genus and differentiae definition
 - Circular definition
 - Synonym definition
- The nature of the lemma and its relation to other lemmata determine the type of definition.

Rufus Gouws

14

The lexicographic definition (2)

- Definition by genus and differentiae.
- Move from general to specific.
- Different types of semantic features
 - General features
 - Diagnostic features
 - Supplementary features.

Rufus Gouws

15

The lexicographic definition (3)

The meaning of the word *chair*:

- Piece of furniture [table, bed, sofa ..]
GENERAL, DIAGNOSTIC
- To sit on [sofa ..]
GENERAL, DIAGNOSTIC
- For one person.
DIAGNOSTIC
- Back, four legs ... SUPPLEMENTARY

Rufus Gouws

16

The lexicographic definition (3)

- **chair** (er) *n.* [[ME & OFr *chaire* < L *cathedra* : see CATHEDRA]] 1 a piece of furniture for one person to sit on, having a back and, usually, four legs 2 a seat of authority or dignity 3 the position of a player in an instrumental section of a symphony orchestra 4 an important or official position, as a professorship or chairmanship 5 a person who presides over a meeting; chairman [*address your remarks to the chair*]

- The Complete Reference Collection. Copyright © 1994, 1995, 1996, 1997
The Learning Company, Inc. All Rights Reserved.

Rufus Gouws

17

Circular definition

Lexicographer Person working in the field of lexicography.

tablecloth *n.* cloth spread over a table, esp. for meals. (*The Pocket Oxford Dictionary*)

- In a dictionary a circular definition to define a complex word is permissible if the components are treated in their own alphabetical slot.

Rufus Gouws

18

Synonym Definition

wireless n. radio; radio receiving set.

radio —n. (pl. -s) 1 (often attrib.) a transmission and reception of sound messages etc. by electromagnetic waves of radio frequency. b apparatus for receiving, broadcasting, or transmitting radio signals.

The Pocket Oxford Dictionary

Rufus Gouws

19

Translation Equivalents (1)

- A translation dictionary offers a coordination of source and target language lexical items. This coordination should assist the user to convey the meaning of a source language item in the target language.
- The coordination links a lemma with one or more translation equivalents.

Rufus Gouws

20

Translation Equivalents (2)

The translation equivalents given for one lemma constitute a translation equivalent paradigm.

lek'ker³, (b) dainty, nice, sweet, palatable, delectable, savoury, delicate, delicious; pleasant; cosy; tipsy; luscious, exquisite; cushy; merry;

Eksteen, L.C., editor. *Groot Woordeboek / Major Dictionary*. (Cape Town: Pharos Dictionaries) 1997.

Rufus Gouws

21

Translation Equivalents (3)

A translation equivalent does not represent the meaning of a word but it is a target language lexical item which can substitute the source language item in specific contexts.

Translation Equivalents (4)

Different relations of equivalence hold between source and target language entries:

- One to one, i.e. an equivalent relation of congruence
- One to more than one, i.e. an equivalent relation of divergence
- One to zero, i.e. a relation of surrogate equivalence.

Congruence

macaro'ni, macaroni.

Eksteen, L.C., editor, Groot Woordeboek / Major Dictionary, (Cape Town: Pharos Dictionaries) 1997.

Where an equivalent relation of congruence holds, little additional information is needed to ensure the correct use of the target language form.

Divergence (1)

pamperlang', (ge-) flatter, cajole, adulate, coax, fawn.

pasel'la, (-s), gift; cheap sweets.

– Eksteen, L.C., editor, *Groot Woordeboek / Major Dictionary*. (Cape Town: Pharos Dictionaries) 1997.

A distinction is made between

- Lexical divergence
- Semantic divergence

Rufus Goewu

25

Divergence (2)

- Lexical divergence
 - The translation equivalents are target language synonyms.
- Semantic divergence
 - The translation equivalents represent different polysemous senses of the lemma.

Rufus Goewu

26

Divergence (3)

- Without additional, complementing entries the user will not be able to choose the correct translation equivalent.
- The semantic comment must include entries indicating the relevant context and cotext.

Rufus Goewu

27

Surrogate equivalence

- Lexical gaps
 - Linguistic gaps
 - Referential gaps

ubuntu ... the quality embodying the values and virtues of essential humanity, or of Africanness; ..

Silva, P. et al. (eds.) *A Dictionary of South African English on Historical Principles*. Oxford: Oxford University Press.
Rufus Gouws

28
