

Seleman Simon Sewangi

**COMPUTER-ASSISTED EXTRACTION
OF
TERMS IN SPECIFIC DOMAINS:
THE CASE OF SWAHILI**

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Arts
of the University of Helsinki, for public criticism in auditorium XIII,
Fabianinkatu 33, on the 15th of December, 2001, at 10 o'clock.

Institute for Asian and African Studies
Publications, 1
University of Helsinki
2001

ISBN 952-10-0253-0 (printed)
ISBN 952-10-0254-9 (pdf)

ISSN 1458-5359

Helsinki University Printing House,
Helsinki 2001

ABSTRACT

This dissertation proposes the method of computer-assisted extraction of terms in domain-specific corpora. The method is developed based on constraints of structures of Swahili terms, primarily with the focus on the extraction of Swahili terms. The title of the dissertation reflects the dual nature of the tasks involved in the implementation of the method: both machine and manual tasks are employed in the implementation. The method incorporates techniques for the formulation of term patterns appropriate for the extraction of terms in their domains. The techniques concern the unique analysis of terms in a domain-specific corpus and formulation of term-patterns on the term structural constraints discovered, based on the terms analysed in the corpus. For the unique analysis of terms, the techniques introduce the term-domain feature, among others, for analysing words in a domain-specific corpus. Tags for the domain-feature are introduced into a lexicon of the morphological analyser or into the rule file for the BETA system, where they are used for marking term base-forms according to their domains. The morphological analyser or the BETA system applies the tags in the lexicon or in the rule file to analyse terms uniquely in the corpus by their domains. The formulation of term-patterns involves manual identification of compound terms in the analysed corpus, where words analysed as terms are used as searching words. Then the identified terms are specified as sequences of tags selected from the annotation. Thereafter, the tags and their relationships in the sequences are employed to derive the possible term formation constraints on which the patterns are formulated for the extraction of terms. The effectiveness of the proposed method in the extraction of terms is evaluated with respect to the extraction of Swahili terms in the domains of health care and literature, and the results obtained are encouraging. The terms compiled from the evaluation are indexed at the end of the dissertation.

ACKNOWLEDGEMENTS

I am most grateful to my supervisors, Professor Arvi Hurskainen and Professor Kimmo Koskenniemi for initiating me in the field of computational linguistics and for their untiring encouragement and guidance which made this research a success.

My sincere gratitude is due to the Norwegian Council of Universities' Committee for Development Research and Education (NUFU) for awarding me a scholarship, and the leadership of the University of Dar es Salaam for granting me study leave.

I cordially thank Professor Sean O'Fahey of the University of Bergen for the fatherly support and advice that he has always given me, and Mr Ove Stocknes of the University of Bergen for friendly cooperation throughout my period of study.

Special thanks go to my wife Felister for her encouragement and support, which have been the source of strength for the completion of this work.

My personal debt is very great to friends who have contributed in one way or another to the success of this study. I deeply acknowledge the contribution of Jussi Piitulainen who wrote the pattern-matching program for this work. I sincerely appreciate the material and moral support I have received from: Harry Halén, Juri Ahlfors, Traute Stude, Mr. & Mrs. Tapio Pitkänen, Sinikka Tuovinen, Helena Pykälämäki, Mr. & Mrs. Leif Packalén, Eeva Uusitalo, Mr. & Mrs. Eljas Suikkanen, Riikka Halme, Lotta Harjula, Tommi Räihä, Marjukka Pajunen, Karin Lund, Mr. & Mrs. Arnold Chivalala, Ibrahim Ngereza and Wanjiku Ng'ang'a. My English has been corrected by Mr. Michael Cox.

My thanks do not count for much if I fail to mention my mother, Mwagha wa Mghanga, who was my first instructor.

I dedicate this work to the loving memory of my father, Mrema Kiangi.

Helsinki, 2001

S.S.S.

CONTENTS

Abstract	2
Acknowledgements	3
CHAPTER 1	7
Introduction	7
1.1 Definition of key terms	7
1.2 Key issues in the study	8
1.3 Computational terminography.....	9
1.4 Computational methods for term compilation	9
1.4.1 Term compilation by the comparative statistical method.....	10
1.4.2 Term compilation by the information extraction method	12
1.4.3 Term compilation by combination of statistics and the information extraction method.....	13
1.5 Statement of the problem	14
1.6 Presuppositions	15
1.6.1 Identification and unique tagging of term-words.....	15
1.6.2 Discovery of multi-word term structures	16
1.6.3. Mapping terms to tag sequences and discovering term formation constraints	17
1.7 Data Source	18
1.7.1 Annotation system for the data.....	18
CHAPTER 2.....	20
Corpus Analysis And Information Extraction.....	20
2.1 Information.....	20
2.2 Corpus analysis	20
2.2.1 Basic concepts in corpus analysis	21
2.2.2 Factors underlying corpus analysis	22
2.2.3 Description of words by different features.....	23
2.2.4 Description of term-words in a corpus.....	24
2.2.4.1 Identification of term-words.....	24
2.2.4.2 Unique description of term-words.....	26
2.2.4.3 Relationship between words in multi-word terms	26
2.3 Extraction of information from an analysed corpus.....	27
2.3.1 Formulation of a query language	28
2.3.2 A program for extracting information.....	28
2.4 Aim and scope.....	28
2.5 Methodology	30
CHAPTER 3.....	33
Tools For Analysing Terms In Corpus.....	33
3.1 The need for comprehensive morphological annotation for term description	33
3.2 The model for a morphological analyser.....	34
3.2.1 The core of the two-level model	34
3.2.2 Morphemic structures and definitions in the two-level lexicon.....	35
3.2.3 Performing morphological analysis	36
3.2.4 Morphological heuristics.....	37
3.3 The model for disambiguating morphological analysis	38
3.3.1 Morphological analysis as a core for the CGP.....	38
3.3.2. Rule formalism for morphological disambiguation	38
3.3.2.1 The structure of a rule file.....	39

3.3.2.2 The structure of rules	39
3.3.2.3 The arrangement of rules.....	40
3.3.3 Heuristic disambiguation constraints	41
3.4. Tools for analysing the Swahili corpus	41
3.4.1 Background to the tools	41
3.4.2 The SWATWOL morphological analyser.....	41
3.4.2.1 SWATWOL annotation scheme.....	41
3.4.2.2 The SWATWOL lexicon.....	47
3.4.3. SWATWOL rules.....	51
3.4.4 SWATWOL morphological heuristics.....	52
3.5. Analysing a Swahili corpus by means of the Swatwol	52
3.6 Disambiguating Swatwol analysis	54
3.7 Tagging term-words in the morphologically analysed corpus	56
CHAPTER 4.....	58
Corpus-Based Development Of Term Formation Patterns	58
4.1 Structural conventions for Swahili compounding and derivation.....	58
4.2 Incorporating a health-care domain-tag into the morphological analyser	62
4.2.1 Identifying health care term-words in the corpus	62
4.2.1.1 Extraction of nouns and verbal nouns in the corpus	62
4.2.1.2 Sorting out term-words from the extracted nouns and verbal nouns	64
4.2.2 Marking term-words in the lexicon with a health care domain-tag	66
4.3 Unique tagging of health care term-word in the analysed corpus.....	67
4.4 Discovering multi-word terms in the analysed corpus.....	69
4.5 Describing words in term collocations.....	70
4.6 Representing term collocations by Linguistic specifications.....	72
4.6.1 Tags for representing categories of words in term collocations	72
4.7 Formulating Linguistic specifications for term collocations	73
4.8 Constraints of structures of Swahili terms	77
CHAPTER 5.....	79
Evaluation And Findings.....	79
5.1 The program for extracting terms.....	79
5.2 Tagging term-words in the extraction corpus	80
5.2.1 Updating the lexicon or the rule file	80
5.3 Term extraction	80
5.3.1 Clearing out the extraction corpus	81
5.3.2 Extraction	83
5.4 Assessing the effectiveness of the extraction method.....	84
5.4.1 Recall assessment.....	84
5.4.2 Precision assessment	85
5.4.2.1 The non-term items	88
5.4.2.2 The term variants.....	91
5.5 Weighing the terms by their statistical and conceptual behaviour.....	93
5.5.1 Statistical behaviour of the terms.....	93
5.5.2 Conceptual characteristics of the terms.....	94
CHAPTER 6.....	97
Conclusion.....	97
6.1 Brief summary of the study.....	97
6.2 Summary of previous methods for term compilation.....	97
6.3 Summary of the proposed method for term extraction.	98
6.4 Summary of the evaluation	98

6.5 Recommendations	99
6.5.1 Recommendations for the annotation of terms in a domain specific corpus	99
6.5.2 Recommendations for the formulation of term patterns	99
6.5.3 Recommendations for the extraction of terms	100
6.5.4 Recommendations for the prospects of term compilation by the method of information extraction.....	100
6.6 Implications for future research	101
6.6.1 Creation of lexicons or BETA rule files for specialised domains.....	101
6.6.2 Creation of the system for automatic identification of term-words	101
Bibliography.....	102
Appendix I.....	109
Appendix 2	113
Appendix 3	116
Appendix 4	148

CHAPTER 1

INTRODUCTION

This study is concerned with methods of computer-assisted compilation of terms selected from a domain-specific corpus, i.e. texts on specific subject domains stored in a computer-readable form. The study focuses on term compilation by using the method of information extraction. The method involves: (a) the identification and analysis of terms as unique entities in a specialised corpus, (b) formulation of term patterns or templates based on the constraints derived from term structures analysed in the corpus, (c) extraction of term candidates as analysed in a specialised corpus, by term patterns and a template matching program. This study proposes the feature 'term-domain' among features for analysing a specialised corpus for term extraction. The feature facilitates the unique tagging of term-words in the corpus, which is deemed necessary for the extraction of terms in domains.

1.1 DEFINITION OF KEY TERMS

This study is entitled **Computer-assisted extraction of terms in domains: The case of Swahili**. Below are definitions of some key terms:

- A **domain** denotes a particular field of knowledge, such as health care or biology.
- A **term** is a word-form used in a communicative setting to represent a concept in a domain. A term consists of one or more words (compound terms or term collocations). A term may represent different concepts in different domains whereby it can be called a multi-disciplinary term, but a term may not represent two different concepts in a domain.
- A **base-term word** is the base element in a string structure, which constitutes a term. A base-term word may be a term in itself (single-word term) or it may be a constituent part of a multi-word term.
- A **corpus** is a computer-readable collection of texts that are selected to represent a particular language for specific goals in linguistic research.
- A **domain-specific corpus** is a body of texts on a particular domain stored in a computer-readable form to be processed and used for the intended linguistic investigation.
- **Extraction of terms** is the task of recovering possible term candidates in a domain-specific text. The task may be carried out manually, automatically or both manually and automatically. The automatically implemented task applies a computer system developed for extracting terms by matching term patterns with the actual term candidates in the text corpus.
- **Computer-assisted extraction of terms** is the task of recovering term candidates from a text corpus by both machine and manual tasks.
- **Description of domain-specific terms** means analysing term-words in a domain-specific corpus by tags which describe the terms as entities of a particular domain. Incorporating the feature 'term-domain' among features of morphological annotation facilitates the description. The feature can be represented in appropriate places in the lexicon of a morphological analyser or it can be inserted into the analysed text by a rewriting program, such as the BETA system .
- **Term formation patterns** are sequences of tags on the basis of which term candidates

are extracted from the corpus. In term formation patterns tags represent relations between categories of headwords and their modifiers. Term formation patterns are used as templates for the extraction of terms from the corpus by a template-matching program. This study focuses on the method for formulating term formation patterns that are appropriate for the extraction of terms, particularly Swahili terms, in their domains.

1.2 KEY ISSUES IN THE STUDY

The fundamental questions of this study are (a) how should terms in a domain-specific corpus be identified and analysed uniquely, (b) how should constraints of term structures for formulating term patterns be discovered based on terms analysed in the corpus, (c) how should the extraction of terms in a specialised corpus by the term patterns and the template-matching program be carried out and evaluated. Answering these questions involves the following:

a. Identification and description of terms in a domain-specific corpus

Traditionally, terms are identified by the standardization principle, which qualifies a word as a term only if it is officially standardized. In addition, the standardized term is supposed to represent only one concept in a domain. Likewise, a concept in a domain is supposed to be represented by only one term. In this study, identification of terms in a domain-specific corpus relies on the meaning of a word in the specialised text corpus, i.e., if a word represents a concept in a domain of the corpus then it qualifies as a term. Moreover, the identification presupposes that one concept can have as many terms as can be identified in the corpus. The approaches for identifying terms in a domain-specific corpus are presented in Chapter 2.

b. Unique description of terms in the corpus

So far, the use of part-of-speech or grammatical tags, tags that cannot uniquely describe terms in any way, has typically provided the description of terms for term extraction. This study introduces the use of the feature 'term-domain' for tagging term-words in a domain-specific corpus to facilitate the unique description of term-words in the morphologically analysed corpus. Tools for morphological analysis, including term tagging, are introduced in Chapter 3.

c. Discovery of term formation constraints from terms analysed in the morphologically annotated corpus

The discovery of constraints of term structures involves the identification of the structures of multi-word terms as annotated in the corpus. This is carried out on the basis of term-word candidates tagged in the corpus. Represented as sequences of tags (linguistic specifications) the structures identified constitute a basis for the discovery of constraints of term structures for the formulation of term formation patterns. The techniques for carrying out this task are illustrated in Chapter 4.

d. Extracting terms in domains by term formation patterns and a template-matching program

The effectiveness of the method developed for extraction of terms in domains is evaluated in Chapter 5.

1.3 COMPUTATIONAL TERMINOGRAPHY

ISO CD 704 Terminology work -Principles and Methods (1997: 5) defines terminography as “the activity concerned with the recording and presentation of terminological data, principally in print and electronic media”. Terminography and lexicography, i.e. the process of compiling lexical data for general-language dictionaries, differ in a number of ways. Table 1.1 below summarises some of the differences:

Table 1.1 Some differences between terminography and lexicography.

Item of investigation	Lexicography	Terminography
Language	general language	special language
Entities	single-words and frozen collocations	single-word and multi-word terms
Meaning	(mostly) general meaning	domain-specific meaning.
Method	-word-based -descriptive	-concept-based -(largely) prescriptive
Presentation	-alphabetical -polysemes/homonyms presented together -synonyms presented separately	-conceptual -polysemes/homonyms presented separately -synonyms in the same domain presented together

For a long period of time computational terminography has been confined to the storage of terms in term banks (Meyer 1992). The recent development in computer technology, particularly in the storage capacity and speed of computers, has motivated expansion of computational terminography to other areas of application. One such area is the systematisation of terms in relation to concepts. Computational terminologists have focused on this area in collaboration with computer scientists in the field of artificial intelligence (AI) in knowledge engineering. CODE (Conceptually Oriented Design Environment) is a computational tool, which is being developed for this application by the Artificial Intelligence Laboratory at the University of Ottawa (Meyer 1992: 193-204).

Another new area of the application is the compilation of terms from corpora. The application targets terms that have been spontaneously developed and used by field specialists in their respective areas of specialization. Such terms are normally not available in the collections of officially developed terms but in texts such as books, journals and articles written by the field specialists. The purpose of this application is to compile the terms and include them in term banks of individual languages or to use them for writing technical dictionaries for a particular language. Manual compilation of the terms would require a long period of time in order to pick up the terms from texts and possibly, the compilation would not be exhaustive. Thus, computational compilation is developed as a viable alternative.

1.4 COMPUTATIONAL METHODS FOR TERM COMPILATION

Research on the application of computers for term compilation has been carried out in a number of institutions, such as the University of Surrey where, the application is carried out as part of the

Translator's Workbench projects (Ahamed *et al.* 1994), and at the University of Birmingham (Yang 1986). Individual researchers have also been engaged in studies of automatic compilation of terms. According to Pearson (1998: 122), Beatrice Daille has done research on the extraction of terms from telecommunications texts for her PhD thesis, Jacquemin and Royaute worked with the MEDICS' corpus, and Gonzalez-Mullier and Gros developed a tool for terminology extraction, the LEXTER. Pearson, too, investigated the automatic extraction of terms (Pearson 1998: 121-134).

The application of computers for the compilation of terms relies on three methods: the comparative statistic method (Yang 1986; Ahmad *et al.* 1994), the information extraction method (Pearson 1998), and a combination of statistical and linguistic or information extraction methods (Daille 1994) The three methods employ different techniques and tools for the compilation. But both methods are concerned with terms as used in texts written by field specialists and not with terms as developed and standardized by language officials.

1.4.1 Term compilation by the comparative statistical method.

The comparative statistical approach for term compilation relies on the statistical behaviour of words and collocations in a domain-specific text. The approach uses frequencies of occurrence of a word or a collocation in a domain-specific corpus as a criterion for judging the word or the collocation as a term. Counting of the frequencies is normally done on comparative parameters, such as frequency rank order (Ahmad *et al.* 1994), and frequencies of occurrence and distribution (Yang 1986). This method is based on the assertion that the relative frequencies of word classes differ between text types. On that basis, Ahmad *et al.* expound the philosophy which underlies the comparative statistical method as follows:

The possibility that there are differences in the relative frequencies of word classes between text types gives us a first indication that significant differences are likely to emerge between the distribution of word classes and of tokens of lexical types in general purpose and special-purpose texts. Since special language is said to be highly nominalised (Sager *et al.* 1980: 234), it seemed reasonable to assume that a corpus of domain-specific texts would contain not only a higher proportion of nouns than a general-purpose corpus but that certain nouns might be very frequent indeed in relation to words from other classes. Hence the technique... is comparative (1994: 270-271).

The same hypothesis is held by Yang (1986), who writes:

... since scientific/technical terms are sensitive to subject matter, they should have fairly high frequencies of occurrence in texts where they occur, but vary dramatically from one subject matter area to another. It is therefore possible to identify scientific/technical terms solely on the basis of their statistical behaviour (1986: 94-97).

The techniques for the comparative statistical method treat single-word terms and multi-word terms separately. Normally, single-word terms are compiled first and used as the basis for the compilation of multi-word terms. Ahmad *et al.* (1994) applied this approach, which they explain as follows:

So far, our attention has been confined to the identification of potential terms according to the assumption that all terms are composed of a single word. We can now take the output of our computations based on this assumption a step further in order to investigate whether the potential terms identified by means of the relative frequency ratio calculation occur as a part of multi-word

terms or as a common collocation (1994: 273).

The process of compiling single-word terms starts by producing frequency lists¹ of the corpora which should include, at least, a general reference corpus for purposes of comparison. Thereafter, the comparative parameters are applied to determine the behaviour of words in the frequency lists where words that are marked by very low distribution and by very high peak ratio and range ratio² are judged as potential single-word term candidates. On the other hand, the compilation of multi-word terms begins with the identification of collocations in the corpus which contain the identified single-word terms. Function words are excluded in the identification because they rarely occur in compound terms. The collocations are repeatedly identified, starting with two-word collocations, then three-word collocations until the collocations in a corpus are exhausted. Then the criterion of frequencies of occurrence is applied to determine the collocations which qualify as multi-word terms.

Yang (1986) identifies term collocations not by the use of the identified single-word terms but by counting the frequencies of occurrence of the collocations in the domain-specific corpus. Yang applies the following presuppositions to limit the collocations that should be weighed to identify possible multi-word terms.

- a. Multi-word terms are mainly nominals;
 - b. Multi-word terms cannot go across punctuation marks;
 - c. Verbs may be terms by themselves but not part of a multi-word term because multi-word terms should be nominals;
 - d. Function words should be excluded with the exception of prepositions, because prepositions may be part of multi-word terms;
 - e. Adverbs may be part of a multi-word term ... but adverbs for text cohesion should be excluded;
 - f. No multi-word term can end with an adjective or adverb.
 - g. The S-ending should be removed for the purpose of frequency counting.
- (Yang 1986: 100).

Based on the above assumptions, the compilation of term collocation is carried out after the removal of word categories that cannot be part of multi-word terms, i.e. verbs and function words, except prepositions, and of collocations that end with adjectives or adverbs. Their removal is implemented by grammatical tagging of the words that are to be removed and systems of rule for the removal of collocations which end with adjectives or adverbs. After the removal of the unwanted words and collocations, term collocations are identified 'in an iterative way: first, two word combinations, then, 3-word combinations until the maximum possible multi-word combinations (Yang 1986: 101). Then the collocations are weighed by the criterion of frequencies of occurrence to distinguish term collocations from non-term collocations.

Compilation of terms by the comparative statistic method relies on computational tools that are associated with two tasks: production of frequency lists of corpora and identification of possible term collocations in corpora. The first task is implemented by a frequency list (Barnbrook 1996: 43-64; Sinclair 1991: 30-31). The frequency list identifies word forms and produces their frequencies of occurrence in the corpus. This work underlies the frequency and distribution counting. The identification of collocations in corpora is done by concordance programs

¹ Barnbrook (1996: 43) defines a frequency list as a list that shows the words, which make up the texts in the corpus, together with their frequencies of occurrence.

² Peak ratio is the maximum frequency of occurrences divided by the average frequency, while range ratio is the maximum frequency divided by the minimum frequency (Yang 1986: 98).

(Barnbrook 1996: 65-85; Sinclair 1991: 42-51). The identification relies on terms that are identified by frequency counting as search words.

Using the frequency criterion as the basis for term compilation lacks terminological motivation because the criterion overlooks the fact that words acquire terminological status only by being associated with concepts in a particular domain. Additionally, as Pearson points out:

When the frequency criterion is used, this means that a term candidate must occur a certain number of times in a corpus before it is considered. The problem with this approach is that it ignores the fact that it is common for terms to occur infrequently. This may be because the corpus being used for the search is not sufficiently large or because the term in question is more usually referred to on a variant or abbreviated form. We would argue that low frequency should not preclude a term candidate from being considered (1998: 123).

1.4.2 Term compilation by the information extraction method

Compilation of terms by the information extraction method relies on the principles of corpus analysis and information extraction (cf. Chapter 2). The compilation process involves description of terms in a corpus in such a way that the terms can uniquely be represented in term formation patterns and retrieved by a template-matching program. The method is based on the assumption that terms in a language are characterised by unique formation constraints which can be formalised in term formation patterns and used for the extraction of the terms from the corpus. The extraction is done by matching the term formation patterns with term descriptions in the corpus where words and collocations whose descriptions exactly match the patterns are extracted as terms. Proper description of terms in a corpus is fundamental to the compilation of terms by the information **extraction** method because the descriptions underlie the formulation of term formation patterns and the **extraction** of the terms. The descriptions should enable the formulation of term formation patterns that are unique to terms only, or else the patterns match words and collocations in the corpus that are not necessarily terms, i.e., the patterns over-retrieve. This is possible only if terms in the corpus can be described as unique entities. How terms in a domain-specific corpus can be uniquely described is a fundamental question in this study.

So far, compilation of terms by the information **extraction** method relies on part-of-speech tags for the description of terms, where the same tags are used for the formulation of term formation patterns for the **extraction** of terms. Consequently, the extraction of terms by the template-matching program has been characterised by over-extracting. This problem is pointed out by Pearson (1998), who writes:

While the approach [extraction of terms by term formation patterns] provides a useful starting point, it also allows for the inclusion of words or phrases which are not actually terms. For example, if the pattern *adj + noun* has been specified as a term pattern, all occurrences of modified nouns, regardless of their status, will be included, resulting in the extraction of a far greater set than is desirable. Pearson (1998: 123).

Pearson suggests that the problem of over-extracting should be dealt with by filtering the falsely retrieved items from the output by using linguistic clues or signals. According to Pearson, the linguistic signals should be used in accordance with the generic reference criterion which associates term structures with generic concepts. Pearson argues that the criterion rules out the possibility of term structures being preceded by a definite determiner. Therefore, the criterion filters out the retrieved items that are proved by a concordance to be preceded by a definite

determiner and leaves only the items with either an indefinite or zero article as terms. However, the criterion should not apply to cases of term structure that are preceded by definite determiners due to anaphoric reference because the generic forms of the terms are already attended to by the indefinite criterion. Pearson believes that the generic reference criterion is a powerful filter, but also admits that this criterion alone cannot solve the problem. She writes:

... even when the generic reference criterion has been applied, the remaining set of term candidates will still contain many non terms, demonstrating that the criterion is not sufficient on its own (Pearson 1998: 130).

In order to supplement the generic reference criterion, Pearson suggests a number of linguistic signals to be used on condition that "all term structures must satisfy the generic reference criterion and must co-occur at least once with at least one of the specified linguistic signals"³.

Even if the criterion proposed by Pearson were firm enough to solve the problem, the use of linguistic signals is confined only to languages which use the signals for marking generic reference, such as English. Thus, the use of linguistic signals cannot be assumed as providing a suitable solution to the problem, which is caused by improper description of terms in the corpus.

1.4.3 Term compilation by combination of statistics and the information extraction method

Combining statistical and linguistic methods is another method that has been developed for the compilation of terms from a corpus (Daille 1994: Justeson *et al.* 1995). The method is well summarised by Daille, in her paper "Study and Implementation of Combined Techniques for Automatic Extraction of Terminology". According to the paper, the linguistic method recovers the possible linguistic co-occurrences in the extraction corpus and the statistic method filters out non-term co-occurrences among the co-occurrences recovered. In other words, the statistic method is utilised to solve the problem of over-extracting in the linguistic method. Daille summarises the goal of the combined techniques as follows:

Our goal is to use statistical scores for extracting domain-specific collocation only and to forget about the other types of collocations. We proceed in two steps: first, by applying a linguistic filter which selects candidates from the corpus; then by applying statistical scores ranking these candidates and selecting the scores that fit our purpose best, in other words, scores that concentrate their high values to terms and their low values to co-occurrences that are not terms (Klavans *et al.* 1995: 50).

This study proposes an alternative solution to the problem. It is obvious that part-of-speech tags are not unique to terms but to grammatical categories of words including term-words. Thus the linguistic specifications formulated by only part-of-speech tags, such as those used by Pearson (1998: 222-225) for English terms, e.g., DT JJ NN (determiner + adjective + noun) and DT NN NN (determiner + noun + noun), are not unique to term structures. Rather, the patterns represent structures of different types of English nominal phrases. Thus, the problem lies in how terms are described and represented as linguistic specifications for the extraction of the terms. The solution proposed by this study is based on the contention that base-term words should be described uniquely in a corpus so that the unique descriptions can be mapped to indicate term formation

³ Examples of signals that are given include: i.e., e.g., called, known as, and the term (Pearson 1998: 131-132).

patterns that are unique to terms only. Additionally, since terms are specific to domains, the patterns should be adaptable to terms in different domains.

1.5 STATEMENT OF THE PROBLEM

Different methods for computer-aided extraction of terms from a specialised corpus have been developed and tested. Implementation techniques for most of the methods apply terminological constraints that have been based on the structures of English terms. For example, the use of the indefinite article as a cue for the identification of terms in a specialised text (Pearson 1998), and the omission of prepositions and possessives from possible categories of term structures (Justeson *et al.* 1995). Unfortunately, constraints like these are not applicable for structures of Swahili terms because Swahili does not use articles and, contrary to English, Swahili relies heavily on a category of genitive connectors, which fall under the categories of the English preposition and possessive, for the formation of multi-word terms. Moreover, the techniques rely solely on tags obtained from the shallow morphological analysis i.e. on part-of-speech tags only. As has already been discussed, the extractions based on such tags fall short of being capable of identifying terms as unique entities in a specialised corpus. This study addresses the problem of how to develop and test a domain-independent method for computer-assisted extraction of terms in domains. It proposes techniques for the implementation. The techniques are proposed based on constraints of structures of Swahili terms and on tags obtained from the comprehensive morphological analysis, including terminological tagging. The techniques are developed through the following procedures:

a. Initial searching of term candidates

In the development phase of the system, words likely to be terms in a domain-specific corpus are identified. The identification involves automatic extraction of nouns and verbal nouns in a specialised corpus and manual searching for the terms among the nouns and verbal nouns. Such words may be single-word terms in themselves or they may be headwords in multi-word constructions.

b. Tagging of term-words in the morphologically analysed corpus

Information on selected term candidates is transferred to the lexicon of the morphological analyser or to the BETA rule file in such a way that the analyser or the BETA system marks term candidates with domain-tags.

c. Formulation of term formation patterns

Since the extraction of terms in the corpus relies on term patterns formulated on tags, rather than on concrete word-forms, in the morphologically analysed corpus, generalised sets of tag combinations are compiled based on different term lengths. The combinations are the basis of the search for actual terms in the text corpus. Each combination consists of at least one domain-tag and one or more part-of-speech tags.

The techniques and procedures are introduced and illustrated by (1) discovery and unique tagging of health care term-words in the Swahili text corpus, and (2) formulation of term

formation patterns which are compatible with the constraints of Swahili terms and, at the same time, unique to Swahili health care terms. The effectiveness of the techniques and procedures is evaluated by extracting Swahili health care and literature terms. Though developed and evaluated on Swahili terms, the techniques and procedures proposed in this study should be applicable to the extraction of terms in languages for which a comprehensive morphological analyser and disambiguator are available.

1.6 PRESUPPOSITIONS

The techniques developed in this study are concerned with procedures for discovering and analysing terms in a domain-specific corpus and for mapping term descriptions to discover term formation constraints for the formulation of term formation patterns that are appropriate for extracting terms in domains. The procedures involve a number of tasks, such as the following:

- a. Identification and analysis of term-words in a morphologically analysed domain-specific corpus
- b. Discovery of multi-word term structures as analysed in the corpus.
- c. Mapping of the analysed term structures to sequences of tags selected for different categories of words in the structures, and discovering term structural constraints on the basis of the sequences of tags for the formulation of term patterns.

Each of the above tasks and its underlying assumptions are described below.

1.6.1 Identification and unique tagging of term-words

In this study, identification of terms in a domain-specific corpus is based on the assumption that terms in a domain, both single-word and multi-word terms, are derived from a finite set of domain-specific term base forms by the processes of derivation and compounding. Thus, identification of domain-specific term base forms underlies term identification in the corpus. But since term base forms are available in single-word terms in the corpus, their identification is preceded by the identification of single-word terms. However, the identification of single-word terms is problematic because there is no conventional criterion for differentiating term-words from non-term-words in a text. This study employs the principle of concept naming for the identification of terms in the corpus. The principle qualifies a word or a collocation as a term only if it denotes a concept in the domain (Sager 1990: 19). Based on this principle, the identification of term-words is restricted to the category of words that can be names of concepts, i.e., nouns, including verbal nouns. This restriction is supported by Sager (1990: 58), who argues that concepts are predominantly expressed by the linguistic form of nouns, and some theoreticians deny the existence of adjective and verb concepts. The identification of single-word terms in this study involves two tasks:

a. Identification of nouns, including verbal nouns, in a text corpus

Identification of nouns and verbal nouns in the corpus underlies the identification of term-words in this study because the study restricts terms to words of the two categories. The nouns are identified in the morphologically analysed text corpus. In the corpus, the 'N' tag and the 'INF' tag uniquely tag nouns and verbal nouns respectively. The two tags are used for the extraction of the nouns from the corpus by a template-matching program.

b. Sorting out term-words among the extracted nouns

Term-words are manually sorted out among the nouns extracted from the corpus. The nouns that are associated with concepts in the domain are picked as terms. This task is supposed to be done by language experts in co-operation with domain specialists experienced in the use of a language in the domain.

c. Unique tagging of term-words

Since concept representation in a domain is what qualifies words as terms in the corpus, terms should be uniquely tagged according to the domains of their concepts. The tagging is facilitated by the use of a domain-tag. The term-words identified are marked as term, or rather as term candidates, of a particular domain, by the domain-tag in the lexicon of the morphological analyser or in the BETA rule file. The morphological analyser, through the lexicon, or the BETA system, through the rule file, insert the domain tag into the analysis of all words in the corpus that are derived from the base-forms tagged by the domain tag.

1.6.2 Discovery of multi-word term structures

The multi-word structures in the analysed corpus are discovered on the basis of the assumption that structures of multi-word terms contain base term-words as headwords or as components of the modifying elements. The modifying elements may consist of single words or a number of words chained together by connectors. Thus, the base-terms, which are uniquely tagged in the corpus, should be used as searching words for the multi-word terms. The searching should be carried out manually by tracing word clusters where base-term words are included. The process is illustrated by the following example of identifying term structures in the annotated Swahili corpus, where term words are described by the domain tag HC-N (health care noun term) or HC-V (health care verbal term).

Before annotation the text appeared as follows:

magonjwa ya minyoo kama vile safura, kichocho, tegu, minyoo askari, ugonjwa wa ameba na kadhalika. Worm diseases such as hookworm, bilharzia, tapeworm, ascaris, amoeba disease, and others.

After annotation the text is in the following form:

```
"<magonjwa>"
  "ugonjwa" 11/6-PL N HC-N
"<ya>"
  "ya" 5/6-PL GEN-CON
"<minyoo>"
  "minyoo" 3/4-PL N HC-N
"<kama_vile>"
  "kama_vile" ADV COLLOC @ADVL
"<safura>"
  "safura" 9/10-0-SG N AR HC-N
"<>"
"<kichocho>"
  "kichocho" 9/10-0-SG N HC-N
"<>"
"<tegu>"
  "tegu" 9/10-0-SG N HC-N
```

"<_>"
"<minyoo>"
"minyoo" 3/4-PL N HC-N
"<askari>"
"askari" 5a/6-SG N AR AN HC-N
"<_>"
"<ugonjwa>"
"ugonjwa" 11/6-SG N HC-N
"<wa>"
"wa" 11-SG GEN-CON
"<ameba>"
"ameba" 9/10-0-SG N HC-N
"<na_kadhalika>"
"na_kadhalika" ADV COLLOC @ADVL
"<.\$>"

The identification of term structures in the form annotated above would involve manual tracing of compositions of words where the tag HC-N or HC-V is found. Such words are:

magojwa	“diseases”
minyoo	“worms”
safura	“hookworm”
kichocho	“bilharziasis”
tegu	“tapeworm”
askari	“ascaris”
ameba	“amoeba”

The tracing would reveal that some terms function as single-word terms whereas others are modified to form multi-word terms or term collocations. Such collocations would be:

magonjwa ya **minyoo** “worm diseases”
(diseases of worms)
minyoo **askari** “ascaris lumbriocoides”
(worms ascaris)
ugonjwa wa **ameba** “amoeba disease”
(disease of amoeba)

1.6.3. Mapping terms to tag sequences and discovering term formation constraints

Formulation of term formation patterns in this study is based on the assumption that creation of terms in a language is guided by constraints that are specific to structures of terms in the language. Discovering such constraints underlies the formulation of patterns for the extraction of terms in the language. This task begins with mapping of categories and subcategories of words in the analysed term structures to the tags selected for such categories. The selected tags should be the same as those used to mark the categories in the morphologically analysed corpus. A relationship of tags in a sequence represents a constraint that is unique to a structure of particular terms in a given language. Thus, different sequences of tags represent different constraints of term structures in the language. Such constraints, when discovered, should be the basis for deriving general constraints of structures of terms on which the term patterns are formulated. In

this study, the constraints of structures of Swahili terms are discovered and used for the formulation of term patterns for the extraction of Swahili terms. The discovery is based on Swahili term structures as obtained in the morphologically analysed health care text corpus. The HC-N tag or HC-V tag is used to mark the category of health care term-words in the corpus.

1.7 DATA SOURCE

The formulation of term patterns for Swahili terms and the extraction of Swahili terms in the domains of health care and literature elaborate the techniques and procedures developed in this study. The patterns have been formulated based on structural constraints of Swahili terms as discovered from term structures in the morphologically analysed Swahili health care corpus. Since the corpus used in this study was not available in a computer-readable form, it had to be compiled first in the form of books and journals. The health care books and journals for the corpus were written by medical doctors for public education, and those on literature were written by experts in the domain for institutions for higher education. The books and journals were converted into computer-readable forms by the scanning and editing method. The health care corpus consists of texts on the sub-domains of mother-child health, public health, and diseases. The size of the corpus is 173,666 words compiled from the following books and journals: Swahili translation of Arkutu (?) *Afya ya wanawake* "Women's health"; African Medical and Research Foundation AMREF (1989) *Kitabu cha magonjwa ya zinaa* "A book on sexually transmitted diseases"; AMREF (?) *Elimu ya Afya* "Health education"; Kaisi (1976) *Ukunga na utunzaji wa watoto vijijini* "Midwifery and baby care in villages"; Mwita (1992) *Magonjwa ya kuambukiza* "Contagious diseases" and *HAM* (volumes 1992-1995) *Jarida la Huduma ya Afya ya Msingi* "Journal of Primary Health Care". The literature corpus consists of 161,689 words compiled from the following books: Institute of Kiswahili Research (1981) *Makala za Semina: Fasihi*. "Seminar papers: Literature" and Kimani Ngogu and Rocha Chimerah (1999) *Ufundishaji wa Fasihi: Nadharia na Mbinu* "Literature teaching: Theory and Techniques". The corpus of each domain was divided into two portions: the training and the extraction corpus for the health care corpus, and the monitoring and the extraction corpus for the literature corpus. The training corpus consists of 57,856 words and was kept for the discovery of structural constraints of Swahili terms for the formulation of the patterns for extracting Swahili terms. The monitoring corpus consists of 56,681 words and was reserved for checking for the possible literature-specific term pattern that could have not been represented in the patterns developed based on the health care training corpus. The monitoring is to be carried out before the patterns are used for the extraction of literature terms. The health care and the literature extraction corpus consist of 173,666 words and 161,689 words respectively, and they were set aside for evaluation extraction. The corpus files were analysed by the Swahili morphological analyser (SWATWOL) and disambiguated by the Swahili constraint grammar disambiguator (SWACG).

1.7.1 Annotation system for the data

According to Leech (1997b: 20), an annotation system is the specification of procedures for analysing a text corpus on the basis of the following questions:

How to divide the text into individual word tokens (or words)?

How to choose a tag set (or a set of word [and inflectional] categories to be applied to the word [and morphological] tokens)?

How to choose which tag is to be applied to which word [or morphological] token?

Additionally, the specification should reveal the techniques and software employed for the analysis.

The analysis of the data in this study involved three tasks: pre-processing of the texts, morphological analysis and morphological disambiguation. The pre-processing was carried out by the program known as PREPROC:SWA, written in BETA by Hurskainen (1992: 109-110). This program is capable of performing a number of tasks, such as changing of upper-case letters to sequences consisting of '*' character and the corresponding lower-case letter, e.g. *Serengeti* to **serengeti*, dissociation of punctuation marks from strings of words without affecting the marks which are part of the word strings such as compressions, and marking of sentence boundaries. Additionally, the program merges multi-word collocations into single orthographic word tokens by underscoring '_'. For example, the program glues together inseparable adverbial phrases such as *mara kwa mara*, and *moja kwa moja* as *mara_kwa_mara* and *moja_kwa_moja*.

CHAPTER 2

CORPUS ANALYSIS AND INFORMATION EXTRACTION

Techniques for extracting terms from a corpus are based on principles for corpus analysis and extraction of information as analysed in the corpus. This chapter introduces principles of analysing and extracting information from a corpus. The chapter starts by defining *information* within the context of corpus analysis. A brief definition of corpus and corpus analysis is given thereafter. The basic concepts in corpus analysis and the factors underlying the analysis are then introduced. Next, the possibility of analysing words in a corpus using different features is discussed, and a discussion of the description of term-words in a corpus follows. The extraction of information from the corpus is then discussed in relation to the query language and information-extracting program. Finally, the aim, scope, and methodology of the study are specified.

2.1 INFORMATION

The definition of information is complicated by the problem of distinguishing between meaning and the representation of meaning. Depending on the communicative context, information could be defined either as the meaning that is conveyed or as the symbols that represent the meaning. Soergel (1985: 17) points out that information could be defined in a content-oriented sense or in a form-oriented sense. In the content-oriented sense, different symbols with the same meaning are regarded as the same information whereas, in the form-oriented sense, different symbols with the same meaning are treated as different information. In the context of corpus analysis, information is associated with representation of linguistic features in a corpus, by the selected symbols to define word forms in the corpus as different categories of the feature. In this context the meaning of a symbol is the category of the feature it represents. Thus, the content-oriented interpretation of information is applicable when different linguists use different symbols to describe a category of the same feature. In such environment the symbols, though different, represent the same information i.e., the same category of the feature. Information extraction involves labels that are used to describe the linguistic features and category of words that they describe in a corpus, such as *nouns*, *verbs*, and *adjectives*.

2.2 CORPUS ANALYSIS

Pearson (1998: 43) defines a corpus as “a collection of transcribed spoken or written pieces of language which is selected according to explicit criteria and stored in electronic form”. A corpus represents a given language in its actual form that is to be used as a source of information on the language. The representative attribute should underlie the selection of a corpus because it allows conclusions to be drawn concerning the large body, which it represents rather than the corpus itself (Barnbrook 1996: 24). The representative attribute has been used to categorise corpora into general corpora, i.e., corpora selected to represent the general characteristics of a language, and special corpora, i.e., corpora selected to represent the language of a specialised communication

(Sinclair 1995). Although the purpose of a corpus is to provide empirical information in language studies, most such information is normally not explicit in the corpus. Thus, the information has to be added to the corpus before it can be obtained.

Corpus analysis is the process of describing words in a corpus by means of symbols of one or more selected linguistic features. The process is also known as ‘corpus enrichment’ or ‘corpus annotation’ (Leech 1997a; 1997b). The analysis is normally done automatically by systems developed for such a purpose. The systems (analysing or annotation systems), are developed in line with the types of linguistic features which they are intended to add to a corpus. Based on the linguistic features that are described in a corpus, corpus analysis has been divided into categories such as those ones in the following table:

Table 2.1 Some categories of corpus analysis

Category of analysis	Information analysed	Analysing system
Part-of-speech or grammatical tagging	Tags for grammatical classes or part-of-speech categories	Grammatical or Part-of-speech tagger (i.e. tagger)
Comprehensive morphological analysis	Tags for categories of inflection, derivation, part-of-speech, word-based syntactic features, such as infinitive and finite verbal features, other word-based features, such as etymological and terminological features	Morphological analyser and disambiguator
Syntactic analysis (parsing)	Tags for categories of syntactic-functions such as subject and object .	Syntactic parser (i.e. Parser)

2.2.1 Basic concepts in corpus analysis

Three concepts underlie the processes of corpus analysis: *entities*, *attributes* and *attribute values*. Entities are the elements that are described in a corpus, such as word forms and sentences. Attributes are features by which the entities are described, such as grammatical features, morphological features, and syntactic features. Attribute values are symbols, such as part-of-speech tags and inflectional tags, for representing attributes in a corpus. The following table provides examples of entities, attributes and attribute values according to different systems:

Table 2.2 Cases of entities, attributes and attribute values by analysing systems

Analysing system	Entities of analysis	Analysing attribute or feature	Values or tags for the attribute
Tagger	Word forms	Part-of-speech category	Part-of-speech tags such as 'N' (noun category) and 'V' (verb category)
Comprehensive morphological analyser	Word forms	Categories of: inflection; derivation; part-of-speech; word-etymology; and term-domain	Inflectional tags; derivational tags; part-of-speech tags; etymological tags; term-domain tags
Parser	Words as syntactic entities	Syntactic categories	Syntactic tags such as 'SBJ' (subject) and 'OBJ' (object)

2.2.2 Factors underlying corpus analysis

Different categories of corpus analysis are determined by: first, the linguistic attributes that are analysable by the analysing system, the underlying goal of the analysis and the attribute values for representing the features in a corpus. Normally, features that are described by the analysing systems are language-general; for example the grammatical feature. Analysing a particular feature in a corpus is possible if the analysing system is capable of handling the symbols of the feature. The goal of the analysis determines the analysable features in a corpus. The analysis can cover all or some of the analysable features in the system. Values for representing the selected features in the analysis of a corpus are corpus-specific and they are developed and elaborated by linguists as analysing schemes⁴ for their corpus. The analysing scheme for a corpus reflects how the developer of the scheme understands entities that are described by the scheme. “There is no purely objective, mechanistic way of deciding what label or labels should be applied to a given linguistic phenomenon” (Leech 1997a: 2). However, the usefulness of the analysis depends considerably on the information provided by the analysing scheme. Leech (1997) proposes six principles as standard guides for the developers of the analysing schemes. The principles may be paraphrased as follows:

1. It should be possible to remove the annotations from the annotated corpus, and to revert to the raw corpus. The raw corpus should be recoverable.
2. It should be possible to extract the annotations separately from the text, to be stored independently if there is need.
3. The annotation schemes should have accompanying guidelines for their users. The guidelines should cover information on the developer of the scheme, the corpus on which the scheme has been developed, procedures for applying the scheme, entities described in the scheme and the symbols used to represent them, and the performance of the scheme in the analysing system.
4. Users of the annotation scheme should not regard the scheme as representing “God’s truth”. Rather, they should see the scheme as a potentially useful tool offered on the assumption that

⁴ The analysing scheme (also known as a parsing scheme or annotation scheme) comprises (i) a list of annotation symbols used (ii) their definitions and (iii) the guidelines for using them (Sampson 1995).

many users may find it valuable, rather than to have to develop their own schemes from scratch (a task which could take years to accomplish).

5. Annotation schemes should be based as far as possible on widely agreed and theory-neutral analysis of data. Developers of the scheme should select symbols that are as widely accepted and understood as is feasible.

6. No single annotation scheme should claim authority as an absolute standard. Rather, the usefulness of the annotation scheme should be judged on the basis of the goals of the annotation because the purpose for which the annotations are primarily intended may give priority to certain kinds of information.

2.2.3 Description of words by different features

A description of a word in a corpus is based on the selected feature or features. The goal of the description and the capability of the describing system underlie the selection of the features for the description. Different values or symbols for a description feature subdivide, uniquely, word forms into sub-categories of the selected feature. For instance, values for the feature GRAMMATICAL CATEGORY subcategorise word forms into unique subcategories of the GRAMMATICAL CATEGORY by associating words of each subcategory with a unique symbol, such as: N (noun) V (verb) Adj. (Adjective) Adv. (adverb) GEN-CON (genitive connector) and PREP (preposition). Word forms that are described by a particular symbol of the describing feature are named after the symbol, e.g. words that are described by the N (noun) symbol and by the V (verb) symbol are respectively known as *nouns* and *verbs*. Separate features can further describe words of a subcategory of the description feature. For example, *nouns* can be described further by the feature NOUN-CLASS, and uniquely subdivided into subcategories of NOUN CLASS by noun class tags. Following this patterns, words in a corpus can be described by a general feature, i.e., the feature that covers all words in a corpus or by a specific feature, i.e., the feature that covers only certain type of words in a corpus. For example, the feature GRAMMATICAL CATEGORY is a general feature, whereas the future NOUN-CLASS is a specific feature. Description of words by general and specific features creates a hierarchical relation between the words described, where those described by the specific feature are a subset of the ones described by the general feature.

The fundamental requirement in the description of words by a given feature or features is to make a decision as to which subcategories should be associated with the feature, and which words should be identified with each category. Features, and their symbols that are used to describe words in a corpus form a relational model, which has been defined by Salton *et al.* (1983: 366) as:

... a table representing the records in a given file in such a way that each record also known as a *tuple* is identified by an ordered set of attribute values. In other words, each record corresponds to a particular row of the table and all rows are assumed to be distinct and of the same length. The order in which the rows appear in the table is immaterial and so is the order of the columns. Each column corresponds to a particular attribute characterising the records, the column entries being the values of that attribute for the various records.

Following is an example of a relation model for the features BASE FORM, INFLECTIONAL CATEGORY, GRAMMATICAL CATEGORY and SYNTACTIC CATEGORY. Values of the features are not given but merely stated because different languages are described by different symbols for the same values.

Table 2.3 Case of a relational model

Description features:			
Base form	Inflectional category	Grammatical category	Syntactic category
Base forms for nouns	Tags for nominal inflections	Tag for noun category	Tags for subject and object nouns
Base forms for verbs	Tags for verbal inflections	Tag for verb category	Tags for finite and infinitive verbs
Base forms for adjectives	Tags for adjectival inflections	Tag for adjective category	Tag for adjectives as nominal modifiers
Base forms for adverbs	Tags for adverbial inflections	Tag for adverb category	Tags for adverbs as verbal specifiers

Each row in the above sample represents a record, and each column corresponds to a particular attribute or feature characterising the record. Elements of a record are values or symbols for specific features used in the description. An element of a record is known as a 'field'. The names for features specified in a relational model are known as a 'relation scheme'. For example, the relation scheme for the above sample comprises `BASE FORM`, `INFLECTIONAL CATEGORY`, `GRAMMATICAL CATEGORY`, and `SYNTACTIC CATEGORY`. Relation schemes are specified in corpus annotation schemes. Record fields of a relational scheme underlie the process of extracting information from the corpus analysed.

2.2.4 Description of term-words in a corpus

Three assumptions underlie the question as to how words in a specialised corpus should be described as terms. First, all terms are words but not all words are terms. On this assumption, the process of describing terms in a corpus should first involve the identification of the terms to be described. Second, though terms are generally assumed to be nouns, i.e., names of concepts, not all nouns qualify as terms. This assumption implies that nouns and terms should be described differently, i.e., by different features. Third, terms are more than single word forms, i.e. terms are in both single-word and multiword forms. Therefore the description of terms should involve establishing relationships between words in a corpus.

2.2.4.1 Identification of term-words

After completing a comprehensive investigation of the possible criteria for distinguishing term-words from non-term-words, Pearson (1998) made the following observation:

... in spite of extensive research in the field of terminology and in the field of sub-languages, there is no usable definition of term and no adequate communication model, which allows us to identify when words are being used as terms.
(Pearson 1998: 8)

In addition to the absence of structural features for the distinction of term-words from non term-words, the behaviour of word forms in different communicative settings is another source of the problem. Apart from relatively few technical vocabularies, most term-words are polysemous in

the sense that they have different references in different communicative situations. For example, the Swahili word *kukojoa*, which in general use means 'urinating' means 'ejaculating' in the domain of human reproduction.

Among the methods that have been proposed for the identification of term-words in a domain-specific corpus is the inclusive approach, Pearson (1998: 8), by which all words in the domain-specific corpus are regarded as potential terms unless otherwise demonstrated. However, the immediate problem with this approach is that it overlooks the fact that a domain-specific corpus contains function words, which cannot be terms by any criterion. Moreover, apart from linguistic signals which are more or less language specific, Pearson does not offer any language independent criterion for demonstrating that a given word in a domain-specific corpus is not a term.

Another approach that has been widely applied for the identification of term-words in a text corpus is frequency counting and distribution. By this criterion, a word is judged as a term on the basis of its frequency of occurrence in a domain-specific corpus (cf. Chapter 1). But, as has been pointed out, frequency counting is a misleading criterion because it discriminates against genuine terms that occur rarely in the text.

Standardisation has been the traditional criterion for distinguishing terms from non-term words where only words that are officially approved as a term and published in lists of standards, are recognised as legal terms. Yet this criterion is a major drawback in the corpus-based approach for term compilation because the criterion disregards terms that are used by domain specialists and practitioners but are not found in the standard lists. There are a number of reasons for the use of non-standardised terms by specialists, including the following ones given by Pearson (1998: 24):

- 1) there are many terms in use which do not appear in the standards because they have either not yet been considered by the standardising bodies or have been considered and rejected by the bodies, but continue to be used in subject fields.
- 2) many subject specialists have no access to the standardised terms, as some are not even aware of the existence of such terms for their domain while others are not financially capable of purchasing the normally expensive standards. These subject specialists continue to use terms, which are available to them regardless of their status.
- 3) it is most likely that subject specialists would prefer to use the type of language which their community accepts rather than the one dictated to them by the standardisation bodies. Specialists may therefore deliberately choose to use deprecated terms instead of standardised ones as long as their community accepts such terms.

Thus, the standardisation criterion cannot cover all terms that are available in the domain-specific corpus.

According to Pearson (1998: 19), Trimble and Trimble, on the basis of teaching special language to foreign students, suggest the familiarity criterion by which, words in a domain-specific text that are unfamiliar to the students are judged as subject-specific terms and those which are familiar to them are judged as words of general vocabulary. However, this criterion overlooks the fact that there is the possibility that words of general vocabulary function as terms in a domain-specific environment. This happens when a word form of general vocabulary is used in a particular domain to represent a concept in the domain. Moreover, according to Pearson this criterion is unworkable because it is not possible to measure familiarity in any objective way.

Sager (1990: 61) suggests a subject knowledge criterion for the distinction of term-words from non-term-words in a domain-specific corpus. The criterion is based on the understanding of how concepts in concept fields of a particular domain, are represented by words in the corpus. Such knowledge underlies the recognition of words in the corpus that are associated with particular concepts in the domain. The subject knowledge criterion relies on the contention that words acquire their term status when they are used to represent concepts in a particular communicative setting.

2.2.4.2 Unique description of term-words

It is generally accepted that term-words in a corpus belong to a subclass of nouns. This implies that term-words should be described uniquely as a subclass of nouns. Unfortunately, this is not the case so far because terms are described as general nouns, i.e., by the symbol 'N' that belongs to the grammatical feature rather than to the terminological feature. A unique description of terms in a corpus requires a descriptive feature that is unique to terms. This feature should be incorporated in the corpus description to facilitate the description of terms by terminological, rather than grammatical features. Since terms belong to different domains, this feature should be capable of subdividing terms according to their domains, possibly by means of domain specific tags.

The incorporation of a terminological feature in a corpus annotation scheme is only possible if the annotation system can accommodate the feature and its values. It is obvious that a system whose performance is limited to part-of-speech tagging can not be used for such a purpose, whereas a system capable of performing comprehensive morphological annotation, such as the two-level based morphological analyser, can accommodate the feature. Analysing terms by the domain feature will divide them into subcategories of term domains where terms of a domain will be named after their respective domain symbol. For example, if the HC symbol has been selected for describing terms in a domain of health care, such terms will be named as HC terms (health care terms). Words in a corpus described with more than one term domain symbol should be regarded as multidisciplinary terms, i.e., words that function as terms in more than one domain.

2.2.4.3 Relationship between words in multi-word terms

Multi-word term structures represent regular relationships (constraints) between words that build the structures. The relationships are language specific and involve headwords and modifiers. The process of identifying these relationships requires the identification of categories of words in a language that are possible headwords and modifiers in the structures. Traditionally, the discovery relies on the intuition of terminologists who, first, specify the categories and then present a number of terms to back the specifications. Temu (1984: 113-116) applied this method to describe linguistic specifications for Swahili compound terms. His description revealed that there are six term formation patterns for Swahili compound terms. In other words, head-modifier relationships between words in Swahili compound terms can be specified in six different ways.

Contrary to the traditional method of discovery, is the corpus method, which relies on evidences from corpora. The method involves, first, identification of actual terms and then representation of words that form the terms by symbols of their respective categories. Ohly (1982) applied this approach to describe term formation patterns for Swahili textile terms. He first collected the terms from the textile industry and then used part-of- speech tags to describe the term formation patterns.

The description revealed the following six patterns:

- I *VN N* (nominalised verb phrase, e.g. *kukaza uzi* “stretching thread?”)
 - II *DV N* (deverbative head with a noun complement, e.g. *kiweka ngalawa* “?”)
 - III *NN* (construction of two nouns in juxtaposition, e.g. *haidrogeni peroksaidi* “hydrogen peroxide”)
 - IV *N Adj* (combination of noun and adjective e.g. *kostiki soda* “caustic soda”)
 - V *N -a N* (construction with a connector *-a* (e.g. *rangi za moto* “?”)
 - VI *N -a VN* (construction with connector *-a* followed by a verb noun qualifier (e.g., *sindano ya kutungia* “boring needle?”)
- (Ohly 1982: 76-77).

Although Ohly used actual terms in his study, his linguistic specifications are not distinctive to terms, particularly to textile terms. Rather, the specifications represent structures of the Swahili noun phrase. Subsequently, such specifications fail to distinguish between formation constraints for compound terms and for general noun phrases. Daille (1996) argues that multi-word terms should be considered a subclass of nominal compounds. However, her linguistic specifications of French compound terms are not distinctive to terms as a unique subclass of nominal compounds. Rather the specifications represent general structures of French noun phrases, such as *N ADJ, N1 de N2* where the *N* and *N1* modify *N2*. In other words, specification of relationships between words in compound terms by part-of-speech tags only fails to capture the relationships that are unique to compound terms.

The specification of word relations in compound terms should involve both major categories and subcategories of words as evidenced in actual term structures. Words that function as members of a subcategory of the major category, such as *noun term* as a *noun* subcategory, and *verbal noun* as a *verb* subcategory, should be specified uniquely in the relationships. By this style, a major category and a subcategory can be specified as different head-modifier relationships. For example, a major category *noun*, and a subcategory *noun term* can be specified as two head-modifier relationships: (1) a head *noun* with a *noun term* modifier and (2) a head *noun term* with a *noun* modifier. Moreover, the presence of subcategories in the relationship filters out members of the major categories that are not required in the relationships. For example, in the relationship ‘a *verb noun* with a *noun term* modifier’, the subcategories *verb noun* and *noun term* exclude from the relationship all verbs that are not verb nouns and all nouns that are not noun terms respectively. In this way, the specification of word relationships (constraints) for the formation of multi-word terms in a language should involve the discovery of the possible categories and subcategories of words that form the terms, and the possible relationships that characterise the categories. This discovery of possible word categories and subcategories for multi-word terms is possible only if the corpus annotation system is capable of analysing words in the corpus by both their categories and subcategories. As has already been pointed out, the two-level based morphological analyser is capable of performing such a task.

2.3 EXTRACTION OF INFORMATION FROM AN ANALYSED CORPUS

The task of extracting information from an analysed corpus depends on two things: (1) a query language to represent feature values of the information requested and (2) a program to extract the requested information by matching the feature values as represented in the query language with the same values as marked in the information requested.

2.3.1 Formulation of a query language

The formulation of a query language for information extraction relies on feature values or symbols that describe the information to be extracted. For the extraction of linguistic information, such as terms, in the analysed corpus, the query language is formulated based on the tags that describe the information in the corpus. The target information for such language could be words in a particular category or subcategory, or chains of words in specific relationships of categories. Where the target is to extract words in a particular category or subcategory, the symbol for the category or subcategory of the words is used as the query language. For example, if the extraction concerns words in the *noun* category, the symbol that is used to describe the category in the corpus, e.g. the N symbol, is the query language. In cases where the focus of the extraction is a chain of words in a particular relationship, the formulation of the query language involves specification of the relationship as a sequence of symbols. Each symbol in the specification represents a unique category or subcategory of words as described in the corpus. For example, if the extraction targets a chain of two words in a head-modifier relationship with the headword in the category *noun* and the modifying-word in the subcategory *noun term*, the sequence of the symbols for the category *noun* and subcategory *noun term* forms the query language for the extraction.

Because features that are used to describe words in a corpus are represented by different symbols, these symbols can be used as a query language to represent a request for the information of interest. The requested information is extracted from the corpus by matching the query language symbol with the same symbol in the annotated corpus. Extracting the right information depends on whether the query language symbol represents the right symbol to match the information requested.

2.3.2 A program for extracting information

Programs for extracting information are developed to compile the query language in the machine language that is used by the machine to match and extract information from the corpus as requested by the query language. Such programs are commonly known as template-matching programs .

2.4 AIM AND SCOPE

The purpose of this study is to develop a computational terminography method for the extraction of terms in domains. The focus of the method is the extraction of Swahili terms, both in single-word and multi-word forms, regardless of the frequency of occurrences in the domain of discourse. Thus, the method does not incorporate the frequency of occurrences as its criterion for the identification of terms in the extraction. Instead, the method relies on a term-domain feature, which is represented by a domain-tag in the term patterns and in the morphologically analysed domain-specific corpus for the extraction of terms as analysed in the corpus. The method is confined to the extraction of terms for terminography rather than for indexing purposes. However, the raw frequency of the extracted terms can be monitored, and terms with a high frequency of occurrences can be evaluated for indexing purposes. The method is intended to be used for the following purposes: (1) A unique description of term-words in a specialised corpus that is morphologically analysed (2) A corpus-based discovery of structural constraints of terms

for the formulation of term formation patterns and (3) The extraction of terms in domains by the term patterns and a template matching program. The method is based on structures of terms as morphologically analysed in a domain-specific corpus. Two factors underlie the development of the method: (1) the availability of tools capable of providing a comprehensive morphological analysis of Swahili specialised corpus and (2) introduction of a term-domain feature among the features for analysing Swahili words in a specialised corpus. The tools for morphological annotation have been developed at the University of Helsinki. The introduced term-domain feature enables the unique tagging of terms-words in a morphologically analysed corpus. The feature tags are either inserted in the lexicon of a morphological analyser or they are added to the analysed text by using a rewriting program. Words in the analysed corpus that are marked by the domain-tag are then used as cues for the discovery of multi-word terms in the corpus. Structures of the discovered multi-word terms are represented as sequences of tags and employed for the discovery of constraints of Swahili term structures. The constraints are then used for the formulation of the possible Swahili term formation patterns that are used as templates for the extraction of Swahili terms by the template-matching program. Extraction of terms in a domain is facilitated by applying the appropriate domain-tag in the term formation patterns and in the morphologically analysed extraction corpus.

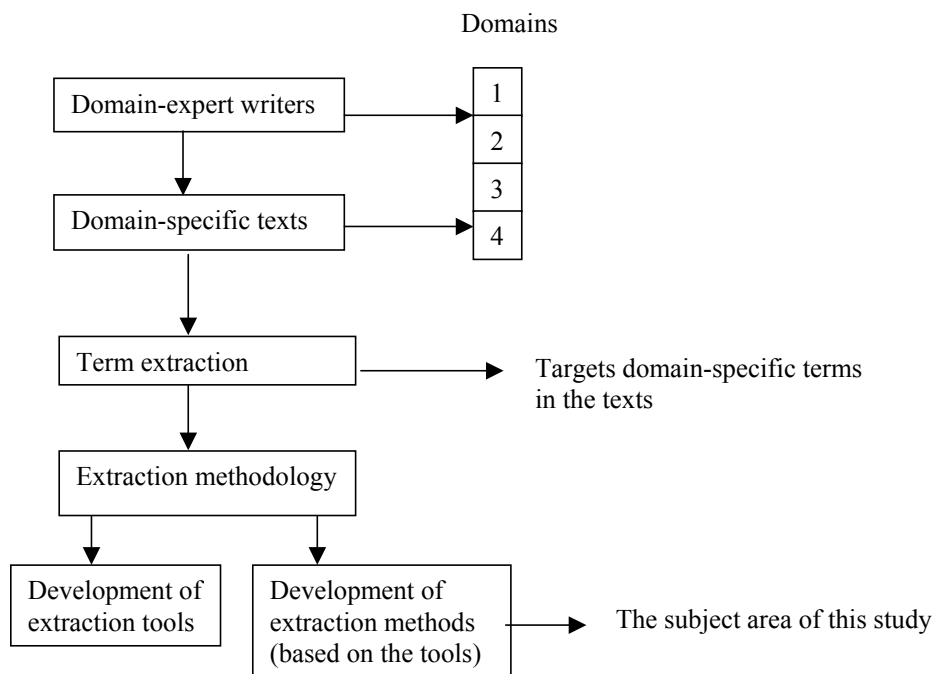
The method has been developed and evaluated on the basis of a Swahili corpus in the domains of health care and literature. The development and evaluation procedures involved the following tasks:

(1) Identification and tagging of term-words in the corpus. The prospective terms are manually identified among the nouns and verbal nouns extracted from the morphologically analysed corpus by the template-matching program. Roots of the term-words are marked with the appropriate domain-tag in the lexicon of the morphological analyser or the tag is inserted into the analysed text with the help of a rewriting program. In this approach, the morphological analyser carries out the tagging of health care terms, whereas the tagging of the literature terms is done by the rewriting program of the BETA system. The domain-tag HC-N or HC-V is used for tagging the health care terms and the tag LT-N or LT-V for the literature terms.

(2) Discovery of structural constraints of Swahili terms and formulation of Swahili term formation patterns based on the constraints. The constraints are discovered on the basis of Swahili term structures as analysed in the training corpus. The resultant patterns, though formulated on constraints discovered on structures of health care terms, represent general constraints of structures of Swahili terms. The patterns were unique to health care terms simply because of the presence of the health care domain-tag in the patterns. The patterns can be applied to the extraction of Swahili terms in different domains by replacing the health care domain-tag with the appropriate domain-tag for the extraction. This is tested by using the patterns for the extraction of literature terms where the health care domain-tag is replaced by the literature domain-tag.

(3) Evaluation of the effectiveness of the extraction method. The effectiveness of the formulated term patterns in the extraction of Swahili terms is tested on the extraction of terms from the morphologically analysed health care and literature extraction corpus. Term-words in the analysed extraction corpus are marked with the relevant domain-tags, i.e. the HC-N tag or HC-V tag for the health care terms and the LT-N tag or LT-V tag for the literature terms.

Chart 1.1 The area of the study

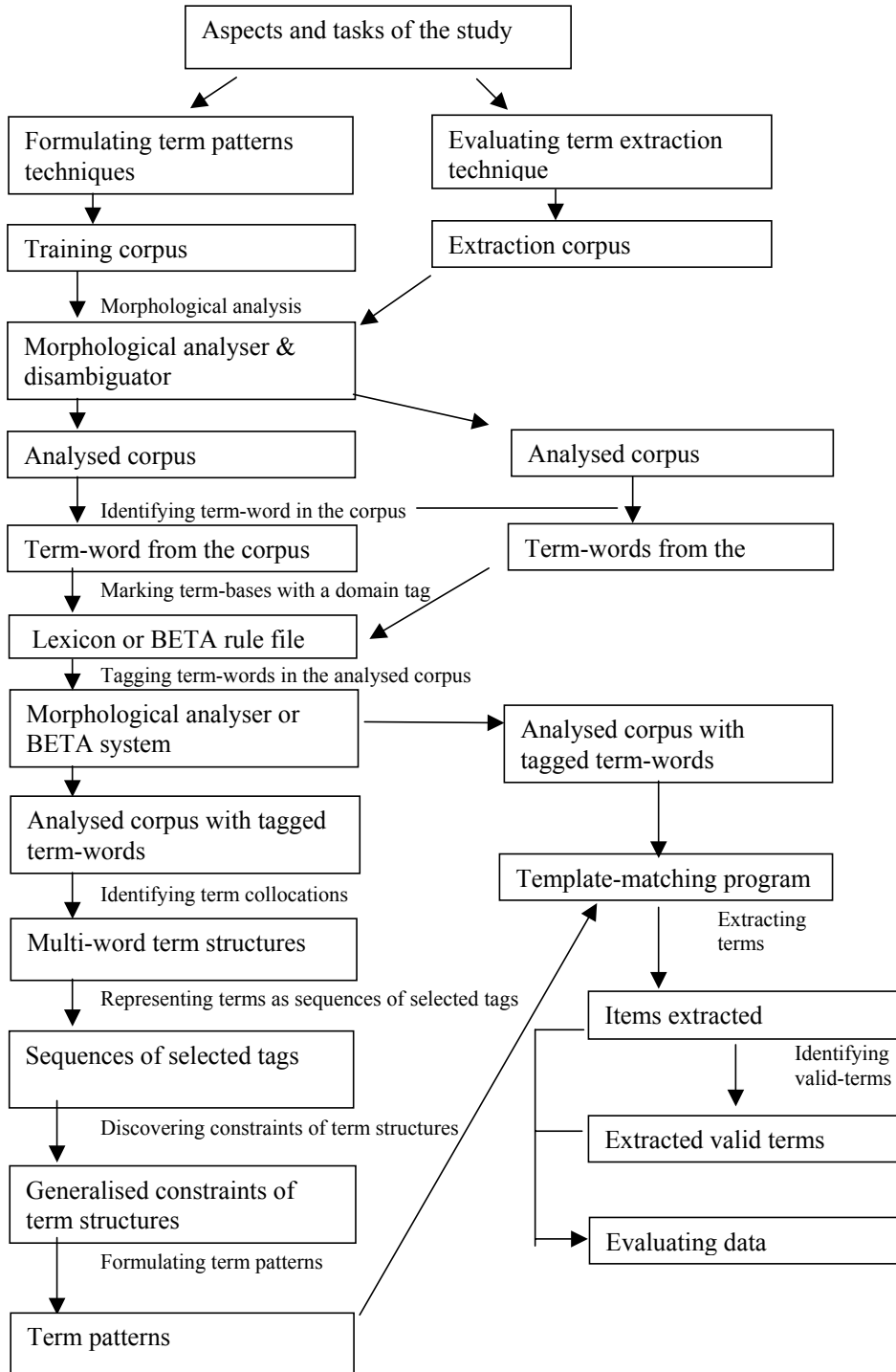


2.5 METHODOLOGY

This study applies the corpus-based approach for developing the method for computer-assisted extraction of terms in a morphologically analysed domain-specific corpus. The development of the method involves two main aspects: (1) techniques for formulating term formation patterns for the extraction of terms in domains (2) techniques for evaluating the effectiveness of the term patterns in the extraction of the targeted terms. The two aspects are implemented on the basis of term data obtained from Swahili corpus in the domains of health care and literature. Tasks in the two aspects involved two categories of corpus: the training and the extraction corpora. The training corpus was used for the tasks in the first aspect, and the extraction corpus for the tasks in the second aspect. Tasks in both aspects involved machine and manual jobs. The machine jobs were carried out by the following tools: (1) the Swahili morphological analyser and disambiguator performed morphological analysis and disambiguation of the corpus, and marked a health care domain-tag in the analysis of term-words in the health care corpus. (2) the BETA system marked a literature domain-tag in the analysis of term-words in the literature corpus. The BETA system was used as a substitute for the morphological analyser. This is because updating the lexicon for the marking of domain-tags by the morphological analyser requires that it be done by a person, usually the owner of the analysing system, who has the right to access and write in the lexicon. Thus, the use of the BETA system for performing the task eliminates the problem of relying on the owner of the lexicon. For the tagging of the terms in the analysed corpus (3) the template matching program, TERMER, performed the extraction tasks. Tasks in the first aspect that were carried out manually were: (1) identification of term-words by sorting through nouns and verbal nouns extracted from the analysed corpus, (2) identification of term collocation in the analysed training corpus, (3) representation of the term collocations as sequences of selected tags

for the discovery of structural constraints of Swahili terms and formulation of Swahili term formation patterns based on the constraints. In the second aspect, the manual job applied for the identification of the valid term candidates among the items extracted. A total of 2,501 health care terms and 2,418 literature terms were identified among the total of 3,665 and 3574 items extracted from the health care extraction corpus and from the literature extraction corpus respectively. The extracted health care and literature terms are presented in Appendices 3 and 4 respectively.

Chart 2.2.



CHAPTER 3

TOOLS FOR ANALYSING TERMS IN CORPUS

The use of a corpus in linguistic investigation is possible only if the corpus can be analysed into a form which favours the goals of the investigation. Corpus analysis (annotation) is normally done by computer tools, and it can involve different levels of linguistic description, such as phonology, morphology, syntax, semantics, terminology and even pragmatics, depending on the goals of the analysis. However, the analysis at any level relies on the availability of computer tools that are capable of performing the required description.

The goal of this study requires analysing tools that are capable of describing words in a domain-specific corpus in such a way that term-words can be described uniquely by the term-domain tag. This chapter presents formalisms for such tools: the two-level-based morphological analyser (Koskeniemi 1983) and the Constraint Grammar-based morphological disambiguator (Karlsson *et al.* (1995); Tapanainen (1996). Moreover, the chapter presents the Swahili-based version of the tools, i.e., Hurskainen's (1992) Swahili two-level morphological analyser (SWATWOL) and (1996) Swahili Constraint Grammar-based morphological disambiguator (SWACGP).

3.1 THE NEED FOR COMPREHENSIVE MORPHOLOGICAL ANNOTATION FOR TERM DESCRIPTION

Terms are characterised by single-word and multi-word forms. But terms in a corpus have no unique structures in the sense that they share the same structures with non-term words. Thus, a unique description of term-words in a corpus requires a describing feature or attribute that is applicable to term-words only. Because both term-words and non-term words (in the head-modifier relationships) form multi-word terms, description of multi-word terms should also cover describing features for categories of non-term words that can be part of the terms. Such description requires computer tools that are capable of performing a comprehensive morphological analysis, i.e., analysing words by the possible word-based features.

Traditionally, the analysis of words in a corpus has been restricted to part-of-speech or grammatical tagging (Garside *et al.* 1997). However, the restriction has never been a deliberate option, rather it is rooted in the use of annotation tools that could not handle comprehensive morphological information. Such tools had been developed for English, the language that lacks challenging morphological structures (Leon 1997). The tools were developed on probabilistic models⁵ and they relied on lexicons that were designed to handle, alphabetically and uniquely, the listing of word forms and suffixes with their potential tags. Such lexicons could not

⁵ With the exception of the rule-based TAGGIT that was used for tagging the Brown corpus at Brown University (Green and Rubin 1971), tools for tagging an English corpus are based on a probabilistic approach. Cases of such tools are the CLAWS developed by UCREL at the University of Lancaster (Marshall 1983; Garside *et al.* 1987) and the Xerox Tagger (Cutting *et al.* 1992; Cutting and Pedersen 1993).

accommodate full morphological information because the information requires a large amount of storage space for lemmas and their associated derived forms.

The need for tools that are capable of handling full morphological information has prompted corpus linguists, particularly those who have been working on the morphologically dominant languages such as Finnish, to work out an alternative model that is capable of handling comprehensive morphological analysis beyond the boundaries of languages. The development of the two-level model for morphological analysis and the Constraint Grammar model for morphological and syntactic disambiguation, reflects such efforts. The following two objectives are among the objectives for the development of the two-level model for morphological analysis:

- to find out what properties a general theory of natural language processing should have in order to cope with the properties of the morphologically dominant languages
- to do the previous with the explicit aim of constructing models accounting for unedited running text, rather than confining the scope of the models to micro-words or small subsets of the potential text classes (Koskeniemi 1985: 111).

Karlsson (1995a, Chapter 1), states twenty-four goals for the development of the Constraint Grammar model with the central objective of designing a language-independent rule-based model for a robust disambiguator which produces appropriate analyses in terms of perfect recall and precision.

Analysis of a corpus by the tools developed on the two-level model and the Constraint Grammar model involves two separate processes: the processes of morphological analysis and morphological disambiguation. The morphological analysis is context independent whereas the disambiguation is context dependent.

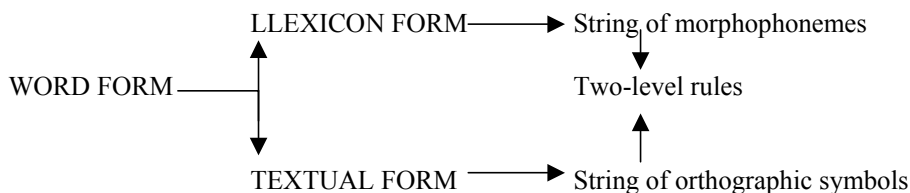
3.2 THE MODEL FOR A MORPHOLOGICAL ANALYSER

The two-level model for morphological analysis has been designed for morphological description as well as for computational application to morphological analysis. The model comprises two components: the lexicon and the rule system. The lexicon handles morphemic constraints and descriptions and the rules describe possible alternations between morphemic strings in the lexicon and word forms texts.

3.2.1 The core of the two-level model

The two-level model assumes morphological processes in the mental lexicon during the process of word recognition. Thus, the model represents possible morphological processes for word recognition in a language. The lexicon and the rules facilitate the process. The lexicon describes morphemic constraints in a language by the mechanism of continuation classes. On the other hand the rules represent relations between string of morphemes in the lexicon and word-forms in the texts.

The relationship can be represented as follows:



In the relations expressed by the two-level rules, the symbol 0 (zero) is used to counterbalance the missing symbols on either level. For example, the symbol is used in the textual form as a substitute for characters that are only used in the lexicon, such as ‘+’ and ‘#’. The following representation of relations between the Swahili textual word *anasoma* “he or she is reading” and its lexicon counterpart “# a + na + som + A #” elaborates how the symbol 0 is used as a substitute in the relations.

a + na + som + A #
 0 a 0 na 0 som 0 a 0

The relations between individual symbols in the above case are expressed as:

< #:0, a:a, +:0, n:n, s:s, o:o, m:m, A:a >

Moreover, meta characters, such as ‘A’ in the above case, are used in the relations. Such characters are only used in the lexicon and they can occur in more than one relation. The textual value of a meta character is determined by the rules.

3.2.2 Morphemic structures and definitions in the two-level lexicon

According to Sproat (1992: 65), a morpheme is an ordered pair construct. The first element of the pair represents the morphological categories, i.e. the morphotactic and semantic features, and the second element of the pair represents the linguistic form and its attachment constraints.

Morphemic entries in the two-level lexicon have three components: the linguistic forms or lexical representation of a morpheme, the attachment constraints and the morphotactic and semantic descriptions. The model of the lexicon provides provisions for the three components. Koskeniemi (1983: 28-29) identifies the provisions as: the lexicon entry component, the continuation class component and the entry definition component. Morphemes are represented in the lexicon by the style of sub-lexicons where a sub-lexicon stands for a morpheme category, such as ‘tense’, ‘subject agreement’, ‘verb root’ or ‘object marker’. Morphemic constraints are represented by the mechanism of continuation class. The following simple lexicon structure for the Swahili verb *anasoma* “he/she is reading” exemplifies the style of sub-lexicons and the mechanism of continuation class in the lexicon.

LEXICON	Verb
#	Sp ;
LEXICON	Sp
a	Pr ;
LEXICON	Pr
na	Vr ;

LEXICON	Vr
<i>som</i>	Vfv -a ;
LEXICON	Vfv-a
<i>a</i>	# ;

The above representation of the verb *anasoma* in the lexicon structures the verb to a number of sub-lexicons beginning with the LEXICON Verb, followed by the LEXICONS; Sp (subject prefix); Pr (present tense); Vr (verb root); and Vfv-a (verb final vowel *a*). The symbol # in the entry of the LEXICON Verb, represents the beginning of the word while the Sp symbol stands for the continuation class. Each entry in the sub-lexicon has the components for lexical representation and for the continuation class. With the exception of the continuation class #, which denotes the end of the word, the continuation classes form the sub-lexicons of the lexicon system.

The presence of three provisions in the two-level lexicon permits the full handling of morphological structures and information. The provision for lexical entry handles linguistic forms for different categories of morpheme whereas the continuation class mechanism takes care of morphological constraints. The constraints filter the possible attachments for the lexical forms in the lexical entry component. In that sense, the lexicon describes a word as a string of morphemic lexical forms that form a chain by the mechanisms of sub-lexicons and continuation classes.

The provision for defining lexical entry facilitates the representation of different kinds of word-based descriptions, such as morphological, terminological and etymological descriptions. The provision serves as an open basket for any information, which can be associated with a word form. This is because the information that is given in this provision does not affect the process of word analysis. The information, though, is retrieved when a word is analysed.

3.2.3 Performing morphological analysis

The lexicon and rule components of the two-level formalism enable the description of natural language morphology, especially concatenative morphology, for automatic analysis and generation of words. The lexicons describe morphotactics and the rules model relations between the lexical and textual forms of a word. The analysis is carried out by first compiling the lexicon and rules into finite a state automaton and transducers, respectively. In the compiled lexicon, words are represented as strings of base forms and tags. For example, the compiled lexicon form of the Swahili word *anasoma* becomes:

somA +1/2-SG3-SP +VFIN +PR:na

The type and number of tags in the string of a compiled lexicon word depend on the description of a word in the lexicon.

The compiled rules, the transducers, facilitate the analysis by mapping between textual words and the compiled lexicon strings. When the matching is complete, lexicon strings are returned as the analysis of textual words. For example, when the Swahili two-level morphological analyser, SWATWOL, analyses words in the text: *vimelea vya uambukizi hutolewa kwenye kinyesi* “infectious germs are produced in faeces”, the analyser returns the following analysis:

"<vimelea>"
 "lea" 7/8-PL-SP VFIN PERF:me V SV SVO

```

    "kimelea" 7/8-PL N ' parasitic plant '
"<vya>"
    "vya" 7/8-PL GEN-CON
"<uambukizi>"
    "uambukizi" 11-SG N
    "uambukizi" 11-SG N
"<hutolewa>"
    "tolea" HABIT-SP VFIN V SV SVO APPL PASS
    "lewa" NEG SG2-SP VFIN FUT:to NON-STD V SV
    "lea" NEG SG2-SP VFIN FUT:to NON-STD V SV SVO PASS
"<kwenye>"
    "kwenye" ' in, at, about ' PREP @ADVL
    "enye" 15-SG POSS: ote
    "enye" 17-SG POSS: ote
"<kinyesi>"
    "kinyesi" 7/8-SG N ' excretum '
    "kinyesi" 9/10-0-SG N
    "kinyesi" 9/10-0-PL N

```

Textual words in the above analysis are surrounded by angle brackets “<>”. The words are followed by their morphological analysis as performed by the morphological analyser. A morphological analysis is known as a reading and a textual word with its readings is called a cohort. In most cases, cohorts contain more than one reading. Every reading that is returned by the analyser is a complete relation between a textual word and the compiled lexicon string as mapped by the two level transducers. The presence of more than one reading signals a textual word that has been matched more than once. The matching could be due to a number of factors, including the interpretation of a textual word as a derivative of more than one base form. For instance, in the above analysis the textual word *vimelea* “bacteria”, has been interpreted as a possible derivative of two base forms: a verb base form *lea* “nature” and a nominal base form *kimelea* “bacteria”. In other words the analysis indicates that in the text the word *vimelea* can be either a verb that is derived from the base form “*lea*” or the plural form of the noun *kimelea*.

Textual words in cohorts that contain more than one reading, are said to be morphologically ambiguous. Each reading in the analysis stands for a unique contextual possibility and there should be only one reading for a context. The task of selecting the readings that are contextually appropriate is not done by the two-level-based morphological analyser but by the constraint grammar-based morphological disambiguator.

3.2.4 Morphological heuristics

Ideally the two-level lexicon represents a full morphological description of a language, which implies that the lexicon is appropriate for the analysis of an unrestricted text. However, there may be some words, in the real text, that are not represented in the lexicon. Subsequently, there are chances that some textual words will escape the analysis. This problem is overcome by providing the analyser with the morphological heuristics or guessing rules for analysing textual words that are not represented in the lexicon. The morphological heuristics enables the analyser to perform morphological analysis for every word in the unrestricted input text. This analysis is a prerequisite for the application of the morphological disambiguator.

3.3 THE MODEL FOR DISAMBIGUATING MORPHOLOGICAL ANALYSIS

The presence of morphological ambiguity, as Voutilainen (1995a) argues, signifies the necessity of contextual knowledge in morphological annotation. The Constraint Grammar Parser (CGP) is a context-based rule model for morphological disambiguation and syntactic annotation. According to Karlsson (1995b: 45), the central goal of the CGP, as a computational tool, is the perfect handling of an ordinary unrestricted text. The CGP morphological disambiguator provides strategies for representing context-based constraint rules for morphological disambiguation. As a program for morphological disambiguation, the CGP morphological disambiguator operates on two inputs: the morphologically analysed text file and the constraint rule file. Thus the proper morphological analysis and description of constraint rules forms the basis of the CGP as a morphological disambiguator.

3.3.1 Morphological analysis as a core for the CGP

Morphological analysis is the basic step of parsing language (Karlsson 1995a: 11). This is because the analysis produces morphological ambiguities that are resolved by the parser. Additionally, the analysis provides morphological information and an environment for the description of the disambiguation rules. The CGP model presupposes full morphological analysis of unrestricted text. Such morphological analysis is possible only with the analyser that has a lexicon system capable of handling the full vocabulary of a language and its morphological description. So far, the two-level-based morphological analyser is the feasible option for the CGP implementation. Among the strategies for optimising morphological analysis in the CGP model is the optimisation of master lexicon and morphological description (Karlsson 1995a: 27-32). According to Karlsson (1995a: 11), “the lexicon is supposed to be large enough, containing at least 30,000 lexemes, one per lemma (lexical entry) behind the assumptions that the morphological rules account for all non-lexicalised inflected and derived forms”. Moreover, Karlsson advocates precise representation of morphological information in the lexicon to avoid unnecessary ambiguities. He argues that “a sloppy lexicon and morphology might cause excessive overgeneration [whereas] a stringent lexicon speeds up disambiguation by minimising the amount of spurious readings” (Karlsson 1995a: 28). Karlsson suggests feature underspecification as a strategy for optimising morphological description, even though he acknowledges its effect of decreasing information. However, it is our opinion that underspecification should not be a general strategy in morphological analysis as it ignores the fact that annotation of a corpus is normally goal-bound and the correct specification is the one which is compatible with the goals of the annotation.

Optimal morphological analysis of an unrestricted text produces full and perfect morphological analysis of the input text. The analysis involves individual word forms in a context-free environment and normally produces more than one possible analysis. When the analysed words are placed in their textual environment only one analysis is acceptable in that environment. The CGP model for morphological disambiguation is a rule formalism for the selection of the analysis that is contextually acceptable in the environment of the input text.

3.3.2. Rule formalism for morphological disambiguation

The CGP rule formalism for morphological disambiguation provides a format for structuring the input rule file and for formulating the constraint-based rules. Moreover, the formalism allows the use of heuristic disambiguation constraints where the ordinary rules fail to disambiguate.

3.3.2.1 The structure of a rule file

The rule file contains a number of sections of different information that is necessary for the functioning of the parser (Karlsson 1995b: 49-54; Tapanainen 1996: 13-24). A header under which is the information associated with the header represents each section. Tapanainen categorises the input rule file headers as follows:

DELIMITERS
PREFERRED-TARGET
SETS
MAPPING
CONSTRAINTS
END

The number of sections in the rule file depends on the task to be performed by the parser. However, every rule file should contain the section DELIMITERS. This section contains information that enables the parser to recognise the end of sentences in the input text as the parser operates on sentences.

The section MAPPING should be in the file only when the parser is to perform syntactic mapping; thus the section is not necessary for morphological disambiguation. The SETS section contains tags that are used in the rules where the tags are defined to facilitate representation of general information in the rules because each set name stands for one or more tags that are defined in the section. Every rule file should contain this section for the purpose of rule generalisation. The information in the section PREFERRED-TARGET restricts certain information in the input cohort from being removed by the parser in cases where the rule system tries to remove all readings of the cohort. The section CONSTRAINTS contains the actual rules that are used by the parser in the processes of disambiguation. The rules are either for morphological or for syntactic disambiguation, depending on the elements of the rules, and they can be represented in more than one section in a rule file. The section END marks the end of the rule file, and the parser ignores whatever follows after that section.

3.3.2.2 The structure of rules

The structure of disambiguation rules is the same for all functions of the parser: morphological disambiguation, syntactic mapping and syntactic disambiguation. The rules have three obligatory constituents and one optional constituent. The obligatory constituents are: operator, target and context condition, whereas the optional constituent is the word form, which is needed only in the rules that are specific to a particular word form. The rule has the following schema:

(WORD FORM) OPERATION TARGET CONTEXT CONDITION(s)

The element that occurs in the constituent WORD FORM particularises an item to be focussed on by the rule. Such an item is a word in a running text and the rule applies only to cohorts with such a word form. In the OPERATION constituent are operators for the kind of operation to be performed on the constituent TARGET. The operators are REMOVE and SELECT. With the REMOVE operator, the parser removes the reading which contains a tag that is given in the target, if all the contextual conditions are satisfied. The SELECT operator directs the parser to select the reading, which contains a tag given in the target if all the context conditions are satisfied (see Tapanainen 1996: 36). In the TARGET position are tags or set names which are

defined in the SET to represent readings in the rule which the parser should operate on. Tag names in the target constituents are surrounded by parentheses while set names are not. Contextual conditions that are necessary for the application of the operation of the rule are defined in the CONTEXT CONDITION(S). The definitions are relative to the target reading. The target reading is given the position zero and the context is defined from the referent zero where elements before the zero are represented by negative numbers and elements after the zero by positive numbers. The context condition has two elements, the position number which is relative to the referent zero and the name of a tag or set which defines a reading or a word form of a cohort in that relative position. Below are cases of relative positions in the context conditions, as given by Tapanainen (1996: 27)

- . (-1 A)
The previous cohort contains a reading which has a tag belonging to set A.
- . (1 A)
The following cohort contains a reading which has a tag belonging to set A.
- . (2 A)
The second cohort to the right contains a reading which has a tag belonging to set A.
- . (3 A)
The third cohort to the right contains a reading which has a tag belonging to set A.

The position in the context condition can be unbound to any particular number. An asterisk ‘*’ is prefixed to the number to unbound the position, from the number, rightward and leftward to the final and initial sentence boundaries for positive and negative numbers respectively. For example:

- . (*-1 A)
There is a cohort to the left that contains a reading which has a tag belonging to set A
- . (*1 A)
There is a cohort to the right that contains a reading which has a tag belonging to set A
- . (*2 A)
Two or more cohorts to the right, there is a cohort that contains a reading which has a tag belonging to set A

The character ‘C’, for “careful” is suffixed to the number to restrict the context condition to be satisfied only in unambiguous contexts (Tapanainen 1996: 28).

It is also possible to negate the context condition. This is done by starting the context condition with the key word NOT, e.g. (NOT 1 N), which means that the following cohort should not contain a reading which has a tag belonging to set N.

3.3.2.3 The arrangement of rules

In the rule file, rules can be arranged in more than one section, each section under the same header “CONSTRAINTS”. “The CONSTRAINT sections are applied in the order in which they appear in the rule file. First, the first section is applied alone, and when it can no longer disambiguate the first and the second sections are applied together, etc” (Tapanainen 1996: 34). In spite of the possibility of ordering rules according to the CONSTRAINT sections in the rule file, representation of rules in the section does not assume any rule order. The rules are independent from each other and can be applied in any order. As Karlsson (1994: 122) puts it, “The constraints do not have the usual task of telling which sentences are correct ... [but] are ... more directly geared towards the basic task of parsing...”.

3.3.3 Heuristic disambiguation constraints

Although the main goal of the CGP framework is to provide a formalism for the constraints that are capable of attaining perfect analysis, the framework allows the use of guessing rules in addition to the ordinary rules. The guessing rules, are however, not obligatory in the rule file and they are only added to the file on the wishes of the grammarian. Such rules are represented under the section HEURISTIC-DISAMBIGUATION- CONSTRAINTS. According to Karlsson (1995b: 68), “these constraints are formulated just like ordinary constraints and they are applied only when the ordinary constraints find no more applications”. Where the heuristics are not needed, the section under which they are represented should be omitted or represented by NIL.

3.4. TOOLS FOR ANALYSING THE SWAHILI CORPUS

This section presents tools for morphological annotation of the Swahili corpus: the two-level-based Swahili morphological analyser (SWATWOL) and the Constraint Grammar-based Swahili morphological disambiguator (SWACGP). Hurskainen has developed the tools at the University of Helsinki (Hurskainen 1992; 1996). The description of terms in the Swahili corpus for this study relies on these tools.

3.4.1 Background to the tools

The journey towards the annotation of an unrestricted Swahili text, particularly morphological annotation, began in 1987 when Hurskainen began to develop the Swahili-based two-level morphological analyser (Hurskainen 1992). The acronym SWATWOL, which stands for ‘Swahili Two-level’, literally means application of the Two-level model in morphological analysis of Swahili texts. The development of the tools began with building up the corpus of Swahili text, which did not exist previously. Then followed the creation of the two-level-based Swahili lexicon and rules. The development of the Swahili Constraint Grammar parser (SCGP) came later as the tool for handling morphological ambiguities produced by the morphological analyser. After a process of testing and refining, the performance of the tools in annotating an unrestricted Swahili corpus has proved a high level of effectiveness, both in recall and precision. According to Hurskainen (1996: 568), the tools achieved 99.8% recall and close to 100% precision in annotating an unrestricted text corpus of one million words.

3.4.2 The SWATWOL morphological analyser

SWATWOL is a composition of three components: the annotation scheme, the lexicon and the rules.

3.4.2.1 SWATWOL annotation scheme

The annotation scheme (cf. Chapter2) for SWATWOL applies a number of symbols for coding various word-based information. The information which is coded is mainly morphological, including inflectional, derivational and parts of speech features. The symbols are largely based on a number of Swahili morphological works (Ashton, 1944; Polome 1967; Mukama, 1976; Khamisi, 1988). Moreover, there are symbols for coding other word-based features, such as etymological, semantic terminological features and, to a lesser extent, word-particular syntactic features. Additionally, there are symbols for defining clause boundaries in the input text. The

SWATWOL annotation scheme is presented below, as produced for us by Hurskainen, the developer of the scheme, on 19th December, 1999.

Clause boundary tag

**CLB clause boundary. ; ? ! ili, kwamba

Tags for marking noun class features

There are the following tags indicating the classes of nouns. Note that most of the tags show a pair of noun classes, which means that the noun takes one class prefix in the singular and another in the plural. Tags with one class only refer to such nouns as do not have a singular/plural distinction. Among such nouns are abstract nouns (uhuru 11-SG, 'freedom'), mass nouns (maji 6-PL, 'water') and verbal nouns (kusoma 15-SG, 'reading').

1/2-PL	noun class 1/2 plural, noun prefix or concord	watu, walimu, wezi
1/2-SG	noun class 1/2 singular, noun prefix or concord	mtu, mwalimu, mwizi
11-SG	noun class 11 singular, noun prefix or concord	umuhimu, wizi
11/10-PL	noun class 11/10 plural, noun prefix or concord	nyimbo, funguo
11/10-SG	noun class 11/10 singular, noun prefix or concord	wimbo, ufunguo
11/6-PL	noun class 11/6 plural, noun prefix or concord	magonjwa, makucha
11/6-SG	noun class 11/6 singular, noun prefix or concord	ugonjwa, ukucha
15-SG	noun class 15 singular, noun prefix or concord	kusoma
16-SG	noun class 16 singular, concord	pa
17-SG	noun class 17 singular, concord	ku
18-SG	noun class 18 singular, concord	mu
3/4-PL	noun class 11/10 plural, noun prefix or concord	miti
3/4-SG	noun class 3/4 singular, noun prefix or concord	mti
5/6-PL	noun class 5/6 plural, noun prefix or concord	mawe
5/6-SG	noun class 5/6 singular, noun prefix or concord	jiwe
5a/6-PL	noun class 5a/6 plural, noun prefix	masomo
5a/6-SG	noun class 5a/6 singular, noun prefix	somo
6-PL	noun class 6 plural, noun prefix	madaraka
7/8-PL	noun class 7/8 plural, noun prefix or concord	viti, vyakula
7/8-SG	noun class 7/8 singular, noun prefix or concord	kiti, chakula
9/10-0-PL	noun class 9/10 plural, no class marker, noun prefix	taa
9/10-0-SG	noun class 9/10 singular, no class marker, noun prefix	taa
9/10-NI-PL	noun class 9/10 plural, class marker n-, m-, or ny-, noun prefix	nta, mbwa, nyota
9/10-NI-SG	noun class 9/10 singular, class marker n-, m-, or ny-, noun prefix	nta, mbwa, nyota
9/10-PL	noun class 9/10 plural concord	anazisoma
9/10-SG	noun class 9/10 singular concord	anaisoma
9/6-0-PL	noun class 9/6 plural, noun prefix	masomo
9/6-0-SG	noun class 9/6 singular, noun prefix	somo
LOC	locative	hapa, huko
LOC-16	locative of the type 'pa'	hapa, pale
LOC-17	locative of the type 'ku'	huku, kule
LOC-18	locative of the type 'mu'	humu, mle

Tags related to part of speech features, excluding features related to noun class concords

:enye	pronoun with pronominal prefixes	mwenye mali
:hV	demonstrative pronoun of the type hV(C)V	hiki, haya, hii
:le	demonstrative pronoun of the type (C)Vle	kile, yale, ile
:ote	pronoun with pronominal prefixes	siku zote
A	adjective	zuri

A-INFL	inflecting adjective	viti vizuri
A-UNINFL	uninflecting adjective	viti haba
AD-ADJ	qualifier of an adjective	kabisa, sana
ADJ	adjective	zuri, haba
ADV	adverb	tena
ADV:ki	adverbial prefix 'ki' of nouns	ki ujanja
CARD	cardinal numeral	tatu
CC	coordinating conjunction	na
CONJ	conjunction	ili
DEM	demonstrative pronoun	ile, hiki, huyo
EXCLAM	exclamation	ala!
GEN-CON	genitive connector	kitabu cha mtoto
INTERROG	interrogative	lini? mbona?
N	noun	kitu
NUM	numeral	ishirini
NUM-INFL	inflecting numeral	wanne
NUM-UNINFL	uninflecting numeral	sita
ORD	ordinal numeral	mtu wa sita
PERS	personal pronoun	mimi
PL	plural	watu
PL1	1. person plural	tunasoma
PL2	2. person plural	mnasoma
PL3	3. person plural	wanasoma
POSS	possessive pronoun	kiti changu
PRON	pronoun	mimi, hiki
PROP	proper name	Ali, Mombasa
PROP-AN	proper name, animate	Ali
SELFSTANDING	selfstanding subject particle	yu, zi
SG	singular	mtu
V	verb	kusoma

Tags referring to features in verbs

1/2-PL1	noun class 1/2 1. person plural	anatu f ikiri
1/2-PL1-SP	noun class 1/2 1. person plural, subject prefix	tun asoma
1/2-PL2	noun class 1/2 2. person plural	anawa f ikiri
1/2-PL2-SP	noun class 1/2 2. person plural, subject prefix	m nasoma
1/2-PL3	noun class 1/2 3. person plural	anawa f ikiri
1/2-PL3-SP	noun class 1/2 3. person plural, subject prefix	w anasoma
1/2-SG1	noun class 1/2 1. person singular	anani f ikiri
1/2-SG1-SP	noun class 1/2 1. person singular, subject prefix	n in f ikiri
1/2-SG1-na-PAST	noun class 1/2 1. person singular, past tense with subject prefix 'ni' changed as 'na'	n alisoma
1/2-SG2	noun class 1/2 2. person singular	anaku f ikiri
1/2-SG2-SP	noun class 1/2 2. person singular, subject prefix	u nasoma
1/2-SG3	noun class 1/2 3. person singular	anam f ikiri, anam w omba
1/2-SG3-SP	noun class 1/2 3. person singular, subject prefix	anasoma
11-SG-SP	noun class 11 singular, subject prefix	uzuri u mezidi
15-SG-SP	noun class 15 singular, subject prefix	kusoma kun afurahisha
16-SG-SP	noun class 16 singular, subject prefix	hapa p amechafuka
17-SG-SP	noun class 17 singular, subject prefix	huko k umechafuka
18-SG-SP	noun class 18 singular, subject prefix	mle m mechaguka
3/4-PL-SP	noun class 3/4 plural, subject prefix	miti i naota
3/4-SG-SP	noun class 3/4 singular, subject prefix	mti u naota

5/6-PL-SP	noun class 5/6 plural, subject prefix	macho yanaona
5/6-SG-SP	noun class 5/6 singular, subject prefix	jicho linaona
7/8-PL-SP	noun class 7/8 plural, subject prefix	viti vimevunjika
7/8-SG-SP	noun class 7/8 singular, subject prefix	kiti kimevunjika
9/10-PL-SP	noun class 9/10 plural, subject prefix	taa zinawaka
9/10-SG-SP	noun class 9/10 singular, subject prefix	taa inawaka
ARCHAIC-PR-yua-SP	subject prefix of an archaic construction	yuaenda
ASSOC:na	associative or genitive structure based on the root 'na'	nina, ninacho anaweza kusaidia anasomesha,
AUXMOD	auxiliary modus	
CAUS	causative	anatembeza anasomesha, anasimamisha
CAUS:ish	causative with marker 'ish'	anapisha anawekeza amekwishasoma
CAUS:sh	causative with marker 'sh'	keshasoma
CAUS:z	causative with marker 'z'	kishasoma
COMPL	completed action	nisingalisoma
COND-COLLOQ:kesh	colloquial conditional form indicating completed action, marker 'kesh'	nisingesoma
COND-COLLOQ:kisha	colloquial conditional form indicating completed action, marker 'kisha'	nisingelisoma
COND-NEG:singali	conditional negative form, marker 'singali'	nijaposoma
COND-NEG:singe	conditional negative form, marker 'singe'	nikisoma
COND-NEG:singeli	conditional negative form, marker 'singeli'	nikiishasoma
COND:japo	conditional form, marker 'japo'	nikishasoma
COND:ki	conditional form, marker 'ki'	ningalisoma
COND:kiisha	conditional form indicating completed action, marker 'kiisha'	ningesoma
COND:kisha	conditional form indicating completed action, marker 'kisha'	ningelisoma
COND:ngali	conditional form, marker 'ngali'	ningeshasoma
COND:nge	conditional form, marker 'nge'	kasome
COND:ngeli	conditional form, marker 'ngeli'	nitasoma
COND:ngesha	conditional form indicating completed action, marker 'ngesha'	nitakachosoma
CONSEQ:ka-e	consecutive form	sitosoma
FUT:ta	future tense with marker 'ta'	nisomaye
FUT:taka	future tense with marker 'taka'	husoma
FUT:to	future tense with marker 'to'	Nenda kasome!
G-REL	general relative	soma! someni!
HABIT-SP	subject prefix 'hu' of the habitual form	kusoma
HORT-COLLOQ:ka-e	colloquial hortative form	kutosoma
IMP	imperative	kutokusoma
IMP-PL2	imperative, 2. person plural	amepotea
INF	infinitive	unakwendapi?
INF-NEG	negative infinitive	kula
INF-NEG:ku	negative infinitive with 'ku' repeated	kasoma
INTR	intransitive	kaishasoma
Interrog:pi	interrogative with suffix 'pi'	akasoma
MONOSLB	monosyllabic verb	
NARR-COLLOQ:ka-a	colloquial narrative past tense	
NARR-COLLOQ:kaisha-a	colloquial narrative perfect tense	
NARR:ka	narrative form with subject prefix	

NEG	negative	sisomi, hasomi
NEG-REL	negative relative form	nisiyesoma
NEG-SG1	negative, 1. person singular	sisomi, sitasoma
NEG-SG2	negative, 2. person singular	husomi, hutasoma
NEG-i	negative, with verb-final 'i'	sisomi, hasomi
OBJ	object prefix	nakisoma
PASS	passive	kitabu kinasomwa
PAST	past tense	nilisoma
PAST-NEG:ku	past tense negative	sikusoma
PAST:liisha	past tense with the marker 'liisha'	niliishasoma
PAST:lisha	past tense with the marker 'lisha'	nilishasoma
PERF-NEG:ja	perfect tense negative with the marker 'ja'	sijasoma
PERF:me	perfect tense with the marker 'me'	nimesoma
PERF:meisha	perfect tense with the marker 'meisha'	nimeishasoma
PERF:mekwisha	perfect tense with the marker 'mekwisha'	nimekwishasoma
PERF:mesha	perfect tense with the marker 'mesha'	nimeshasoma
PERF:sha	perfect tense with the marker 'sha'	nishasoma
PL1-SP 1.	person plural, subject prefix	tunasoma
PL2-SP 2.	person plural, subject prefix	mnasoma
PL3-SP 3.	person plural, subject prefix	wanasoma
PR:a	present tense with the marker 'a'	asoma
PR:na	present tense with the marker 'na'	anasoma
PREP	prepositional verb form	ninakusomea kitabu
REC	reciprocal verb form	tunapendana
REFL	reflexive	anajjipenda
REL	relative marker in verb	ninayesoma
SBJN	subjunctive form of verb	nisome
SG1	1. person singular	mimi
SG1-SP	1. person singular, subject prefix	ninasoma
SG2	2. person singular	wewe
SG2-SP	2. person singular, subject prefix	unasoma
SG3	3. person singular	yeye
SG3-SP	3. person singular, subject prefix	anasoma
SP	subject prefix	ni, u, yu
STAT	stative verb form	kinasomeka
STAT:ik	stative verb form with the marker 'ik'	kinapikika
SV	verb with the argument structure subject+verb	kitabu kimepotea
SVO	verb with the argument structure subject+verb+object	mtoto anasoma kitabu
SVOO	verb with the argument structure subject+verb+object+indirect object	mama anansomea mtoto kitabu
TEMP	temporal	hapo awali
TEMP/LOC	temporal or locative	baada ya hapo
UNINFL-V:ni	uninflected verb 'ni'	mimi ni mtu
VB	verb base	soma
VFIN	finite verb	anasoma

Tags indicating the type of derivation in nouns

DER:a	derived form, extension -a	msoma
DER:ano	derived form, extension -ano	mabishano
DER:e	derived form, extension -e	lishe
DER:eo	derived form, extension -eo	upendeleo
DER:fi	derived form, extension -fi	ulafi

DER:fu	derived form, extension -fu	ulaanifu
DER:i	derived form, extension -i	usomi
DER:ia	derived form, extension -ia	demokrasia
DER:io	derived form, extension -io	mahudhurio
DER:ji	derived form, extension -ji	usomaji
DER:kio	derived form, extension -kio	mafanikio
DER:ko	derived form, extension -ko	kusanyiko
DER:ni	derived form, extension -ni	machungani
DER:o	derived form, extension -o	mwaliko
DER:sha	derived form, extension -sha	mtapisha
DER:shi	derived form, extension -shi	mcheshi, mpishi
DER:shio	derived form, extension -shio	tishio
DER:sho	derived form, extension -sho	tisho
DER:si	derived form, extension -si	mfuasi
DER:so	derived form, extension -so	utakaso
DER:u	derived form, extension -u	fungu, fonimu
DER:uo	derived form, extension -uo	ufunguo
DER:vi	derived form, extension -vi	mlevi
DER:vu	derived form, extension -vu	maumivu
DER:wa	derived form, extension -wa	mwajiriwa
DER:zi	derived form, extension -zi	ulanguzi, mlezi
DER:zo	derived form, extension -zo	mgandamizo

Tags referring to semantic features

AN	animate	punda, ndugu
AN/INAN	animate or inanimate	ajenti, edita, nyanya
HUM	human	afisa
IDIOM	idiom	punde si punde
INAN	inanimate noun	kiti

Etymological tags

AR	Arabic origin	kufikiri
IND	Indian origin	chapa
PO	Portuguese origin	mpira

Tags for term-domain feature

HC	Health Care term	
HC-V	Health Care Verbal noun term	kuambukiza
HC-N	Health Care Noun term	utapiamlo

Miscellaneous tags

COLLOC	collocations joined together in pre-processing	hapa_na_pale
COMP	comparative construction	kuliko, kupita, kushinda
DIM	diminutive	kijitabu
EMPH	emphatic form	mimi ndimi
NON-STD	non-standard form	sitofanya, muakilishi
NOUN+POSSESSIVE	noun plus possessive	mwenzangu
REDUP	reduplication	anasomasoma
RHET	rhetorical	Unakwenda, sio?

Syntactic tags inserted into the lexicon

@<A-AD	additional adjectival qualifier, head on the left	kitu kizuri kabisa
@<NQ	numerical qualifier, head on the left	viti viwili
@ADVL	adverbial	yaani, kadhalika
@ CC	co-ordinating conjunction	na
@CS	subordinating conjunction	ili, kwamba

The above scheme is open-ended in the sense that it can accommodate more symbols as the need arises. The scheme is meant to cover full morphological information for the Standard Swahili. However, the scheme does not stand for the absolute standard for the representation of Swahili morphology. Rather, the scheme is a reflection of symbols and their interpretation as used by the developer of the morphological analyser. This is in accordance with the following assertion by Leech:

“... there can be no claim that the annotation scheme represents ‘God’s truth’. Rather, the annotation scheme is made available to a research community on a *caveat emptor* principle. It does not come with any ‘gold standard’ guarantee, but is offered as a matter of practical usefulness only, on the assumption that many users will find it valuable to use a corpus with annotations already built in, rather than to have to devise and apply their own annotations and annotation schemes from scratch (a task which could take years to accomplish)” (Leech 1997a: 6-7)

Moreover, the scheme is open for adaptation without affecting the performance of the analyser.

3.4.2.2 The SWATWOL lexicon

The SWATWOL lexicon is based on Koskenniemi’s (1983) two-level lexicon system. The lexicon is structured into sub-lexicons and continuation classes through which the description of Swahili morphology is carried out. Since Swahili morphology is characterised by both prefixation and suffixation, the description places sub-lexicons for word base-forms or lemma in the middle position rather than in the initial position as is the case with languages which inflect only by suffixation.

The LEXICON Start forms the beginning of the lexicon structure with the following entries:

```
LEXICON Start
#      NORM;
#*     NORM;
#      NUM;
#*     PROPNAME;
#      COLLOC;
#*     COLLOC
```

The entries of the LEXICON Start have no lexical form but the symbols ‘#’ and ‘#*’. The two symbols indicate, respectively, that a word can begin with either a small letter or a capital letter. The continuation classes of the entries serve as gates for accessing morphological description of different words in Swahili as follows:

NORM opens to words of different part-of- speech categories, such as, nouns , verbs, adjectives and adverbs;
NUM leads to words that can be categorised as numerals;

PROPNAME opens to words, which function as proper names, such as Dar-es- Salaam and Kilimanjaro;

COLLOC branches to words that are defined as collocations, such as *mara_kwa_mara* and *moja_kwa_moja*.

The above continuation classes, with their entries, have the following representation in the lexicon structure:

LEXICON NORM

Verb;
Noun;
Adjective;
Adverb;
Pronoun;
Preposition;

LEXICON NUM

Numeral;

LEXICON PROPNAME

Kilimanjaro;
Dar -es- Salaam;
Matata;

LEXICON COLLOC

mara_kwa_mara;
moja_kwa_moja;

The entries in the above sub-lexicons do not contain lexical forms, but they contain only continuation classes which open the description to various categories of words. From here, the description continues with the same style of continuation classes and sub-lexicons according to the specific categories of words. The symbol ‘#’, which signals the end of a word form, is the final continuation class in the description.

3.4.2.2.1 Description of verb morphology

The description of Swahili verbs in the lexicon is accessed via the sub-lexicon ‘Norm’. The verb structure is specified on three dimensions: inflectional prefixes, root and derivational suffixes. Description of the inflectional prefixes covers features such as negation, subject concord, tense and aspect, relative marker and object concord. Normally, the features are constrained in a linear chain of negative marker, subject concord, tense, relative marker, and object marker. However, this is not always the case. In some cases such as in imperatives, verb forms can function without any of these features. Moreover, the feature for negation is only present when negation is used, and when it is not present, the feature for subject concord occupies the initial position. The SWATWOL lexicon is structured to accommodate these variations by “creating parallel sub-lexicons for facilitating routes through the array of prefixes ... all [such] routes meet in the sub-lexicon of verb roots ... There is a total of 22 sub-lexicons of verbal prefixes facilitating various combinations of verbal inflections” (Hurskainen 1992: 99).

Verb roots are described in a number of sub-lexicons to facilitate all the possible

combinations of derivation. Only basic verb roots are included in the sub-lexicons of verb roots. The roots of monosyllabic verbs are described in three sub-lexicons: *+kuR*, *-kuR* and *-kuRd*. The entries of the sub-lexicons have the same lexical form, the monosyllabic verb roots, but different continuation classes, as follows:

LEXICON *+kuR*

l A;
 f A;
 nyw A;
 ny A;
 j A;
 p A;
 ch A;
 chw A;
 w A;

LEXICON *-kuR*

l E;
 f E;
 nyw E;
 ny E;
 j E;
 p E;
 ch E;
 chw E;
 w E;

LEXICON *-kuRd*

l /d;
 f /d;
 nyw /d;
 ny /d;
 j /d;
 p /d;
 ch /d;
 chw /d;
 w /d;

The continuation class ‘A’ leads the description of the monosyllabic verb to the final vowel ‘a’, which is always the case with such verbs when their roots contain the prefix ‘ku’, such as in *kula* “to eat”, *kufa* “to die”, *kuja* “to come”, *kunya* “to shit”, *kunywa* “to drink”. The continuation class ‘E’ takes the description to the verb final vowel ‘e’ in imperative forms, such as: *ale* “1/2SG3 eat”, *afe* “1/2SG3 die”, *1/2SG3 aje* “come”. The continuation class /d opens the description to verbal derivations.

Description of sub-lexicons of verbal derivation is a continuation from the sub-lexicons of basic verb roots. The descriptions involve a total of 13 different formal derivation patterns with a total of 97 sub-lexicons. (Hurskainen 1992: 100). It is hoped, however, that with more research and testing with real language material the sub-lexicons of verbal derivation can be simplified.

3.4.2.2.2 Description of nouns and adjectives

The description of nouns is not as complex as that of verbs. Apart from derived nouns, the morphology of Swahili nouns is characterised by two morphological components, the prefix morpheme for class and number, and the root. In most cases the class markers are in pairs of singular and plural. There are more than 15 class markers in the description.

From the sub-lexicon 'NORM' the description of nominal morphology is accessed through the continuation class 'NOUN' whose entries contain continuation classes for different classes of nouns, as follows:

```
LEXICON Noun
M/WA;
M/MI;
JI/MA;
KI/VI;
N/N;
0/0;
U;
```

In the next step, each of the above continuation classes is represented as a sub-lexicon. The entries of such sub-lexicons contain number prefixes as their lexical forms, as in the following example of a sub-lexicon for the continuation class M/WA.

```
LEXICON M/Wa
mU   Nr1;
wA   Nr1;
```

The number prefix in the entries of each sub-lexicon, such as 'm' and 'wa' in the above sub-lexicon is directed to its respective nominal roots by the given continuation class. For example, the above continuation class 'Nr1' connects the nominal prefixes 'm' and 'wa' to their possible nominal roots. The connection is established through the sub-lexicon 'Nr1' below.

```
LEXICON Nr1
tu   # ;
nyama # ;
zee  # ;
ganga # ;
temi # ;
```

Certain nominal prefixes take roots from more than one class. Enlarging the number of sub-lexicons for nominal roots to 20 has accommodated such a tendency (Hurskainen 1992: 102). The description of derived nouns has taken a similar style of nominal prefix followed by a nominal root. Moreover, such nouns are defined as derived nouns. Proper nouns are accessed through the sub-lexicon '*PROP. Thus they have their own pattern. This is because proper nouns are not characterised by the nominal prefixes. Morphologically, they are indivisible and are accessed.

The description of adjectives categorises the adjectives into non-inflecting adjectives and inflecting adjectives. The inflecting adjectives are divided into numeral and non-numeral adjectives and they are described as prefixes and roots. The lexicon structure for the adjectives

starts with the sub-lexicon ‘Adjective’, which contains two entries, as follows:

LEXICON Adjective

A-INFL;
A-UNINFL;

The two entries of the above sub-lexicon contain continuation classes only. Such continuation classes form the next sub-lexicons in the description. The entries of the sub-lexicon ‘A-INFL’ contain continuation classes that open to the inflecting numeral and non-numeral adjectives, and the entries of the sub-lexicon ‘A-UNINFL’ contain lexical forms of the uninflecting adjectives with the word final tag ‘#’ as their continuation class. Below is the sub-lexicon ‘A-INFL’ with the continuation classes for inflecting numeral and non-numeral adjectives:

LEXICON A-INFL
NUM;
NON-NUM;

The description continues by representing the continuation classes ‘NUM’ and ‘NON-NUM’ as the next sub-lexicons and assigns their entries with the lexical forms of their respective prefixes. The lexical forms are given continuation classes, which connect them with their respective roots.

The description of words of other categories, such as adverbs and demonstratives, is done following the same mechanism of sub-lexicons and continuation classes.

3.4.3. SWATWOL rules

A rule section is an obligatory component of the Two-level-based morphological analyser. The rules denote relations between lexical and textual representations. Moreover, they describe variations that may occur between the two representations. SWATWOL (1992) has 18⁶ rules. The rules are based on variations between representations in the SWATWOL lexicon and textual forms. One case of variations that are described by the rules is the mapping of the lexical form of the noun class 9/10 prefix, which is represented as ‘NI’, to the surface forms ‘n’, ‘m’, ‘ny’ and ‘0’, as the following examples indicate.

Lexical form:	NIgao	NIumba	NIbao	NIpombe
Surface form:	n0gao “shield”	nyumba “house”	m0bao “timber”	00pombe “alcohol”

The above mapping of the lexical form NI to the surface form shows the following variations: N:n, I:y, I:0, N:0, and N:m. Some of the SWATWOL rules are formulated to describe these variations. Other variations that are described by the rules include devocalisation of U in front of vowels, and change of the final verb vowel ‘A’ to ‘e’ and to ‘i’ in subjunctive and negative forms. However, not all variations are touched upon by the rules. Certain variations are difficult to handle by the rules because they occur in environments that cannot be exclusively defined. Such variations are handled in the lexicon.

⁶ The full SWATWOL rule schema is set out in Hurskainen (1992: 116-118).

3.4.4 SWATWOL morphological heuristics

The SWATWOL is designed for analysing an unrestricted text, i.e., analysing every word in any input text file. However, this goal can be achieved only if the lexicon contains all the words in the language. Since the language is growing and new words are constantly being introduced there is always a chance of missing some words in the lexicon. To overcome this problem, the SWATWOL has an optional component of heuristics for handling words that are missing in the lexicon. The heuristics, i.e. rules for predicting morphological analysis for words that cannot be analysed through the lexicon, supplement the lexicon during the process of word analysis.

3.5. ANALYSING A SWAHILI CORPUS BY MEANS OF THE SWATWOL

Analysing a text with SWATWOL involves the pre-processing of the text to be analysed⁷, and the compilation of the SWATWOL lexicon and rules. The morphological analyser, which uses the compiled lexicon and rules, takes the processed text as its input and outputs the same text in full morphological analysis. The following analysis exemplifies the analysis of the sentence:

Baadhi ya magonjwa yanayotokana na kinyesi ni kipindupindu, kuhara, kuhara damu, homa ya matumbo, polio, na homa ya ini, “some of the diseases which originate from faeces are cholera, diarrhoea, typhoid, polio and hepatic fever”

```
"<*baadhi>"
  "baadhi" 9/10-0-SG N AR
  "baadhi" 9/10-0-PL N AR
"<ya>"
  "ya" 3/4-PL GEN-CON
  "ya" 9/10-SG GEN-CON
  "ya" 5/6-PL GEN-CON
  "ya" 5/6-PL
"<magonjwa>"
  "gonjwa" 5a/6-PL N
  "ugonjwa" 11/6-PL N
"<yanayotokana>"
  "tokana" 5/6-PL-SP VFIN PR:na 3/4-PL REL V SV SVO STAT REC
  "tokana" 5/6-PL-SP VFIN PR:na 5/6-PL REL V SV SVO STAT REC
  "tokana" 5/6-PL-SP VFIN PR:na 9/10-SG REL V SV SVO STAT REC
"<na>"
  "na" CC @CC
"<kinyesi>"
  "kinyesi" 7/8-SG N ' excretum '
  "kinyesi" 9/10-0-SG N
  "kinyesi" 9/10-0-PL N
"<ni>"
  "ni" ADV:ni
  "ni" SG1
```

⁷ The preprocessing process prepares the text for the analysis. It involves marking of sentence boundaries, recognition of fixed structures, changing of all upper case letters to lower case letters by prefixing them with the character ‘*’ and segmentation of the text into word tokens. All these processes are not done by the SWATWOL but by different BETA programs (Karlsson and Koskeniemi 1990).

```

"<kipindupindu>"
  "kipindupindu" 9/10-0-SG N
  "kipindupindu" 9/10-0-PL N
"<,>"
"<kuhara>"
  "hara" INF V AR SV
"<,>"
"<kuhara>"
  "hara" INF V AR SV
"<damu>"
  "damu" 9/10-0-SG N AR
  "damu" 9/10-0-PL N AR
"<,>"
"<homa>"
  "homa" 9/10-0-SG N AR
  "homa" 9/10-0-PL N AR
"<ya>"
  "ya" 3/4-PL GEN-CON
  "ya" 9/10-SG GEN-CON
  "ya" 5/6-PL GEN-CON
  "ya" 5/6-PL
"<matumbo>"
  "tumbo" 5a/6-PL N ' belly, womb '
"<,>"
"<polio>"
  "polio" 9/10-0-SG N
  "polio" 9/10-0-PL N
"<na>"
  "na" CC @CC
"<homa>"
  "homa" 9/10-0-SG N AR
  "homa" 9/10-0-PL N AR
"<ya>"
  "ya" 3/4-PL GEN-CON
  "ya" 9/10-SG GEN-CON
  "ya" 5/6-PL GEN-CON
  "ya" 5/6-PL
"<ini>"
  "ini" 5a/6-SG N ' liver '
  "ini" 9/10-0-SG N
  "ini" 9/10-0-PL N
"<.$>"

```

In the above analysis, with the exception of *kuhara*, *matumbo* and *na*, words are ambiguously analysed, i.e. they are assigned more than one reading. Hurskainen (1996) gives a number of causes of ambiguity in the morphological analysis of the Bantu languages, particularly Swahili, as follows:

“First, nouns are grouped into more than ten-marked noun classes. The marking of these classes extends across the noun phrase whereby the noun governs the choice of the markers in dependent constituents. Second, verbs inflect stem-initially and mark the subject, object, and relative referent by prefixes, whereby the actual form of each prefix is governed by the noun class of the noun it refers to. In addition, verb derivation also adds to the complexity of verbal morphology” (Hurskainen 1996: 568).

According to Hurskainen the ambiguity rate in Swahili is almost 50%, which means that half of word-form tokens are ambiguous in at least two ways. Some word tokens, though very few, are assigned even more than 60 readings. Each reading in an ambiguous analysis represents a contextual possibility. Morphological annotation is complete only when each word form is assigned with a single contextually appropriate reading. The Swahili morphological disambiguator, the SWACGP, performs such a task.

3.6 DISAMBIGUATING SWATWOL ANALYSIS

The Swahili Constraint Grammar Parser (SWACGP) performs the task of disambiguating the SWATWOL morphological analyses. Moreover, the SWACGP performs Swahili syntactic mapping and disambiguation. The disambiguator functions according to Constraint Grammar formalism (cf. 2.3.3). The SWATWOL morphological analyses provide the environment and information for formulating the constraint rules for morphological disambiguation. According to Hurskainen (1996: 571), the SWACGP had to date a total of 656 rules in four different sections. The rules are formal statements which constrain the disambiguator to select or remove certain readings in cohorts according to the given context. Contexts for the rules are readings in cohorts that occur before or after a cohort from which the disambiguation is effected. For example, the disambiguation of the analysis of *ya* in the following analysis of *homa ya ini* ‘‘hepatic fever’’.

```
"<homa>"
    "homa" 9/10-0-SG N AR
    "homa" 9/10-0-PL N AR
"<ya>"
    "ya" 3/4-PL GEN-CON
    "ya" 9/10-SG GEN-CON
    "ya" 5/6-PL GEN-CON
    "ya" 5/6-PL
"<ini>"
    "ini" 5a/6-SG N 'liver '
```

is facilitated by the following rule;

```
SELECT (9/10-SG)
    (-1 9/10-0-SG);
```

By this rule, the disambiguator selects, from the analyses of ‘*ya*’, the reading that contains the tag 9/10-SG, because there is the reading in the preceding cohort (-1) which contains the tag 9/10-0-SG. The same mechanism is used to constrain the disambiguator for removing unwanted readings by using the operator ‘REMOVE’. For example, REMOVE can be used in a rule for disambiguating the analysis of ‘*homa*’ ‘‘fever’’ in the following analysis.

```
"<homa>"
    "homa" 9/10-0-SG N AR
    "homa" 9/10-0-PL N AR
"<ya>"
    "ya" 9/10-SG GEN-CON
"<ini>"
    "ini" 5a/6-SG N 'liver '
```

Such a rule becomes:

REMOVE (9/10-0-PL)
(1 (9/10-SG) AND (GEN:CON) ;

This rule constrains the disambiguator to remove a reading that contains the tag 9/10-0-PL if immediately followed by a cohort with a reading which contain the tags 1 9/10-SG, and GEN-CON.

The morphological analyser produces all the tags that are used in the rules.

According to Hurskainen (1996), the performance of the SWACGP in morphological disambiguation was not yet 100% - it was less by 4.9%, however, its output was considered promising given that the disambiguator was still in process of development. Hurskainen points out that the common types of ambiguity that still remained included: noun vs. adverb, adjective vs. adverb, noun vs. conjunction, verb (imperative) vs. noun and verb (infinitive) vs. noun. Also, to a lesser extent, the genitive connectors *ya* and *wa* and possessive pronouns.

The SWACGP morphological disambiguator takes the output of SWATWOL as its input. The disambiguator, by using the constraint rules, removes the readings that are contextually inappropriate and outputs the analysed text with cohorts, which contain only contextually legitimate readings. "It is easy to view the [disambiguator] as a textual machine that removes some lines from the input but does not produce new ones" (Tapanainen 1996: 6). Below is the disambiguated form of the analysis of the above sentence, *Baadhi ya magonjwa yanayotokana na kinyesi ni kipindupindu, kuhara, kuhara damu, homa ya matumbo, polio, na homa ya ini.*

"<*baadhi>"
 "baadhi" 9/10-0-SG N AR
"<ya>"
 "ya" 9/10-SG GEN-CON
"<magonjwa>"
 "ugonjwa" 11/6-PL N
"<yanayotokana>"
 "tokana" 5/6-PL-SP VFIN PR:na 9/10-SG REL V SV SVO STAT REC
"<na>"
 "na" CC @CC
"<kinyesi>"
 "kinyesi" 7/8-SG N
"<ni>"
 "ni" ADV:ni
"<kipindupindu>"
 "kipindupindu" 9/10-0-SG N
"<,>"
"<kuhara>"
 "hara" INF V AR SV
"<,>"
"<kuhara>"
 "hara" INF V AR SV
"<damu>"
 "damu" 9/10-0-SG N AR
"<,>"
"<homa>"
 "homa" 9/10-0-SG N AR
"<ya>"
 "ya" 9/10-SG GEN-CON


```

"<matumbo>"
  "tumbo" 5a/6-PL N '
"<_>"
"<polio>"
  "polio" 9/10-0-SG N
"<na>"
  "na" CC @CC
"<homa>"
  "homa" 9/10-0-SG N AR
"<ya>"
  "ya" 9/10-SG GEN-CON
"<ini>"
  "ini" 5a/6-SG N
"<.$>"

```

Each cohort in the above disambiguated text has one reading that is contextually legitimate. This means that all ambiguities have been effectively resolved, though this is not always the case. As was mentioned above, in most cases, a few cases of unresolved ambiguities cannot be ruled out. However, the above output represents our presupposition of a morphologically annotated corpus, whereby each word form in the corpus is described by a number of features or tags associated with different attributes used for morphological description of the lexicon.

3.7 TAGGING TERM-WORDS IN THE MORPHOLOGICALLY ANALYSED CORPUS

Morphological annotation describes word forms in a corpus by tags that are used in the lexicon of the analyser to represent features by which the words are analysed. During the analysis processes the tags are ordered in the readings of the analysed corpus according to the features described in the lexicon. For example, in the previous annotated Swahili corpus, tags in the readings represent the features: BASE FORM, INFLECTIONS, GRAMMATICAL CATEGORY, ETYMOLOGY and SYNTACTIC CATEGORY. Selection of features and their values for analysing words in a corpus depends on the purpose of the analysis. Since the focus of this study is that terms in the annotated domain specific corpus should be tagged uniquely as entities of a particular domain, the study introduces the feature TERM-DOMAIN to facilitate the description. The feature is represented by domain-tags and it is discriminative in the sense that it is concerned only with term-words in the corpus.

The study proposes two ways by which term-words in the morphologically analysed corpus can be tagged with the domain-tag. The first way is by incorporating the domain-tag into the morphological analyser. This is done by identifying term-words in the corpus to be analysed and marking the terms in the lexicon of the analyser by the relevant term-domain tag. The marking is carried out in the sub-lexicon root of the lexicon. During the analysis, the analyser inserts the domain-tag into all words in the corpus that are derived from the roots tagged by the domain tag in the lexicon. The second way is by incorporating the domain-tag in the BETA rule file⁸ system and inserting the tag into the morphologically analysed corpus by the system. This approach is carried out through the following procedures: (1) morphological analysis of the corpus (2) identification of term-words in the corpus (3) morphological analysis of the identified term words to obtain their base-forms as analysed in the corpus (4) representation of the term base-forms in the rule file in a way that base-form representation corresponds to a rewrite rule for the BETA

⁸ Structure and features of the BETA rule file are well elaborated in Hurskainen (1992).

system. The following is a portion of base-forms of literature terms represented as rules in the BETA rule file and used for the tagging of literature term-words in the analysed corpus by the BETA system:

```
RULES
! Supply base-forms of literature term-words with a literature domain ! tag
(LT-N for noun base forms or LTV for verb base forms
!
lc rc sc rs mv md
0 B 1 0 5 1

"ala"; "ala" LT-N ;
"arudhi" ; "arudhi" LT-N ;
"bahari"; "bahari" LT-N ;
"data"; "data" LT-N ;
"dhamira"; "dhamira" LT-N ;
"dhana"; "dhana" LT-N ;
```

As rules, each of the above representations gives the command to rewrite, in the morphologically analysed corpus, the given base-form as the same form plus the domain-tag. Through the rule file, the BETA system tags term-words in the analysed corpus by rewriting the base-forms in the corpus as stated by the rules.

CHAPTER 4

CORPUS-BASED DEVELOPMENT OF TERM FORMATION PATTERNS

This chapter demonstrates techniques for the formulation of term patterns for the extraction of terms in domains. The techniques are developed based on a morphologically analysed, including term tagging, domain-specific corpus. The demonstration involves two aspects: First, discovery and unique tagging of term-words in the morphologically analysed corpus, and second, discovery of structural constraints of terms based on the analysed corpus and derivation of term formation patterns from the constraints. The two aspects are implemented on three tasks. First, the identification and tagging of health care term-words in the morphologically analysed Swahili health-care corpus. Second, the discovery of structural constraints of Swahili terms based on the terms analysed in the corpus. Third the formulation of term patterns on the basis of the constraints for the extraction of Swahili terms. The chapter begins with a short introduction to the principles of Swahili derivation and compounds. The principles underlie the structural constraints of Swahili terms. The rest of the chapter covers six tasks involved in the implementation of the techniques. First, incorporation of a domain tag, a health-care domain-tag for this case, in the Swahili morphological analyser for the tagging of health-care term-words in the analysed corpus. Second, tagging health-care term-words in the corpus. Third, identification of health-care multi-word terms as analysed in the corpus. Fourth, representation of terms identified as sequences of selected tags for the possible categories and subcategories of words in the terms. Fifth, specification of structural constraints of Swahili terms based on the categories of words and their relations as represented by the tags in the sequences represented. Sixth, formulation of Swahili term-patterns on the basis of the constraints specified.

4.1 STRUCTURAL CONVENTIONS FOR SWAHILI COMPOUNDING AND DERIVATION

Compounding and derivation are two processes of forming new terms by changing the existing words. The newly formed terms designate concepts that are more specific when compared with the ones represented by the original words. The processes are associated with concept specification, also known as concept determination (Sager 1990). Concept specification by derivation involves changing word categories, such as adjectives and verbs to nouns. On the other hand, concept determination by compounding involves combining two or more words into a head-modifier relationship to form a new unit with a new concept independent of the constituent parts. The expressions ‘compound term’ and ‘term collocation’ are commonly used interchangeably. Compounding consistently carries out specification of concepts by narrowing a wide meaning of words with modifiers indicative of a specific property of the wide meaning. One or more words can function as modifiers of the wide meaning. Words whose meanings are modified are the heads of the collocations. Depending on the language, the modifying elements can precede the heads (pre-modifiers) or follow the heads (post-modifiers). Swahili compounding applies post-modifiers; i.e. heads of collocations come first. A term collocation must contain at least one term-word as its head or as a component of its modifying element.

To date, a number of studies have been carried out on Swahili term formation by compounding and derivation (Batibo, 1988; Tumbo-Masabo, 1990; Mwansoko, 1993a,b). Swahili term formation by derivation involves a set of prefixes and suffixes, which are attached to the

derivative stems, normally verb stems. The prefixes belong to different noun class markers, whereas the suffixes belong to a finite set of the final nominal derivational markers. The process produces terms that are associated, regularly and systematically, with the derivative root, where each nominal suffix is associated with particular nominal prefixes. Batibo (1988) establishes the relationship between the prefixes and suffixes for Swahili nominal derivation as follows:

Nominal prefixes	corresponding nominal suffixes
I) 1/2(m-/wa-); 5/6 (ji-/ma-); 7/8 (ki-/vi); 14 (u-)	-i
II) 1/2 (m-/wa-); 14 (u-).	-aji
III) 3/4 (m-/mi-); 5/6 (ji-/ma-); 7/8 (ki-/vi); 14/11 (u- 0)	-o
IV) 14/6 (u- ma-)	-(v)u
V) 1/2(m-/wa-); 14 (u-).	-e

Based on the above relationship, Batibo categorises, conceptually, the resultant terms in the following groups:

Terms which are formed by the relation in (I) represent concepts that are associated with the accomplished or established state of a person, object or phenomenon. Examples of terms from this relationship are (all examples are taken from the field of primary health care):

muuguzi	(a verb stem -uguz-)	“nurse”
mzazi	(a verb stem -za-)	“delivered woman”
kizazi	(a verb stem -za-)	“uterus”
uzazi	(a verb stem -za-)	“birth”

Terms created from the relationship (II) represent concepts that are related to actions which are done regularly. Cases of such terms include:

mzalishaji	“midwife” (a verb stem -zalish-)
muokoaji	“rescuer” (a verb stem -oko-)
uponaji	“recovery” (a verb stem -pon-)
nyonyeshaji	“breast feeding” (a verb stem -nyonyesh-)
udungaji	“injecting” (a verb stem -dung-)

Terms developed by the relationship in (III) denoted objects, states, activities or phenomena resultant from, or associated with an action or process. The following are a few cases of such terms:

unyeo	(a verb stem -ny-)	“anus”
utando	(a verb stem -tand-)	“membrane”
koromeo	(a verb stem -korom-)	“larynx”
kichujio	(a verb stem -chuj-)	“filter”
mdungo	(a verb stem -dung-)	“injection”

Terms that are formed by the relationship in (IV) represent characteristics or qualities of an object or state resulting from an action or process, for example:

maumivu	(a verb stem -umi-)	“pain”
upungufu	(a verb stem -pungu-)	“deficiency”

The association in (V) is not very productive, though the resultant terms denote a person or a state, which has suffered the consequences of the action. One example of such resultant terms is:

uvimbe	(a verb stem -vimb-)	“swelling”
--------	----------------------	-------------------

The above relations of affixes in Swahili derivation and the categorisation of the resultant terms according to their concept representation reveal how general and systematic is the process of term derivation in the language. The possibility of deriving many terms from a single stem underscores the assertion that the stem or base form represents a nuclear element of single-word terms that are derived from a common base form.

The formation of Swahili terms by compounding involves modifying the existing words, especially nouns, with selected determinants to form new units for naming concepts. The modified words precede the modifying elements. Mwansoko (1993b) categorises the Swahili compounding process into compounding by juxtaposition and compounding by conjunction. Compounding by juxtaposition, also known as composition, involves attaching two words, usually both nouns without a conjunction. The following are cases of compounding by juxtaposition in the field of primary health care:

kifua ‘chest’ + kikuu ‘big’ > kifuakikuu	“tuberculosis”
mja ‘comer’ + mzito ‘heavy’ > mja mzito	“expectant mother”
pepo ‘evil spirit’ + punda ‘donkey’ > pepopunda	“tetanus”
kiini ‘nucleus’ + tete ‘soft’ > kiini tete	“embryo”
utando ‘spread’ + seli ‘cell’ > utandoseli	“cell membrane”
kidaka ‘catcher’ + tonge ‘food ball’ > kidakatonge	“epiglottis”
mfuga ‘keeper’ + nyuki ‘bee’ > mfuganyuki	“bee keeper”
uzazi ‘birth’ + salama ‘safe’ > uzazisalama	“safe delivery”

Some rare cases of compounding by juxtaposition are not motivated by the principle of nucleus determination. Examples of such cases are *pepo*+ *punda* > *pepopunda* “tetanus” and *kifua* + *kikuu* > *kifuakikuu* “tuberculosis”. In these cases the constituent nouns are not decomposable to the head-modifier relation. Such cases are not productive. However, compounding by juxtaposition is very productive in Swahili term formation. The following example demonstrates how productive the process is. The noun *Elimu* ‘education’, functions as a nucleus word and takes a number of modifiers to form compound terms, which designate different fields of study.

elimu ‘education’ + viumbe ‘living things’ > elimuviumbe	“biology”
+ uyoga ‘mushroom’ > elimuuyoga	“mycology”
+ mimea ‘plants’ > elimumimea	“botany”
+ wanyama ‘animals’ > elimuwanyama	“zoology”

Compounding by juxtaposition is largely used in the formation of verb-based compound terms where a transitive verb that is prefixed with a nominal class marker, functions as the head word and the object of the verb functions as a modifier. The following are examples of the verb-based juxtapositions:

transitive verbs with object juxtaposition

1st person singular

noun class marker

mpima ‘examine’ + *damu* ‘blood’ >
mpiga ‘beat’ + *sindano* ‘injection’ >
mtoa ‘give out’ + *dawa* ‘medicine’ >
mpima ‘examine’ + *maiti* ‘corpse’ >

noun compounds

mpimadamu “**blood examiner**”
mpigasindano “**injection person**”
mtoadawa “**pharmacist**”
mpimamaiti “**autopsist**”

The noun class marker that is attached to the transit verbs functions as a modifier that changes the verbs to a person who performs the action expressed by the verbs and the object noun specifies the type of action performed by the person. The two elements are combined in a head-modifier relation to form juxtaposition compound terms. Replacement of the noun class marker changes the meaning of the term. For example, the replacement of the above nominal class maker *m*, ‘first person singular’, by ‘*ki*’ ‘class 3 singular’ *changes* the meaning of the resultant compound term from the ‘doer’ of the action to the ‘instrument’ for doing the action.

Orthographically, Swahili juxtaposition compounds are represented in three different styles:

1) single word style, for example in the following:

<i>mja mzito</i>	(<i>mja</i> + <i>mzito</i>)	“ expectant mother ”
<i>pepopunda</i>	(<i>pepo</i> + <i>punda</i>)	“ tetanus ”
<i>kifuakikuu</i>	(<i>kifua</i> + <i>kikuu</i>)	“ tuberculosis ”
<i>mpimamaiti</i>	(<i>mpima</i> + <i>maiti</i>)	“ autopsist ”

2) two words with a connective hyphen, such as follows:

<i>mja-mzito</i>	“ expectant mother ”
<i>pepo-punda</i>	“ tetanus ”
<i>kifua-kikuu</i>	“ tuberculosis ”
<i>mpima-maiti</i>	“ autopsist ”

3) two unconnected words, such as:

<i>mja mzito</i>	“ expectant mother ”
<i>pepo punda</i>	“ tetanus ”
<i>kifua kikuu</i>	“ tuberculosis ”
<i>mpima maiti</i>	“ autopsist ”

However, the three styles do not affect the semantic representation of the terms. Mwansoko (1993b: 67) suggests that the style of single word form should be used for writing juxtaposition compounds. The reason for the suggestion is that juxtaposition compounds are pronounced as a single word with one primary-stressed syllable. However, the restriction might not be very useful in the corpus-based approach because the reality relies on the evidence gained from data. In that case representation of such terms in our work is left open to all the three styles as long as they occur in the text corpus.

Chaining modifiers of the headword by connectors effects compounding by conjunction. A genitive connector is commonly used to chain modifiers in Swahili conjunction compounding. The connector is formed by the subject concord of the modified word and the connective *-a*. Other connectors that are rarely used include prepositions, possessives, and, very rarely, co-

ordinators. A connector creates the head-modifier relationship between the connected words where the headword precedes the connectors. The headword is normally a noun or verbal noun whereas the modifying word can belong to any category of open words. Compounding by conjunction is very productive because it can be carried out progressively. The resultant compound term can further be specified or used as a modifier of another headword. The following are examples of Swahili conjunction compound terms chained by the genitive connector. The terms belong to the domain of primary health care:

mganga wa kienyeji (mganga ‘doctor’ + wa ‘of’ + kienyeji. ‘traditional’)	“ traditional healer ”
mji wa mimba (mji ‘town’ + wa ‘of’ + mimba ‘pregnancy’)	“ uterus ”
mrija wa mayai (mrija ‘tube’ + wa ‘of’ + mayai ‘eggs’)	“ fallopian tube ”
uvungu wa pango la kike (uvungu ‘cavity’ + wa ‘of’ + pango ‘cave’ + la ‘of’ + kike ‘female’)	“ female pelvic cave ”

4.2 INCORPORATING A HEALTH-CARE DOMAIN-TAG INTO THE MORPHOLOGICAL ANALYSER

As has been pointed out before, this study introduces a term-domain feature for the unique description of term-words in a domain specific corpus. The feature is represented in the morphologically analysed corpus with a domain tag that is inserted by a morphological analyser or by the BETA system. In this section we demonstrate how the feature is introduced into the morphological analyser through the lexicon of the analyser. The illustration involves tagging of health care term-words through the lexicon of the Swahili morphological analyser. The process involves two tasks: manual identification of health care term-words in the corpus, and second, marking the terms identified in the lexicon with the health care domain-tag.

4.2.1 Identifying health care term-words in the corpus

Health care term-words were manually identified in the training corpus on the basis of the criterion of concept naming (cf. 2.2.4.1). First, all nouns and verbal nouns in the corpus were automatically identified. Then, the nouns and verbal nouns were manually sorted out, where those which represented concepts in the domain were picked as term-words.

4.2.1.1 Extraction of nouns and verbal nouns in the corpus

The nouns and verbal nouns in the training corpus were identified after the corpus had been morphologically analysed by the Swahili morphological analyser and disambiguator (cf. 3.4). In the analysed corpus, among other symbols, the descriptions for nouns and verbal nouns carried the symbols “N” and “INF” respectively. The two symbols were used separately as a query language for extracting the nouns and verbal nouns by a template-matching program. The program compiled the symbols and matched them with their counterparts in the analysed corpus in order to extract the nouns and verbal nouns. The output of the extraction was sorted uniquely and a total of 1,724 nouns and 900 infinitive verbs was obtained. The following are cases of the nouns and infinitive verbs, which were extracted from the corpus:

555 "0"	mimba	“pregnancy”
507 "0"	mtoto	“baby/child”
383 "0"	mama	“mother”
300 "0"	damu	“blood”

286 "0"	ujauzito	“pregnancy”
253 "0"	mwanamke	“woman”
238 "0"	uzazi	“reproduction”
225 "0"	afya	“health”
202 "0"	njia	“way/passage”
196 "0"	mji	“town”
161 "0"	matatizo	“problems”
145 "0"	muda	“time/period”
136 "0"	ugonjwa	“disease”
133 "0"	hedhi	“menstruation”
125 "0"	dalili	“symptom”
123 "0"	uke	“vagina”
119 "0"	sehemu	“part”
110 "0"	maumivu	“pain”
108 "0"	siku	“day”
107 "0"	magonjwa	“diseases”
107 "0"	hatari	“danger”
106 "0"	yai	“egg”
106 "0"	hospitali	“hospital”
103 "0"	hali	“condition”
101 "0"	uchungu	“pain”
100 "0"	kipindi	“period”
93 "0"	mwili	“body”
92 "0"	tatizo	“problem”
85 "0"	uambukizo	“infection”
80 "0"	maji	“water”
80 "0"	kliniki	“clinic”
78 "0"	wiki	“week”
78 "0"	mbegu	“seed”
78 "0"	kazi	“job/function”
75 "0"	uzito	“weight”
75 "0"	uwezekano	“possibility”
75 "0"	mwanaume	“man”
5 "0"	ugumba	“barrenness”
5 "0"	*shingo	“neck”
5 "0"	*lishe	“nutrition”
5 "0"	*kiwango	“level”
5 "0"	*kitovu	“umbilical-cord”
5 "0"	*kansa	“cancer”
4 "0"	warts	“warts”
2 "0"	yeast	“yeast”
5 "0"	cell	“cell”
212 "v0"	kujifungua	“delivering/giving birth”
181 "v0"	kupata	“getting”
156 "v0"	kujamiiiana	“sexing”
93 "v0"	kutoka	“coming out”
93 "v0"	kufanya	“doing”
89 "v0"	kuzuia	“preventing”
79 "v0"	kutumia	“using”
71 "v0"	kuzaa	“bearing/delivering”
64 "v0"	kusababisha	“causing”
50 "v0"	kutoa	“giving out/discharging”
47 "v0"	kutokea	“happening”
45 "v0"	kutosha	“satisfying”
39 "v0"	kuweza	“managing”

39 "v0"	kuhakikisha	“ensuring”
37 "v0"	kusaidia	“helping”
34 "v0"	kuanza	“starting”
30 "v0"	kunyonyesha	“breast-feeding”
24 "v0"	kuongezeka	“increasing”
23 "v0"	kupevuka	“maturing”
22 "v0"	kula	“eating”
22 "v0"	kujisikia	“feeling-like”
21 "v0"	kutambua	“understanding”
21 "v0"	kupunguza	“reducing”
20 "v0"	kuepuka	“avoiding”
20 "v0"	kuambukiza	“infecting”
19 "v0"	kukojoa	“urinating/ejaculating”
16 "v0"	kukua	“growing”
16 "v0"	kuelewa	“understanding”
16 "v0"	kuchunguza	“investigating”
15 "v0"	kutibu	“curing”
15 "v0"	kukosa	“lacking”

Each of the extracted items above has three components, a number, a “0” or “v0” and a word form. The number represents the frequency of occurrence of the word form in the corpus and the “0” and “v0” stands for the respective symbols ‘N’ and ‘INF’ that were matched to extract the word form. The extracted word forms could be placed into three categories: nouns which begin with low case letters, such as *mimba* “pregnancy”; nouns which begin with an asterisk, such as **lishe*; “nutrition” and nouns which are in foreign orthography, such as “warts” and “yeast”. The asterisk at the beginning of a word form indicates that the word starts with the upper-case letter in the text. The words in foreign orthography belong to foreign languages, in particular English. The nouns which started with an asterisk and matched nouns that began with low case letters were counted as one. Verbal nouns which contained an object inflectional marker, such as, *kumnyonyesha* “to breast feed him/her”, were not counted on the pretext that infinitive verbs with object markers function as verbs rather than as verbal nouns.

4.2.1.2 Sorting out term-words from the extracted nouns and verbal nouns

Deciding whether a non-technical word qualifies as a valid term in a domain is, to some extent, subjective. As has been suggested before, this task should involve a team of linguists and domain-experts who have experience with the use of the language in the domain. However, for the identification carried out here, the researcher carried out the task alone, using his linguistic knowledge and his experience with the use of Swahili in the domain of health care. This was because subject domain-experts, experienced in the use of Swahili in the domain, could not be obtained in the environment of the task. Concept representation in the domain of health care was used as the basis for qualifying a word as a health care term-word. The sorting was carried out manually by picking up all possible term-words from the list. A total of 626 single-word terms were picked. Some of the terms are given below, and a full list is provided in Appendix 1.

afya	“health”	anofelina	“anopheline”
aleji	“allergy”	antibiotiki	“antibiotic”
ambukizo	“infection”	antibodi	“antibody”
ameba	“amoeba”	aorta	“aorta”
anemia	“anaemia”	askari	“Ascaris”
anofelesi	“anopheles”	bakteria	“bacteria”

balehe	“adolescent”	chembe	“cell”
bamia	“okra”	choo	“stool”
bandama	“spleen”	chuchu	“nipple”
bangi	“marijuana”	chumvi	“salt”
baridi	“fever”	chungwa	“orange”
bega	“pectoral girdle”	chunjua	“wart”
biringanya	“eggplant”	chunusi	“pimple”
busha	“scrotum elephantiasis”	chupa	“pre- birth membrane”
buu	“larva”	sisti	“cyst”
chakula	“food”	daktari	“doctor”
chandarua	“mosquito net”	dalili	“symptom”
chanjo	“vaccination”	damu	“blood”
chankroidi	“chancroid”	darubini	“telescope”
chati	“chart”	dawa	“medicine”
chawa	“lice”		

Most of the identified term-words represented concepts for names of diseases, parts of the body, medicines, disease-causing agents, tools and actions for treating diseases, reproductive organs and actions, foods and nutrients, and states of the body. A good number of words that were identified as health care term-words belong to words of the general vocabulary. Such words were qualified as terms because they have been used in the corpus to represent specific concepts in the domain of health care. The following are a few cases of such words with their general and specific references:

word form	general reference English equivalent	specific reference English equivalent
askari	“soldier”	“ascar”
baridi	“cold”	“fever”
choo	“toilet”	“stool”
chupa	“bottle”	“pre-birth membrane”
hamu	“appetite”	“sexual urge”
kukojoa	“urinating”	“ejaculating”
kukutana	“meeting each other”	“sexual intercourse”
kumwaga	“spilling”	“ejaculating”
mpira	“rubber”	“condom”

Where two or more different word forms were identified as representing the same concept, the words were sorted out as different terms. The following are a few cases of such words:

pumbu
hasua
korodani
kende
“testicles”

kondomu
mpira
“condom”

ashki
hamu

“sexual urge”

choo
kinyesi
“stool”

kujifungua
kuzaa
“delivering/giving birth”

kukojoa
kumwaga
“ejaculating”

kansa
sarakani
“cancer”

Juxtaposition compound terms that were written in single word forms, such as the following cases, were sorted as term-words:

kidolegumba (kidole ‘finger’ + gumba ‘?’)	“thumb”
kidoletumbo (kidole ‘finger’ + tumbo ‘stomach’)	“appendicitis”
kifuakikuu (kifua ‘chest’ + kikuu ‘big’)	“tuberculosis”
kisononogono (kisonono + gono)	“?”
pepopunda (pepo ‘devil’ + punda ‘donkey’)	“tetanus”

The sorting did not involve juxtapositions that were written as two separate word forms because the extraction covered only single-word forms.

4.2.2 Marking term-words in the lexicon with a health care domain-tag

The 626 term-word compiled were marked in the lexicon by a health care domain tag. The tag [HC] was selected for the marking where the ‘HC-N’ tag and the ‘HC-V’ tag marked noun terms and verbal noun terms respectively. However, before the terms were added to the lexicon, terms in foreign orthography were adapted to Swahili orthography. The following are a few cases of the adapted terms:

terms in foreign orthography	terms adapted to Swahili orthography
pelvic	pelviki
cell	seli
yeast	yisti
cancer	kansa
acid	asidi
oxygen	oksijeni
varicose	varikose
thyroid	thairoidi
progesterone	projesteroni
pneumonia	nyumonia
cyst	sisti

The adaptation was automatically performed by the program, which is based on the sorted list of word forms in non-Swahili orthography. The program developed by Hurskainen recognises words in non-Swahili orthography in the input text file and adapts them to Swahili orthography. The adaptation does not affect word forms in the input text. After the adaptation, the 625 term words were automatically added to the lexicon by cross-checking words in the list of the terms against words in the lexicon where words which were in the list but missing in the lexicon were added to the lexicon. In the lexicon all roots of the 625 term-words were marked as health care terms by either the 'HC-N' tag or the 'HC-V' tag.

4.3 UNIQUE TAGGING OF HEALTH CARE TERM-WORD IN THE ANALYSED CORPUS

The SWATWOL morphological analyser carried out the tagging of health care term-words in the corpus. The analyser inserted the tag HC-N or HC-V into the analysis of words in the corpus derived from the roots marked by such symbols in the lexicon. The following portion of the analysis demonstrated how the tagging was effected in the corpus.

The description assigned the tag 'HC-N' or 'HC-V' to all word forms in the corpus derived from the roots marked by such symbols in the lexicon. The following portion of the analysis demonstrates how the description was carried out in the corpus. The portion is the analysed form of the following text:

Matiti ya watoto wengine hivimba kidogo. Hii ni athari ya vichocheo vya mama nayo baadaye hupotea. Watoto wengine wa kike hutokwa na damu kidogo ukeni. Hii ni kama hedhi kidogo husababishwa na vichocheo vya mama.

“Breasts of some babies swell a bit. This is due to the effect of maternal hormones and it disappears afterwards. Some female babies show a little bleeding. This resembles minor menstruation and it is caused by the maternal hormones”.

```
"<*matiti>"
  "titi" 5a/6-PL N HC-N
"<ya>"
  "ya" 5/6-PL GEN-CON
"<watoto>"
  "mtoto" 1/2-PL N HC-N
"<wengine>"
  "ingine" 1/2-PL A-INFL
"<huvimba>"
  "vimba" HABIT-SP VFIN V SV HC-V
"<kidogo>"
  "dogo" 7/8-SG A-INFL
  "kidogo" AD-ADJ @<A-AD
"<.$>"
"<*hii>"
  "hii" PRON DEM: hV 9/10-SG
"<ni>"
  "ni" ADV:ni
"<athari>"
  "athari" 9/10-0-SG N AR
"<ya>"
```

"ya" 9/10-SG GEN-CON
 "<vichocheo>"
 "kichocheo" 7/8-PL N DER:eo HC-N
 "<vya>"
 "vya" 7/8-PL GEN-CON
 "<mama>"
 "mama" 9/10-0-SG N AN
 "<,>"
 "<nayo>"
 "nayo" it also 9/10-SG PRO
 "<baadaye>"
 "baadaye" AR ADV @ADVL
 "<hupotea>"
 "potea" HABIT-SP VFIN V SV
 "<.\$>"
 "<*watoto>"
 "mtoto" 1/2-PL N
 "<wengine>"
 "ingine" 1/2-PL A-INFL
 "<wa>"
 "wa" 1/2-PL GEN-CON
 "<kike>"
 "kike" ADV @ADVL
 "<hutokwa>"
 "toka" HABIT-SP VFIN V SV SVO STAT PASS
 "<damu>"
 "damu" 9/10-0-SG N AR HC-N
 "<kidogo>"
 "dogo" 7/8-SG A-INFL
 "kidogo" AD-ADJ @<A-AD
 "<ukeni>"
 "uke" 11-SG N HC-N LOC
 "<.\$>"
 "<*hii>"
 "hii" PRON DEM: hV 9/10-SG
 "<ni>"
 "ni" ADV:ni
 "<kama>"
 "kama" AR ' like , such as ' ADV @ADVL
 "<hedhi>"
 "hedhi" 9/10-0-SG N AR HC-N
 "<kidogo>"
 "dogo" 7/8-SG A-INFL
 "kidogo" AD-ADJ @<A-AD
 "<husababishwa>"
 "sababisha" HABIT-SP VFIN V AR SV CAUS:sh PASS
 "<na>"
 "na" CC @CC
 "<vichocheo>"
 "kichocheo" 7/8-PL N DER:eo HC-N
 "<vya>"
 "vya" 7/8-PL GEN-CON
 "<mama>"
 "mama" 9/10-0-SG N AN HC-N
 "<.\$>"

In the above analysis, words in cohorts that contain the ‘HC-N’ tag and the HC-V tag have been tagged as health care term-words. Such words are:

```
"<*matiti>"
    "titi" 5a/6-PL N HC-N
    (matiti “breasts”)
"<huvimba>"
    "vimba" HABIT-SP VFIN V SV HC-V
    (huvimba *)
"<vichocheo>"
    "kichocheo" 7/8-PL N DER:eo HC-N
    (vichocheo “hormones”)
"<damu>"
    "damu" 9/10-0-SG N AR HC-N
    (damu “blood”)
"<ukeni>"
    "uke" 11-SG N HC-N LOC
    (ukeni “vagina + loc.”)
"<hedhi>"
    "hedhi" 9/10-0-SG N AR HC-N
    (hedhi “menstruation”)
```

Since the tagging affects all words derived from the roots marked in the lexicon by the ‘HC-N’ or ‘HC-V’ tag, it included even non-term words, especially verbs, derived from such roots. For example in the above case, the verb *huvimba* is described as a health care term simply because it is derived from the base form *vimb(a)* “swell”, which is marked in the lexicon by the ‘HC-V’ tag as the base of the verbal noun term *kuvimba* “swelling”. Such words were filtered out by using the criterion of concept naming which restricts term-words to names of concepts, i.e. nouns and verbal nouns.

4.4 DISCOVERING MULTI-WORD TERMS IN THE ANALYSED CORPUS

The discovery of multi-word terms was carried out manually in the analysed corpus, where valid term-words in cohorts that contain the term-domain tag, ‘HC-N’ or ‘HC-V’, were used as searching cues. The search involved checking the searching words if they functioned as individual words or as part of collocations. Where they functioned as part of collocations, the collocations were picked up as compound terms. However, the collocations were not supposed to go across punctuation marks. The application of this process to the discovery of the terms in the portion of the analysed corpus given in the previous section will reveal that the following searching words function as parts of collocations.

Searching Word	Collocation	
matiti	> matiti ‘breasts’ ya ‘of’ mtoto ‘baby’	“ baby breasts ”
vichocheo	> vichocheo ‘activators’ vya ‘of’ mama ‘mother’ (x 2)	“ maternal hormones ”
damu	> damu ‘blood’ kidogo ‘little’	“ a little blood ”
hedhi	> hedhi ‘menstruation’ kidogo ‘little’	“ little menstruation ”

However, for a collocation to qualify as a compound term it should represent a concept in the domain. Thus, collocations which were found not to represent particular concepts, such as *damu kidogo* “a little blood” and *hedhi kigogo* “a little menstrual” were not regarded as compound terms. Where the same collocation, such as *vichocheo vya mama* “maternal hormones”, occurred more than once, only one collocation was picked. For collocations which occurred in both singular and plural forms, such as *matiti ya mtoto* ‘baby breasts’ and *titi la mtoto* “baby breast”, the collocation in the singular form was picked up. After the search, a total of 203 term collocations were identified as multi-word terms. The following are a few cases of the collocations. A full list is given in Appendix II.

kupevuka kwa yai (maturing of egg)	“egg maturing”
njia ya uzazi (passage of birth)	“birth passage”
njia ya uzazi ya msichana (passage of birth of girl)	“girl birth passage”
kujifungua salama (delivering safe)	“safe delivery”
kukwama kwa mtoto (sticking of baby)	“baby sticking”
lishe duni (nutrition poor)	“poor nutrition”
shinikizo la damu (pressure of blood)	“blood pressure”
uchungu wa muda mrefu (pain of long period)	“prolonged pain”
uchungu wa uzazi (pain of birth)	“labour pain”
misuli ya mji wa mimba (muscles of town of pregnancy)	“uterus muscles”
kujikaza na kuachia kwa misuli ya mji wa mimba (tightening and relaxing of muscles of town of pregnancy)	“uterus muscles tightening and relaxing”
mlango wa mji wa mimba (door of town of pregnancy)	“uterus opening”
tendo la kujamiiana (action of sexing)	“sexual intercourse”
viungo vya nyonga (organs of pelvis)	“pelvic organs”
kutunga mimba nje ya mji wa mimba (conceiving pregnancy out of town of pregnancy)	“outside uterus conceiving”
mfuko wa chakula (bag of food)	“stomach”
misuli ya mji wa mimba (muscle of town of pregnancy)	“uterus muscles”
hatua ya kwanza ya uchungu (stage of first of pain)	“first stage of pain”
mzunguko wa hedhi (round of menstruation)	“menstrual circle”
siku za rutuba (days of fertility)	“fertility days”
njia za asili za uzazi wa mpango (ways of tradition of birth of planning)	“family planning traditional method”
pima joto (measure temperature)	“thermometer”
njia za kijadi za kuzuia mimba (ways of tradition of preventing pregnancy)	“pregnancy preventing traditional method”
vichocheo vya kuzuia yai kupevuka (activators of preventing egg maturing)	“egg maturing prevention hormones”
mirija ya mayai (tubes of eggs)	“fallopian tubes”
magonjwa ya kujamiiana (diseases of sexing)	“sexually transmitted diseases”
ugonjwa sugu (disease chronic)	“chronic disease”
homa ya mapafu (fever of lungs)	“pneumonia”
kutoka usaha ukeni (discharging pus vaginal + loc.)	“vaginal pus discharging”
magonjwa ya via vya uzazi (diseases of organs of birth)	“reproductive organs diseases”
matatizo ya hedhi (problems of menstruation)	“menstrual problems”
dalili za uambukizo wa via vya uzazi (symptoms of infection of organs of birth)	“reproductive organs infection symptoms

4.5 DESCRIBING WORDS IN TERM COLLOCATIONS

The 203 term collocations that were compiled from the corpus were saved as a text file and then,

analysed to capture their linguistic descriptions. The following examples represent the analysed form of the collocation:

"<kupevuka>"
"pevuka" INF V SV SVO HC-V STAT
"<kwa>"
"kwa" 15-SG GEN-CON
"<yai>"
"yai" 5a/6-SG N 'egg' HC-N

kupevuka kwa yai “**egg maturing**” (maturing of egg)

"<njia>"
"njia" 9/10-NI-SG N
"<ya>"
"ya" 9/10-SG GEN-CON
"<uzazi>"
"uzazi" 11-SG N DER:zi HC-N

njia ya uzazi “**birth passage**” (passage of reproduction)

"<njia>"
"njia" 9/10-NI-SG N
"<ya>"
"ya" 9/10-SG GEN-CON
"<uzazi>"
"uzazi" 11-SG N DER:zi HC-N
"<ya>"
"ya" 9/10-SG GEN-CON
"<msichana>"
"msichana" ½-SG N

njia ya uzazi ya msichana “**girl birth passage**” (passage of reproduction of girl)

"<kupata>"
"pata" INF V SV SVO
"<ujauzito>"
"ujauzito" 11-SG N HC-N

kupata ujauzito “**conceiving**” (getting pregnant)

"<lishe>"
"lishe" 9/10-0-SG N HC-N
"<duni>"
"duni" AR A-UNINFL

lishe duni “**poor nutrition**” (nutrition poor)

"<shinikizo>"
"shinikizo" 5a/6-SG N HC-N
"<la>"
"la" 5/6-SG GEN-CON
"<damu>"
"damu" 9/10-0-SG N AR HC-N

shinikizo la damu “**blood pressure**” (pressure of blood)

"<uchungu>"

"uchungu" 11/6-SG N HC-N

"<wa>"

"wa" 11-SG GEN-CON

"<muda>"

"muda" 3/4-SG N

"<mrefu>"

"refu" 3/4-SG A-INFL

uchungu wa muda mrefu “**prolonged pain**” (pain of long period)

"<uchungu>"

"uchungu" 11/6-SG N HC-N

"<wa>"

"wa" 11-SG GEN-CON

"<uzazi>"

"uzazi" 11-SG N DER:zi HC-N

uchungu wa uzazi “**labour pain**” (pain of reproduction)

"<misuli>"

"msuli" 3/4-PL N HC-N

<ya>"

"ya" 3/4-PL GEN-CON

"<mji>"

"mji" 3/4-SG N'

"<wa>"

"wa" 3/4-SG GEN-CON

"<mimba>"

"mimba" 9/10-0-SG N HC-N

misuli ya mji wa mimba “**uterus muscles**” (muscles of town of pregnancy)

As the above analysis shows, each form of the collocations analysed had at least a cohort that contained a reading with the tag ‘HC-N’ or ‘HC-V’, among other morphological tags. Some forms contained more than one cohort with such readings, for example, the above forms for the collocations *uchungu wa uzazi* “labour pain” and *misuli ya mji wa mimba* “uterus muscles”.

4.6 REPRESENTING TERM COLLOCATIONS BY LINGUISTIC SPECIFICATIONS

After the collocations were described, the next task was to represent words in the collocations by linguistic tags for their respective categories and subcategories to form linguistic specifications for the collocations. The selection of appropriate tags for the categories and subcategories of words in the collocations underlies the representation.

4.6.1 Tags for representing categories of words in term collocations

Tags for representing categories of words in the collocations analysed were selected from the readings of the words. The selection focussed on tags that were specific to categories and subcategories of words as used in the collocations. The following are the selected tags with the categories and subcategories they represent for words in the collocations:

Selected tag	Category/subcategory represented
HC-N	Health care noun term
HC-V	Health care verbal noun term
INF	Verbal noun
N	Noun
GEN-CON	Genitive connector
A-INFL	Inflecting adjective
A-UNINFL	Non-inflecting adjective
NUM-INFL	Inflecting number
ADV	Adverb
LOC	Locative
PREP	Preposition
CARD	Cardinal number
ORD	Ordinal number
POSS	Possessive
CC	Co-ordinating conjunction

The above tags were selected and used according to the SWATWOL annotation scheme (cf 3.4.2.1). In terms of coverage of word representation, the tags are of two types, category tags such as N and ADV, and subcategory tags such as, HC-N, HC-V and INF. The subcategory tags are more restrictive than the category tags. For example, the subcategory tag HC-N represents only noun terms whereas the category tag N represents all nouns including the noun terms. Function-wise the tags are also divided into two types, tags for categories of co-ordinating words, such as GEN-CON and POSS, and tags for categories and subcategories for main words.

4.7 FORMULATING LINGUISTIC SPECIFICATIONS FOR TERM COLLOCATIONS

The formulation of linguistic specifications for the collocations was carried out manually by chaining together the selected tags for words in each collocation by the character '[' to form sequences of tags. Each sequence of tags was assigned a number where sequences of the same tags were given the same number. The following cases elaborate how the representation was carried out.

```
"<kupevuka>"
    "pevuka" INF V SV SVO HC-V STAT
"<kwa>"
    "kwa" 15-SG GEN-CON
"<yai>"
    "yai" 5a/6-SG N 'egg' HC-N
```

01 HC-V + GEN-CON + HC-N (kupevuka kwa yai “egg maturing”)

```
"<njia>"
    "njia" 9/10-NI-SG N
"<ya>"
    "ya" 9/10-SG GEN-CON
"<uzazi>"
    "uzazi" 11-SG N HC-N
```

02 N + GEN-CON + HC-N (njia ya uzazi “birth passage”)

"<njia>"
"njia" 9/10-NI-SG N
"<ya>"
"ya" 9/10-SG GEN-CON
"<uzazi>"
"uzazi" 11-SG N HC-N
"<ya>"
"ya" 9/10-SG GEN-CON
"<msichana>"
"msichana" 1/2-SG N

03 N + GEN-CON + HC-N + GEN-CON + N (njia ya uzazi ya msichana “**girl birth passage**”)

"<kujifungua>"
"fungua" INF REFL-SG OBJ V SV SVO HC-V
"<salama>"
"salama" AR ADV @ADVL

04 HC-V + ADV (kujifungua salama [safe delivery])

"<kupata>"
"pata" INF V SV SVO
"<ujauzito>"
"ujauzito" 11-SG N HC-N

05 INF + HC-N (kupata ujauzito “**conceiving**”)

"<lishe>"
"lishe" 9/10-0-SG N HC-N
"<duni>"
"duni" AR A-UNINFL

06 HC-N + A-UNINFL (lishe duni “**poor nutrition**”)

"<shinikizo>"
"shinikizo" 5a/6-SG N HC-N
"<la>"
"la" 5/6-SG GEN-CON
"<damu>"
"damu" 9/10-0-SG N AR HC-N

07 HC-N + GEN-CON + HC-N (shinikizo la damu “**blood pressure**”)

"<uchungu>"
"uchungu" 11/6-SG N HC-N
"<wa>"
"wa" 11-SG GEN-CON
"<muda>"
"muda" 3/4-SG N
"<mrefu>"
"refu" 3/4-SG A-INFL

08 HC-N + GEN-CON + N + A-INFL (uchungu wa muda mrefu “**prolonged pain**”)

"<uchungu>"
 "uchungu" 11/6-SG N HC-N
 "<wa>"
 "wa" 11-SG GEN-CON
 "<uzazi>"
 "uzazi" 11-SG N DER:zi HC-N

07 HC-N + GEN-CON + HC-N (uchungu wa uzazi “labour pain”)

"<misuli>"
 "msuli" 3/4-PL N HC-N
 "<ya>"
 "ya" 3/4-PL GEN-CON
 "<mji>"
 "mji" 3/4-SG N
 "<wa>"
 "wa" 3/4-SG GEN-CON
 "<mimba>"
 "mimba" 9/10-0-SG N HC-N

09 HC-N + GEN-CON + N + GEN-CON + HC-N (misuli ya mji wa mimba “uterus muscles”)

In the above representation, the specification “HC-N + GEN-CON +HC-N”, which represents the following collocations, occurred twice. However, the two occurrences are given the same number, 07.

uchungu wa uzazi	“labour pain”
shinikizo la damu	“blood pressure”

After all the collocations were represented by sequences of tags, the sequences were counted where sequences that were assigned the same number were counted only once. Subsequently, 88 different sequences of tags were obtained. The sequences were then sorted and marked according to their length; the number of tags for categories and subcategories of main words in a sequence determined the length of the sequence. Length-wise, the sequences had the following distribution:

20 sequences of length 2
 39 sequences of length 3
 25 sequences of length 4
 03 sequences of length 5

The following are the sequences marked with the number of their length.

HC-N + A-INFL "2"
 HC-N + GEN-CON + ADV "2"
 HC-N + GEN-CON + HC-N "2"
 HC-N + GEN-CON + HC-V "2"
 HC-N + GEN-CON + N "2"
 HC-N + HC-N "2"
 HC-N + INF "2"
 HC-N + N "2"
 HC-N + POSS + N "2"
 HC-N + PREP + HC-N "2"
 HC-V + ADV "2"
 HC-V + GEN-CON + HC-N "2"
 HC-V + GEN-CON + N "2"
 HC-V + HC-N "2"

HC-V + N "2"
 INF + GEN-CON + HC-N "2"
 INF + HC-N "2"
 N + GEN-CON + HC-N "2"
 N + GEN-CON + HC-V "2"
 N + HC-N "2"
 HC-N + A-INFL + GEN-CON + N "3"
 HC-N + A-UNINFL + GEN-CON + N "3"
 HC-N + GEN-CON + HC-N + GEN-CON + ADV "3"
 HC-N + GEN-CON + HC-N + GEN-CON + N "3"
 HC-N + GEN-CON + HC-N + PREP + HC-N "3"
 HC-N + GEN-CON + HC-V + GEN-CON + N "3"
 HC-N + GEN-CON + HC-V + HC-N "3"
 HC-N + GEN-CON + INF + GEN-CON + HC-N "3"
 HC-N + GEN-CON + INF + HC-N "3"
 HC-N + GEN-CON + INF + N "3"
 HC-N + GEN-CON + N + A-INFL "3"
 HC-N + GEN-CON + N + CC + N "3"
 HC-N + GEN-CON + N + GEN-CON + HC-N "3"
 HC-N + GEN-CON + N + HC-V "3"
 HC-N + GEN-CON + N + PREP + N "3"
 HC-N + GEN-CON + PREP + N "3"
 HC-N + GEN-CON + HC-N + GEN-CON + HC-N "3"
 HC-N + N + GEN-CON + HC-N "3"
 HC-N + POSS + HC-N + CARD "3"
 HC-N + PREP + HC-N + GEN-CON + HC-N "3"
 HC-N + PREP + N + GEN-CON + HC-N "3"
 HC-V + GEN-CON + HC-N + A-INFL "3"
 HC-V + GEN-CON + N + GEN-CON + ADV "3"
 HC-V + GEN-CON + N + GEN-CON + HC-N "3"
 HC-V + HC-N + PREP + HC-V "3"
 HC-V + N + GEN-CON + HC-N "3"
 HC-V + N + GEN-CON + N "3"
 INF + HC-N + GEN-CON + INF "3"
 INF + HC-N + GEN-CON + N "3"
 INF + HC-N + LOC "3"
 INF + N + HC-N "3"
 N + A-INFL + GEN-CON + HC-N "3"
 N + GEN-CON + ADV + GEN-CON + HC-N "3"
 N + GEN-CON + HC-N + A-INFL "3"
 N + GEN-CON + HC-N + GEN-CON + N "3"
 N + GEN-CON + N + GEN-CON + HC-N "3"
 N + GEN-CON + ORD + GEN-CON + HC-N "3"
 N + POSS + HC-N + GEN-CON + N "3"
 N + PREP + N + GEN-CON + HC-N "3"
 HC-N + A-UNINFL + GEN-CON + N + GEN-CON + HC-N "4"
 HC-N + A-UNINFL + GEN-CON + N + GEN-CON + HC-V "4"
 HC-N + GEN-CON + HC-N + GEN-CON + N + GEN-CON + HC-N "4"
 HC-N + GEN-CON + INF + GEN-CON + HC-N + GEN-CON + ADV "4"
 HC-N + GEN-CON + N + GEN-CON + N + GEN-CON + HC-N "4"
 HC-N + GEN-CON + N + N + GEN-CON + HC-N "4"
 HC-N + INF + PREP + N + GEN-CON + HC-N "4"
 HC-N + GEN-CON + HC-N + PREP + N + GEN-CON + HC-N "4"
 HC-N + GEN-CON + HC-V + HC-N + HC-V "4"
 HC-V + ADV + GEN-CON + HC-N + GEN-CON + HC-N "4"
 HC-V + GEN-CON + N + GEN-CON + N + GEN-CON + HC-N "4"
 HC-V + HC-N + PREP + N + GEN-CON + HC-N "4"
 HC-V + N + GEN-CON + ADV + ADV "4"
 HC-V + N + HC-N + GEN-CON + N "4"
 HC-V + ADV + N + GEN-CON + ADV "4"
 INF + ADV + PREP + HC-N + GEN-CON + HC-N "4"
 INF + GEN-CON + N + N + GEN-CON + HC-V "4"
 INF + HC-N + N + GEN-CON + INF "4"
 INF + HC-N + POSS + HC-N + LOC "4"
 INF + HC-N + PREP + HC-N + GEN-CON + HC-V "4"

N + GEN-CON + ADV + GEN-CON + HC-V + HC-N "4"
 N + GEN-CON + HC-N + INF + CC + HC-N "4"
 N + GEN-CON + N + GEN-CON + HC-N + GEN-CON + N "4"
 N + GEN-CON + N + GEN-CON + HC-N + N "4"
 N + GEN-CON + N + GEN-CON + N + GEN-CON + HC-N "4"
 HC-N + GEN-CON + HC-N + GEN-CON + HC-N + A-INFL + GEN-CON + HC-N "5"
 HC-V + CC + HC-V + GEN-CON + HC-N + GEN-CON + N + GEN-CON + HC-N "5"
 HCV + HC-N + ADV + HC-N + A-UNINFL "5"

4.8 CONSTRAINTS OF STRUCTURES OF SWAHILI TERMS

The categories and subcategories of words with their relationships in the sequences were observed to determine the possible structural constraints for the formation of Swahili compound terms. The observation revealed that:

1. Two types of processes are involved in the formation of Swahili compound terms: juxtaposition and connection. The juxtaposition process involves no connectors whereas the connection processes involve different categories of connectors, with the dominant use of the genitive connector (GEN-CON). The co-ordinating conjunction (CC) is used but very rarely.
2. Words of the following categories are the possible headword in Swahili compound terms:
 - a *noun term* category (represented by a domain-tag for noun terms, such as the HC-N tag)
 - a *verbal noun term* category (represented by a domain-tag for verbal noun terms, such as the HC-V tag)
 - a *verbal noun* category (represented by the INF tag)
 - a *noun* category (represented by the N tag)
3. Where the headword belongs to either the *verbal noun* or the *noun* category, the modifying element should contain at least a word of the *noun term* or *verbal noun term* category. The modifying element can be without a word of these two categories if the headword belongs to one of the categories.
4. Compound terms should not contain words of the *finite verb* category or *infinitive verb* category that function as verbs rather than as verbal nouns, i.e. inflected with verbal inflections such as negative and object markers.

Based on these observations the structure for compound terms was constrained to the follows definition:

A composition, by juxtaposition or connection, of two or more non-function words of categories and subcategories other than the *finite verb* category, headed by a word of one of the categories given under observation 2, and containing at least one word of the *noun term* or the *verbal noun term* category.

The constraining definition was then used as a basis for formulating other possible sequences based on the sequences represented. For example, by this definition, the following two sequences were the two possible sequences generated from the length 2 sequence 'HC-N + HC-N':

1. N + HC-N (length 2 juxtaposition composition headed by a word of *noun* category and containing a word of *noun term* category)

2. HC-N + N (length 2 juxtaposition composition with a word of *noun term* category as its head.)

The generated sequences that already existed in the compiled list were omitted. The new sequences generated by the mechanism raised the number of sequences for different lengths to:

- 30 sequences of length 2 (from the previous 20)
- 144 sequences of length 3 (from the previous 39)
- 204 sequences of length 4 (from the previous 25)
- 39 sequences of length 5 (from the previous 3)

The sequences were saved in separate files according to their length. In the files, tags in a sequence were in the form of TAG + TAG + ... + TAG, where the tags were separated by space and connected by the character '+'. Each sequence of tags occupied its own line, which may extend over a number of lines. Information surrounded by the inverted commas was assigned to the end of each sequence to express the following:

1. the length of a sequence, represented by the letter 'l' and the number for the length, e.g. 'l2' for length 2.
2. the type of term represented by the sequence, where all sequences that began with the symbol N or HC-N were marked with 'nt' for noun terms and sequences that began with the symbol INF or HC-V were marked with 'vt' for verbal noun terms.
3. the number assigned to each sequence.

For example, if the sequence had length 2 and it began with either N or HC-N and was assigned number 4, the information at its end appeared as "l2nt:4" (length 2, noun term, number 4).

Though formulated on evidence obtained from terms in the domain of health care, the sequences are bound to term formation constraints in Swahili. The presence of the health care (HC) domain tag in the sequences is what makes them unique to the health care terms. The sequences can be adapted to terms of different domains by using domain-tags appropriate to the domains. For example, the replacement of the health care domain-tag by the motor mechanics domain-tag, would adapt the same sequences to the domain of motor mechanics. However, the adaptation should involve the whole process of incorporating a domain-tag to the lexicon of the morphological analyser.

Though the number of sequences formulated in this study is relatively large, we cannot claim that it is exhaustive because there is the possibility of discovering more sequences with the increased use of a domain-specific corpus. What is equally important is to see the effectiveness of the formulated sequences in the extraction of terms. In the next chapter the sequences are applied as a query language for extracting terms from the morphologically annotated corpus by the template-matching program.

CHAPTER 5

EVALUATION AND FINDINGS

This chapter evaluates the effectiveness of the patterns formulated in the previous chapter in the extraction of terms by a template-matching program. The evaluation undergoes the processes involved in the extraction and in the findings obtained from the extraction. The chapter commences with a brief description of the program used for the extraction. Thereafter, it presents procedures for preparing the extraction corpus for the extraction, i.e., the tagging of term-words in the corpus and clearing out the corpus from unwanted items in the extraction. Then, the chapter evaluates the performances of the extraction by assessing the degree of coverage or recall and validity or precision in the extraction. Finally, the chapter surveys the statistical and conceptual characteristics of the terms recovered in the extraction.

5.1 THE PROGRAM FOR EXTRACTING TERMS

The extraction of terms was carried out by TERMER, the program written by Jussi Piitulainen of the Department of General Linguistics at the University of Helsinki.

TERMER is a language independent program for extracting information from a corpus that is morphologically analysed and disambiguated. The program runs on two input files: a pattern file, which contain values or tags for linguistic features that are specified as patterns in the file, and a file of corpus whose words are analysed by the feature values that are specified in the patterns of the pattern file. The program uses the feature values, as specified in the patterns, for the extraction of items that are analysed with the same values in the corpus file. During the extraction, the program is first given the pattern file. The program compiles patterns in the file into two operational modules, a tokeniser and a pattern matcher. Then, the file of the analysed corpus is given to the operational modules, which carry out the matching process sequentially. The analysed corpus is given first to the tokeniser, and the output of the tokeniser is given to the pattern matcher. The tokeniser reads through the cohorts of the analysed corpus and assigns an arbitrary symbol to word forms in cohorts, which contain tags that are matched. To avoid repetition during the process of reading, the tokeniser reads leftwards from the longest to the shortest sequences. The pattern matcher goes through the output of the tokeniser, removing the arbitrary symbol and labelling the beginning and the end of the marked words or sequences of words with symbols to identify them as the matched items to be extracted. The actual extraction of the items is carried out by the catching program, which extracts the words and sequences of words that are marked by the pattern matcher. The retrieved items can then be sorted and uniquely counted by using the Unix utilities.

Based on the patterns developed in Chapter 4, TERMER has been adapted for the extraction of Swahili information. This was facilitated by developing a Swahili tokeniser (swatok) and pattern matcher (swapt) that TERMER should compile from the pattern files intended for extracting the information. It is necessary that tags that are specified in the pattern file should be available in the Swahili corpus from which the information extraction is carried out.

5.2 TAGGING TERM-WORDS IN THE EXTRACTION CORPUS

The extraction of terms carried out in this implementation involved the extraction corpus in the domains of health care and literature (cf. 1.7). Since a domain-tag is among the tags specified in the pattern files for extracting terms, the tag should also be available in the extraction corpus. In other words, all term-words in the extraction corpus should be tagged by the domain-tag that is appropriate to the domain of the terms. The morphological analyser tagged term-words in the health care extraction corpus, whereas the BETA system carried out the same task in the literature extraction corpus. The two systems perform the tagging by using different techniques. While the morphological analyser performs the task through the lexicon, the BETA system carries out the task through the rule file. Thus, updating the lexicon or the rule file should be carried out before the tagging.

5.2.1 Updating the lexicon or the rule file

Updating of the lexicon was carried out to ensure that term-words in the health care extraction corpus whose roots were not yet marked in the lexicon, during the development of the patterns, could be added to the lexicon and marked by a health care domain tag. The process involved a number of tasks. First, the corpus was analysed and words in cohorts containing the HC-N or HC-V tag were removed from the analysed corpus. This separated term-words that were already marked in the lexicon from the analysed corpus. Then all nouns and verbal nouns in the corpus were compiled and the terms were identified among them. The identified term-words were added to the lexicon and marked by the HC-N tag, for the noun terms, and the HC-V tag, for the verbal noun terms. The morphological analyser tagged the terms in the analysed corpus during the analysis, through the updated lexicon.

Since it was the first time that the BETA rule file was used for the tagging of literature term-words in the corpus, the rule file for the task had to be developed. The development of the rule file involved a number of tasks. First, the literature extraction was morphologically analysed and disambiguated. Then, nouns and verbal nouns were compiled from the analysed corpus and among them the literature term-words were identified. The term-words were then morphologically analysed and disambiguated to determine their base-forms as produced by the analyser. The term base-forms as obtained from the analyser were then formulated as rewriting rules in the BETA rule file. Each base-form in the rule file was formulated as an independent rule for replacing the base-form with the same base form plus a domain-tag in the analysed corpus. The LT-N tag, for noun terms, and LT-V tag, for verbal noun terms, was used in the rule file. The BETA system performed the tagging of term-words in the analysed literature corpus by rewriting the term base-forms in the corpus as specified by the rules in the rule file.

5.3 TERM EXTRACTION

After the tagging of term-words, with the relevant domain-tag, in the morphologically analysed extraction health care and literature corpus, the extraction of terms was carried out separately from each corpus. However, certain items had to be cleared from the corpus extraction before the extraction.

5.3.1 Clearing out the extraction corpus

The following small portion of the analysed health care extraction corpus provides cases of items that had to be cleared from the corpus before the extraction was carried out.

```
"<*magonjwa>"
  "ugonjwa" 11/6-PL N HC-N
"<ya>"
  "ya" 5/6-PL GEN-CON
"<kuambukiza>"
  "ambukiza" INF V SV SVO HC-V
"<yanayotokana>"
  "tokana" 5/6-PL-SP VFIN PR:na 5/6-PL REL V SV SVO STAT REC
"<na>"
  "na" CC @CC
"<kinyesi>"
  "kinyesi" 7/8-SG N HC-N
"<ni>"
  "ni" ADV:ni
"<magonjwa>"
  "ugonjwa" 11/6-PL N HC-N
"<gani>"
  "gani" INTERROG
"<?>"
"<*haya>"
  "haya" PRON DEM: hV 5/6-PL
"<ni>"
  "ni" ADV:ni
"<magonjwa>"
  "ugonjwa" 11/6-PL N HC-N
"<tofauti>"
  "tofauti" AR ADV @ADVL
"<tofauti>"
  "tofauti" AR ADV @ADVL
"<ambayo>"
  "amba" REL PRON 5/6-PL G-REL
"<vimelea>"
  "kimelea" 7/8-PL N HC-N
"<vya>"
  "vya" 7/8-PL GEN-CON
"<uambukizi>"
  "uambukizi" 11-SG N HC-N
"<hutolewa>"
  "tolea" HABIT-SP VFIN V SV SVO APPL PASS
"<kwenye>"
  "kwenye" ' in, at, about ' PREP @ADVL
"<kinyesi>"
  "kinyesi" 7/8-SG N' HC-N
"<na>"
  "na" CC @CC
"<mara>"
  "mara" 9/10-0-SG N AR
"<nyingine>"
  "ingine" 9/10-SG A-INFL
```

"<kwenye>"
 "kwenye" PREP @ADVL
 "<mkojo>"
 "mkojo" 3/4-SG N HC-N
 "<wa>"
 "wa" 3/4-SG GEN-CON
 "<mgonjwa>"
 "mgonjwa" 1/2-SG N HC-N
 "<au>"
 "au" CC CONJ AR
 "<mbebaji>"
 "mbebaji" 1/2-SG N SV SVO DER:ji
 "<.\$>"
 "<*vimelea>"
 "kimelea" 7/8-PL N HC-N
 "<hivi>"
 "hivi" PRON DEM: hV 7/8-PL
 "<huwafikia>"
 "fikia" HABIT-SP VFIN 1/2-PL3 OBJ V SV APPL
 "<watu>"
 "mtu" 1/2-PL N
 "<wengine>"
 "ingine" 1/2-PL A-INFL
 "<pale>"
 "pale" 16-SG LOC PRON DEM
 "<kinyesi>"
 "kinyesi" 7/8-SG N HC-N
 "<kinapozagaa>"
 "zagaa" 7/8-SG-SP VFIN PR:na 16-SG REL V AR SV
 "<kwenye>"
 "kwenye" PREP @ADVL
 "<mazingira>"
 "mazingira" 6-PL N
 "<hususani>"
 "susana" NEG-a 1/2-SG2-SP VFIN V SV REC
 "<udongo>"
 "udongo" 11-SG N HC-N
 "<na>"
 "na" CC @CC
 "<maji>"
 "maji" 6-PL N' HC-N
 "<.\$>"

The following is the portion of the corpus before the analysis:

Magonjwa ya kuambukiza yanayotokana na kinyesi in magonjwa gani? Haya ni magonjwa tofauti tofauti ambayo vimelea vya uambukizo hutolewa kwenye kinyesi na mara nyingine kwenye mkojo wa mgonjwa au mbebaji. Vimelea hivi huwafikia watu wengine pale kinyesi kinapozagaa kwenye mazingira hususan udongo na maji. “Which infectious diseases are caused by excrement? These are different kinds of diseases, which are spread by the infectious bacteria that reproduce in excrement and sometimes in the urine of the infected patient or carrier. The bacteria spread to other people when the excrement scatters in the environment, particularly in the soil and water.”

The analysed corpus is verticalised, where a word-form and a punctuation mark occupy their own

line. Capital letters have been changed to an asterisk * followed by a small letter, such as in the following words: **magonjwa*, **haya* and **vimelea*. Each cohort contains a word form and a reading containing a base form, and other tags for the features used to describe words in the annotation. Words in cohorts with readings that contain the HC-N or HC-V tag have been described in the analysis as health care terms. The following are such words in the analysed text:

*magonjwa	“diseases”
kuambukiza	“infecting”
kinyesi	“faeces”
magonjwa	“diseases”
vimelea	“parasites”
*vimelea	“parasites”
uambukizi	“infection”
mkojo	“urine”
mgonjwa	“patient”
udongo	“soil”
maji	“water”

The items removed from the corpus included the following:

- (1) The asterisk *, whose removal was necessary to eliminate the difference it creates between words with the same form, such as *ugonjwa (disease) and ugonjwa (disease).
- (2) The analysis or readings that contain one of the following tags: VFIN (for finite verbs); INTERROG (for interrogatives); DEM (for demonstratives); PRON (for pronouns) and OBJ (for verb forms, including infinitive verbs, inflected with object marker). Word forms analysed by these tags in the extraction corpus are deemed inapplicable in the structures of Swahili terms, as specified by the developed patterns.

5.3.2 Extraction

Terms were extracted according to their length, i.e., the number of non-function words in a term. The tags HC-N (health care noun term) and HC-V (health care verbal noun term) and the tags LT-N (literature noun term) and LT-V (literature verbal noun term) were used for the extraction of length 1 health care and literature terms respectively. Each pattern file for a particular length was separately applied for the extraction of terms of the respective length. The TERMER program compiled the patterns in the files into the Swahili tokeniser (the swatok) and pattern matcher (the swapt) (cf. 5.1) by means of the following command:

```
$ termer -v tokker=swatok -v patter=swapt [pattern file name]
```

After the pattern compilation, the extraction, sorting and counting of items from the extraction corpus file was carried out by the following chain of commands:

```
$ swatok [corpus file name] | swapt | catch | sort | uniq -c | sort -nr > output file.
```

In this chain of commands, the ‘swatok’; ‘swapt’; and ‘catch’ enabled the extraction of items from the corpus file, whereas the commands ‘sort’ and ‘uniq -c’ facilitated the sorting and unique counting of the items. In the output files, each of the extracted items had three components, as given in the following cases:

320 "11nt:1"	mimba	“pregnancy”
13 "12nt:8"	mfumo wa afya (system of health)	“health system”
1 "13nt:88"	choo chenye bomba la hewa (latrine with pipe of air)	“ventilated latrine”
1 "14nt:3"	dalili kubwa ya athari za ugonjwa (symptom big of effects of disease)	“disease effect main symptom”

The first component is the number of occurrences of the item in the corpus. The second one concerns the length of the matched sequence, the term-category of the matched item and the number of the matched sequence. The third component is the matched word or collocation. For example, the components of the first item state that the extracted item occurred in the corpus a total of 320 times; the item was matched by the sequence of length 1 as a noun term; the number of the matched sequence was 1; and the word form matched was *mimba* “pregnancy”.

Since the extraction patterns were developed based on the training health care corpus, they had to be monitored before they were used for the extraction of literature terms. The monitoring was carried out by using the monitoring literature corpus (cf. 1.7) and it was intended for the discovery of structures based on literature terms that were not captured by the training corpus. However, no distinct structure was discovered from the monitoring corpus. Before the extraction of literature terms, the health care domain-tag was replaced with the literature domain-tag in the patterns.

5.4 ASSESSING THE EFFECTIVENESS OF THE EXTRACTION METHOD

The assessment carried out in this section is intended to provide answers to the following two questions:

- (1) What is the percentage ratio of the valid items that the extraction method recovers to all valid terms available in the extraction, regardless of what else the method recovers? Computing the percentage degree of *recall* for the method provides the answer to this question.
- (2) What is the percentage ratio of the valid items that the extraction method recovers to all items that the method recovers from the extraction? Evaluating the percentage degree of *precision* for the method provides the answer to this question.

The terms *recall* and *precision* are used in the evaluation tasks for information retrieval and for information extraction. Sundhein (1996) defines the two terms by the following formulas:

$$\text{Recall} = \# \text{ relevant-returned} / \# \text{ relevant}$$

$$\text{Precision} = \# \text{ relevant-returned} / \# \text{ returned}$$

Where, in information extraction, *relevant* refers to relevant facts in extraction and *returned* to extracted facts in extraction.

5.4.1 Recall assessment

As defined in the above formulas, the denominator for computing *recall* is derived from the number of relevant facts (terms, in our case) available in the extraction corpus. Determining all

terms available in the extraction corpus for computing the *recall* for our extraction method requires manual identification of all valid terms, in different lengths, available in the extraction corpus. Considering the large size of our extraction corpus, 173,666 and 161,689 words for the health care and literature corpus respectively, this task proved to be too demanding to be carried out within the limited time available for this research. Thus, the recall was evaluated based on a portion of the health care corpus consists of 25,000 words, and it involved only terms of length 2. The use of length 2 terms for the evaluation was based on the fact that, as will be evidenced later, the terms are dominantly used in the corpus. Manual identification of the terms in the portion of corpus was carried out and 948 distinct terms were recovered, with singular and plural forms treated as distinct terms. The identification was done based on the researcher's experience with the use of Swahili in the domain of health care. Then the extraction was carried out in the same portion of the corpus, yielding 923 valid terms. All of the extracted terms were in the list of the terms identified manually. The percentage ratio of the valid terms extracted to the terms identified manually, i.e. valid terms in the extraction corpus, rates the *recall* of the extraction at 97.3%. Terms extracted manually that were not recovered in the extraction contained relativised verbal modifiers, such as in the following cases:

ugonjwa usiotibika "incurable disease"
homa isiyoiisha "interminable fever"
ugonjwa usiopona "unrecoverable illness"

The structure of Terms modified by relativised verbs were omitted in the formulation of term patterns because relativised verbs belong to the category of infinite verbs. Moreover, in most cases, the terms have synonyms which have structures that are compatible with certain patterns developed. For example, the following terms are the synonyms for the above terms:

ugonjwa sugu "incurable disease"
homa ya kusedeka "interminable fever"
ugonjwa wa kusedeka "unrecoverable illness"

The high recall of the extraction signifies that almost all structures of valid terms available in the extraction corpus are compatible with the constraints on which the patterns were developed.

5.4.2 Precision assessment

According to the precision formula (cf. 5.4), the denominator for computing the precision of our extraction method should consist of the number of all valid and invalid term candidates recovered in the extraction and the numerator should consist of the number of all valid term candidates extracted. The number for the valid and invalid items extracted was obtained by compiling the contents of the output files. The following tables show the numbers as compiled from each output file. Also, the tables show the number of matched and unmatched patterns in the extraction.

Table 5.1 Number of extracted items with the matched and unmatched patterns

(1) From the extraction of health care terms

Pattern length	Patterns developed	Patterns matched	Patterns unmatched	No. of extracted items
1	2	2	0	752
2	30	24	06	2082
3	144	72	72	771
4	204	21	183	60
5	39	00	39	00
Total number of items extracted				3665

(2) From the extraction of literature terms

Pattern length	Patterns developed	Patterns matched	Patterns unmatched	No. of extracted items
1	2	2	0	381
2	30	22	08	2314
3	144	77	67	816
4	204	35	169	63
5	NIL	NIL	NIL	NIL
Total number of items extracted				3574

Comparing the numbers of unmatched sequences with those of the developed sequences, the number of unmatched sequences increases in proportion to the length of the sequences. This suggests that sequences of long length, i.e., above length 3, can be used, but infrequently. The numbers for the extracted items indicate that items of length 2 are the most dominant.

The number of the valid terms extracted was obtained manually by going through the extracted items in each output file and identifying items that qualified as valid term candidates. The identification was carried out based on the criterion of concept representation in the relevant domain. Moreover, our knowledge of the domains of health care and literature and of the Swahili language facilitated the determination of the status of the items. Non-term items and some term variants were removed from the extracted items as follows:

1. All plural forms of single-word terms that had singular forms in the list were removed, whereas certain plural forms of multiword terms were retained. The forms represented generic concepts for their singular counterparts. The following are a few cases of the retained plural forms with their singular counterparts:

**Plural form
(generic)**

magonjwa ya tumbo “stomach diseases”
(diseases of the stomach)
mifupa ya nyonga “pelvic bones”
(bones of the pelvis)
vimelea vya magonjwa “diseases germs”
(germs of diseases)

**Singular form
(specific)**

ugonjwa wa tumbo “stomach disease”
(disease of the stomach)
mfupa wa nyonga “pelvic bone”
(bone of the pelvis)
kimelea cha ugonjwa “disease germ”
(germ of disease)

2. The variants caused by synonyms were left as interchangeable terms. Likewise, the variants caused by the use of alternate prepositions were left as synonyms.

3. The variants caused by minor orthographic variation and by the positioning of the complement of the verbal nouns were removed on the basis of intuitive knowledge of the language. For instance, we retained the item *mbegu ya mwanaume* “sperm” and removed the other two items from the following matched items.

mbegu ya mwanaume
mbegu ya mwanamume
mbegu ya mwanamme

However, it would also be possible to retain the variants as synonyms because their removal interferes with the living evidence of language use as provided by the corpus.

4. The clearing of the variants caused by the pre-positioning and post-positioning of the verbal noun complement affected the items with the pre-positioned complements, such as:

yai kupevuka
(egg maturing)
mimba kutunga
(pregnancy composing)
kichwa kuuma
(head aching)

After the clearing out of the non-term items and the term variants, the valid terms in each output file were counted. The following table presents the total number of the terms as counted in the output files.

Table 5.2 Number of valid health care and literature term candidates extracted

Term length	No. of health care terms	No. of literature terms
1	520	295
2	1,397	1,577
3	524	506
4	60	40
Total of valid terms	2,501	2,418

The percentage ratio of the number of valid term candidates identified to the number of items recovered in the extraction rates the precision of the extraction method at 68.2% for health care terms and 67.7% for literature terms. By the nearest tenth point, the precision in the two extractions is 68%. The following table gives the numbers and the precision rate.

Table 5.3 Precision rate for the extraction

Extraction domain	No. of valid terms extracted	No. of items extracted	Rate of precision
Health care	2,501	3,665	68.2%
Literature	2,418	3,574	67.7%

Compared to the recall, the precision of the extraction is low. This implies that, although the extraction captures almost all terms from the extraction corpus, it also captures a number of invalid items. However, because the extraction should be concluded by manual sorting of valid terms among the extracted items, the higher recall is preferred over precision.

Given that the proposed method has a precision rate of 68%, how does this compare with the rate of the extraction method that is based on part-of-speech tags only? In other words, supposing that the valid terms recovered in this extraction had been obtained by the extraction method that is based on part-of-speech tags only, what would have been the rate of precision of that method? To determine this, the precision of the method was evaluated on the extraction of length 2 terms. For the extraction, the domain-tag in the length 2 extraction patterns was replaced with the part-of-speech tag, 'N' tag or 'INF' tag for the noun and verbal noun terms respectively. The replacement reduced the patterns from 30 to 12. The 12 patterns were then used for the extraction of length 2 items from the health care and the literature extraction corpus. The percentage ratio of the extracted items to the valid length 2 terms recovered by the proposed method was evaluated to determine the rate of precision. The evaluation puts the rate of precision at 21%, as represented in the following table.

Table 5.4 Precision rate of the extraction based on part-of -speech tags only

Extraction domain	No. of valid length 2 terms extracted	No. of length 2 items extracted	Rate of Precision
Health care	1,397	6,638	21.04%
Literature	1,577	7,382	21.36%

It is obvious that the precision rate of 21% does not compare with that of 68% achieved by the proposed extraction method. The higher precision rate of the proposed method signifies the effectiveness of the domain-tag as a term filtering cue in the extraction.

Causes of the extraction of invalid items were established on the basis of observation of the invalid items extracted. The observation revealed that the invalid items fell into two categories: the non-term and the term variant items.

5.4.2.1 The non-term items

The causes of the extraction of non-term items were established based, on the extracted single-word and multi-word items. It was discovered that the single-words extracted belonged to the categories of: locativised nouns, adverbialised nouns, passivised verbal nouns, and imperative (command) forms. All such words were derived from term base-forms. The following are a few cases of such forms:

Term base	Derived non-term word
	(locativised form)
utosi “crown of head”	utosini “at the crown of the head”

tumbo “stomach”	tumboni “in the stomach”
uke “vagina”	ukeni “in the vagina”
jicho “eye”	jichoni “in the eye”
unyeo “anus”	unyeoni “in the anus”
	(adverbialised form)
(u)gonjwa "sickness"	kigonjwa "sickly"
maumivu "pain"	kimaumivu "painfully"
kichaa "insane"	kikichaa "insanely"
(u)dhaifu "wickness"	kidhaifu "wickly"
	(passivised form)
kuzaa “giving birth”	kuzaliwa “being born”
kuchanja “vaccinating”	kuchanjwa “being vaccinated”
kuzuia “preventing”	kuzuiwa “being prevented”
kupima “testing”	kupimwa “being tested”
	(imperative (command) form)
zaa	zaa “give birth”
chanja	chanja “vaccinate”
zuia	zuia “prevent”
pima	pima “test”

The extraction of words of the above categories is attributed to the ambiguity caused by the tagging of term-words in the extraction corpus from their roots. The tagging affects all words derived from the roots in the corpus, regardless of their terminological status. Removing such categories of words from the extraction corpus before the extraction is carried out can solve this problem.

The observation revealed that the multi-word items extracted belonged to the following categories:

1. Items that contained the forms that were falsely extracted as single-word items, i.e., the locative, adverbial, passive and imperative forms of term-words.
2. Items that contained homonyms for general and for specific meanings.
3. Items that were extracted in incomplete forms.

The following are cases of homonyms contained by some non-term items with their general and health care-specific meanings:

Homonym	General meaning	Specific meaning
ziwa	lake	breast
kiungo	spice	organ
haja	need	excretion
koo	clans	throat
chupa	bottle	pre-natal membrane

When observed individually, the above items are qualified as health care terms because of their specific meanings. But when analyzed as part of multi-word collocations, the items are easily judged as homonyms on the basis of their contextual meanings. The following cases demonstrate how the contextual meanings of the items can be different. The items marked with an asterisk ‘*’ represent the general meanings, and their collocations did not qualify as terms.

kansa ya ziwa (ziwa “ breast ”)	“ breast cancer ” (cancer of breast)
maji ya *ziwa (ziwa “ lake ”)	“ lake water ” (water of lake)
kiungo cha uzazi (kiungo “ organ ”)	“ reproductive organ ” (organ of birth)
*kiungo cha mboga (kiungo “ spice ”)	“ soup spice ” (spice for soup)
chupa ya uzazi (chupa “ membrane ”)	“ pre-natal membrane ” (bottle of birth)
*chupa ya maji (chupa “ bottle ”)	“ water bottle ” (bottle of water)
mkereketo wa koo (koo “ throat ”)	“ itching throat ” (itching of throat)
*koo nyingi (koo “ clan ”)	“ many clans ” (clan many)
*haja ya kufanya utafiti (haja “ need ”)	“ need for doing research ” (need of doing research)
"njia ya haja ndogo (haja “ excretion ”)	“ urinary passage ” (passage for excretion small)

The non-term collocations, such as the above ones, were recovered because the extraction is carried out based on context independent patterns.

The extraction of items that were in incomplete forms was due to two reasons:

1. The presence in the analysed corpus of two length juxtapositions analysed as separate word forms, such as:

kifua kikuu (analysed as a *noun term* (kifua “chest”) and an *adjective* (kikuu “big”)
 pepo punda (analysed as a *noun* (pepo “devil”) and a *noun* (punda “donkey”)
 mja mzito (analysed as a *noun* (mja “?”) and an *adjective* (mzito “heavy”)
 dege dege (analysed as a *noun* (dege “big bird”) and the *noun* dege.

Such separate analysis caused the extraction of false multiword items, such as:

*chanjo ya pepo “devil vaccination” instead of chanjo ya pepopunda “tetanus vaccination”
 *mama mja “?” instead of mama mjamzito “expectant mother”
 *ungojwa wa dege “big bird disease” instead of ungojwa wa degedege “convulsion”

The incomplete items ‘chanjo ya pepo’ and ‘ugonjwa wa dege’ were matched by the sequence HC-N + GEN-CON + N, i.e., *noun term* (chanjo “vaccination”; ugonjwa “disease”) + *genitive connector* (ya; wa “of”) + *noun* (pepo “devil”; dege “bird”). But their complete form, ‘chanjo ya pepopunda’ and ‘ugonjwa wa degedege’, would have been matched by the sequence HC-N + GEN-CON + HC-N, i.e., *noun term* (chanjo; ugonjwa) + *genitive connector* (ya; wa) + *noun term* (pepopunda “tetanus”; degedege “convulsion”). Like-wise, the item ‘mama mja’ was matched by the sequence HC-N + N, i.e., *noun term* (mama “baby bearer”) + *noun* (mja “?”), but the pattern for its complete form would have been HC-N + HC-N (*noun term* + *noun term*).

This type of false extractions was commonly observed in length two items because the sequences of that length are restricted to two main words.

2. Another reason for the presence of incomplete forms was the extraction of portions of longer items that were not fully modified. The portions were extracted by the sequences of shorter length. The following are a few cases of incomplete length 2 items extracted as portions of length 3 terms. Sequences for the respective items are also given.

Length 3 term

nguvu za kukinga maradhi “disease-preventing power”
(power of preventing disease)
'N + GEN-CON + HC-V + HC-N'
dalili za ukosefu wa damu “blood deficiency symptoms”
(symptoms of deficiency of blood)
'HC-N + GEN-CON + N + GEN-CON + HC-N'
mrija wa kutolea mkojo “urinary tube”
(tube for withdrawing urine)
'HC-N + GEN-CON + INF + HC-N'

Incomplete length 2 items

*nguvu za kukinga
“preventing power”
'N + GEN-CON + HC-V'
*dalili za ukosefu “deficiency
symptoms”
'HC-N + GEN-CON + N'
*mrija wa kutolea
“withdrawing tube”
'HC-N + GEN-CON + INF'

5.4.2.2 The term variants

It was observed that there were both linguistic and stylistic reasons for the presence of term variants in the lists. Linguistically, the variants were due to the presence of singular and plural forms as well as synonyms. Stylistically, the variants resulted from variations in writing certain words, the use of certain prepositions, and the positioning of complements for verbal nouns.

The singular and plural forms were observed in both single-word and multiword items. The following are a few cases of the single-word items:

mayai	“eggs”
yai	“egg”
matiti	“breasts”
titi	“breast”
mitoki	“groins”
mtoki	“groin”

Plural forms in multiword terms resulted where one or more words in the terms were in plural forms. The plural forms could be in the head-position, in the modifying position, or in both positions as illustrated below.

1. Plural forms in the head-position:

magojwa ya zinaa “STDs” ← **ugonjwa** wa zinaa “STD”

(diseases of adultery)

mifupa ya kinena “groin bones” ← **mfupa** wa kinena “groine bone”

(bones of the groin)

mishipa wa damu “blood vessels” ← **mshipa** wa damu “blood vessel”

(vessels of blood)

2. Plural forms in the modifying position

mfuko wa **mbegu za uzazi** “testicle/ovary” ← mfuko wa **mbegu ya uzazi**

(bag for seeds of birth)

huduma ya **kinga za magonjwa** “diseases prevention service” ← huduma ya **kinga**

(service of prevention of diseases)

ya ugonjwa “diease prevention service”

3. Plural forms in both head and modifying positions

huduma za afya kwa watoto “children health services” ← **huduma ya afya kwa**

(services of health for children)

mtoto “child healthcare service”

Synonyms were commonly observed as the cause of multiword term variants. The following are a few cases of the variants:

kifafa cha mimba “pregnancy epilepsy”
(epilepsy of pregnancy)
kifafa cha uzazi “birth eplepdy”
(epilepsy of birth)
kifafa cha ujauzito “pregnancy epilepsy”
(epilepsy of pregnancy) “**eclampsia**”

magonjwa ya zinaa “adultery diseases”
(diseases of adultery)
magonjwa ya kujamiiana “sexual diseases”
(diseases of sexing)
magonjwa ya uasherati
(diseases of adultery) “**sexually transmitted diseases**”

damu ya mwezi “monthly blood”
“blood of month/moon”
damu ya hedhi “menstrual blood”
(blood of menstruation) “**menstruation**”

mrija wa kupitishia mkojo “urine withdrawing tube”
(tube for withdrawing urine)
mfereji wa kupitishia mkojo “urine withdrawing pipe”
(pipe for withdrawing urine) “**urinary tube**”

Minor orthographic variations caused the extraction of term variants such as:

mbegu ya **mwanaume**
mbegu ya **mwanamume**
mbegu ya **mwanamme**
(seed of man) “**sperm**”

The use of the alternating prepositions *kwenye* and *katika* (meaning: ‘in’) caused the extraction of term variants such as:

maumivu kwenye uume
maumivu katika uume
(pain in penis) “**penis pain**”

damu kwenye mkojo
damu katika mkojo
(blood in urine) “**bloody urine**”

The style of positioning complements for verbal nouns, before or after the verbal noun, led to the extraction of the following types of term variants:

yai kupevuka (pre-positioned complement)
(egg maturing)
kupevuka **yai** (post-positioned complement)
(maturing egg) “**egg maturing**”

mimba kutunga (pre-positioned complement)
(pregnancy making)
kutunga **mimba** (post-positioned complement)
(making pregnancy) “**conceiving**”

kichwa kuuma (pre-positioned complement)
(head aching)
kuumwa **kichwa** (post-positioned complement)
(“**headache**”)

In the normal syntax of Swahili, the complements should come after the verbal nouns, but this was not always the case in the texts, as the above evidence shows.

5.5 WEIGHING THE TERMS BY THEIR STATISTICAL AND CONCEPTUAL BEHAVIOUR

In this section, we examine characteristics of the extracted terms in relation to their statistical and conceptual behaviour. The statistical characterisation intends to throw light on how the extraction method developed in this study is capable of capturing terms that are normally rejected if judged based on their raw frequency of occurrences. On the other hand, the conceptual characterisation aims to expose the incompatibility of the terms with certain naming criteria and principles that are propagated in the traditional approach to terminology.

5.5.1 Statistical behaviour of the terms

As has already been pointed out (cf. Chapter 1), the statistical behaviour of word forms and collocations in a domain-specific corpus is the major criterion of the statistical method for the compilation of terms from the text corpus. The frequency of occurrences of the word forms and collocations is applied where the items with high values of occurrences are picked up as terms. However, the minimal occurrences for qualifying a word form or a collocation as a term are not clearly stated. In her implementation, Dille (1996) applied two occurrences as the minimal value for qualifying a collocation as a term. Based on the same value, we computed the proportion of relevance for the extracted terms and the findings are as set out in Table 5.5 below:

Table 5.5 Rate of valid terms by the statistical criterion

(1) Health care terms

Term length	No of extracted terms	No. of terms with 2+ occurrences	Percentage of statistically relevant terms
1	520	414	79.6%
2	1,397	383	27.4%
3	524	56	10.7%
4	40	06	15.0%

(2) Literature terms

Term length	No of extracted terms	No. of terms with +2 occurrences	Percentage of statistically relevant terms
1	295	219	74.2%
2	1577	267	16.9%
3	506	15	03%
4	39	00	00

Judged by these findings, the opinion that terms should have fairly high frequencies of occurrence in the texts where they occur, becomes very questionable. It is evidenced from the findings that most of the terms extracted by the extraction method, particularly terms of length 2 and above, had only one value of occurrence in the corpus, a value that is not 'fairly high' by any measure.

5.5.2 Conceptual characteristics of the terms

According to Sager (1991: 89-90), the traditional criteria and principles for naming concepts are highly idealised requirements which can only be useful in a strictly controlled environment. Sager lists the criteria and principles as follows:

1. The term must relate directly to the concept [and] must express the concept clearly.
2. The term must be lexically systematic [and] must follow an existing lexical pattern.
3. The term must conform to the general rules of word-formation of the language.
4. Terms should be capable of providing derivatives.
5. Terms should not be pleonastic (i.e. no redundant repetitions).
6. Without sacrificing precision, terms should be concise and not contain unnecessary information.
7. There should be no synonyms, whether absolute, relative or apparent.
8. Terms should not have morphological variants.
9. Terms should not have homonyms.
10. Terms should be monosemic.
11. The content of terms should be precise and not overlap in meaning with other terms.
12. The meaning of terms should be independent of context.

Observation of the extracted terms revealed that the terms violated most of the above

requirements, particularly criteria 7, 8, 9, 10 and 12.

1. Contrary to requirement number 7, the terms are characterised by many cases of synonyms such as:

ngozi ya hasua
ngozi ya korodani
ngozi ya pumbu
(skin of testicle) **“testicle skin”**

kuongezwa damu
kuongezewa damu
kupewa damu
kuwekewa damu
[to add (passive) blood] **“blood transfusion”**

kumwaga mbegu
kumwaga shahawa
(spilling sperm) **“ejaculation”**

damu ya mwezi (blood of month)
damu ya hedhi (blood of menstruation)
“menstrual blood”

kondo la nyuma (? of behind)
kondo la uzazi (? of birth)
“placenta”

ugonjwa wa uasherati
ugonjwa wa zinaa
ugonjwa wa kujamiiiana
“sexually transmitted diseases”

kifafa cha mimba (epilepsy of pregnancy)
kifafa cha ujauzito (epilepsy of pregnancy)
kifafa cha uzazi (epilepsy of birth)
“eclampsia”

2. The presence of morphological variants in the terms, such as those given below, violated requirement number 8.

kuongezewa damu
kuongezwa damu ,
(blood transfusion)

3. Criterion number 10 is violated by the presence of homonyms⁹ such as the following:

⁹ According to (*ISO CD 704, Terminology Work-Principles and Methods 1997: 31*), homonyms are terms with identical forms but naming different concepts.

kupasuka kwa chupa
“**pre-natal membrane rupturing**” ; “**thermos breaking**”

vipimo vya choo
“**stool tests**” ; “**latrine measurement**”

matatizo ya kukojoa
“**ejaculation problems**” ; “**urination problems**”

kujisikia kukojoa
“**reaching orgasm**” ; “**feeling of passing urine**”

4. Most of the terms, especially non-technical terms, violated requirement number 12 because the terminological status of the terms depended on the context of the health care domain. However, there were some technical terms which met the requirement, such as the following:

degedege
“**convulsion**”
kaswende
“**syphilis**”
vipele vya kaswende
“**syphilis rushes**”
tauni ya tezi
“**lymphatic plague**”
shahawa
“**sperms**”
maumivu ya tumbo
“**stomach pain**”

Generally, therefore, the conceptual behaviour of the terms did not totally comply with the traditional principles of concept naming. Rather, the terms manifest varieties of conceptual characteristics, most likely due to the fact that the terms covered both officially and privately developed terms. The mixture may be attributed to the fact that specialists use both the officially and privately coined terms.

The violation of certain traditional requirements by the corpus-based terms has been the basis of the rejection of these corpus-based terms by traditional terminographers. However, the terms are the priority of corpus-based terminography where the terminological status of a word is determined by the communicative environment of the word, rather than by traditional criteria and principles. Computer-assisted compilation of such terms, to facilitate the creation of term banks, technical dictionaries and other types of terminography-based research, underlies the method of corpus-based terminography. The extraction method developed in this study contributes to the achievement of the underlying goal of the corpus-based terminography.

CHAPTER 6

CONCLUSION

6.1 BRIEF SUMMARY OF THE STUDY

This study has proposed a method for computer-assisted extraction of terms from a domain-specific corpus. The method is based on techniques for formulating term patterns for the extraction of terms in domains, by a template-matching program, from morphologically analysed domain-specific corpus. The development of the method is based on: (1) Computational tools for morphological annotation and for information extraction developed at the University of Helsinki. (2) Constraints of Swahili term structures discovered on the basis of structures of the terms as morphologically analysed in the corpus. The implementation techniques for the method are systematically presented with respect to the formulation of Swahili term patterns and the evaluation of the effectiveness of the patterns in the extraction of Swahili terms in the domains of health care and literature.

The study has been motivated by the need to develop a method for automatic assisted extraction of terms that is capable of extracting terms according to their respective domains. Developing the term extraction method based on constraints of structures of Swahili terms has been another motivation. The underlying assumption for the first motivation has been that previous term extraction methods have failed to extract terms in domains due to improper annotation of the terms in the domain-specific corpus. The second motivation has been based on the assumption that the previous term extraction methods have been developed based on term structural constraints that are not compatible with the structural constraints of Swahili terms. The proposed techniques for the implementation of the method handle the unique annotation of terms in a corpus by incorporating a term-domain attribute among the features for annotating words in a domain-specific corpus. The attribute can be introduced through a lexicon of the morphological analyser or through the rule file of a rewriting program, such as the BETA system, and it is represented by domain-tags for terms in different domains. The morphological analyser and the rewriting system utilise the relevant domain-tag in the lexicon and in the rule file, respectively, to annotate term-words in the corpus. Then, based on the annotated term-words, the structures of multi-word terms are discovered in the analysed corpus and used as the basis for the discovery of the structural constraints for the formulation of patterns for the extraction of terms from the corpus. The availability of computational tools capable of performing proper analysis of terms in a domain-specific corpus has been the basic impetus for this study.

6.2 SUMMARY OF PREVIOUS METHODS FOR TERM COMPILATION

It was pointed out in Chapter 1 that the computer-assisted compilation of terms has relied on the methods of comparative statistics, information extraction and a combination of comparative statistics and information extraction. The compilation by comparative statistics relies on the criterion of frequency counting, whereby term-words in a domain-specific corpus are supposed to show a higher frequency of occurrences than non-term-words. It was pointed out that the criterion of frequency counting overlooks the fact that a term-word may occur a few times in a domain-

specific corpus, depending on a variety of factors. The information extraction method, which has been the focus of this study, applies the annotation of terms in a domain-specific corpus in such a way that the terms may be represented as term-formation patterns for the extraction of terms by a template-matching program. It was pointed out that the previous techniques for this method lacked the capability of annotating terms in such a way that the terms could be represented as term-formation patterns unique to terms in a domain. Reliance on part-of-speech tags alone for the annotation of terms was indicated as the source of this problem. The combination method applies comparative statistics to solve the problems inherent in the method of information extraction.

6.3 SUMMARY OF THE PROPOSED METHOD FOR TERM EXTRACTION.

The proposed method encompasses techniques for the unique analysis of term-words in a domain-specific corpus and for the formulation of term patterns based on constraints of term structures discovered on terms analysed in the corpus. The techniques are developed from the morphological annotation tools described in Chapter 3, and from constraints of structures of Swahili terms elaborated in Chapter 4. The development of the techniques considered two aspects as necessary for the annotation of terms in a domain-specific corpus and for the formulation of term patterns. The four aspects have been determined as: (1) incorporation of a term-domain feature into the morphological analysis of a corpus for term extraction (2) the use of structures of terms analysed in the corpus as a basis for the discovery of constraints of term structures for the formulation of term patterns. The two aspects were illustrated by incorporating the domain-feature into the analysis of Swahili corpus in the domains of health care and literature and by formulating term patterns based on structural constraints of Swahili terms discovered, based on the terms analysed in the health care corpus.

6.4 SUMMARY OF THE EVALUATION

The effectiveness of the developed method was evaluated with respect to the extraction of Swahili health care and literature terms. The evaluation assessed the rate of recall and the precision of the method. The extraction of terms involved separate pattern files of lengths 1 to 5, morphologically analysed health care and literature corpus with term-words tagged by the relevant domain-tag and the template-matching program. The output assessment revealed that the number of unmatched patterns increased in proportion to the length of the sequences, and the majority of extracted items belonged to the sequences of length 2. The recall assessment involved length 2 terms in a portion of the health care extraction corpus. The assessment rated the recall of the method at 97.3%. The precision evaluation showed that 68% of the extracted items qualified as valid terms. Observation of the extracted items that did not qualify as valid terms revealed that the items were in the categories of non-terms and term variants. The non-term items of length 1 belonged to certain locativised, passivised adverbialised and imperative forms of the term-words. The multi-word non-term items belonged to: (1) items containing the falsely extracted single-word items (2) items containing homonyms for general and for specific meanings (3) items extracted in incomplete forms. The presence of term variants was associated with graphic, orthographic, semantic and stylistic factors. The non-term and certain term variant items were removed from the extracted items. Assessment of the raw frequency of occurrences of the extracted valid terms showed that not all the terms could have been compiled by statistic techniques because of the requirement that a term should have a fairly high frequency of

occurrence in the domain-specific corpus. Moreover, characterisation of the terms by their conceptual behaviour showed that most of the terms did not comply with the traditional principles for concept designation.

6.5 RECOMMENDATIONS

Our recommendations are associated with three aspects: the linguistic annotation of terms in a domain-specific corpus, the formulation of term-formation patterns that are unique to terms in a domain, and prospects of computer-assisted compilation of terms by the information extraction method.

6.5.1 Recommendations for the annotation of terms in a domain specific corpus

The technique for term annotation proposed here relies on the term domain-tag and on the comprehensive morphological analysis of the corpus. The two aspects prompt us to make the following recommendations in relation to the annotation of terms in domains:

- 1) The description of terms in a domain-specific corpus should be carried out based on a comprehensive morphological analysis, i.e. the analysis that can accommodate all the necessary features for the description of words in a corpus.
- 2) Term-words in a domain-specific corpus should be analysed as entities of a unique category rather than as nouns. This should be facilitated by introducing the feature 'term domain' among the features for corpus analysis and by using the appropriate domain-tag in the tagging of the terms.

6.5.2 Recommendations for the formulation of term patterns

This study proposes a technique for the corpus-based discovery of term structural constraints for the formulation of term patterns. The constraints are discovered based on terms analysed in the corpus and are used for the formulation of term patterns for term extraction. The following recommendation is reflected from this technique:

- 3) The formulation of term patterns for the extraction of terms in a language should be preceded by the discovery of the possible constraints of structures of the language terms. The discovery should be based on the structures of multi-word terms as analysed in the morphologically annotated domain-specific corpus. Term-words in the analysed corpus, marked by the term-domain tag, should be used as the searching guide for the discovery of the multi-word terms in the corpus. After the terms have been identified, the selected category tags for each word in the terms should be joined together to form sequences of tags for the terms. Then the sequences should be observed in order to determine the possible relationships between all categories represented in the sequences. The tags and their relationships in the sequences provide the possible constraints of term structures that can be represented as term patterns of the language. The presence in the patterns of a domain-tag for a particular domain makes the patterns specific to the domain.

6.5.3 Recommendations for the extraction of terms

It has been pointed out that the annotated term-words in the corpus underlie the extraction of term in domains. The extraction is intended to capture only words that are marked as terms or collocations containing at least one word that is marked as a term. Since the tagging of term-words is effected in the roots, it covers non-term words, such as finite verbs, derived from the roots. The coverage causes over-extraction, i.e. extraction of the non-term words. However, the unmarking of certain term-words leads to under-extraction, i.e. extraction of fewer terms. The following recommendations are proposed for the prevention of over-generation and under-generation during the extraction of terms:

4. The lexicon of a morphological analyser or the rule file for the BETA system should be updated before the extraction of terms. This will ensure that all term-words in the corpus are represented and marked in the lexicon or in the rule file to avoid unmarking of such terms during the tagging of terms in the corpus.
5. The analysed corpus should be cleared out before the extraction of terms in order to remove all words that may not be terms or part of term collocations, e.g. function words and finite verbs.
6. Where cases of over-extraction occur, the erroneously extracted items should be removed manually, but cases of synonyms should not be removed because they reflect the actual use of terms by different specialists.
7. Since much of the terminological status of the extracted items is determined by the criterion of concept representation in a particular domain, the focus of term extraction should be the domain of a corpus. This is important because what might qualify as a term in a corpus of one domain might be a non-term in a corpus of another domain.

6.5.4 Recommendations for the prospects of term compilation by the method of information extraction

The proposed method is expected to facilitate the compilation of corpus-based terms in such a way that the terms can be organised in the term bank according to their domains. The following recommendations are important for the compilation of terms in domains.

8. Corpora for term extraction should be available in domain-specific forms. Where the corpora are not available, they should be created by compiling domain-specific texts and storing them in a machine-readable form. The creation of the corpora should involve the updating of the lexicon of the morphological analyser or the rule file of the BETA system.
10. The compilers of terms, i.e. computational terminographers, and the developers of the relevant computational tools for the compilation, i.e. computer experts, should work together. This is because the compilation requires the appropriate computer tools, and these are normally not developed by terminographers but by computer experts.
11. The traditional attitude opposed to corpus-based terms should change. It has been pointed out that most of the terms extracted did not conform to the traditional principles of terminology. In the traditional terminology, such terms have been negatively considered as causing terminological confusion (Massamba 1997: 87). This attitude needs to change in order for the

terms to be accepted as an important input for the development of term banks and technical dictionaries.

6.6 IMPLICATIONS FOR FUTURE RESEARCH

As has been noted, the focus of the method proposed in this study is the use of the feature ‘term domain’ in order to facilitate the description and extraction of terms in domains. The lexicon of a morphological analyser or the rule file for the BETA system underlies the description. This is expected to have impact in the creation of lexicons or rule files for specialised domains. Additionally, the identification of term-words in the corpus which lies behind the description is expected to have an impact on the creation of an automated system for the recognition of the terms.

6.6.1 Creation of lexicons or BETA rule files for specialised domains

Since the description of terms in a domain-specific corpus is facilitated by the incorporation of the term-domain feature into the lexicon of a morphological analyser or into the rule file for the BETA system, the creation of lexicons or the rule files for specialised domains is necessary for the tagging of terms in different domains. The task involves identification of term-words in different domains and their representation in the lexicon or in the rule files by marking their base forms with the selected domain-tag. Such term-words are available in domain-specific texts and in other sources, such as technical dictionaries and standard lists. For languages that have no domain-specific texts in a computer-readable form, such as Swahili, the creation of the lexicons and the rule files should commence with the building up of the corpora in different domains.

6.6.2 Creation of the system for automatic identification of term-words

The technique proposed in this study for the identification of term-words in the corpus relies on manual sorting of the terms among the nouns and verbal nouns extracted from the corpus. However, development of a computer system for performing the task provides a possible area for future research. The research falls in the area of artificial intelligence (AI) and requires the development of a database of systematised fields and sub-fields of concepts with their respective term-words and the system for the recognition of the terms based on the database.

BIBLIOGRAPHY

Ahmad, K., Davies A., Fulford, H. and Rogers, M. 1994a. "The elaboration of special language terms; the role of contextual examples, representing samples and normative requirements." *Euralex '92 Proceedings I*. Tampere: Studia Translatologica, 139-150.

1994b. "What is a term? The semi-automatic extraction of terms from text." In Snell-Hornby, M., Pöchhacker, F. and Kaindl, K. (eds.), *Translation Studies: an Interdiscipline*. Amsterdam: John Benjamins Publishing Company, 267-277.

Anadiadou, S. 1994. "A methodology for automatic term recognition." In *COLINGS-94. Papers presented to the 15th International Conference on Computational Linguistics*. Kyoto. Vol. 1, 1034-1038.

Arppe, A. 1995. "Term extraction from unrestricted text." In *Nodalida1995*.
<http://www.lingsoft.fi/doc/nptool/term-extraction.html>

Ashton, E. O., 1944. *Swahili Grammar* (including intonation). London: Longman.

Baker, M., Francis G. and Tognini-Bonelli, E. (eds.), 1993. *Text and Technology*. Amsterdam. John Benjamins Publishing Company.

Banbrook, G. 1996. *Language and Computers*. Edinburgh: Edinburgh University Press.

Batibo, H.M. 1988. "Root Affixation Rules in Zairean Kiswahili as Evidence for Earlier Bantu Rules." In *Kiswahili. Journal of the Institute of Kiswahili Research*. Dar es Salaam. University of Dar es Salaam, 58-70.

Ben and Company Ltd. 1995. *Kamusi Awali ya Sayansi na Teknolojia (Elementary Dictionary of Science and Technology)*. Dar es Salaam.

Biber, D. 1993. "Representativeness in corpus design." *Literary and Linguistic Computing*, 8(4). Oxford: Oxford University Press, 243-257.

Bogurajev, B. and Kennedy, C. 1999. "Application of term identification technology: domain description and content characterisation." In *Natural Language Engineering* 5(1). Cambridge University Press, 17-44.

Bourigault, D. 1992. "Surface grammatical analysis for the extraction of terminological noun phrases." In *Proceedings of the 15th International Conference on Computational Linguistics*. Nantes, France.

Braasch, A. 1992. "Text based dictionary work for a domain-specific language." In Kiefer, F. *et al.* (eds.), 71-19.

Butler, C. 1985. *Statistics in Linguistics*. London: Basil Blackwell.

- Dagan, I. and Church, K. 1995. "Termight: identifying and translating technical terminology." In *Proceedings of the 4th Conference on Applied Natural Language Processing*. Stuttgart, German.
- Daille, B. 1996. "Study and Implementation of Combined Techniques for Automatic Extraction of Terminology." In Klavans, J.L. and Resnik, P. (eds.), *The Balancing Act. Combining Symbolic and Statistical Approaches to Language*. Massachusetts & London: The MIT Press Cambridge.
- Daille, B., Gaussier, E. and Lange, J. 1994. "Towards automatic extraction of monolingual and bilingual terminology." In *COLINGS-94. Papers presented to the 15th International Conference on Computational Linguistics*. Kyoto. Vol. 1, 515-521.
- Damerou, F. 1993. "Generating and evaluating domain-oriented multi-word terms from texts." In *Information Processing and Management* 29 (4), 433-447.
- Frantzi, K. 1998. *Automatic recognition of multi-word terms*. Ph.D. Thesis, Manchester Metropolitan University, England.
- Frantzi, K. and Anadiadou, S. 1997. "Automatic term recognition using contextual cues. In *Proceedings of the 3rd DELOS Workshop*, Zurich, Switzerland.
- Garside, R., Leech, G. and McEnry T. (eds.), 1997. *Corpus Annotation*. London & New York: Longman.
- Gerzymisch-Aebogast, H. 1994. "Identifying term variants in context: The SYSTEXT approach." In Snell-Hornby, M., Pöchhacker, F. and Kaindl K. (eds.), *Translation Studies: an Interdiscipline*. Amsterdam: John Benjamins Publishing Company, 279-289.
- Grefenstette, G. 1994. *Explorations in Automatic Thesaurus Discovery*. Kluwer, Dordrecht.
- Hindle, D. 1990. "Noun classification from predicate-argument structures." In *Proceedings of the 28th Annual Meetings of the Association of Computational Linguistics, Pittsburgh, Penna*, 268-275.
- Hodges, J. , Yie, S. , Reighart, R. and Boggess, L. 1996. "An automatic system that assists in the generation of document indexes." In *Natural Language Engineering* 2 (2). Cambridge University Press, 137-160.
- Hopcroft, J. E. and Ullman, J. D. 1979. *Introduction to Automata Theory, Languages, and Computation*. Reading & Massachusetts: Addison-Wesley Publ. Co.
- Hurskainen, A. 1992. BETA: A Tool for a Linguist. Original version written by Karlsson, F. and Koskenniemi, K. in Finnish. Manuscript in English Translation.
- 1992a. "A two-level computer formalism for the analysis of Bantu morphology: An application to Swahili." In *Nordic Journal of African Studies*. 1(1): 87-122.
- 1992b. "Computer Archives of Swahili Language and Folklore - What is it ?" In *Nordic Journal of African Studies*. 1(1): 123-127.
1994. "Kamusi ya Kiswahili Sanifu in test: A computer system for analyzing dictionaries and

- for retrieving lexical data.” In *Afrikanistische Arbeitspapiere 37 (Swahili Forum 1)*, 169-179.
1996. “Disambiguation of morphological analysis in Bantu languages.” In *COLINGS-96: The 16th International Conference on Computational Linguistics. Proceedings*. Vol. 1. Copenhagen, 568- 573.
- (-) 1992. *ISO WD 12 200 Computational aids in terminology - Terminology interchange format (TIF) - An SGML Application*.
- (-) 1992. *ISO 10241 International terminology standards - Preparation and layout*.
- (-) 1994. *ISO TR 12618 Computational aids in terminology - Creation and use of terminological databases and text corpora*.
- (-) 1997. *ISO CD 704 Terminological work - Principles and Methods*.
- Johnston, M., Bogurajev, B. and Pustefovsky, J. 1994. "The structure and interpretation of compound nominals." In *AAI Spring Symposium on Generativity and Lexicon*. Stanford, Ca.
- Justeson, L. and Katz, S. 1995. "Technical terminology: some linguistic properties and an algorithm for identification in text" In *Natural Language Engineering 2 (2)*. Cambridge University Press, 9-27.
- Karlgren, J. 1998. *The Basics of Information Retrieval*. Stockholm: Swedish Institute of Computer Sciences.
- Karlsson, F. 1990. “Constraint Grammar as a framework for parsing running text.” In Karlgren H.S. (ed.), *COLING-90. Papers presented to the 13th International Conference on Computational Linguistics*. Vol. 3. Helsinki, 168-173.
1994. “Robust parsing of unconstrained text.” In Oostdijk, N. *et al.* (eds.), 121-142.
- 1995a. “Designing a parser for unrestricted text.” In Karlsson, F. *et al.* (eds.), 1-40.
- 1995b. “The formalism and environment of Constraint Grammar Parsing.” In Karlsson, F. *et al.* (eds.), 41-88.
- Karlsson, F. and Koskenniemi, K., 1990. *Beta-ohjelma kielentutkijan apuvälineenä*. Helsinki: Yliopistopaino.
- Karlsson, F., Voutilainen, A., Heikkilä, J. and Anttila, A. (eds.), 1995. *Constraint Grammar: A language-Independent System for Parsing Unrestricted Text*. Berlin & New York: Mouton de Gruyter.
- Karttunen, L. 1994. “Constructing Lexical Transducers.” In *COLINGS-94. Papers presented to the 15th International Conference on Computational Linguistics*. Kyoto. Vol. 1, 406-411.
- Khamis, A. M. 1985. *Swahili Verb Derivation*. Ph.D. Dissertation. University of Hawaii.
- Kiefer, F., Kiss, G. and Paiz, J. (eds.), 1992. *Papers in Computational Lexicography Complex*

'92. Budapest: Linguistics Institute of the Hungarian Academy of Sciences.

Klavans, J., Tzoukerman, E. and Jacquemin, C. 1997. "Effective use of NLP techniques for automatic conflation of multi-word terms: the role of derivational morphology, POS tagging, and shallow parsing." In Belkin, N., Narasimhalu, D., Willett, P. (eds.) *Proceedings of the 20th Annual International ACM SIGIR Conference on R&D in IR, SIGIR'97, July 1997*, Philadelphia, Penn, 148-155.

Koskenniemi, K. 1983. *Two-level Morphology: A General Computational Model for Word-Form Recognition and Analysis*. (Publications of the Department of General Linguistics, University of Helsinki, No. 11.) Helsinki: University of Helsinki, Department of General Linguistics.

1985. "A General two-level computational model for word-form recognition and production." In Karlsson, F. (ed.), *Computational Morphosyntax. Report on Research 1981-84*. (Publications of the Department of General Linguistics, University of Helsinki, No.13.) Helsinki: University of Helsinki, Department of General Linguistics, 1-18.

Leech, G. 1993. "Corpus Annotation Schemes." In *Literary and Linguistic Computing*, 8 (4). Oxford: Oxford University Press. 275-281.

1997a. "Introducing corpus annotation." In Garside, R. *et al.* (eds.), 1-18.

1997b. "Grammatical tagging." In Garside, R. *et al.* (eds.), 19-33.

Leonhardt, C. 1990. "Overview of Terminology and Language Planning in Canada." In *NORDTERM 4*. Stockholm, 255-282.

Lynch, L. 1997. "Medical Terminology Management." In Wright, S. E. and Budid, G. (eds.), 160-170.

Massamba, D. P. B. 1997. "Problems in terminology development: The case of Tanzania." In *Kiswahili. Journal of the Institute of Kiswahili Research*. Dar es Salaam. University of Dar es Salaam, 86-98.

Maynard, D. and Ananiadou, S. (1999) "Incorporating Linguistic Information for Multi-Word Term Extraction", in *Proc. of 2nd Computational Linguistics UK. Research Colloquium (CLUK-2)*, University of Essex, UK, 1999.

(2000a). "Term extraction using a similarity based approach." In D. Bourigault, C. Jacquemin and M.-C. L'Homme (eds.), *Recent Advances in Computational Terminology*. John Benjamins, Amsterdam.

(2000b). "Identifying Terms by their Family and Friends" In *Proc. of COLING 2000*, Saarbrücken, Germany, 2000, pp. 530-536.

McEnery, T. and Wilson, A. 1996. *Corpus Linguistics*. Edinburgh University Press.

Meyer, I. 1990. "Computer-assisted Concept Analysis for Terminology" In *ORDTERM 4*. Stockholm, 163-212.

- Mukama, R. G. 1976. "Towards the anatomy of the object selection prohibiting rules in Swahili." In *Kiswahili. Journal of the Institute of Kiswahili Research*. Dar es Salaam. University of Dar es Salaam, Vol. 1. 6-25.
- Mulokozi, M. M. 1991. "English versus Kiswahili in Tanzania secondary education." In Blommaert, J. (ed.), *Swahili Studies: Essays in Honour of Marcel Van Spaandonck*. Ghent. Academic Press, 7-16.
- Mwansoko, H. J. M. 1993a. "Acceptability of terms: The case of the Swahili linguistics and literature terms in Tanzania." In *Multilingua 12-2*. Berlin. Walter de Gruyter. 177-188.
- 1993b. "Compounding as a Method of Swahili Terminology Development" In *Messenger. Journal of Oriental Studies*. University of Moscow. 65-70.
1994. *The Modernization of Swahili Technical Terminologies: An Investigation of the Linguistic and Literature Terminologies*. Ph.D. Dissertation. University of York.
- Nicoletta, C. and Remo, B. 1990. "Acquisition of lexical information from a large textual Italian corpus." In *Proceedings of the Thirteen International Conference on Computational Linguistics*, Helsinki.
- Ohly, R. 1982. "Report on Lexicographic Research at the Friendship Textile Mill." In *Kiswahili. Journal of the Institute of Kiswahili Research*. Dar es Salaam. University of Dar es Salaam, 73-86.
- Oostdijk, N. and Haan, P. (eds.), 1994. *Corpus-Based Research into Language, In Honour of Jan Aarts*. Amsterdam - Atlanta, GA.
- Pearson, J. 1998. *Terms in Context*. Amsterdam & Philadelphia: Benjamins Publishing Company.
- Polome, E. C. 1967. *Swahili Language Handbook*. Washington, D.C.: Center for Applied Linguistics.
- van Rijsbergen, C. J. 1979. *Information Retrieval*. London: Butterworths.
- Sager, J. C. 1990. *Practical Course in Terminology Processing*. Amsterdam & Philadelphia: Benjamins Publishing Company.
- Sampson, G. 1995. *English for the computer: The SUSANNE Corpus and analytic scheme*. Oxford: Clarendon Press.
- Salton, G. 1988 "Syntactic approaches to automatic book indexing." In *Proceedings of the 26th Annual Meeting of the Association of Computational Linguistics*. Buffalo, New York.
- Salton, G. and McGill, M.J. 1983. *Introduction to Modern Information Retrieval*. Singapore: McGraw-Hill, Inc.
- Salton, G., Zhao, Z. and Buckley, C., 1990. "A simple syntactic approach for generation of indexing phrases." In *Technical Report*. Department of Computer Science, Cornell University, 90-1137.

- Sinclair, J. M. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
1992. "The Automatic Analysis of Corpora" In Svartvik, J. (ed.), 379-397.
- Slartvik, J. (ed.), 1992. *Directions in Corpus Linguistics: Proceedings of the Nobel Symposium 82*. Berlin: Mout de Gruyter.
- Smadja, F. A. 1993. "Retrieving collocations from texts: Xtract." In *Computational Linguistics*, 19(1): 143-177.
- Smadja, F. A. and Kathleen, Mc. R. 1990. "Automatically extracting and representing collocations for language generation. In *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*, 252-259.
- Soergel, D. 1985. *Organising Information: Principles of Data Base and Retrieval Systems*. New York: Academic Press, Inc.
- Sproat, R. 1992. *Morphology and Computation*. Cambridge, Massachusetts: The MIT Press.
- Sundheim, B. 1996. "Task oriented text analysis evaluation" In Cole, R. A. , Marian, J. , Uszkoreit, H. , Zeanen, A. and Zue, V. (eds.) *Survey of the State of the Art in Human Language Technology*. <http://www.cse.ogi.edu/CSLU/HLTsurvey>
- Taasisi ya Uchunguzi wa Kiswahili. 1990. *Kamusi Sanifu ya Biolojia, Fizikia, na Kemia (Standard Dictionary of Biology, Physics and Chemistry)*. Dar es Salaam. University of Dar es Salaam.
- Tapanainen, P. 1996. *The Constraint Grammar Parser CG-2*. (Publications of the Department of General Linguistics, University of Helsinki, No. 27.) Helsinki: University of Helsinki, Department of General Linguistics.
- Temu, C. W. 1984. "Kiswahili Terminology: Principles Adopted for the Enrichment of the Kiswahili Language." In *Kiswahili. Journal of the Institute of Kiswahili Research*. Dar es Salaam. University of Dar-es Salaam, 112-127.
- Tumbo-Masabo, Z. 1989. *The Development of Neologisms in Kiswahili: A Diachronic and Synchronic Approach with Special Reference to Mathematical Terms*. Ed.D. Dissertation. Columbia University.
- Voutilainen, A. 1995a. "Morphological disambiguation." In Karlsson *et al.* (eds.), 165-284.
- 1995b. "NP-tool, a detector of English noun phrases." In *Proceedings of Workshop on Very Large Corpora*, June 22, 1993 at Ohio State University.
- Wright, S. E., and Budin, G. (eds.), 1997. *Handbook of Terminology Management, Vol.1, Basic Aspects of Terminology Management*. Amsterdam & Philadelphia: Benjamins Publishing Company.
- Yang, H. Z. 1986. "A new technique for identifying scientific and technical terms and describing science texts." In *Literary and Linguistic Computing*, 1 (2). Oxford: Oxford

University Press, 93-103.

APPENDIX I

Single-word terms identified from the training corpus

afya	dondakoo	jipu
aleji	dondora	joto
ambukizo	dozi	juisi
ameba	embe	kachumbari
anemia	endokrini	kadi
aneurisimu	enema	kalori
anofelesi	fangasi	kalsiam
anofelina	fibroidi	kandida
antibiotiki	figo	kandidiasisi
antibodi	filari	kansa
aorta	filaria	karantini
askari	fistula	karoti
ashki	fundo	kaswende
asidi	funza	katheta
bakteria	gamba	kauri
balehe	ganglia	kemikali
bamia	ganzi	kende
bandama	gego	kiazi
bangi	jenetiki	kibofu
baridi	gimbi	kichefuchefu
bega	glovu	kichocheo
biringanya	gonorea	kichocho
borelia	goti	kichomi
bujuu	gugu	kichwa
busha	haemoglobini	kidole
buu	haemoreji	kidolegumba
chakula	haidrosefalusi	kidoletumbo
chandarua	haja	kidonda
chanjo	hamira	kidonge
chankroidi	hamu	kifaduro
chati	hanithi	kifafa
chawa	hapataitisi	kifereji
chembe	harufu	kifo
choo	hasua	kifua
choroko	hedhi	kifuakikuu
chuchu	hepesi	kifuko
chumvi	heroini	kifumba
chungwa	hewa	kifundo
chunjua	hitilafu	kiganja
chunusi	homa	kiharusi
chupa	homoni	kiini
sisti	hospitali	kiinilishe
daktari	inzi	kijeraha
dalili	jasho	kijidudu
damu	jaundisi	kijimelea
darubini	jeli	kijinyama
dawa	jene	kijipu
degedege	jeraha	kijiuavisumu
disfasia	jicho	kikohozi

kilema
kilengelenge
kiluwiluwi
kimelea
kimeo
kimwili
kinena
kinga
kinyesi
kinywa
kinyweleo
kiotea
kipandikizi
kipele
kipimo
kipindupindu
kipofu
kiporo
kiroboto
kirusi
kirutubisho
kisigino
kisima
kisimi
kisonono
kisononogono
kisukari
kitanzi
kitovu
kiu
kiumbe
kiungo
kiungulia
kiuno
kivimbe
kiwambo
kiwiko
kiwiliwili
kizazi
kizuizi
kizunguzungu
klamidia
kliniki
klorini
klorokwini
kokeni
kokwa
kolestro
kondo
kondomu
konokono
koo
korodani
koromeo
kovu

krimu
kromozomu
kuambukiza
kuangamiza
kubaka
kubalehe
kuchanganyikiwa
kuchanja
kuchoka
kuchunguza
kudhoofu
kudoea
kudumaa
kuenea
kufa
kufahamu
kufyonza
kuhara
kuharisha
kuhasi
kuishi
kujamiiiana
kujeruhi
kujifungua
kujikinga
kujikuna
kukakamaa
kukamua
kukanda
kukinga
kukohoa
kukojoa
kukomaa
kukua
kulala
kulawiti
kulegea
kulemaa
kulikiza
kunajisi
kunde
kunguni
kunyonya
kuota
kuoza
kupasua
kupe
kupevuka
kupima
kupofuka
kupona
kupooza
kupumua
kurithi
kurutubisha

kusambaa
kusedeka
kushona
kushukashtuka
kusionsimua
kusukuma
kutahiri
kutakasa
kutarajia
kutetemeka
kuthuru
kutibu
kuua
kuugua
kuuma
kuvimba
kuvu
kuwasha
kuweweseka
kuyeyuka
kuzaa
kuzimia
kuzirai
kuzuia
kwapa
kyuleksi
kyulisina
lava
leishimania
lengelenge
limfatiki
limfu
lishe
maabara
mabaka
mabingobingo
machozi
madafu
madini
mafua
mafundofundo
mafuta
maharage
mahatama
maji
majimaji
makohozi
malale
malaria
malendalenda
malengelenge
mapacha
mapele
mapigo
maradhi

matapishi	mkereketo	nguvu
matarajio	mkojo	nipo
mate	mkono	nitrojeni
matekenya	mkunga	njiti
matibabu	mlemavu	nyama
matiti	mlenda	nyanya
matumbwitumbwi	mlipuko	nyege
matunzo	mlo	nyonga
maudhi	mmomonyoko	nyonyo
maumbile	mninga	nyuki
maumivu	mnyoo	nyumonia
mavi	moyo	nyurosis
mawese	mpanuko	ogani
maziwa	mpira	oksijeni
mazoezi	mrija	oksotsini
mba	mringaringa	onchoserea
mbaazi	msamba	operesheni
mbazo	msedeko	ostrojeni
mbegu	mshindo	pafu
mboga	mshipa	paja
mbolea	mshono	pamba
mboni	mshtuko	pangusa
mbu	msisimko	panya
mbung'o	mstuko	papai
mchango	msuguano	papasi
mchicha	msukumo	pelviki
mchoko	msuli	penisilini
mchovu	mtaalam	pepopunda
mchuzi	mtindio	pituitari
mchwa	mtishamba	plasenta
mdomo	mtoki	plasmodia
mdudu	muonzi	plastiki
mende	muuguzi	polio
menopozi	muwasho	pombe
metranidazole	mvuke	poju
mfadhaiko	mwanga	projesterone
mfereji	mwata	projestojeni
mfumo	mwayo	prostate
mfupa	mwelimishaji	protini
mgando	mwili	protozoa
mganga	mzizi	pua
mgongo	mzunguko	pumba
mgonjwa	nafaka	pumbu
mgumba	nagana	pumu
mguu	nanasi	pumzi
mharo	nazi	riketisia
mhogo	ndago	rutuba
mimba	ndizi	sabuni
mirasidia	ndorobo	safura
mjamzito	neiseria	salinonela
mkaa	neva	samadi
mkandamizo	nge	samaki
mkao	ngiri	sangare
mkasi	ngozi	sarakani
mkazo	nguru	sekaria

sekeneko
seli
serebela
serebra
shabu
shahawa
shambulio
shankroidi
shavu
shigela
shingo
shinikizo
sifisiotomi
sikio
simuliumu
sindano
siri
sksitosini
spektinomaisini
spiriti
sponji
strongiloidi
sugu
sumu
surua
simfisiotomi
taahira
taenia
tahiri
tako
taksorinkati
tasa
tauni
taya
tegu
tekenya
teknolojia
tende
tetani
tetanasi
tetrasaiklini
tezi
thairoidi
thalamasi
tiba
tindikali
tishu
titi
togwa

tohara
toksakara
toksoplasma
toksorinkati
treponema
trichomonas
trikomonasi
tripanosoma
tumbo
tunda
uambukizaji
uambukizi
uambukizo
uasherati
ubavu
ubongo
ubongombe
ubongonkati
ubongonyuma
uchafu
uchovu
uchungu
uchunguzi
udevu
udhaifu
udongo
udowezi
ufizi
ugonjwa
ugumba
uhanithi
uharo
uhusiano
ujauzito
uke
ukimwi
ukoma
ukunga
ukungu
ukurutu
ulemavu
ulevi
ulimi
ungo
unvyevunyevu
unyayo
unyeo
unyevu
unyonyeshaji

unywele
uoto
upandikizaji
upasujaji
upele
upevukaji
upimaji
upofu
usaha
ushaurinasaha
usingizi
ustawi
utafiti
utakasaji
utando
utapiamlo
ute
utelezi
utindio
utosi
utumbo
utunzaji
uuguzi
uume
uvimbe
uvuguvugu
uwonjo
uwozo
uzazi
usinzi
uzito
uzuiaji
varikosi
vasektomi
veni
via
vitamini
vumbi
wengu
wuchereria
yai
yatima
yisti
yersinia
zeruzeru
zezeta
zinaa
zingizi
ziwa

APPENDIX 2

Term collocations identified in the training corpus

kupevuka kwa yai	kliniki maalum za magonjwa ya kujamiiana
njia ya uzazi	uambukizo katika viungo vya uzazi
njia ya uzazi ya msichana	tumbo la chini
kujifungua salama	maumivu ya kudumu ya tumbo la chini
kukwama kwa mtoto	nyama laini ya njia ya mkojo
kupata ujauzito	kuota vidonda sehemu za siri
lishe duni	vidonda sehemu za siri
shinikizo la damu	vidonga vya sehemu za siri
uchungu wa muda mrefu	chembechembe nyeupe
uchungu wa uzazi	uwezo wa mwili kupambana na magonjwa
misuli ya mji wa mimba	kinga ya mwili dhidi ya maradhi
kujikaza na kuachia kwa misuli ya mji wa mimba	upungufu mkubwa wa uzito
mlango wa mji wa mimba	upungu wa uzito
tendo la kujamiiana	kuharisha kusikoisha
viungo vya nyonga	kikohozi kisichoisha
kutunga mimba nje ya mji wa mimba	maradhi ya ngozi
mfuko wa chakula	kupoteza hamu ya kula
mfuko wa mtoto	kupoteza hamu ya kujamiiana
misuli ya mji wa mimba	uti wa mgongo
hatua ya kwanza ya uchungu	ugonjwa wa uti wa mgongo
mzunguko wa hedhi	homa ya uti wa mgongo
kupevuka kwa yai	kuchanika kwa shingo ya mji wa mimba
siku za rutuba	kuanzisha uchungu
njia za asili za uzazi wa mpango	matarajio ya kujifungua
pima joto	kupasuka mapema kwa chupa ya uzazi
njia ya ute	njia ya haja kubwa
mlango wa shingo ya mji wa mimba	mapigo ya moyo wa mtoto
upevukaji wa yai	kuzalisha kwa kutumia koleo ya uzazi
kurudi kwa hedhi	kujiambia kwa kondo kwenye kuta za mji wa mimba
kuzuia mimba kwa njia ya kufunga uzi kiunoni	upasuaji mkubwa
njia za kijadi za kuzuia mimba	kuzalisha kwa upasuaji mkubwa
kuzuia mimba kwa njia za viziuzi	kuvunjika mapema kwa chupa ya uzazi
sponji ya kuzuia mimba	kitangulizi cha matako
vichocheo vya kuzuia yai kupevuka	kitangulizi cha bega
mirija ya mayai	mlalo wa mtoto
magonjwa ya kujamiiana	mlalo wa ubavu
magonjwa ya zinaa	mlalo wa kuingama
ugonjwa sugu	kukwama kwa mototo
wakati wa ujauzito	pango la nyonga
homa ya mapafu	pango la nyonga la mama
kutoka usaha ukeni	mfuko wa uzazi
uambukizo wa ugonjwa	via vya uzazi
magonjwa ya via vya uzazi	kibofu cha mkojo
matatizo ya hedhi	utumbo mpana
kutoka uchafu wenye harufu ukeni	utumbo mkubwa
dalili za uambukizo wa via vya uzazi	utumbo wa haja kubwa
uambukizo wa via vya uzazi	uchungu wa muda mrefu
mirija ya uzazi	kuta za mji wa mimba
kliniki za afya	tumbo la uzazi

kliniki za uzazi wa mpango
 mshauri wa uzazi wa mpango
 kumwaga mbegu
 kumwaga shahawa
 kumwaga mbegu nje
 upasuaji mdogo
 dalili za hatari za ujauzito
 lishe duni kwa watoto
 afya ya mama na mtoto
 kuzuia mimba
 kuzuia ujauzito
 kuaa kizazi kwa upasuaji
 muhudumu wa afya ya uzazi wa mpango
 afya ya uzazi wa mpango
 kituo cha afya
 kutunga kwa mimba
 mbegu za kiume na kike
 dawa ya povu
 njia ya vizuizi
 mpira wa mama
 mbegu za uzazi za mwanaume
 kiwambo mama
 vidonge vya kumeza vya kuzuia mimba
 vidonge vyenye vichocheo viwili
 kiliniki ya uzazi wa mpango
 kusikia maumivu ya tumbo
 kuongezeka uzito
 kujaa matiti
 mgando wa damu
 ugonjwa wa moyo
 kansa ya ziwa
 sarakani ya ziwa
 kansa ya shingo ya mji wa mimba
 kutokwa damu ukenu
 kinga ya uambukizo
 matumizi ya vidonge
 upevushaji wa yai
 sindano ya kuzuia mimba
 vijiti vyenye kichocheo cha projestojeni
 via vya ndani vya uzazi
 maumivu ya mgongo
 kufunga kizazi
 sindano ya ganzi
 ngozi ya korodani
 chanjo za mtoto
 njia za uzazi wa mpango
 vidonge vya chuma
 kichocheo cha ostrojini
 mpara wa mwanamke
 damu ya hedhi
 kifafa cha uzazi
 kifafa cha mimba
 kutoka majimaji ukenu
 mhudumu wa afya
 mtoto mchanga

uzi wa kitovu
 maji ya sukari
 mapigo ya moyo
 kadi ya maendeleo ya ukuaji wa mtoto
 viganja vya mtoto
 sehemu za siri za mtoto
 vichocheo vya mama
 chunusi ndogondogo
 watoto mapacha
 magonjwa ya kuambukiza
 homa ya manjano
 upungufu wa damu
 ugonjwa wa kuharisha
 chanjo ya surua
 chanjo ya pepopunda
 chanjo ya kifaduro
 maziwa ya mama
 maziwa ya chupa
 maziwa ya n'gombe
 chembechembe za protini
 kujiminywa kwa mji wa mimba
 maji mazito ya njano
 chuchu zenye nyufa
 mzingo wa chuchu
 chembechembe za asili
 matiti yenye uambukizo
 jipu la titi
 chuchu mbonyeo
 kusukuma kondo la nyuma nje
 kusukuma nje kondo la nyuma
 kusisimua mji wa mimba
 kunyonyesha mtoto maziwa ya mama
 kichwa cha chuchu
 kufunguka kwa mlango wa mji wa mimba
 kulegea kwa misuli ya mji wa mimba
 kupasuka chupa mapema
 yai kujishikiza nje ya mji wa mimba
 mimba nje ya kizazi
 hitilafu ya maumbile ya chembechembe
 nyekundu za damu
 kuvunjika kwa chupa
 uchungu wa kuzaa
 mfuko wa maji ya mtoto
 arobaini ya uzazi
 upasuaji wa mfupa wa kinena
 misuli ya tumbo
 dalili za mtoto kuchoka
 mfupa wa kinena
 jinsia ya mtoto
 jukumu la kibiolojia
 kuongeza njia ukenu
 operesheni ya kuongeza njia
 kumpa mama enema
 kuambua kondo la nyuma
 pango la mji wa mimba

ugonjwa wa akili
mabadiliko ya kimaumbile
kidonda cha upasuaji

yai lilitunga mimba
sindano iliyotakaswa

APPENDIX 3.

Health care terms recovered in the extraction

HC-N	"llnt"		29	"llnt:1"	kipindupindu
			28	"llnt:1"	ubongo
406	"llnt:1"	afya	28	"llnt:1"	pombe
328	"llnt:1"	maji	27	"llnt:1"	sumu
320	"llnt:1"	mimba	27	"llnt:1"	joto
313	"llnt:1"	uzazi	26	"llnt:1"	uchafu
269	"llnt:1"	damu	26	"llnt:1"	kipimo
266	"llnt:1"	ugonjwa	24	"llnt:1"	safura
164	"llnt:1"	dawa	24	"llnt:1"	mapacha
160	"llnt:1"	mwili	24	"llnt:1"	kinga
160	"llnt:1"	chakula	23	"llnt:1"	uasherati
152	"llnt:1"	dalili	23	"llnt:1"	polio
131	"llnt:1"	ngozi	22	"llnt:1"	upofu
130	"llnt:1"	vijidudu	22	"llnt:1"	siri
113	"llnt:1"	nyonga	22	"llnt:1"	koo
109	"llnt:1"	kliniki	22	"llnt:1"	kiungo
95	"llnt:1"	chanjo	22	"llnt:1"	kibofu
91	"llnt:1"	maziwa	22	"llnt:1"	chupa
89	"llnt:1"	malaria	21	"llnt:1"	utumbo
84	"llnt:1"	homa	21	"llnt:1"	udongo
83	"llnt:1"	mkojo	21	"llnt:1"	sukari
73	"llnt:1"	ukimwi	21	"llnt:1"	mgongo
73	"llnt:1"	kondo	21	"llnt:1"	mapafu
70	"llnt:1"	mgonjwa	21	"llnt:1"	jeraha
69	"llnt:1"	hospitali	20	"llnt:1"	utapiamlo
63	"llnt:1"	kitovu	20	"llnt:1"	surua
62	"llnt:1"	maradhi	20	"llnt:1"	mikazo
60	"llnt:2"	utunzaji	20	"llnt:1"	mdomo
58	"llnt:1"	tumbo	20	"llnt:1"	mafuta
58	"llnt:1"	mbu	19	"llnt:1"	viluwiluwi
57	"llnt:1"	kifua	19	"llnt:1"	mrija
54	"llnt:1"	kinyesi	19	"llnt:1"	mpira
52	"llnt:1"	uzito	18	"llnt:1"	vipele
51	"llnt:1"	matibabu	18	"llnt:1"	ulemavu
50	"llnt:1"	utafiti	18	"llnt:1"	nyama
49	"llnt:1"	kaswende	18	"llnt:1"	mbegu
48	"llnt:1"	usafi	18	"llnt:1"	madini
47	"llnt:1"	zinaa	17	"llnt:1"	tegu
47	"llnt:1"	maumivu	17	"llnt:1"	sabuni
47	"llnt:1"	hewa	17	"llnt:1"	mbung'o
46	"llnt:1"	nguvu	17	"llnt:1"	mboga
46	"llnt:1"	choo	16	"llnt:1"	ziwa
45	"llnt:1"	madhara	16	"llnt:1"	zahanati
44	"llnt:1"	moyo	16	"llnt:1"	utosi
44	"llnt:1"	kichocho	16	"llnt:1"	utomwili
41	"llnt:1"	tauni	16	"llnt:1"	usaha
40	"llnt:2"	ukuaji	16	"llnt:1"	upele
40	"llnt:1"	tezi	16	"llnt:1"	pepopunda
39	"llnt:1"	mikono	16	"llnt:1"	hedhi
38	"llnt:1"	mfupa	15	"llnt:1"	daktari
37	"llnt:1"	msamba	15	"llnt:1"	ameba
37	"llnt:1"	mfumo	14	"llnt:1"	yai
37	"llnt:1"	lishe	14	"llnt:1"	pamba
36	"llnt:1"	kinena	14	"llnt:1"	kichaa
34	"llnt:1"	malale	13	"llnt:2"	upimaji
33	"llnt:1"	sindano	13	"llnt:2"	uchanjaji
33	"llnt:1"	kisonono	13	"llnt:1"	viinilishe
32	"llnt:1"	kidonda	13	"llnt:1"	viini
31	"llnt:1"	tiba	13	"llnt:1"	majimaji
30	"llnt:1"	uke	12	"llnt:1"	viotea

12	"llnt:1"	unyeo	6	"llnt:1"	chunjua
12	"llnt:1"	umbile	6	"llnt:1"	bega
12	"llnt:1"	uchunguzi	5	"llnt:1"	vitamini
12	"llnt:1"	mzunguko	5	"llnt:1"	viazi
12	"llnt:1"	mlo	5	"llnt:1"	unyevunyevu
12	"llnt:1"	darubini	5	"llnt:1"	unyevu
11	"llnt:1"	viroboto	5	"llnt:1"	tende
11	"llnt:1"	usingizi	5	"llnt:1"	sugu
11	"llnt:1"	pua	5	"llnt:1"	sodo
11	"llnt:1"	protini	5	"llnt:1"	shashi
11	"llnt:1"	kuvu	5	"llnt:1"	pumzi
11	"llnt:1"	kemikali	5	"llnt:1"	oksijeni
11	"llnt:1"	figo	5	"llnt:1"	nyuki
11	"llnt:1"	baridi	5	"llnt:1"	mganga
10	"llnt:1"	wanga	5	"llnt:1"	kifaduro
10	"llnt:1"	vena	5	"llnt:1"	karanga
10	"llnt:1"	ulevi	5	"llnt:1"	jipu
10	"llnt:1"	ujauzito	5	"llnt:1"	hanamu
10	"llnt:1"	udhaifu	5	"llnt:1"	asidi
10	"llnt:1"	uchungu	4	"llnt:2"	ushonaji
10	"llnt:1"	karoti	4	"llnt:2"	upasujaji
10	"llnt:1"	kansa	4	"llnt:2"	unyonyeshaji
9	"llnt:1"	uhusiano	4	"llnt:2"	umwagiliaji
9	"llnt:1"	mguu	4	"llnt:1"	zoezi
9	"llnt:1"	maambukizo	4	"llnt:1"	usichana
9	"llnt:1"	jino	4	"llnt:1"	paja
9	"llnt:1"	haja	4	"llnt:1"	pafu
9	"llnt:1"	chumvi	4	"llnt:1"	njiti
8	"llnt:1"	ute	4	"llnt:1"	nishati
8	"llnt:1"	ulimi	4	"llnt:1"	musuli
8	"llnt:1"	plastiki	4	"llnt:1"	msongo
8	"llnt:1"	papasi	4	"llnt:1"	mfadhaiko
8	"llnt:1"	nywele	4	"llnt:1"	matiti
8	"llnt:1"	mshipa	4	"llnt:1"	mahusiano
8	"llnt:1"	mkasi	4	"llnt:1"	chungwa
8	"llnt:1"	konokono	4	"llnt:1"	maabara
8	"llnt:1"	chembe	4	"llnt:1"	kunguni
7	"llnt:1"	utando	4	"llnt:1"	kunde
7	"llnt:1"	tishu	4	"llnt:1"	kizazi
7	"llnt:1"	sikio	4	"llnt:1"	kiwiliwili
7	"llnt:1"	samaki	4	"llnt:1"	kinywa
7	"llnt:1"	njaa	4	"llnt:1"	kifo
7	"llnt:1"	neva	4	"llnt:1"	harufu
7	"llnt:1"	mnyoo	4	"llnt:1"	dutu
7	"llnt:1"	malengelenge	4	"llnt:1"	chawa
7	"llnt:1"	mafua	4	"llnt:1"	ambukizo
7	"llnt:1"	mabaka	3	"llnt:1"	yatima
7	"llnt:1"	kiuno	3	"llnt:1"	vumbi
7	"llnt:1"	kilema	3	"llnt:1"	uvuguvugu
7	"llnt:1"	kikohozi	3	"llnt:1"	uume
7	"llnt:1"	hamu	3	"llnt:1"	ustawi
7	"llnt:1"	chirwa	3	"llnt:1"	ukurutu
6	"llnt:2"	uzalishaji	3	"llnt:1"	ukungu
6	"llnt:2"	uenezaji	3	"llnt:1"	ukunga
6	"llnt:2"	uambukizaji	3	"llnt:1"	uchovu
6	"llnt:1"	ukoma	3	"llnt:1"	trikomnasi
6	"llnt:1"	sigara	3	"llnt:1"	seli
6	"llnt:1"	ndizi	3	"llnt:1"	pumbu
6	"llnt:1"	mtoki	3	"llnt:1"	nundu
6	"llnt:1"	kombe	3	"llnt:1"	ngono
6	"llnt:1"	kokwa	3	"llnt:1"	msukumo
6	"llnt:1"	kiembe	3	"llnt:1"	mkazo
6	"llnt:1"	kano	3	"llnt:1"	mjamzito
6	"llnt:1"	jicho	3	"llnt:1"	mentari
6	"llnt:1"	hasua	3	"llnt:1"	mchango

3	"llnt:1"	mapigo	2	"llnt:1"	jasho
3	"llnt:1"	maboga	2	"llnt:1"	inzi
3	"llnt:1"	kupe	2	"llnt:1"	ini
3	"llnt:1"	kohozi	2	"llnt:1"	ganzi
3	"llnt:1"	kivimbe	2	"llnt:1"	gango
3	"llnt:1"	kisima	2	"llnt:1"	floridi
3	"llnt:1"	kipuku	2	"llnt:1"	fizi
3	"llnt:1"	kipele	2	"llnt:1"	dondora
3	"llnt:1"	kimelea	2	"llnt:1"	degedege
3	"llnt:1"	kidonge	2	"llnt:1"	bendeji
3	"llnt:1"	kidole	2	"llnt:1"	begani
3	"llnt:1"	kichefuchefu	2	"llnt:1"	balehe
3	"llnt:1"	kauri	2	"llnt:1"	ateri
3	"llnt:1"	karantini	2	"llnt:1"	antijeni
3	"llnt:1"	homoni	2	"llnt:1"	antibodi
3	"llnt:1"	hitilafu	2	"llnt:1"	alkali
3	"llnt:1"	enema	1	"llnt:2"	mzalishaji
3	"llnt:1"	busha	1	"llnt:1"	utoto
3	"llnt:1"	anemia	1	"llnt:1"	utineja
2	"llnt:2"	uzuiaji	1	"llnt:1"	utasa
2	"llnt:2"	utakasaji	1	"llnt:1"	usumushaji
2	"llnt:2"	usambazaji	1	"llnt:1"	unene
2	"llnt:1"	vioevu	1	"llnt:1"	ulegevu
2	"llnt:1"	vinyweleo	1	"llnt:1"	ukavu
2	"llnt:1"	vidudu	1	"llnt:1"	uchachu
2	"llnt:1"	toksorinkati	1	"llnt:1"	tunda
2	"llnt:1"	toksoiditetanasi	1	"llnt:1"	trakoma
2	"llnt:1"	tindikali	1	"llnt:1"	togwa
2	"llnt:1"	tetanasi	1	"llnt:1"	tetekuwanga
2	"llnt:1"	teknolojia	1	"llnt:1"	tako
2	"llnt:1"	taenia	1	"llnt:1"	taahira
2	"llnt:1"	pumba	1	"llnt:1"	soya
2	"llnt:1"	operesheni	1	"llnt:1"	simsim
2	"llnt:1"	ogani	1	"llnt:1"	siagi
2	"llnt:1"	nge	1	"llnt:1"	shinikizo
2	"llnt:1"	ndui	1	"llnt:1"	shingo
2	"llnt:1"	nafaka	1	"llnt:1"	shambulio
2	"llnt:1"	mvuke	1	"llnt:1"	shabu
2	"llnt:1"	mshtuko	1	"llnt:1"	sekeneko
2	"llnt:1"	mshono	1	"llnt:1"	rutuba
2	"llnt:1"	mmomonyoko	1	"llnt:1"	rika
2	"llnt:1"	mlipuko	1	"llnt:1"	pumu
2	"llnt:1"	mkundu	1	"llnt:1"	protozoa
2	"llnt:1"	mkao	1	"llnt:1"	pituitari
2	"llnt:1"	mkaja	1	"llnt:1"	pini
2	"llnt:1"	mhogo	1	"llnt:1"	pedi
2	"llnt:1"	mfereji	1	"llnt:1"	nyanya
2	"llnt:1"	mgando	1	"llnt:1"	njugu
2	"llnt:1"	mbolea	1	"llnt:1"	njegere
2	"llnt:1"	mbavu	1	"llnt:1"	nguru
2	"llnt:1"	mavuzi	1	"llnt:1"	ndorobo
2	"llnt:1"	maungo	1	"llnt:1"	mzizi
2	"llnt:1"	matapishi	1	"llnt:1"	mwata
2	"llnt:1"	maharagwe	1	"llnt:1"	msuli
2	"llnt:1"	maharage	1	"llnt:1"	monoksidi
2	"llnt:1"	magimbi	1	"llnt:1"	mbaazi
2	"llnt:1"	mabuu	1	"llnt:1"	mavumbi
2	"llnt:1"	limfu	1	"llnt:1"	maunzi
2	"llnt:1"	krimu	1	"llnt:1"	maunganio
2	"llnt:1"	klorini	1	"llnt:1"	matunzo
2	"llnt:1"	kitambi	1	"llnt:1"	matubwitubwi
2	"llnt:1"	kisononogono	1	"llnt:1"	matekenya
2	"llnt:1"	kipofu	1	"llnt:1"	panya
2	"llnt:1"	kimo	1	"llnt:1"	mango
2	"llnt:1"	jinsia	1	"llnt:1"	limau

1	"llnt:1"	makombe	1	"llnt:1"	kaboni
1	"llnt:1"	makamasi	1	"llnt:1"	juisi
1	"llnt:1"	goti	1	"llnt:1"	goti
1	"llnt:1"	madafu	1	"llnt:1"	goita
1	"llnt:1"	konjaktiva	1	"llnt:1"	ganzi
1	"llnt:1"	kondomu	1	"llnt:1"	gamba
1	"llnt:1"	klorokwini	1	"llnt:1"	fundu
1	"llnt:1"	kizunguzungu	1	"llnt:1"	filari
1	"llnt:1"	kizuzi	1	"llnt:1"	fangasi
1	"llnt:1"	kiziwi	1	"llnt:1"	eksirei
1	"llnt:1"	kisukari	1	"llnt:1"	dhara
1	"llnt:1"	kirusi	1	"llnt:1"	choroko
1	"llnt:1"	kiporo	1	"llnt:1"	chokaa
1	"llnt:1"	kiinitete	1	"llnt:1"	chati
1	"llnt:1"	kiini	1	"llnt:1"	chandarua
1	"llnt:1"	kigutu	1	"llnt:1"	bandama
1	"llnt:1"	kigasha	1	"llnt:1"	aorta
1	"llnt:1"	kiganja	1	"llnt:1"	antiseptiki
1	"llnt:1"	kifadungu	1	"llnt:1"	alizeti
1	"llnt:1"	kichomi	1	"llnt:1"	aleji
1	"llnt:1"	kalori	1	"llnt:1"	aiskrimu

HC-V "llvt"

147	"llvt:2"	kuzuia	7	"llvt:2"	kuauni
71	"llvt:2"	kuhara	6	"llvt:2"	kusambaa
60	"llvt:2"	kuzaa	6	"llvt:2"	kufyonza
54	"llvt:2"	kuvimba	6	"llvt:2"	kudhoofika
50	"llvt:2"	kunyonya	6	"llvt:2"	kuchanganyikiwa
46	"llvt:2"	kukua	5	"llvt:2"	kuzuilika
46	"llvt:2"	kuambukiza	5	"llvt:2"	kuzirai
44	"llvt:2"	kuenea	5	"llvt:2"	kuumia
40	"llvt:2"	kuumwa	5	"llvt:2"	kutetemeka
38	"llvt:2"	kupima	5	"llvt:2"	kusedeka
36	"llvt:2"	kutapika	5	"llvt:2"	kusambaza
34	"llvt:2"	kukojoa	5	"llvt:2"	kulikiza
34	"llvt:2"	kuishi	5	"llvt:2"	kulazwa
29	"llvt:2"	kupumua	5	"llvt:2"	kuchoka
26	"llvt:2"	kulevya	4	"llvt:2"	kuwasha
25	"llvt:2"	kunyonyesha	4	"llvt:2"	kutibika
24	"llvt:2"	kupevuka	4	"llvt:2"	kupasua
22	"llvt:2"	kuua	4	"llvt:2"	kulemaa
20	"llvt:2"	kuzalisha	4	"llvt:2"	kukonda
20	"llvt:2"	kuuma	4	"llvt:2"	kukomaa
19	"llvt:2"	kukohoa	4	"llvt:2"	kukamua
19	"llvt:2"	kueneza	3	"llvt:2"	kuumbuka
17	"llvt:2"	kupungua	3	"llvt:2"	kutakasa
16	"llvt:2"	kuona	3	"llvt:2"	kupofuka
15	"llvt:2"	kuchunguza	3	"llvt:2"	kuoza
14	"llvt:2"	kuzaliana	3	"llvt:2"	kumwagilia
14	"llvt:2"	kupooza	3	"llvt:2"	kulegea
13	"llvt:2"	kunawa	3	"llvt:2"	kukaza
12	"llvt:2"	kuugua	3	"llvt:2"	kufahamu
12	"llvt:2"	kuoga	3	"llvt:2"	kudumaa
11	"llvt:2"	kuungua	2	"llvt:2"	kuzaana
11	"llvt:2"	kusukuma	2	"llvt:2"	kuunguza
10	"llvt:2"	kutunza	2	"llvt:2"	kushona
9	"llvt:2"	kutunga	2	"llvt:2"	kurutubisha
9	"llvt:2"	kumeza	2	"llvt:2"	kurithi
9	"llvt:2"	kulala	2	"llvt:2"	kuogecha
9	"llvt:2"	kukinga	2	"llvt:2"	kunywea
9	"llvt:2"	kujamiana	2	"llvt:2"	kunenepa
8	"llvt:2"	kuchanja	2	"llvt:2"	kukanda
7	"llvt:2"	kutibu	2	"llvt:2"	kukakamaa
7	"llvt:2"	kugundua	2	"llvt:2"	kudoea

2	"11vt:2"	kudhuru	1	"11vt:2"	kutetema
2	"11vt:2"	kuathirika	1	"11vt:2"	kusisimua
1	"11vt:2"	kuzimua	1	"11vt:2"	kushtuka
1	"11vt:2"	kuzimia	1	"11vt:2"	kupevusha
1	"11vt:2"	kuzibua	1	"11vt:2"	kumwaga
1	"11vt:2"	kuweweseka	1	"11vt:2"	kumeng'enya
1	"11vt:2"	kuvimbiwa	1	"11vt:2"	kulawiti
1	"11vt:2"	kuvia	1	"11vt:2"	kukuza
1	"11vt:2"	kuulisha	1	"11vt:2"	kuathiri
1	"11vt:2"	kutoga	1	"11vt:2"	kuambukizana
HC-N + HC-N "12nt:1"					
3	"12nt:1"	donda koo	1	"12nt:1"	maziwa mgando
1	"12nt:1"	viini lishe	1	"12nt:1"	madini chokaa
1	"12nt:1"	monoksidi kaboni	1	"12nt:1"	antijeni polio
HC-N + N "12nt:2"					
7	"12nt:2"	minyoo askari	1	"12nt:2"	uchanjaji watoto
4	"12nt:2"	minyoo	1	"12nt:2"	uchanjaji vijijini
		strongyloides	1	"12nt:2"	tiba haraka
2	"12nt:2"	minyoo trichuris	1	"12nt:2"	minyoo enteroblus
1	"12nt:2"	vimelea gonorrhoea	1	"12nt:2"	mgonjwa mwanamke
1	"12nt:2"	uzazi salama	1	"12nt:2"	maji salama
1	"12nt:2"	upele mdogomdogo	1	"12nt:2"	maji lita
1	"12nt:2"	unyevu kiasi	1	"12nt:2"	madini chuma
1	"12nt:2"	ugonjwa hatari	1	"12nt:2"	kemikali babuzi
1	"12nt:2"	uzalishaji	1	"12nt:2"	dawa makini
1	"12nt:2"	viwandani	1	"12nt:2"	dalili bayana
N + HC-N "12nt:3"					
11	"12nt:3"	uja uzito	1	"12nt:3"	usalama mwilini
5	"12nt:3"	kinga mwilini	1	"12nt:3"	mtoto njiti
2	"12nt:3"	wanawake	1	"12nt:3"	mama mjamzito
		wajawazito	1	"12nt:3"	magonjwa sugu
2	"12nt:3"	nyuzi joto	1	"12nt:3"	mabadiliko kimwili
2	"12nt:3"	mtoto tumboni	1	"12nt:3"	akina mama
2	"12nt:3"	kiongeza nguvu			wajawazito
1	"12nt:3"	watoto yatima			
HC-V + HC-N "12vt:5"					
10	"12vt:5"	kuhara damu	1	"12vt:5"	kupona maradhi
6	"12vt:5"	kunawa mikono	1	"12vt:5"	kupasua ngozi
5	"12vt:5"	kunywa pombe	1	"12vt:5"	kuota nywele
4	"12vt:5"	kuvimba tezi	1	"12vt:5"	kuota kuvu
4	"12vt:5"	kuvimba miguu	1	"12vt:5"	kuoga mwili
4	"12vt:5"	kupima ukeni	1	"12vt:5"	kuoga majini
4	"12vt:5"	kunywa maji	1	"12vt:5"	kumeng'enya
3	"12vt:5"	kuua vijidudu			chakula
3	"12vt:5"	kunywa sumu	1	"12vt:5"	kulemaa miguu
3	"12vt:5"	kukojoa damu	1	"12vt:5"	kulazwa
2	"12vt:5"	kuumwa tumbo			hospitalini
2	"12vt:5"	kusambaa mwilini	1	"12vt:5"	kulala usingizi
1	"12vt:5"	kuwashwa unyeoni	1	"12vt:5"	kueneza vijidudu
1	"12vt:5"	kuwashwa macho	1	"12vt:5"	kuenea mwilini
1	"12vt:5"	kuvimba mitoki	1	"12vt:5"	kudoea damu
1	"12vt:5"	kuumwa viungo	1	"12vt:5"	kuchanja ndui
1	"12vt:5"	kuuma miguuni	1	"12vt:5"	kuchanika msamba
1	"12vt:5"	kuugua kansa	1	"12vt:5"	kuambukizwa
1	"12vt:5"	kupungukiwa maji			vijidudu
1	"12vt:5"	kupungua uzito			

HC-V + N "l2vt:6"

3	"l2vt:6"	kuumwa kichwa	1	"l2vt:6"	kupasua kuni
2	"l2vt:6"	kuvimba matezi	1	"l2vt:6"	kuota ndoto
2	"l2vt:6"	kuungua moto	1	"l2vt:6"	kuenea haraka
2	"l2vt:6"	kuugua magonjwa	1	"l2vt:6"	kuchanganyikiwa
1	"l2vt:6"	kutibiwa nyumbani			akili
1	"l2vt:6"	kutapika kusudi			

INF + HC-N "l2vt:7"

16	"l2vt:7"	kutoka damu	1	"l2vt:7"	kushambulia ogani
5	"l2vt:7"	kula chakula	1	"l2vt:7"	kupitia nyeneoni
4	"l2vt:7"	kuzama majini	1	"l2vt:7"	kupika nyama
4	"l2vt:7"	kutoka hospitali	1	"l2vt:7"	kupelekea
3	"l2vt:7"	kufia tumboni			maambukizo
2	"l2vt:7"	kutoka usaha	1	"l2vt:7"	kupakaa mwilini
2	"l2vt:7"	kupeleka	1	"l2vt:7"	kula udongo
2	"l2vt:7"	kupelekwa	1	"l2vt:7"	kula mlo
		hospitalini	1	"l2vt:7"	kula mboga
2	"l2vt:7"	kupelekwa	1	"l2vt:7"	kukosekana choo
		hospitali	1	"l2vt:7"	kukosa usingizi
2	"l2vt:7"	kula nyama	1	"l2vt:7"	kukauka midomo
2	"l2vt:7"	kukosa maji	1	"l2vt:7"	kukaa hospitalini
2	"l2vt:7"	kuingia majini	1	"l2vt:7"	kujaribu dawa
2	"l2vt:7"	kufika kliniki	1	"l2vt:7"	kuishiwa pumzi
2	"l2vt:7"	kucheza majini	1	"l2vt:7"	kuishiwa nguvu
1	"l2vt:7"	kuvuma sikioni	1	"l2vt:7"	kuingiza puani
1	"l2vt:7"	kutoka	1	"l2vt:7"	kuingiza kinywani
		malengelenge	1	"l2vt:7"	kuimarisha misuli
1	"l2vt:7"	kutanda kinenani	1	"l2vt:7"	kufufua misuli
1	"l2vt:7"	kusia mbegu	1	"l2vt:7"	kupekua mimba
1	"l2vt:7"	kushambulia	1	"l2vt:7"	kueleka mgongoni
		vijidudu	1	"l2vt:7"	kubeba vijidudu

HC-N + GEN-CON + HC-N "l2nt:8"

14	"l2nt:8"	viungo vya uzazi	5	"l2nt:8"	kiungo cha uzazi
14	"l2nt:8"	vijidudu vya	4	"l2nt:8"	vimelea vya ukimwi
		ugonjwa	4	"l2nt:8"	vimelea vya malale
13	"l2nt:8"	mfumo wa afya	4	"l2nt:8"	uzazi wa mapacha
12	"l2nt:8"	dalili za ugonjwa	4	"l2nt:8"	ugonjwa wa moyo
10	"l2nt:8"	homa ya mbu	4	"l2nt:8"	msongo wa pumzi
9	"l2nt:8"	viotea vya ngozi	4	"l2nt:8"	minyoo ya tegu
9	"l2nt:8"	ugonjwa wa upofu	4	"l2nt:8"	minyoo ya kichocho
9	"l2nt:8"	mfupa wa kinena	4	"l2nt:8"	miili ya viumbe
9	"l2nt:8"	maumivu ya tumbo	4	"l2nt:8"	mifupa ya kinena
9	"l2nt:8"	homa ya matumbo	4	"l2nt:8"	mayai ya kichocho
8	"l2nt:8"	mrija wa mpira	4	"l2nt:8"	maji ya baridi
7	"l2nt:8"	ugonjwa wa malale	4	"l2nt:8"	kinga ya maradhi
7	"l2nt:8"	mayai ya minyoo	3	"l2nt:8"	wagonjwa wa kifua
6	"l2nt:8"	virusi vya ukimwi	3	"l2nt:8"	ugonjwa wa zinaa
6	"l2nt:8"	vimelea vya	3	"l2nt:8"	ugonjwa wa kifua
		malaria	3	"l2nt:8"	ugonjwa wa
		mishipa ya damu			kaswende
6	"l2nt:8"	mazoezi ya viungo	3	"l2nt:8"	ugonjwa wa ameba
6	"l2nt:8"	maji ya mimba	3	"l2nt:8"	utunzaji wa kitovu
6	"l2nt:8"	kibofu cha mkojo	3	"l2nt:8"	unywaji wa pombe
5	"l2nt:8"	usafi wa chakula	3	"l2nt:8"	tumbo la uzazi
5	"l2nt:8"	ugonjwa wa tauni	3	"l2nt:8"	tishu za mwili
5	"l2nt:8"	ugonjwa wa malaria	3	"l2nt:8"	mwili wa mbu
5	"l2nt:8"	ugonjwa wa figo	3	"l2nt:8"	mshipa wa vena
5	"l2nt:8"	tauni ya tezi	3	"l2nt:8"	mimba ya mapacha
5	"l2nt:8"	mzunguko wa damu	3	"l2nt:8"	mifupa ya nyonga
5	"l2nt:8"	minyoo ya safura	3	"l2nt:8"	mapigo ya moyo

3	"12nt:8"	maambukizo ya vijidudu	1	"12nt:8"	viroboto vya ugonjwa
3	"12nt:8"	kuvu ya kinywani	1	"12nt:8"	vipimo vya ukimwi
3	"12nt:8"	kombe la kiuno	1	"12nt:8"	vipimo vya uchunguzi
3	"12nt:8"	kipimo cha dawa	1	"12nt:8"	vipimo vya damu
3	"12nt:8"	jipu la ziwa	1	"12nt:8"	vipimo vya choo
3	"12nt:8"	homa ya papasi	1	"12nt:8"	vipele vya viganja
3	"12nt:8"	hamu ya chakula	1	"12nt:8"	vipele vya kaswende
3	"12nt:8"	dawa ya chanjo	1	"12nt:8"	vimelea vya trikomonasi
2	"12nt:8"	vyakula vya wanga	1	"12nt:8"	vimelea vya kipindupindu
2	"12nt:8"	viungo vya mwili	1	"12nt:8"	vimelea vya ameba
2	"12nt:8"	vipele vya ngozi	1	"12nt:8"	viluwiluwi vya safura
2	"12nt:8"	vimelea vya ugonjwa	1	"12nt:8"	viluwiluwi vya minyoo
2	"12nt:8"	vimelea vya tauni	1	"12nt:8"	vilema vya miguu
2	"12nt:8"	vimelea vya homa	1	"12nt:8"	vijidudu vya upele
2	"12nt:8"	vijidudu vya maradhi	1	"12nt:8"	vijidudu vya malaria
2	"12nt:8"	vijidudu vya kifua	1	"12nt:8"	vijidudu vya kichocho
2	"12nt:8"	ulemavu wa viungo	1	"12nt:8"	viinilishe vya utomwili
2	"12nt:8"	ugonjwa wa ukimwi	1	"12nt:8"	viini vya ugonjwa
2	"12nt:8"	ugonjwa wa macho	1	"12nt:8"	viini vya maradhi
2	"12nt:8"	ugonjwa wa fizi	1	"12nt:8"	viini vya kipindupindu
2	"12nt:8"	udongo wa kauri	1	"12nt:8"	viini vya ameba
2	"12nt:8"	uchafu wa maji	1	"12nt:8"	viganja vya mikono
2	"12nt:8"	utunzaji wa wajawazito	1	"12nt:8"	vifereji vya mkojo
2	"12nt:8"	ushonaji wa msamba	1	"12nt:8"	vidonge vya madini
2	"12nt:8"	unyonyeshaji wa maziwa	1	"12nt:8"	vidonda vya tumbo
2	"12nt:8"	nguvu za mwili	1	"12nt:8"	vidonda vya kinywani
2	"12nt:8"	ngozi ya pumbu	1	"12nt:8"	vidonda vya kaswende
2	"12nt:8"	ngozi ya msamba	1	"12nt:8"	vidole vya miguuni
2	"12nt:8"	mshipa wa damu	1	"12nt:8"	ute wa uzazi
2	"12nt:8"	mkojo wa mgonjwa	1	"12nt:8"	utafiti wa vimelea
2	"12nt:8"	misuli ya tumbo	1	"12nt:8"	utafiti wa matibabu
2	"12nt:8"	mishipa ya vena	1	"12nt:8"	utafiti wa dawa
2	"12nt:8"	mimba za usichana	1	"12nt:8"	usumushaji wa chakula
2	"12nt:8"	midomo ya uke	1	"12nt:8"	usafi wa mwili
2	"12nt:8"	mfumo wa mkojo	1	"12nt:8"	usafi wa meno
2	"12nt:8"	mfumo wa damu	1	"12nt:8"	usafi wa maji
2	"12nt:8"	mayai ya tegu	1	"12nt:8"	umbile la yai
2	"12nt:8"	maumivu ya viungo	1	"12nt:8"	umbile la mifupa
2	"12nt:8"	majeraha ya mishipa	1	"12nt:8"	uhusiano wa maji
2	"12nt:8"	koo za mbu	1	"12nt:8"	ugonjwa wa viluwiluwi
2	"12nt:8"	kiungo cha kombe	1	"12nt:8"	ugonjwa wa utotoni
2	"12nt:8"	kichocho cha tumbo	1	"12nt:8"	ugonjwa wa pepopunda
2	"12nt:8"	kichocho cha mkojo	1	"12nt:8"	ugonjwa wa minyoo
2	"12nt:8"	hewa ya oksijeni	1	"12nt:8"	ugonjwa wa konjaktiva
2	"12nt:8"	dawa za usingizi	1	"12nt:8"	ugonjwa wa ugonjwa
2	"12nt:8"	dawa ya meno	1	"12nt:8"	ugonjwa wa kisukari
2	"12nt:8"	dalili za maradhi	1	"12nt:8"	ugonjwa wa homa
2	"12nt:8"	dalili za kaswende	1	"12nt:8"	
2	"12nt:8"	ambukizo la ugonjwa	1	"12nt:8"	
1	"12nt:8"	wagonjwa wa polio	1	"12nt:8"	
1	"12nt:8"	wagonjwa wa minyoo	1	"12nt:8"	
1	"12nt:8"	vyakula vya protini	1	"12nt:8"	
1	"12nt:8"	vivimbe vya ngozi	1	"12nt:8"	
1	"12nt:8"	viungo vya mifupa	1	"12nt:8"	
1	"12nt:8"	virusi vya ugonjwa	1	"12nt:8"	
1	"12nt:8"	virusi vya polio	1	"12nt:8"	
1	"12nt:8"	virusi vya homa	1	"12nt:8"	

1 "12nt:8"	ugonjwa wa chirwa	1 "12nt:8"	mishipa ya limfu.
1 "12nt:8"	udhaifu wa tishu	1 "12nt:8"	mishipa ya ateri
1 "12nt:8"	udhaifu wa mwili	1 "12nt:8"	mirija ya sindano
1 "12nt:8"	udhaifu wa miguu	1 "12nt:8"	mimba za mapacha
1 "12nt:8"	uchunguzi wa	1 "12nt:8"	mifupa ya kombe
	ugonjwa	1 "12nt:8"	mifereji ya maji
1 "12nt:8"	uchunguzi wa ngozi	1 "12nt:8"	mfupa wa utosi
1 "12nt:8"	uchunguzi wa	1 "12nt:8"	mfupa wa paja
	maabara	1 "12nt:8"	mfumo wa uzazi
1 "12nt:8"	uchunguzi wa damu	1 "12nt:8"	mfumo wa neva
1 "12nt:8"	uchungu wa uzazi	1 "12nt:8"	mfumo wa hewa
1 "12nt:8"	uchafu wa mwili	1 "12nt:8"	mdomo wa uke
1 "12nt:8"	uchafu wa mgonjwa	1 "12nt:8"	mbegu za uzazi
1 "12nt:8"	uchafu wa chakula	1 "12nt:8"	mbegu za mafuta
1 "12nt:8"	uzalishaji wa	1 "12nt:8"	mbegu za maboga
	kondo	1 "12nt:8"	mazoezi ya mwili
1 "12nt:8"	uzalishaji wa	1 "12nt:8"	maziwa ya mgando
	chakula	1 "12nt:8"	mayai ya safura
1 "12nt:8"	utunzaji wa	1 "12nt:8"	maumivu ya sikio
	vyakula	1 "12nt:8"	matibabu ya upele
1 "12nt:8"	utunzaji wa afya	1 "12nt:8"	matibabu ya
1 "12nt:8"	usambazaji wa dawa		maradhi
1 "12nt:8"	upunguzaji wa	1 "12nt:8"	matibabu ya
	vimelea		kisonono
1 "12nt:8"	upunguzaji wa	1 "12nt:8"	maradhi ya ngozi
	papasi	1 "12nt:8"	malengelenge ya
1 "12nt:8"	upunguzaji wa mbu		damu
1 "12nt:8"	upimaji wa vidole	1 "12nt:8"	makohozi ya
1 "12nt:8"	upimaji wa uzito		mgonjwa
1 "12nt:8"	upimaji wa tumbo	1 "12nt:8"	maji ya sukari
1 "12nt:8"	upimaji wa maji	1 "12nt:8"	maji ya kunde
1 "12nt:8"	ukuaji wa mwili	1 "12nt:8"	majeraha ya mifupa
1 "12nt:8"	ukuaji wa misuli	1 "12nt:8"	mafigo ya mgonjwa
1 "12nt:8"	ukuaji wa mbu	1 "12nt:8"	madhara ya pombe
1 "12nt:8"	uenezaji wa	1 "12nt:8"	madhara ya mwili
	kipindupindu	1 "12nt:8"	madhara ya mimba
1 "12nt:8"	ueneaaji wa malaria	1 "12nt:8"	madhara ya
1 "12nt:8"	tiba ya ukimwi		kichocho
1 "12nt:8"	sumu ya vimelea	1 "12nt:8"	mabaka ya ukoma
1 "12nt:8"	sumu ya vijidudu	1 "12nt:8"	maambukizo ya
1 "12nt:8"	sindano ya ganzi		maradhi
1 "12nt:8"	shingo la kizazi	1 "12nt:8"	maambukizo ya
1 "12nt:8"	nundu za mifupa		kifua
1 "12nt:8"	nundu za mfupa	1 "12nt:8"	maambukizo ya
1 "12nt:8"	nundu ya mfupa		kibofu
1 "12nt:8"	nguvu za kinga	1 "12nt:8"	konokono wa maji
1 "12nt:8"	nguvu za homoni	1 "12nt:8"	kokwa za mayai
1 "12nt:8"	ngozi ya tumbo	1 "12nt:8"	kiungo cha mwili
1 "12nt:8"	ngozi ya mwili	1 "12nt:8"	kiungo cha mboga
1 "12nt:8"	ngozi ya hasua	1 "12nt:8"	kiungo cha goti
1 "12nt:8"	mzunguko wa	1 "12nt:8"	kisonono cha
	ugonjwa		viungo
1 "12nt:8"	mzunguko wa maji	1 "12nt:8"	kisonono cha macho
1 "12nt:8"	mzunguko wa hewa	1 "12nt:8"	kisonono cha kooni
1 "12nt:8"	mwili wa papasi	1 "12nt:8"	kisima cha maji
1 "12nt:8"	mwili wa konokono	1 "12nt:8"	kipimo cha sukari
1 "12nt:8"	msukumo wa maji	1 "12nt:8"	kipimo cha joto
1 "12nt:8"	msukumo wa damu	1 "12nt:8"	kipimo cha chumvi
1 "12nt:8"	mshipa wa aorta	1 "12nt:8"	kinga ya mwili
1 "12nt:8"	mpira wa miguu	1 "12nt:8"	kimelea cha
1 "12nt:8"	minyoo ya filari		malaria
1 "12nt:8"	mmomonyoko wa	1 "12nt:8"	kimelea cha malale
	udongo	1 "12nt:8"	kimelea cha homa
1 "12nt:8"	mlo wa damu	1 "12nt:8"	kigutu cha kitovu
1 "12nt:8"	mlipuko wa ugonjwa	1 "12nt:8"	kidole cha mgonjwa

1 "12nt:8"	kichocho cha	1 "12nt:8"	dawa ya mafuta
	utumbo	1 "12nt:8"	dawa ya klorini
1 "12nt:8"	kansa ya mapafu	1 "12nt:8"	dawa ya kinga
1 "12nt:8"	kansa ya kibofu	1 "12nt:8"	dalili za uzazi
1 "12nt:8"	joto la mwili	1 "12nt:8"	dalili za mapacha
1 "12nt:8"	jeraha la ubongo	1 "12nt:8"	dalili za malaria
1 "12nt:8"	jeraha la kibofu	1 "12nt:8"	dalili za malale
1 "12nt:8"	homa za malaria	1 "12nt:8"	dalili za kisonono
1 "12nt:8"	homa ya	1 "12nt:8"	dalili za jeraha
	kipindupindu	1 "12nt:8"	dalili za chirwa
1 "12nt:8"	fundo la mguu	1 "12nt:8"	chanjo ya surua
1 "12nt:8"	dawa za vidonge	1 "12nt:8"	chanjo ya polio
1 "12nt:8"	dawa za tiba	1 "12nt:8"	chanjo ya ndui
1 "12nt:8"	dawa za sumu	1 "12nt:8"	chakula cha
1 "12nt:8"	dawa za sindano		protini
1 "12nt:8"	dawa za chanjo	1 "12nt:8"	chakula cha mafuta
1 "12nt:8"	dawa ya vidonge	1 "12nt:8"	ambukizo la sikio
1 "12nt:8"	dawa ya maji		

HC-N + GEN-CON + N "12nt:9"

23 "12nt:9"	maziwa ya mama	3 "12nt:9"	ugonjwa wa hatari
14 "12nt:9"	kichaa cha mbwa	3 "12nt:9"	tegu wa mbwa
13 "12nt:9"	uzazi wa mpango	3 "12nt:9"	nguvu ya mwendo
13 "12nt:9"	utunzaji wa watoto	3 "12nt:9"	mwili wa mdudu
12 "12nt:9"	usafi wa mazingira	3 "12nt:9"	maziwa ya unga
9 "12nt:9"	kliniki za watoto	3 "12nt:9"	maradhi ya watoto
8 "12nt:9"	vimelea vya	3 "12nt:9"	kinyesi cha
	magonjwa		wanyama
8 "12nt:9"	uzazi wa mpasuo	3 "12nt:9"	kinyesi cha paka
8 "12nt:9"	macho ya mtoto	3 "12nt:9"	kinyesi cha
8 "12nt:9"	homa ya vipindi		binadamu
8 "12nt:9"	afya ya msingi	3 "12nt:9"	hedhi ya mwisho
7 "12nt:9"	mfupa wa pingili	3 "12nt:9"	damu ya mwezi
7 "12nt:9"	maziwa ya ng'ombe	3 "12nt:9"	afya ya binadamu
7 "12nt:9"	ukuaji wa mtoto	2 "12nt:9"	vimelea vya
6 "12nt:9"	moyo wa mtoto		filaria
6 "12nt:9"	mboga za majani	2 "12nt:9"	vimelea vya
6 "12nt:9"	kliniki ya watoto		bakteria
6 "12nt:9"	damu ya mama	2 "12nt:9"	vifo vya
6 "12nt:9"	afya ya mtoto		akina_mama
6 "12nt:9"	ukuaji wa haraka	2 "12nt:9"	uzazi wa taabu
5 "12nt:9"	mwili wa mtoto	2 "12nt:9"	uzazi wa majira
5 "12nt:9"	mishipa ya	2 "12nt:9"	mtunzaji wa watoto
	limfatiki	2 "12nt:9"	waenezaji wa
5 "12nt:9"	maji ya moto		magonjwa
5 "12nt:9"	homa ya ini	2 "12nt:9"	utunzaji wa vyanzo
4 "12nt:9"	ugonjwa wa watu	2 "12nt:9"	utunzaji wa
4 "12nt:9"	ugonjwa wa filaria		akina_mama
4 "12nt:9"	mwili wa binadamu	2 "12nt:9"	ukuaji wa watoto
4 "12nt:9"	hospitali ya rufaa	2 "12nt:9"	uzazi wa
4 "12nt:9"	dawa za kinga		kitangulizi
4 "12nt:9"	choo cha shimo	2 "12nt:9"	upofu wa mtoni
4 "12nt:9"	afya ya mama	2 "12nt:9"	uhusiano wa
4 "12nt:9"	afya ya jamii		kimapenzi
3 "12nt:9"	vijidudu vya	2 "12nt:9"	udhaifu wa kinga
	magonjwa	2 "12nt:9"	tumbo la mtoto
3 "12nt:9"	vifo vya watoto	2 "12nt:9"	tezi za kwapani
3 "12nt:9"	uzito wa mtoto	2 "12nt:9"	ngozi ya kichwa
3 "12nt:9"	uzito wa kilo	2 "12nt:9"	mwili wa mtu
3 "12nt:9"	usafi wa watoto	2 "12nt:9"	mishipa ya fahamu
3 "12nt:9"	utunzaji wa mzazi	2 "12nt:9"	mfupa wa kisogo
3 "12nt:9"	utunzaji wa mtoto	2 "12nt:9"	mdomoni mwa mtoto
3 "12nt:9"	utunzaji wa mama	2 "12nt:9"	matibabu ya shida

2	"12nt:9"	malale ya gambia	1	"12nt:9"	ugonjwa wa muda
2	"12nt:9"	mafuta ya uto	1	"12nt:9"	ugonjwa wa mama
2	"12nt:9"	madhara ya	1	"12nt:9"	ugonjwa wa koo
		magonjwa	1	"12nt:9"	ugonjwa wa
2	"12nt:9"	kinyesi cha mifugo			kifuakikuu
2	"12nt:9"	homa ya manjano	1	"12nt:9"	udongo wa
2	"12nt:9"	dawa ya penisilini			mfinyanzi
2	"12nt:9"	damu ya binadamu	1	"12nt:9"	udhaifu wa watoto
2	"12nt:9"	dalili za awali	1	"12nt:9"	udhaifu wa shingo
2	"12nt:9"	afya za watu	1	"12nt:9"	uchunguzi wa
2	"12nt:9"	afya za vijiji			mafanikio
1	"12nt:9"	vyoo vya shimo	1	"12nt:9"	uchafu wa
1	"12nt:9"	vyakula vya mimea			mazingira
1	"12nt:9"	viungo vya mtoto	1	"12nt:9"	ubongo wa mtoto
1	"12nt:9"	viungo vya mbeleni	1	"12nt:9"	uzuiaji wa
1	"12nt:9"	visima vya asili			magonjwa
1	"12nt:9"	virusi vya papo	1	"12nt:9"	uzalishaji wa
1	"12nt:9"	viroboto vya panya			nyumbani
1	"12nt:9"	viotea vya	1	"12nt:9"	utunzaji wa
		condylomata			takwimu
1	"12nt:9"	vioevu vya moto	1	"12nt:9"	utunzaji wa
1	"12nt:9"	vinyesi vya			kumbukumbu
		wanyama	1	"12nt:9"	usambazaji wa
1	"12nt:9"	vinyesi vya			matokeo
		binadamu	1	"12nt:9"	upimaji wa
1	"12nt:9"	vimelea vya			kitangulizi
		uambukizo	1	"12nt:9"	ukuzaji wa mazao
1	"12nt:9"	vimelea vya	1	"12nt:9"	uambukizaji wa
		uambukizi			mfululizo
1	"12nt:9"	vimelea vya	1	"12nt:9"	tumbo la mama
		toxoplasma	1	"12nt:9"	tiba ya magonjwa
1	"12nt:9"	vimelea vya hepesi	1	"12nt:9"	tezi za limfatiki
1	"12nt:9"	vimelea vya	1	"12nt:9"	tezi ya prostate
		donovania	1	"12nt:9"	tegu wa binadamu
1	"12nt:9"	viluwiluwi vya	1	"12nt:9"	taahira ya akili
		anofelina	1	"12nt:9"	surua ya
1	"12nt:9"	vijidudu vya			kijerumani
		uambukizo	1	"12nt:9"	sindano ya
1	"12nt:9"	vifo vya mama			penisilini
1	"12nt:9"	vifo vya kina_mama	1	"12nt:9"	sindano ya musuli
1	"12nt:9"	vidonge vya	1	"12nt:9"	pumba za mahindi
		sulphametrole	1	"12nt:9"	pua ya mtoto
1	"12nt:9"	vidonge vya	1	"12nt:9"	pombe za kiasilia
		pyraziguanter	1	"12nt:9"	nyonga ya
1	"12nt:9"	vidonge vya			mwanadamu
		metronidazole	1	"12nt:9"	nyonga ya mama
1	"12nt:9"	vidonda vya mtoto	1	"12nt:9"	nyama ya ng'ombe
1	"12nt:9"	utumbo wa mnyama	1	"12nt:9"	nyama ya mnyama
1	"12nt:9"	utosi wa mtoto	1	"12nt:9"	nguvu ya mvutano
1	"12nt:9"	utapiamlo wa kiasi	1	"12nt:9"	ngozi za watu
1	"12nt:9"	usafi wa kina_mama	1	"12nt:9"	ngozi ya mtoto
1	"12nt:9"	usafi wa banda	1	"12nt:9"	ngozi ya binadamu
1	"12nt:9"	umbile la vistari	1	"12nt:9"	neva za usoni
1	"12nt:9"	umbile la vinywele	1	"12nt:9"	mzunguko wa maisha
1	"12nt:9"	umbile la uvungu	1	"12nt:9"	mwili wa watoto
1	"12nt:9"	umbile la nukta	1	"12nt:9"	mwili wa nzi
1	"12nt:9"	umbile la mviringo	1	"12nt:9"	mwili wa mama
1	"12nt:9"	umbile la kichwa	1	"12nt:9"	mpira wa wavu
1	"12nt:9"	ulimi wa mtoto	1	"12nt:9"	mnyoo wa filaria
1	"12nt:9"	ulemavu wa msingi	1	"12nt:9"	mlo wa wazee
1	"12nt:9"	ulemavu wa akili	1	"12nt:9"	mkojo wa mama
1	"12nt:9"	ugonjwa wa vijana	1	"12nt:9"	mkao wa mtoto
1	"12nt:9"	ugonjwa wa typhoid	1	"12nt:9"	minyoo ya
1	"12nt:9"	ugonjwa wa panya			strongyloides
1	"12nt:9"	ugonjwa wa nyongea	1	"12nt:9"	minyoo ya filaria

1 "12nt:9"	mimba za wasichana	1 "12nt:9"	mabega ya mtoto
1 "12nt:9"	milo ya watoto	1 "12nt:9"	mabaka ya rangi
1 "12nt:9"	mikono ya mtoto	1 "12nt:9"	kokwa za vijiyai
1 "12nt:9"	miili ya watu	1 "12nt:9"	kliniki ya mama
1 "12nt:9"	miili ya panya	1 "12nt:9"	kitovu cha mtoto
1 "12nt:9"	miili ya paka	1 "12nt:9"	kitovu cha mama
1 "12nt:9"	miili ya mbwa	1 "12nt:9"	kinyesi cha
1 "12nt:9"	miili ya binadamu	1 "12nt:9"	mwanamke
1 "12nt:9"	mifupa ya uti	1 "12nt:9"	kinyesi cha mtoto
1 "12nt:9"	mifupa ya paji	1 "12nt:9"	kinyesi cha baba
1 "12nt:9"	mifupa ya fuvu	1 "12nt:9"	kilema cha mtoto
1 "12nt:9"	mgonjwa ya	1 "12nt:9"	kifua cha mtoto
1 "12nt:9"	mwanauume	1 "12nt:9"	kidole cha shahada
1 "12nt:9"	mgonjwa wa pangusa	1 "12nt:9"	kemikali za
1 "12nt:9"	mganga wa wilaya	1 "12nt:9"	nyumbani
1 "12nt:9"	mfupa wa uti	1 "12nt:9"	kaswende za awamu
1 "12nt:9"	mfupa wa paji	1 "12nt:9"	kansa ya watoto
1 "12nt:9"	mfupa wa fuvu	1 "12nt:9"	hospitali ya
1 "12nt:9"	mfumo wa	1 "12nt:9"	wilaya
1 "12nt:9"	utekelezaji	1 "12nt:9"	hospitali ya bombo
1 "12nt:9"	mfumo wa sheria	1 "12nt:9"	homa ya uti
1 "12nt:9"	mfumo wa elimu	1 "12nt:9"	homa ya kipindi
1 "12nt:9"	mfadhaiko wa	1 "12nt:9"	homa ya filaria
1 "12nt:9"	kisaikolojia	1 "12nt:9"	dawa za vitamin
1 "12nt:9"	mdomo wa mtoto	1 "12nt:9"	dawa za asili
1 "12nt:9"	mbu wa anofelesi	1 "12nt:9"	dawa ya
1 "12nt:9"	mboga za kijani	1 "12nt:9"	thiacetozone
1 "12nt:9"	maungu ya mtoto	1 "12nt:9"	dawa ya tetanus
1 "12nt:9"	maumivu ya kichwa	1 "12nt:9"	dawa ya kinga
1 "12nt:9"	matunda ya rangi	1 "12nt:9"	damu za wanyama
1 "12nt:9"	matumbo ya	1 "12nt:9"	damu ya watu
1 "12nt:9"	binadamu	1 "12nt:9"	damu ya mtoto
1 "12nt:9"	matibabu ya	1 "12nt:9"	dalili za nyongea
1 "12nt:9"	nyongea	1 "12nt:9"	dalili za mwanzoni
1 "12nt:9"	matibabu ya msaada	1 "12nt:9"	dalili za magonjwa
1 "12nt:9"	mapafu ya watoto	1 "12nt:9"	dalili za awali
1 "12nt:9"	malaria ya msimu	1 "12nt:9"	choo cha mtoto
1 "12nt:9"	malaria ya kichwa	1 "12nt:9"	choo cha mama
1 "12nt:9"	malale ya ng'ombe	1 "12nt:9"	choo cha jumua
1 "12nt:9"	maji ya vuguvugu	1 "12nt:9"	chanjo ya msingi
1 "12nt:9"	maji ya mito	1 "12nt:9"	chakula cha watu
1 "12nt:9"	maji ya baharini	1 "12nt:9"	chakula cha mtoto
1 "12nt:9"	maji ya bahari	1 "12nt:9"	bega la mtoto
1 "12nt:9"	mafuta ya wanyama	1 "12nt:9"	afya za wazazi
1 "12nt:9"	mafuta ya taa	1 "12nt:9"	afya za vijana
1 "12nt:9"	mafuta ya mimea	1 "12nt:9"	afya ya watoto
1 "12nt:9"	mafuta ya mahindi	1 "12nt:9"	afya ya mtu
1 "12nt:9"	madini ya	1 "12nt:9"	afya ya mazingira
1 "12nt:9"	potassium	1 "12nt:9"	afya ya familia
1 "12nt:9"	madhara ya tatizo		

HC-N + GEN-CON + ADV "12nt:10"

2 "12nt:10"	kondo ya nyuma	1 "12nt:10"	mbegu ya kiume
1 "12nt:10"	zahanati ya karibu	1 "12nt:10"	matibabu ya
1 "12nt:10"	viungo vya ndani	1 "12nt:10"	mara moja
1 "12nt:10"	viungo vya kike	1 "12nt:10"	malaria ya
1 "12nt:10"	vimelea vya nje	1 "12nt:10"	mara kwa mara
1 "12nt:10"	usafi wa binafsi	1 "12nt:10"	kliniki za nje
1 "12nt:10"	mimba ya sasa	1 "12nt:10"	jinsia ya kike
1 "12nt:10"	mdomo wa nje.	1 "12nt:10"	jeraha la wazi
1 "12nt:10"	mdomo wa ndani	1 "12nt:10"	hospitali ya
1 "12nt:10"	mbung'o wa kike		karibu
1 "12nt:10"	mbu wa kike		

N + GEN-CON + HC-N "12nt:11"

123	"12nt:11"	mji wa mimba	2	"12nt:11"	uso wa maji
71	"12nt:11"	pango la nyonga	2	"12nt:11"	urefu wa mwili
30	"12nt:11"	uzi wa kitovu	2	"12nt:11"	ulezi wa mapacha
28	"12nt:11"	elimu ya afya	2	"12nt:11"	uhaba wa chakula
25	"12nt:11"	upungufu wa damu	2	"12nt:11"	uanzishaji wa
16	"12nt:11"	mfuko wa maji			kliniki
16	"12nt:11"	huduma za afya	2	"12nt:11"	uambukizo wa
14	"12nt:11"	huduma ya afya			kitovu
13	"12nt:11"	wizara ya afya	2	"12nt:11"	takwimu za maradhi
13	"12nt:11"	sehemu za siri	2	"12nt:11"	shida ya njaa
13	"12nt:11"	magonjwa ya zinaa	2	"12nt:11"	rasi ya mfupa
12	"12nt:11"	sekta ya afya	2	"12nt:11"	programu ya lishe
12	"12nt:11"	njia ya hewa	2	"12nt:11"	pango la mifupa
11	"12nt:11"	magonjwa ya	2	"12nt:11"	ncha za vidole
		uasherati	2	"12nt:11"	mwanya wa mkojo
10	"12nt:11"	uti wa mgongo	2	"12nt:11"	mradi wa lishe
10	"12nt:11"	mafunzo ya afya	2	"12nt:11"	mpango wa utafiti
9	"12nt:11"	vyanzo vya maji	2	"12nt:11"	michezo ya mwili
9	"12nt:11"	njia ya uzazi	2	"12nt:11"	mazingira ya joto
9	"12nt:11"	kituo cha afya	2	"12nt:11"	mazao ya chakula
8	"12nt:11"	njia ya mkojo	2	"12nt:11"	matumizi ya maji
8	"12nt:11"	hali ya usafi	2	"12nt:11"	mahitaji ya afya
7	"12nt:11"	ukosefu wa chakula	2	"12nt:11"	mafunzo ya utafiti
7	"12nt:11"	shirika la afya	2	"12nt:11"	maendeleo ya uzazi
7	"12nt:11"	nyumba ya uzazi	2	"12nt:11"	mabadiliko ya
7	"12nt:11"	matokeo ya utafiti			kimwili
7	"12nt:11"	kitangulizi cha	2	"12nt:11"	kundi la kisonono
		matako	2	"12nt:11"	kituo cha chanjo
7	"12nt:11"	bomba la hewa	2	"12nt:11"	kanuni za usafi
6	"12nt:11"	vituo vya afya	2	"12nt:11"	jointi za mifupa
6	"12nt:11"	mwendo wa damu	2	"12nt:11"	idara ya ustawi
6	"12nt:11"	magonjwa ya minyoo	2	"12nt:11"	hali ya mgonjwa
6	"12nt:11"	hali ya afya	2	"12nt:11"	eneo la tumbo
5	"12nt:11"	watumishi wa afya	2	"12nt:11"	chama cha uzazi
5	"12nt:11"	tao la mifupa	2	"12nt:11"	bomba la sindano
5	"12nt:11"	programu za chanjo	2	"12nt:11"	athari za ugonjwa
5	"12nt:11"	nyumonia ya tauni	1	"12nt:11"	zao la mizizi
5	"12nt:11"	mifuko ya mbegu	1	"12nt:11"	wimbo wa afya
5	"12nt:11"	huduma za kliniki	1	"12nt:11"	waziri wa afya
5	"12nt:11"	hali ya uzazi	1	"12nt:11"	wataalamu wa afya
4	"12nt:11"	upana wa hanamu	1	"12nt:11"	wataalam wa
4	"12nt:11"	ujenzi wa choo			mazoezi
4	"12nt:11"	miinuko ya ngozi	1	"12nt:11"	wataalam wa afya
4	"12nt:11"	kiwango cha vifo	1	"12nt:11"	wastani wa mayai
4	"12nt:11"	kilango cha uke	1	"12nt:11"	wanakamati wa afya
4	"12nt:11"	kifafa cha uzazi	1	"12nt:11"	wafanyakazi wa
4	"12nt:11"	jaribio la uzazi			hospitali
4	"12nt:11"	huduma za mazoezi	1	"12nt:11"	vyombo vya chakula
3	"12nt:11"	utoaji wa dawa	1	"12nt:11"	viwango vya vifo
3	"12nt:11"	umuhimu wa chanjo	1	"12nt:11"	vitambaa vya
3	"12nt:11"	sehemu ya kinena			mikononi
3	"12nt:11"	njia ya chanjo	1	"12nt:11"	vitabu vya afya
3	"12nt:11"	mwanya wa utumbo	1	"12nt:11"	vipindi vya afya
3	"12nt:11"	matatizo ya afya	1	"12nt:11"	vipande vya shashi
3	"12nt:11"	magonjwa ya polio	1	"12nt:11"	vipande vya kondo
3	"12nt:11"	kitengo cha	1	"12nt:11"	vilabu vya pombe
		utafiti	1	"12nt:11"	vikundi vya ustawi
3	"12nt:11"	huduma za uzazi	1	"12nt:11"	vidokezo vya afya
3	"12nt:11"	huduma za chanjo	1	"12nt:11"	uwezo wa vimelea
2	"12nt:11"	wafanyakazi wa	1	"12nt:11"	uvimbe wa tezi
		afya	1	"12nt:11"	uvimbe wa ngozi
2	"12nt:11"	uvimbe wa mishipa	1	"12nt:11"	uvimbe wa miguu
2	"12nt:11"	utumiaji wa vyoo	1	"12nt:11"	uvamizi wa viungo

1 "12nt:11"	uvamizi wa viluwiluwi	1 "12nt:11"	uharibifu wa mishipa
1 "12nt:11"	uvamizi wa majeraha	1 "12nt:11"	uharibifu wa mifupa
1 "12nt:11"	uvamizi wa damu	1 "12nt:11"	uharibifu wa mfupa
1 "12nt:11"	utupaji wa kinyesi	1 "12nt:11"	uhaba wa nishati
1 "12nt:11"	utumiaji wa mafuta	1 "12nt:11"	ugunduzi wa vimelea
1 "12nt:11"	utumiaji wa dawa	1 "12nt:11"	ufugaji wa nyuki
1 "12nt:11"	utoaji wa matibabu	1 "12nt:11"	uendeshaji wa kliniki
1 "12nt:11"	utoaji wa homoni	1 "12nt:11"	uendeshaji wa afya
1 "12nt:11"	utoaji wa enema	1 "12nt:11"	udumishaji wa afya
1 "12nt:11"	utoaji wa afya	1 "12nt:11"	uchujaji wa maji
1 "12nt:11"	utengenezaji wa mbegu	1 "12nt:11"	uchimbaji wa choo
1 "12nt:11"	utayarishaji wa maziwa	1 "12nt:11"	uchambuzi wa ugonjwa
1 "12nt:11"	utayarishaji wa kliniki	1 "12nt:11"	ubora wa maziwa
1 "12nt:11"	utayarishaji wa chakula	1 "12nt:11"	uangalizi wa daktari
1 "12nt:11"	utawala wa afya	1 "12nt:11"	uambukizo wa ukimwi
1 "12nt:11"	utaratibu wa kliniki	1 "12nt:11"	tundu la choo
1 "12nt:11"	utangulizi wa mataako	1 "12nt:11"	tofauti za minyoo
1 "12nt:11"	ushirikiano wa wagonjwa	1 "12nt:11"	tendo la ngono
1 "12nt:11"	usalama wa maji	1 "12nt:11"	tatizo la minyoo
1 "12nt:11"	usafishaji wa sindano	1 "12nt:11"	tatizo la maji
1 "12nt:11"	urefu wa sindano	1 "12nt:11"	tatizo la lishe
1 "12nt:11"	upungufu wa vifo	1 "12nt:11"	tathmini ya afya
1 "12nt:11"	upungufu wa uzazi	1 "12nt:11"	tarehe za chanjo
1 "12nt:11"	upungufu wa oksijeni	1 "12nt:11"	taratibu za afya
1 "12nt:11"	upungufu wa maji	1 "12nt:11"	tando za ubongo
1 "12nt:11"	upungufu wa asidi	1 "12nt:11"	tamko la afya
1 "12nt:11"	upungafu wa damu	1 "12nt:11"	tabia ya uchafu
1 "12nt:11"	upoteaji wa maji	1 "12nt:11"	taasisi za afya
1 "12nt:11"	upoteaji wa damu	1 "12nt:11"	taasisi ya utafiti
1 "12nt:11"	upatikanaji wa chakula	1 "12nt:11"	taarifa za wagonjwa
1 "12nt:11"	upana wa vidole	1 "12nt:11"	taarifa za chanjo
1 "12nt:11"	umuhimu wa sabuni	1 "12nt:11"	stoo ya chakula
1 "12nt:11"	umuhimu wa maziwa	1 "12nt:11"	sinia la sindano
1 "12nt:11"	umri wa mgonjwa	1 "12nt:11"	siku za afya
1 "12nt:11"	umbo la mwili	1 "12nt:11"	shughuli za uzazi
1 "12nt:11"	umaarufu wa maziwa	1 "12nt:11"	shughuli za ukunga
1 "12nt:11"	ulinzi wa mwili	1 "12nt:11"	shughuli za lishe
1 "12nt:11"	ulaji wa chakula	1 "12nt:11"	shingo ya uzazi
1 "12nt:11"	ukubwa wa mwili	1 "12nt:11"	shimo la choo
1 "12nt:11"	ukubwa wa mimba	1 "12nt:11"	sera za utafiti
1 "12nt:11"	ukosefu wa damu	1 "12nt:11"	sera ya usafi
1 "12nt:11"	ukingo wa nywele	1 "12nt:11"	sehemu za viungo
1 "12nt:11"	ukingo wa kiganja	1 "12nt:11"	sehemu za pafu
1 "12nt:11"	ukali wa tauni	1 "12nt:11"	sehemu ya mwili
1 "12nt:11"	ukaguzi wa kondo	1 "12nt:11"	sehemu ya mrija
1 "12nt:11"	ujuzi wa dawa	1 "12nt:11"	sehemu ya mfupa
1 "12nt:11"	ujenzi wa visima	1 "12nt:11"	sehemu ya kondo
1 "12nt:11"	uharibifu wa viungo	1 "12nt:11"	sehemu ya kiumbe
1 "12nt:11"	uharibifu wa tishu	1 "12nt:11"	sehemu ya kifua
1 "12nt:11"	uharibifu wa ngozi	1 "12nt:11"	ripoti ya mganga
1 "12nt:11"	uharibifu wa mshipa	1 "12nt:11"	rekodi za hospitali
		1 "12nt:11"	rasi ya nyonga
		1 "12nt:11"	rangi ya ngozi
		1 "12nt:11"	rangi ya maji
		1 "12nt:11"	rangi ya krimu
		1 "12nt:11"	programu ya chanjo

1 "12nt:11"	povu la sabuni	1 "12nt:11"	matumizi ya
1 "12nt:11"	pingili za vidole		utafiti
1 "12nt:11"	pingili za nyonga	1 "12nt:11"	matumizi ya
1 "12nt:11"	pindo la makombe		kemikali
1 "12nt:11"	pembe za mfupa	1 "12nt:11"	matumizi ya dawa
1 "12nt:11"	nyumba ya kliniki	1 "12nt:11"	matokeo ya ulemavu
1 "12nt:11"	nusu ya dawa	1 "12nt:11"	matayarisho ya
1 "12nt:11"	njia ya maji		chakula
1 "12nt:11"	njia ya chakula	1 "12nt:11"	matatizo ya
1 "12nt:11"	njia ya viluwiluwi		nishati
1 "12nt:11"	njia ya vidonda	1 "12nt:11"	matatizo ya moyo
1 "12nt:11"	njia ya usafi	1 "12nt:11"	matatizo ya mimba
1 "12nt:11"	njia ya uchafu	1 "12nt:11"	matatizo ya meno
1 "12nt:11"	njia ya sindano	1 "12nt:11"	matatizo ya
1 "12nt:11"	njia ya ngono		malaria
1 "12nt:11"	njia ya mshtuko	1 "12nt:11"	matatizo ya
1 "12nt:11"	njia ya mrija		mahusiano
1 "12nt:11"	njia ya mishipa	1 "12nt:11"	matatizo ya lishe
1 "12nt:11"	njia ya mazoezi	1 "12nt:11"	maswala ya afya
1 "12nt:11"	njia ya kinyesi	1 "12nt:11"	mashirika ya dini
1 "12nt:11"	ncha ya sindano	1 "12nt:11"	mashambulizi ya
1 "12nt:11"	ncha ya pafu		vimelea
1 "12nt:11"	mzungako wa damu	1 "12nt:11"	mashambulio ya
1 "12nt:11"	mwanzo wa mimba		vimelea
1 "12nt:11"	musuli wa tako	1 "12nt:11"	mashamba ya ndizi
1 "12nt:11"	musuli wa msamba	1 "12nt:11"	mambo ya afya
1 "12nt:11"	muda wa matibabu	1 "12nt:11"	majani ya karoti
1 "12nt:11"	msimu wa machungwa	1 "12nt:11"	maisha ya wagonjwa
1 "12nt:11"	mshiko wa kidole	1 "12nt:11"	maisha ya viumbe
1 "12nt:11"	mshauri wa afya	1 "12nt:11"	maisha ya mgonjwa
1 "12nt:11"	msaada wa darubini	1 "12nt:11"	maisha ya mbu
1 "12nt:11"	msaada wa daktari	1 "12nt:11"	maisha ya kichocho
1 "12nt:11"	mradi wa maji	1 "12nt:11"	maingiliano ya
1 "12nt:11"	mradi wa afya		kinywani
1 "12nt:11"	mpasuko wa ngozi	1 "12nt:11"	mahitaji ya
1 "12nt:11"	mpango wa uzazi		kliniki
1 "12nt:11"	miti ya matunda	1 "12nt:11"	magonjwa ya ukungu
1 "12nt:11"	mipangilio ya	1 "12nt:11"	magonjwa ya surua
	utafiti	1 "12nt:11"	magonjwa ya macho
1 "12nt:11"	mikunjo ya ngozi	1 "12nt:11"	magonjwa ya
1 "12nt:11"	mikakati ya uzazi		kipindupindu
1 "12nt:11"	mifuko ya plastiki	1 "12nt:11"	maganda ya viazi
1 "12nt:11"	mifano ya maradhi	1 "12nt:11"	maganda ya ndizi
1 "12nt:11"	miezi ya uzazi	1 "12nt:11"	mafanikio ya
1 "12nt:11"	mchanganyiko wa		matibabu
	vyakula	1 "12nt:11"	mafanikio ya kinga
1 "12nt:11"	mchanganyiko wa	1 "12nt:11"	maeneo ya ugonjwa
	viinilishe	1 "12nt:11"	maeneo ya nyonga
1 "12nt:11"	mchanganyiko wa	1 "12nt:11"	maeneo ya mwili
	viini	1 "12nt:11"	maeneo ya malaria
1 "12nt:11"	mchanganyiko wa	1 "12nt:11"	maeneo ya malale
	utomwili	1 "12nt:11"	maendeleo ya mwili
1 "12nt:11"	mchanganyiko wa	1 "12nt:11"	madimbwi ya maji
	maziwa	1 "12nt:11"	mabaki ya vyakula
1 "12nt:11"	mchanganyiko wa	1 "12nt:11"	mabadiliko ya
	damu		ngozi
1 "12nt:11"	mbinu za uzazi	1 "12nt:11"	mabadiliko ya
1 "12nt:11"	mazoea ya chakula		mwili
1 "12nt:11"	mazingira ya	1 "12nt:11"	likizo ya uzazi
	kiafya	1 "12nt:11"	lengo la zoezi
1 "12nt:11"	mazingara ya	1 "12nt:11"	kuta za utumbo
	kliniki	1 "12nt:11"	kuta za uke
1 "12nt:11"	matumizi ya	1 "12nt:11"	kunjo za ngozi
	vyakula	1 "12nt:11"	kundi la minyoo
		1 "12nt:11"	kovu la chanjo

1	"12nt:11"	kondo la uzazi	1	"12nt:11"	hatua za uzazi
1	"12nt:11"	klabu ya pombe	1	"12nt:11"	hatua ya ugonjwa
1	"12nt:11"	kiwango cha uzazi	1	"12nt:11"	hatua ya mshtuko
1	"12nt:11"	kiwango cha utapiamlo	1	"12nt:11"	hatari za dawa
1	"12nt:11"	kiwango cha afya	1	"12nt:11"	hatari ya vijidudu
1	"12nt:11"	kitangulizi cha bega	1	"12nt:11"	hali ya viungo
1	"12nt:11"	kitana cha nywele	1	"12nt:11"	hali ya utapiamlo
1	"12nt:11"	kiriya cha mayai	1	"12nt:11"	hali ya ukavu
1	"12nt:11"	kifuakikuu cha mapafu	1	"12nt:11"	hali ya nguvu
1	"12nt:11"	kifafa cha mimba	1	"12nt:11"	hali ya mango
1	"12nt:11"	kichwa cha pini	1	"12nt:11"	haki ya afya
1	"12nt:11"	karatasi za mgonjwa	1	"12nt:11"	gharama ya maziwa
1	"12nt:11"	karatasi za kliniki	1	"12nt:11"	gesi za sumu
1	"12nt:11"	karatasi ya plastiki	1	"12nt:11"	faragha ya choo
1	"12nt:11"	kadi za kliniki	1	"12nt:11"	faida ya karantini
1	"12nt:11"	kadi ya mgonjwa	1	"12nt:11"	elimu ya viumbe
1	"12nt:11"	jina la mgonjwa	1	"12nt:11"	elimu ya lishe
1	"12nt:11"	jamaa ya viumbe	1	"12nt:11"	chimbuko la vimelea
1	"12nt:11"	jamaa ya kupe	1	"12nt:11"	chimbuko la mayai
1	"12nt:11"	idara ya mazoezi	1	"12nt:11"	chimbuko la kichocho
1	"12nt:11"	idara ya maji	1	"12nt:11"	chembe cha moyo
1	"12nt:11"	idara ya afya	1	"12nt:11"	chanzo cha ugonjwa
1	"12nt:11"	huduma za kinga	1	"12nt:11"	chanzo cha maji
1	"12nt:11"	huduma ya mazoezi	1	"12nt:11"	bakteria wa kipindupindu
1	"12nt:11"	huduma ya matunzo	1	"12nt:11"	athari ya ubongo
1	"12nt:11"	hofu ya mimba	1	"12nt:11"	athari ya joto
1	"12nt:11"	hatua za viluwilwi	1	"12nt:11"	athari ya homa
1	"12nt:11"		1	"12nt:11"	asili ya ugonjwa
1	"12nt:11"		1	"12nt:11"	asili wa mbu

HC-V + GEN-CON + HC-N "12vt:12"

6	"12vt:12"	kuchanika kwa msamba	1	"12vt:12"	kuvimba kwa mwili
4	"12vt:12"	kuvimba kwa tezi	1	"12vt:12"	kuvimba kwa miguu
4	"12vt:12"	kuenea kwa ugonjwa	1	"12vt:12"	kuvimba kwa mguu
4	"12vt:12"	kuenea kwa kipindupindu	1	"12vt:12"	kuumizwa kwa kibofu
3	"12vt:12"	kuvimba kwa viungo	1	"12vt:12"	kuumia kwa mwili
2	"12vt:12"	kuzaliwa kwa kondo	1	"12vt:12"	kuuma kwa mitoki
2	"12vt:12"	kuvimba kwa hasua	1	"12vt:12"	kupungua kwa utapiamlo
2	"12vt:12"	kukua kwa mwili	1	"12vt:12"	kupooza kwa viungo
2	"12vt:12"	kukua kwa mimba	1	"12vt:12"	kupooza kwa misuli
2	"12vt:12"	kuenea kwa ukimwi	1	"12vt:12"	kupooza kwa miguu
1	"12vt:12"	kuziba kwa koo	1	"12vt:12"	kupevuka kwa mimba
1	"12vt:12"	kuziba kwa mrija	1	"12vt:12"	kunawa kwa sabuni
1	"12vt:12"	kuziba kwa mishipa	1	"12vt:12"	kulemaa kwa mdomo
1	"12vt:12"	kuzaliana kwa vimelea	1	"12vt:12"	kulegea kwa misuli
1	"12vt:12"	kuzaliana kwa vijidudu	1	"12vt:12"	kukua kwa yai
1	"12vt:12"	kuzaliana kwa mbu	1	"12vt:12"	kukua kwa misuli
1	"12vt:12"	kuzaana kwa konokono	1	"12vt:12"	kukaza kwa kano
1	"12vt:12"	kuvimba kwa tezi.	1	"12vt:12"	kukandwa kwa maji
1	"12vt:12"		1	"12vt:12"	kuhara kwa ameba
1	"12vt:12"		1	"12vt:12"	kuenea kwa kipuku
1	"12vt:12"		1	"12vt:12"	kuauni kwa mpira

HC-V + GEN-CON + N "12vt:13"

6	"12vt:13"	kuenea kwa magonjwa	5	"12vt:13"	kuzaliwa kwa mtoto
			2	"12vt:13"	kuvimba kwa jointi

2	"l2vt:13"	kuvimba kwa ini	1	"l2vt:13"	kutibika kwa
2	"l2vt:13"	kupumua kwa taabu			haraka
2	"l2vt:13"	kukua kwa mtoto	1	"l2vt:13"	kupungua kwa watu
2	"l2vt:13"	kukua kwa akili	1	"l2vt:13"	kupumua kwa mtoto
1	"l2vt:13"	kuziba kwa virija	1	"l2vt:13"	kupevuka kwa
1	"l2vt:13"	kuvimba kwa wengu			kijiyai
1	"l2vt:13"	kuvimba kwa uso	1	"l2vt:13"	kuoza kwa majani
1	"l2vt:13"	kuumbuka kwa	1	"l2vt:13"	kukojoa kwa taabu
		kichwa	1	"l2vt:13"	kuenea kwa
1	"l2vt:13"	kutunza kwa			matumizi
		uangalifu	1	"l2vt:13"	kuchoka kwa mama
1	"l2vt:13"	kutibu kwa			
		penicillin			

INF + GEN-CON + HC-N "l2vt:14"

12	"l2vt:14"	kuharibika kwa	1	"l2vt:14"	kusawazika kwa
		mimba			tumbo
3	"l2vt:14"	kukawia kwa uzazi	1	"l2vt:14"	kurudia kwa
3	"l2vt:14"	kubaki kwa kondo			ugonjwa
2	"l2vt:14"	kutoboka kwa	1	"l2vt:14"	kurudi kwa mganga
		kibofu	1	"l2vt:14"	kupotea kwa maji
2	"l2vt:14"	kutangulia kwa	1	"l2vt:14"	kupoa kwa mwili
		kondo	1	"l2vt:14"	kupapasika kwa
2	"l2vt:14"	kuonekana kwa			utosi
		macho	1	"l2vt:14"	kupanda kwa
2	"l2vt:14"	kulisha kwa chupa			shinikizo
2	"l2vt:14"	kukubalika kwa	1	"l2vt:14"	kupanda kwa joto
		vimelea	1	"l2vt:14"	kuongezeka kwa
1	"l2vt:14"	kuzuka kwa vimelea			nguvu
1	"l2vt:14"	kuzuka kwa ugonjwa	1	"l2vt:14"	kuongezeka kwa
1	"l2vt:14"	kuzuka kwa			mbung'o
		pepopunda	1	"l2vt:14"	kuongezeka kwa
1	"l2vt:14"	kuzuka kwa kansa			joto
1	"l2vt:14"	kuvunjika kwa	1	"l2vt:14"	kula kwa mikono
		mgongo	1	"l2vt:14"	kujaa kwa majimaji
1	"l2vt:14"	kutoweka kwa	1	"l2vt:14"	kujaa kwa hewa
		nywele	1	"l2vt:14"	kuhema kwa mdomo
1	"l2vt:14"	kutoboka kwa	1	"l2vt:14"	kubanwa kwa
		utumbo			mishipa
1	"l2vt:14"	kutoboka kwa jino	1	"l2vt:14"	kubanwa kwa mifupa
1	"l2vt:14"	kusiriba kwa	1	"l2vt:14"	kubanduka kwa
		udongo			kondo
1	"l2vt:14"	kusigana kwa	1	"l2vt:14"	kuanza kwa hedhi
		utumbo	1	"l2vt:14"	kuambata kwa kondo

HC-N + PREP + HC-N "l2nt:15"

2	"l2nt:15"	sugu kwa dawa	1	"l2nt:15"	mkojo katika
2	"l2nt:15"	mayai hadi mbu			kibofu
2	"l2nt:15"	maji katika	1	"l2nt:15"	minyoo kwenye
		mishipa			darubini
2	"l2nt:15"	jeraha katika	1	"l2nt:15"	maumivu kwenye
		ubongo			tezi
2	"l2nt:15"	damu chini_ya	1	"l2nt:15"	maumivu kwenye
		ngozi			nyonga
1	"l2nt:15"	vidonda ndani_ya	1	"l2nt:15"	maumivu katika
		uke			nyonga
1	"l2nt:15"	vidonda katika	1	"l2nt:15"	matibabu kwa
		kiungo			mgonjwa
1	"l2nt:15"	ute katika mkojo	1	"l2nt:15"	malengelenge
1	"l2nt:15"	ugonjwa katika			katika kondo
		damu	1	"l2nt:15"	malaria katika
1	"l2nt:15"	tauni katika mwili			viungo

1 "12nt:15"	malale	katika	1 "12nt:15"	dawa kwa ajili ya
	mwili			malengelenge
1 "12nt:15"	madhara	katika	1 "12nt:15"	chanjo dhidi ya
	moyo			ukimwi
1 "12nt:15"	kupe mwenye ngozi		1 "12nt:15"	chanjo dhidi ya
1 "12nt:15"	kupe mwenye gamba			pepopunda
1 "12nt:15"	jeraha	katika	1 "12nt:15"	chakula kwa afya
	ngozi		1 "12nt:15"	antibodi dhidi ya
				antijeni

HC-V + PREP + HC-N "12vt:18"

12 "12vt:18"	kupevuka kwa mimba	1 "12vt:18"	kuoza kwa jino
3 "12vt:18"	kuishi katika afya	1 "12vt:18"	kuota kwa meno
2 "12vt:18"	kuzuia kwa chanjo	1 "12vt:18"	kunawa kwa sabuni
2 "12vt:18"	kuenea kwa	1 "12vt:18"	kulegea kwa mikazo
	kipindupindu	1 "12vt:18"	kukua ndani ya
1 "12vt:18"	kuzuilika kwa		mwili
	chanjo	1 "12vt:18"	kukua kwa mimba
1 "12vt:18"	kuzuia kutoka kwa	1 "12vt:18"	kukomaa kwa kifua
	damu	1 "12vt:18"	kukojoa katika
1 "12vt:18"	kuzaliana		maji
	ndani ya vyoo	1 "12vt:18"	kukojoa baada ya
1 "12vt:18"	kuzaliana		uzazi
	ndani ya jeraha	1 "12vt:18"	kukakamaa kwa
1 "12vt:18"	kuzaliana kwa		mishipa
	mainzi	1 "12vt:18"	kuhara
1 "12vt:18"	kuzaana kwa mbu		kwa sababu ya
1 "12vt:18"	kuvimba kwa chanjo		uchafu
1 "12vt:18"	kutungwa kwa yai	1 "12vt:18"	kugunduliwa kwa
1 "12vt:18"	kutunga kwa mimba		dawa
1 "12vt:18"	kutibiwa katika	1 "12vt:18"	kugundua kwa
	kliniki		utafiti
1 "12vt:18"	kusambaa kwa	1 "12vt:18"	kueneza kwa
	vijidudu		uasherati
1 "12vt:18"	kusambaa kwa	1 "12vt:18"	kuenea kwa sumu
	kaswende	1 "12vt:18"	kuenea katika
1 "12vt:18"	kupungua kwa uzito		viungo
1 "12vt:18"	kupooza kwa mikono	1 "12vt:18"	kuenea katika damu
1 "12vt:18"	kupona kwa mitoki	1 "12vt:18"	kuchoka kwa moyo
1 "12vt:18"	kupevuka kwa yai		

INF + PREP + HC-N "12vt:20"

3 "12vt:20"	kukawia kwa uzazi	1 "12vt:20"	kuonekana kwa
3 "12vt:20"	kuharibika kwa		macho
	mimba	1 "12vt:20"	kuingiza ndani ya
2 "12vt:20"	kutoka katika		uke
	hospitali	1 "12vt:20"	kuingia katika
2 "12vt:20"	kuonekana kwa		damu
	darubini	1 "12vt:20"	kuhimili kwa
1 "12vt:20"	kuzuka kwa ugonjwa		shashi
1 "12vt:20"	kushambuliwa kwa	1 "12vt:20"	kufa
	mapafu		kwa sababu ya
1 "12vt:20"	kurekebishwa kwa		ugonjwa
	operesheni	1 "12vt:20"	kufa kwa mimba
1 "12vt:20"	kupitia katika	1 "12vt:20"	kueleza juu ya
	unyeo		dawa
1 "12vt:20"	kupitia katika	1 "12vt:20"	kuchovya katika
	kondo		maji
1 "12vt:20"	kupita katika	1 "12vt:20"	kubonyea kwa pua
	nyonga	1 "12vt:20"	kubanwa kwa
1 "12vt:20"	kuongezeka kwa		mishipa
	wagonjwa	1 "12vt:20"	kubanwa kwa kibofu

1 "12vt:20"	kubaki kwa kondo	1 "12vt:20"	kuanza kwa mikazo
HC-N + GEN-CON + INF "12nt:22"			
6 "12nt:22"	mfupa wa kukalia	1 "12nt:22"	mazoezi ya kutosha
2 "12nt:22"	ugonjwa wa	1 "12nt:22"	mafuta ya kupakaa
	kuharisha	1 "12nt:22"	madhara ya kudumu
2 "12nt:22"	meno ya kudumu	1 "12nt:22"	homa za
2 "12nt:22"	hewa ya kutosha		kurudiarudia
2 "12nt:22"	hamu ya kula	1 "12nt:22"	dawa za kupakaa
1 "12nt:22"	uambukizo wa	1 "12nt:22"	dawa za kuchoma
	kudumu	1 "12nt:22"	dawa ya kupulizia
1 "12nt:22"	mbegu za kuotesha	1 "12nt:22"	dawa ya kupakaa
HC-N + HC-V "12nt:24"			
1 "12nt:24"	ugonjwa kukomaa	1 "12nt:24"	bandama kunenepa
1 "12nt:24"	homa kupungua		
HC-N + INF "12nt:25"			
2 "12nt:25"	mimba kutoka	1 "12nt:25"	macho kugeuka
1 "12nt:25"	tumbo kujaa	1 "12nt:25"	kondo kutoka
1 "12nt:25"	mimba kuharibika		
N + HC-V "12nt:26"			
2 "12nt:26"	mtu kuumwa	2 "12nt:26"	mama kuzaa
2 "12nt:26"	mtoto kuzaliwa	2 "12nt:26"	mama kupasuliwa
HC-N + POSS + HC-N "12nt:27"			
2 "12nt:27"	mwili wenye afya	1 "12nt:27"	mikazo yenye nguvu
2 "12nt:27"	mgonjwa mwenye	1 "12nt:27"	mguu wenye kilema
	ukimwi	1 "12nt:27"	mgonjwa mwenye
2 "12nt:27"	damu yenye vimelea		kidonda
2 "12nt:27"	chakula chenye	1 "12nt:27"	meno yenye afya
	protini	1 "12nt:27"	maziwa yenye maji
1 "12nt:27"	wagonjwa wenye	1 "12nt:27"	matunda yenye
	virusi		uchachu
1 "12nt:27"	wagonjwa wenye	1 "12nt:27"	maji yenye vimelea
	tauni	1 "12nt:27"	maji yenye
1 "12nt:27"	vyakula vyenye		uvuguvugu
	sukari	1 "12nt:27"	maji yenye
1 "12nt:27"	vyakula vyenye		konokono
	chumvi	1 "12nt:27"	kondo yenye
1 "12nt:27"	sindano zenye dawa		ugonjwa
1 "12nt:27"	pafu zenye ugonjwa	1 "12nt:27"	kipele chenye
1 "12nt:27"	mishipa wenye		maumivu
	maumivu	1 "12nt:27"	dutu zenye sumu
1 "12nt:27"	minyoo yenye	1 "12nt:27"	dawa zenye madini
	madhara	1 "12nt:27"	dawa yenye madini
HC-N + POSS + N "12nt:28"			
2 "12nt:28"	uzazi wenye	1 "12nt:28"	maji yenye
	matatizo		vuguvugu
2 "12nt:28"	pamba yenye spirit	1 "12nt:28"	dutu zenye hatari
1 "12nt:28"	pamba yenye savlon	1 "12nt:28"	dawa yenye
1 "12nt:28"	nyama yenye ladha		potassium
1 "12nt:28"	mkojo wenye rangi	1 "12nt:28"	chupa zenye
1 "12nt:28"	mkojo wenye		vifuniko
	acetone	1 "12nt:28"	choo chenye bomba
1 "12nt:28"	mgonjwa mwenye		
	nyumonia		

N + POSS + HC-N "12nt:29"							
4	"12nt:29"	maeneo	yenye	1	"12nt:29"	mzazi	mwenye
		malaria				mapacha	
3	"12nt:29"	wanawake	wenye	1	"12nt:29"	mtu mwenye	vimelea
		mimba		1	"12nt:29"	mtu	mwenye
3	"12nt:29"	mnyama	mwenye			vijidudu	
		kichaa		1	"12nt:29"	mtu mwenye	upele
2	"12nt:29"	watu wenye	kinga	1	"12nt:29"	mtoto	mwenye
2	"12nt:29"	nyumba	yenye afya			ugonjwa	
2	"12nt:29"	mtoto	mwenye afya	1	"12nt:29"	mtoto	mwenye
1	"12nt:29"	watu	wenye virusi			kilema	
1	"12nt:29"	watu	wenye	1	"12nt:29"	mama wenye	mimba
		majeraha		1	"12nt:29"	mama mwenye	kifua
1	"12nt:29"	watu wenye	kifua	1	"12nt:29"	maeneo	yenye
1	"12nt:29"	watoto	wenye			unyevenyevu	
		utapiamlo		1	"12nt:29"	maeneo	yenye
1	"12nt:29"	watoto	wenye			ugonjwa	
		ukimwi		1	"12nt:29"	madawa	yenye sumu
1	"12nt:29"	wanyama	wenye	1	"12nt:29"	eneo	lenye ugonjwa
		minyoo		1	"12nt:29"	eneo	lenye malaria
1	"12nt:29"	uvimbe	wenye	1	"12nt:29"	bwawa	lenye
		maumivu		1	"12nt:29"	kichocho	
1	"12nt:29"	taifa	lenye afya	1	"12nt:29"	akina_mama	wenye
1	"12nt:29"	nchi	zenye baridi			mimba	
HC-V + ADV "12nt:30"							
5	"12nt:30"	kutibiwa	mapema	1	"12nt:30"	kusedeka	kwa_mfano
5	"12nt:30"	kukua	vizuri	1	"12nt:30"	kupumua	vizuri
3	"12nt:30"	kunyonya	sawasawa	1	"12nt:30"	kupofuka	taratibu
2	"12nt:30"	kupasuka	mapema	1	"12nt:30"	kuozwa	mapema
2	"12nt:30"	kunyonyesha		1	"12nt:30"	kuoga	vizuri
		mara_kwa_mara		1	"12nt:30"	kuoga	
2	"12nt:30"	kukojoa				mara_kwa_mara	
		mara_kwa_mara		1	"12nt:30"	kuishi	vizuri
1	"12nt:30"	kuzaa	salama	1	"12nt:30"	kugundulika	mapema
1	"12nt:30"	kuumbuka	ghafila	1	"12nt:30"	kuchunguza	vizuri
1	"12nt:30"	kuugua	ghafila	1	"12nt:30"	kuchunguza	
1	"12nt:30"	kutunzwa	vizuri			kwa_makini	
1	"12nt:30"	kutunza	vizuri	1	"12nt:30"	kuchanjwa	tena
1	"12nt:30"	kusukuma	mapema				
HC-N + A-INFL + GEN-CON + N "13nt:1"							
3	"13nt:1"	mganga	mkuu wa	1	"13nt:1"	maziwa	mazito ya
		wilaya				njano	
1	"13nt:1"	vimelea	vingi vya	1	"13nt:1"	dalili	kubwa za
		magonjwa				magonjwa	
1	"13nt:1"	vijidudu	vingi	1	"13nt:1"	dalili	kubwa ya
		vya magonjwa				shankaroidi	
1	"13nt:1"	viini	vingi vya	1	"13nt:1"	chakula	kizuri
		magonjwa				cha sungura	
1	"13nt:1"	utosi	mdogo wa	1	"13nt:1"	afya	nzuri ya
		nyuma				mtoto	
1	"13nt:1"	mbu	wazima wa				
		anofelina					
HC-N + A-INFL + GEN-CON + HC-N "13nt:2"							
3	"13nt:2"	chembe	nyekundu	1	"13nt:2"	ute	mweupe wa
		ya damu				majimaji	
2	"13nt:2"	mishipa	midogo ya	1	"13nt:2"	ugonjwa	mpya wa
		damu				uasherati	

1 "13nt:2"	ugonjwa mkuu wa virusi	1 "13nt:2"	dalili kubwa za polio
1 "13nt:2"	ngozi nyororo ya kitambi	1 "13nt:2"	dalili kubwa za kipindupindu
1 "13nt:2"	mfumo mkuu wa neva	1 "13nt:2"	dalili kubwa za homa
1 "13nt:2"	damu nyingi ya hedhi	1 "13nt:2"	dalili kubwa ya ugonjwa
1 "13nt:2"	dalili kuu za ukimwi	1 "13nt:2"	dalili kubwa ya kansa
1 "13nt:2"	dalili kubwa za ugonjwa	1 "13nt:2"	chembe nyekundu za damu

N + A-INFL + GEN-CON + HC-N "13nt:3"

2 "13nt:3"	vyanzo vidogovidogo vya maji
1 "13nt:3"	vyanzo vizuri vya utomwili
1 "13nt:3"	vyanzo vizuri vya madini
1 "13nt:3"	uwiano mzuri wa madini
1 "13nt:3"	tatizo kubwa la kiafya
1 "13nt:3"	sehemu nyeti za mwili
1 "13nt:3"	njia mojawapo ya matibabu
1 "13nt:3"	mazingira mazuri ya afya
1 "13nt:3"	matumizi mazuri ya choo
1 "13nt:3"	kiwango kikubwa cha viinilishe
1 "13nt:3"	kiasi kikubwa cha utomwili
1 "13nt:3"	kiasi kikubwa cha nishati
1 "13nt:3"	kiasi kidogo cha maji
1 "13nt:3"	huduma nzuri za afya
1 "13nt:3"	hatari kubwa ya maambukizo
1 "13nt:3"	gharama kubwa za matibabu
1 "13nt:3"	adui mkubwa wa meno

HC-N + GEN-CON + HC-N + GEN-CON + ADV "13nt:7"

1 "13nt:7"	viungo vya uzazi vya nje
1 "13nt:7"	viungo vya uzazi vya ndani
1 "13nt:7"	kombe la kiuno la kike

HC-N + GEN-CON + N + GEN-CON + ADV "13nt:8"

1 "13nt:8"	usafi wa hali ya juu
1 "13nt:8"	nguvu za asili za nje
1 "13nt:8"	kliniki za watoto za nje

N + GEN-CON + HC-N + GEN-CON + ADV "13nt:9"

3 "13nt:9"	pango la nyonga la kike
1 "13nt:9"	vituo vya afya vya awali
1 "13nt:9"	utengenezaji wa mbegu za kiume
1 "13nt:9"	ukaguzi wa kondo ya nyuma
1 "13nt:9"	pango la nyonga la kiume
1 "13nt:9"	miinuko ya ngozi ya nje
1 "13nt:9"	kituo cha afya cha karibu

N + GEN-CON + HC-N + GEN-CON + N "13nt:10"

6 "13nt:10"	huduma ya afya ya msingi
2 "13nt:10"	upungufu wa damu ya mama
2 "13nt:10"	rasi ya mfupa wa pingili
2 "13nt:10"	huduma za kliniki ya watoto
1 "13nt:10"	watumishi wa afya ya jamii
1 "13nt:10"	watumishi wa afya wa jumuia
1 "13nt:10"	vikwazo vya kliniki ya watoto
1 "13nt:10"	vikundi vya ustawi wa jamii

1 "l3nt:10" vidokezo vya afya ya mama
 1 "l3nt:10" uzi wa kitovu cha mtoto
 1 "l3nt:10" uvimbe wa ngozi ya kichwa
 1 "l3nt:10" utoaji wa matibabu ya shida
 1 "l3nt:10" utayarishaji wa maziwa ya kopo
 1 "l3nt:10" upungufu wa vifo vya watoto
 1 "l3nt:10" upana wa vidole vya mkono
 1 "l3nt:10" umuhimu wa maziwa ya mama
 1 "l3nt:10" umaarufu wa maziwa ya mama
 1 "l3nt:10" ukingo wa nywele za mtoto
 1 "l3nt:10" ukingo wa kiganja cha mkono
 1 "l3nt:10" uharibifu wa mishipa ya limfatiki
 1 "l3nt:10" ubora wa maziwa ya mama
 1 "l3nt:10" uambukizo wa kitovu cha mtoto
 1 "l3nt:10" takwimu za maradhi ya msimu
 1 "l3nt:10" siku za afya za kijiji
 1 "l3nt:10" sehemu ya kifua cha nzi
 1 "l3nt:10" rangi ya maji ya mchele
 1 "l3nt:10" pango la nyonga ya mama
 1 "l3nt:10" pango la nyonga la mama
 1 "l3nt:10" mshauri wa afya wa kanda
 1 "l3nt:10" matokeo ya ulemavu wa msingi
 1 "l3nt:10" matokeo ya nguvu za asili
 1 "l3nt:10" mafunzo ya afya ya jumla
 1 "l3nt:10" kiwango cha vifo vya wanawake
 1 "l3nt:10" karatasi za kliniki za watoto
 1 "l3nt:10" huduma za uzazi wa mpango
 1 "l3nt:10" huduma za kliniki za watoto
 1 "l3nt:10" huduma za afya ya watoto
 1 "l3nt:10" huduma za afya ya msingi
 1 "l3nt:10" huduma ya afya ya mama
 1 "l3nt:10" chimbuko la vimelea vya magonjwa
 1 "l3nt:10" athari ya homa ya ini

N + GEN-CON + N + GEN-CON + HC-N "l3nt:11"

27 "l3nt:11" shingo ya mji wa mimba
 5 "l3nt:11" mlango wa pango la nyonga
 4 "l3nt:11" uambukizo wa mji wa mimba
 3 "l3nt:11" wafanyakazi wa sekta ya afya
 2 "l3nt:11" wafanyakazi wa idara ya ustawi
 2 "l3nt:11" uvungu wa pango la nyonga
 2 "l3nt:11" utokeo wa pango la nyonga
 2 "l3nt:11" somo la elimu ya afya
 1 "l3nt:11" watumishi wa sekta ya afya
 1 "l3nt:11" watu wa idara ya maji
 1 "l3nt:11" watekelezaji wa sera ya usafi
 1 "l3nt:11" walimu wa huduma za afya
 1 "l3nt:11" wakufunzi wa mpango wa utafiti
 1 "l3nt:11" wadudu wa jamii ya kupe
 1 "l3nt:11" vipengele vya huduma ya afya
 1 "l3nt:11" viongozi wa idara ya afya
 1 "l3nt:11" vilengelenge vya sehemu za siri
 1 "l3nt:11" uzuwio wa upungufu wa damu
 1 "l3nt:11" uvimbe wa mifuko ya mbegu
 1 "l3nt:11" utoaji wa huduma za afya
 1 "l3nt:11" usimamiaji wa shughuli za lishe
 1 "l3nt:11" upeo wa mradi wa maji
 1 "l3nt:11" ukuta wa nyumba ya uzazi
 1 "l3nt:11" ukubwa wa tatizo la minyoo
 1 "l3nt:11" ukubwa wa pango la nyonga
 1 "l3nt:11" uboreshaji wa hali ya usafi
 1 "l3nt:11" uamuzi wa shirika la afya
 1 "l3nt:11" tundu la bomba la hewa
 1 "l3nt:11" tando za uti wa mgongo

1 "13nt:11" shughuli za mafunzo ya utafiti
 1 "13nt:11" robo ya urefu wa mwili
 1 "13nt:11" programu ya huduma ya afya
 1 "13nt:11" njia ya mafunzo ya afya
 1 "13nt:11" mstari wa kati wa pua
 1 "13nt:11" msingi wa pango la nyonga
 1 "13nt:11" mshikamano wa shirika la afya
 1 "13nt:11" msaada wa shirika la afya
 1 "13nt:11" mratibu wa mradi wa lishe
 1 "13nt:11" mratibu wa huduma za afya
 1 "13nt:11" mlango wa nyumba ya uzazi
 1 "13nt:11" mkutano wa shirika la afya
 1 "13nt:11" matumizi ya matokeo ya utafiti
 1 "13nt:11" matokeo ya mashambulio ya vimelea
 1 "13nt:11" majina ya magonjwa ya zinaa
 1 "13nt:11" majarida ya huduma ya afya
 1 "13nt:11" magonjwa ya njia ya hewa
 1 "13nt:11" magonjwa ya njia ya chakula
 1 "13nt:11" magonjwa ya kundi la kisonono
 1 "13nt:11" mafunzo ya wataalam wa mazoezi
 1 "13nt:11" mafunzo ya elimu ya lishe
 1 "13nt:11" mafanikio ya utaalim ya utafiti
 1 "13nt:11" mafanikio ya elimu ya afya
 1 "13nt:11" maeneo ya pango la nyonga
 1 "13nt:11" mabadiliko ya taratibu za afya
 1 "13nt:11" kozi ya mafunzo ya utafiti
 1 "13nt:11" kipindi cha redio cha afya
 1 "13nt:11" kampeni za elimu ya afya
 1 "13nt:11" ishara za upungufu wa damu
 1 "13nt:11" huduma za awali za tiba
 1 "13nt:11" hofu ya magonjwa ya zinaa
 1 "13nt:11" historia ya huduma za mazoezi
 1 "13nt:11" hatua ya mwisho ya ugonjwa
 1 "13nt:11" hatua ya mwanzo ya ugonjwa
 1 "13nt:11" hatua ya mwanzo ya homa
 1 "13nt:11" gharama ya ujenzi wa choo
 1 "13nt:11" athari za upungufu wa damu
 1 "13nt:11" athari za magonjwa ya uasherati

N + GEN-CON + HC-N + GEN-CON + HC-N "13nt:12"

4 "13nt:12" tao la mifupa ya kinena
 3 "13nt:12" huduma za mazoezi ya viungo
 2 "13nt:12" mifuko ya mbegu za uzazi
 1 "13nt:12" viwango vya vifo vya wajawazito
 1 "13nt:12" utoaji wa dawa za matibabu
 1 "13nt:12" utoaji wa dawa za malaria
 1 "13nt:12" uharibifu wa mfupa wa pua
 1 "13nt:12" ugunduzi wa vimelea vya ukimwi
 1 "13nt:12" tofauti za minyoo ya kichocho
 1 "13nt:12" tabia ya uchafu wa mwili
 1 "13nt:12" taasisi ya utafiti wa tiba
 1 "13nt:12" sehemu za viungo vya uzazi
 1 "13nt:12" pango la mifupa ya nyonga
 1 "13nt:12" njia ya uchafu wa mikono
 1 "13nt:12" njia ya mrija wa mpira
 1 "13nt:12" njia ya mishipa ya vena
 1 "13nt:12" mkusanyiko wa majimaji ya limfu
 1 "13nt:12" maisha ya wagonjwa wa ukimwi
 1 "13nt:12" idara ya mazoezi ya viungo
 1 "13nt:12" huduma za kinga ya maradhi
 1 "13nt:12" huduma ya mazoezi ya viungo
 1 "13nt:12" athari za ugonjwa wa ukimwi

HC-N + GEN-CON + HC-N + GEN-CON + N "l3nt:13"

3 "l3nt:13" mapigo ya moyo wa mtoto
2 "l3nt:13" vimelea vya homa ya vipindi
2 "l3nt:13" tauni ya tezi za limfatiki
2 "l3nt:13" majeraha ya mishipa ya fahamu
2 "l3nt:13" unyonyeshaji wa maziwa ya mama
1 "l3nt:13" vyakula vya protini ya wanyama
1 "l3nt:13" vidonge vya madini ya chuma
1 "l3nt:13" utafiti wa matibabu ya binaadamu
1 "l3nt:13" umbile la yai la kuku
1 "l3nt:13" upimaji wa uzito wa watoto
1 "l3nt:13" mwili wa viroboto vya panya
1 "l3nt:13" mishipa ya damu ya ini
1 "l3nt:13" minyoo ya safura ya mbwa
1 "l3nt:13" mifupa ya nyonga ya mama
1 "l3nt:13" mfumo wa damu wa binadamu
1 "l3nt:13" kimelea cha homa ya vipindi
1 "l3nt:13" dalili za ugonjwa wa toxacara
1 "l3nt:13" dalili za ugonjwa wa filaria
1 "l3nt:13" dalili za maradhi ya watoto

HC-N + GEN-CON + N + GEN-CON + HC-N "l3nt:14"

8 "l3nt:14" mikazo ya mji wa mimba
3 "l3nt:14" nguvu ya mwendo wa damu
3 "l3nt:14" mfupa wa pingili za nyonga
2 "l3nt:14" uzazi wa kitangulizi cha matako
2 "l3nt:14" ngozi za mfuko wa maji
1 "l3nt:14" vimelea vya kundi la virusi
1 "l3nt:14" vimelea vya kundi la protozoa
1 "l3nt:14" vena ya uzi wa kitovu
1 "l3nt:14" ulegevu wa mji wa mimba
1 "l3nt:14" tiba ya magonjwa ya uasherati
1 "l3nt:14" mzunguko wa maisha ya minyoo
1 "l3nt:14" mishipa ya mji wa mimba
1 "l3nt:14" mfupa wa uti wa mgongo
1 "l3nt:14" maumivu ya njia ya mkojo
1 "l3nt:14" matunda ya rangi ya machungwa
1 "l3nt:14" matibabu ya upungufu wa damu
1 "l3nt:14" matibabu ya njia ya joto
1 "l3nt:14" madhara ya upungufu wa damu
1 "l3nt:14" madhara ya ukosefu wa chakula
1 "l3nt:14" madhara ya magonjwa ya minyoo
1 "l3nt:14" kliniki za chama cha uzazi
1 "l3nt:14" kiwiliwili cha mji wa mimba
1 "l3nt:14" kipimo cha wingi wa maziwa
1 "l3nt:14" kinga ya upungufu wa damu
1 "l3nt:14" homa ya uti wa mgongo
1 "l3nt:14" dalili za ukosefu wa chakula
1 "l3nt:14" dalili za mkereketo wa koo
1 "l3nt:14" dalili za magonjwa ya uasherati
1 "l3nt:14" dalili za awali za ugonjwa
1 "l3nt:14" dalili za awali za kisonono
1 "l3nt:14" dalili ya uvamizi wa ameba

HC-N + GEN-CON + HC-N + GEN-CON + HC-N "l3nt:15"

2 "l3nt:15" ugonjwa wa upofu wa macho
2 "l3nt:15" kiungo cha kombe la kiuno
2 "l3nt:15" kichocho cha mfumo wa mkojo
1 "l3nt:15" viroboto vya ugonjwa wa upofu
1 "l3nt:15" vipimo vya uchunguzi wa damu
1 "l3nt:15" vijidudu vya ugonjwa wa pepopunda
1 "l3nt:15" vijidudu vya ugonjwa wa malaria

1 "l3nt:15" vijidudu vya ugonjwa wa kaswende
 1 "l3nt:15" upunguzaji wa vimelea vya malaria
 1 "l3nt:15" utafiti wa dawa za tiba
 1 "l3nt:15" umbile la mifupa ya nyonga
 1 "l3nt:15" nguvu za kinga ya mwili
 1 "l3nt:15" msukumo wa maji ya mimba
 1 "l3nt:15" mifupa ya kombe la kiuno
 1 "l3nt:15" mayai ya minyoo ya kichocho
 1 "l3nt:15" kansa ya kibofu cha mkojo
 1 "l3nt:15" dalili za ugonjwa wa malaria
 1 "l3nt:15" dalili za ugonjwa wa homa
 1 "l3nt:15" dalili za ugonjwa wa chirwa
 1 "l3nt:15" dalili za kisonono cha koo

HC-N + GEN-CON + N + GEN-CON + N "l3nt:16"

4 "l3nt:16" vimelea vya kundi la bakteria
 3 "l3nt:16" ugonjwa wa watu wa umri
 1 "l3nt:16" upimaji wa upana wa utokeo
 1 "l3nt:16" ukuaji wa maendeleo ya mtoto
 1 "l3nt:16" vyoo vya shimo vya umma
 1 "l3nt:16" vimelea vya kundi la yeast
 1 "l3nt:16" ngozi ya kichwa cha mtoto
 1 "l3nt:16" mifupa ya paji la uso
 1 "l3nt:16" afya za mamilioni ya watu

HC-N + GEN-CON + HC-N + PREP + HC-N "l3nt:17"

1 "l3nt:17" vimelea vya tauni katika mwili
 1 "l3nt:17" vimelea vya malaria kwenye darubini
 1 "l3nt:17" vimelea vya malaria katika viungo
 1 "l3nt:17" vijidudu vya ugonjwa katika damu
 1 "l3nt:17" sumu ya vimelea kwenye ubongo
 1 "l3nt:17" mazoezi ya viungo kwenye hospitali
 1 "l3nt:17" mayai ya minyoo kwenye darubini
 1 "l3nt:17" kinga ya mwili dhidi ya maradhi
 1 "l3nt:17" dalili za jeraha katika ubongo

HC-N + GEN-CON + N + PREP + HC-N "l3nt:18"

1 "l3nt:18" uzazi wa mpasuo katika mimba
 1 "l3nt:18" homa ya vipindi katika mwili
 1 "l3nt:18" dawa za kinga kwa ugonjwa
 1 "l3nt:18" dawa za kinga kwa kipindupindu

HC-N + GEN-CON + HC-N + PREP + N "l3nt:19"

1 "l3nt:19" ugonjwa wa chirwa katika watoto
 1 "l3nt:19" mfumo wa afya kwa ujumla
 1 "l3nt:19" mfumo wa afya katika wizara
 1 "l3nt:19" mazoezi ya viungo kwenye jamii
 1 "l3nt:19" ambukizo la ugonjwa kwa watoto

N + GEN-CON + HC-N + PREP + HC-N "l3nt:21"

1 "l3nt:21" shughuli za uchanjaji katika kliniki
 1 "l3nt:21" mkusanyiko wa damu chini ya ngozi
 1 "l3nt:21" mchanganyiko wa damu katika mkojo
 1 "l3nt:21" mashambulizi ya vimelea katika mishipa
 1 "l3nt:21" maisha ya kichocho katika maji
 1 "l3nt:21" eneo la tumbo chini ya kitovu
 1 "l3nt:21" elimu ya afya dhidi ya ukimwi

N + GEN-CON + N + PREP + HC-N "13nt:22"

1 "13nt:22" shimo la takataka kwa_ajili_ya uchafu
1 "13nt:22" matatizo ya usafiri kwa wagonjwa
1 "13nt:22" maisha ya udowezi ndani_ya mwili
1 "13nt:22" jukumu la wazazi katika afya
1 "13nt:22" hatua za kinga dhidi_ya minyoo

N + GEN-CON + HC-N + PREP + N "13nt:23"

1 "13nt:23" uvamizi wa ameba kwenye ini
1 "13nt:23" upungufu wa damu kwa mama
1 "13nt:23" umbo la mwili kwa jumla
1 "13nt:23" ulinzi wa mwili dhidi_ya magonjwa
1 "13nt:23" sekta ya afya kwenye jumuiya
1 "13nt:23" programu za chanjo katika jumuiya
1 "13nt:23" magonjwa ya zinaa kwa ujumla
1 "13nt:23" mabadiliko ya kimwili kwa wasichana
1 "13nt:23" huduma za afya kwa mtoto
1 "13nt:23" huduma za afya kwa kina_mama

HC-N + GEN-CON + HC-V + GEN-CON + N "13nt:24"

1 "13nt:24" ogani ya kufyonzea ya mdudu

HC-N + GEN-CON + INF + GEN-CON + HC-N "13nt:26"

1 "13nt:26" madhara ya kudumu ya ugonjwa

N + GEN-CON + HC-V + GEN-CON + N "13nt:28"

1 "13nt:28" hatua ya kuzaliwa kwa mtoto
1 "13nt:28" hali ya kuzuiwa kwa maendeleo

N + GEN-CON + HC-V + GEN-CON + HC-N "13nt:29"

1 "13nt:29" njia ya upunguzaji wa vimelea
1 "13nt:29" mafanikio ya kupungua kwa utapiamlo

N + GEN-CON + INF + GEN-CON + HC-N "13nt:30"

1 "13nt:30" uharibifu wa kudumu wa viungo
1 "13nt:30" tatizo la kutoboka kwa jino
1 "13nt:30" mipango ya kudumu ya afya
1 "13nt:30" mazalio ya kudumu ya mbu
1 "13nt:30" historia ya kuharibika kwa mimba

HC-N + GEN-CON + INF + HC-N "13nt:31"

1 "13nt:31" mirija ya kupitia mayai
1 "13nt:31" hamu ya kula vyakula
1 "13nt:31" dalili za kukosa maji

HC-N + GEN-CON + HC-V + HC-N "13nt:33"

1 "13nt:33" ugonjwa wa kuhara damu
1 "13nt:33" dawa ya kuchanja ndui

N + GEN-CON + HC-V + HC-N "13nt:35"

1 "13nt:35" desturi ya kunawa mikono

HC-N + GEN-CON + N + CC + N "13nt:44"

1 "13nt:44" vidonge vya sulphametrole na trimethoprim

1 "l3nt:44" usafi wa watoto na mazingira
 1 "l3nt:44" usafi wa banda na vyombo
 1 "l3nt:44" miili ya binadamu na wanyama
 1 "l3nt:44" matumbo ya binadamu na wanyama
 1 "l3nt:44" mafuta ya wanyama na mimea
 1 "l3nt:44" kinyesi cha paka na mbwa

HC-N + GEN-CON + HC-N + CC + N "l3nt:45"

1 "l3nt:45" mnyoo ya filari na onchocerea
 1 "l3nt:45" kinga ya maradhi na matatizo
 1 "l3nt:45" kichocho cha tumbo na ini

HC-N + GEN-CON + N + CC + HC-N "l3nt:46"

1 "l3nt:46" virusi vya papo na ukimwi
 1 "l3nt:46" vioevu vya moto na mvuke
 1 "l3nt:46" mboga za majani na matunda
 1 "l3nt:46" mafuta ya uto na madini
 1 "l3nt:46" homa ya ini na polio

HC-N + GEN-CON + HC-N + CC + HC-N "l3nt:47"

1 "l3nt:47" viinilishe vya utomwili na mafuta
 1 "l3nt:47" vifereji vya mkojo na figo
 1 "l3nt:47" uhusiano wa maji na afya
 1 "l3nt:47" ugonjwa wa tauni na homa
 1 "l3nt:47" maumivu ya viungo na misuli
 1 "l3nt:47" mabaka ya ukoma na vidonda
 1 "l3nt:47" maambukizo ya kibofu na figo
 1 "l3nt:47" homa ya mbu na mimba
 1 "l3nt:47" dawa za tiba na chanjo

N + GEN-CON + HC-N + CC + HC-N "l3nt:48"

2 "l3nt:48" magonjwa ya polio na homa
 2 "l3nt:48" elimu ya afya na kinga
 1 "l3nt:48" uvamizi wa damu na viungo
 1 "l3nt:48" sinia ya sindano na dawa
 1 "l3nt:48" sinia la sindano na dawa
 1 "l3nt:48" njia ya mkojo na uzazi
 1 "l3nt:48" njia ya maji na chakula
 1 "l3nt:48" magonjwa ya zinaa na ukimwi
 1 "l3nt:48" magonjwa ya surua na tetanasi
 1 "l3nt:48" hatua za viluwiluwi na mabuu
 1 "l3nt:48" hali ya afya na nguvu

HC-N + GEN-CON + N + PREP + N "l3nt:54"

1 "l3nt:54" uzazi wa mpango kwenye kanisa
 1 "l3nt:54" maziwa ya ng'ombe mwenye kifukakuu
 1 "l3nt:54" afya za watu kwa ujumla

HC-N + GEN-CON + PREP + HC-N "l3nt:62"

2 "l3nt:62" vimelea vya ndani_ya mwili
 1 "l3nt:62" kinyesi cha mwenye tegu
 1 "l3nt:62" damu ya kwenye kibofu

N + GEN-CON + PREP + HC-N "l3nt:63"

1 "l3nt:63" sehemu ya ndani_ya uke
 1 "l3nt:63" sehemu ya ndani_ya jino
 1 "l3nt:63" sehemu ya juu_ya jino

1 "l3nt:63" sehemu ya chini_ya jino

HC-N + PREP + N + GEN-CON + HC-N "l3nt:74"

2 "l3nt:74" kondo katika mji wa mimba
1 "l3nt:74" utafiti katika sekta ya afya
1 "l3nt:74" uhusiano katika uendeshaji wa afya
1 "l3nt:74" maumivu kwenye sehemu za siri
1 "l3nt:74" maumivu katika njia ya mkojo
1 "l3nt:74" maumivu katika mifuko ya mbegu
1 "l3nt:74" kidonda katika uti wa mgongo
1 "l3nt:74" kidonda katika sehemu ya mwili
1 "l3nt:74" damu katika mji wa mimba
1 "l3nt:74" afya kuhusu utupaji wa kinyesi
1 "l3nt:74" afya katika wizara ya afya
1 "l3nt:74" afya katika kituo cha afya

HC-N + PREP + HC-N + GEN-CON + HC-N "l3nt:75"

5 "l3nt:75" utafiti kuhusu mfumo wa afya
1 "l3nt:75" vidonda katika kiungo cha uzazi
1 "l3nt:75" utafiti katika mfumo wa afya
1 "l3nt:75" mfupa karibu_na fundo la mguu
1 "l3nt:75" maumivu kwenye tezi ya nyonga
1 "l3nt:75" malale katika mwili wa mbung'o
1 "l3nt:75" maji katika mishipa ya vena
1 "l3nt:75" jeraha katika ngozi ya msamba

HC-N + PREP + HC-N + GEN-CON + N "l3nt:76"

1 "l3nt:76" vimelea katika mwili wa mdudu
1 "l3nt:76" jeraha katika ubongo wa mtoto
1 "l3nt:76" damu chini_ya ngozi ya kichwa

N + PREP + N + GEN-CON + HC-N "l3nt:78"

4 "l3nt:78" mtoto katika pango la nyonga
2 "l3nt:78" matezi katika mji wa mimba
1 "l3nt:78" ukakamavu kwa njia ya mazoezi
1 "l3nt:78" uendeshaji katika sekta ya afya
1 "l3nt:78" mtoto katika hali ya afya
1 "l3nt:78" mama mwenye upungufu wa damu
1 "l3nt:78" malendalenda katika njia ya mkojo
1 "l3nt:78" mabadiliko katika takwimu za maradhi
1 "l3nt:78" kichwa katika pango la nyonga

N + PREP + HC-N + GEN-CON + HC-N "l3nt:79"

1 "l3nt:79" mwasho katika kiungo cha uzazi
1 "l3nt:79" mtoto katika tumbo la uzazi
1 "l3nt:79" mpasuko katika tishu za mwili
1 "l3nt:79" magonjwa katika miili ya viumbe
1 "l3nt:79" kazi katika mfumo wa afya

N + PREP + HC-N + GEN-CON + N "l3nt:80"

1 "l3nt:80" udowezi ndani_ya mwili wa binadamu
1 "l3nt:80" salama kwa afya ya binadamu
1 "l3nt:80" mbwa mwenye kichaa cha mbwa
1 "l3nt:80" maisha kwenye mwili wa binadamu
1 "l3nt:80" kinywaji mbali_na maziwa ya mama
1 "l3nt:80" hatari kwa afya ya binadamu

N + GEN-CON + ADV + GEN-CON + HC-N "l3nt:81"

3 "l3nt:81" kiwango cha juu cha uzazi
1 "l3nt:81" sehemu za juu za mikono
1 "l3nt:81" mfuniko wa nje wa miili
1 "l3nt:81" mashambulizi ya mara_kwa_mara ya malaria
1 "l3nt:81" kiwango cha chini cha vifo

HC-N + GEN-CON + ADV + GEN-CON + HC-N "l3nt:82"

1 "l3nt:82" kipimo cha chini cha uzito

N + POSS + HC-N + GEN-CON + HC-N "l3nt:84"

2 "l3nt:84" watu wenye msongo wa pumzi
1 "l3nt:84" watu wenye vilema vya miguu
1 "l3nt:84" nchi zenye homa ya mbu
1 "l3nt:84" mtu mwenye vimelea vya ugonjwa
1 "l3nt:84" mama wenye mimba za mapacha

N + POSS + HC-N + GEN-CON + N "l3nt:86"

3 "l3nt:86" mnyama mwenye kichaa cha mbwa
1 "l3nt:86" maeneo yenye malaria ya msimu

HC-N + POSS + HC-N + GEN-CON + HC-N "l3nt:87"

1 "l3nt:87" mgonjwa wenye virusi vya ukimwi
1 "l3nt:87" mgonjwa wenye tauni ya tezi
1 "l3nt:87" karoti zenye viini vya ugonjwa
1 "l3nt:87" damu yenye vimelea vya ukimwi

HC-N + POSS + N + GEN-CON + HC-N "l3nt:88"

1 "l3nt:88" mgonjwa mwenye nyumonia ya tauni
1 "l3nt:88" choo chenye bomba la hewa

HC-N + POSS + N + GEN-CON + N "l3nt:89"

1 "l3nt:89" minyoo yenye urefu wa wastani
1 "l3nt:89" choo chenye umbo la mstatili

HC-N + POSS + HC-N + GEN-CON + N "l3nt:90"

1 "l3nt:90" vijidudu vyenye umbile la vistari
1 "l3nt:90" vijidudu vyenye umbile la vinywele
1 "l3nt:90" vijidudu vyenye umbile la nukta
1 "l3nt:90" maji yenye vimelea vya magonjwa
1 "l3nt:90" dawa zenye madini ya potassium

N + GEN-CON + ORD + GEN-CON + HC-N "l3nt:92"

8 "l3nt:92" hatua ya pili ya uzazi
3 "l3nt:92" hatua ya kwanza ya uzazi
2 "l3nt:92" siku ya kwanza ya hedhi
1 "l3nt:92" wiki ya tatu ya ugonjwa
1 "l3nt:92" hatua ya tatu ya uzazi

HC-N + GEN-CON + ORD + GEN-CON + N "l3nt:93"

1 "l3nt:93" dalili ya kwanza ya tatizo

HC-V + GEN-CON + HC-N + GEN-CON + ADV "l3nt:98"

2 "l3nt:98" kuzaliwa kwa kondo ya nyuma

INF + GEN-CON + HC-N + GEN-CON + HC-N "13vt:100"

1 "13vt:100" kuzuka kwa vimelea vya malaria
1 "13vt:100" kupanda kwa shinikizo la damu
1 "13vt:100" kubanwa kwa mishipa ya damu

INF + GEN-CON + N + GEN-CON + HC-N "13vt:101"

8 "13vt:101" kutangulia kwa uzi wa kitovu
2 "13vt:101" kubanwa kwa uzi wa kitovu
1 "13vt:101" kuongezeka kwa umbo la mwili

INF + GEN-CON + HC-N + GEN-CON + N "13vt:102"

1 "13vt:102" kuzuka kwa kansa ya shingo
1 "13vt:102" kusawazika kwa tumbo la mama
1 "13vt:102" kuondoshwa kwa vimelea vya magonjwa
1 "13vt:102" kubanwa kwa mifupa ya fuvu

HC-V + GEN-CON + HC-N + GEN-CON + HC-N "13vt:103"

1 "13vt:103" kuzaliana kwa vimelea vya malale
1 "13vt:103" kuvimba kwa tezi za limfu
1 "13vt:103" kuumizwa kwa kibofu cha mkojo
1 "13vt:103" kupungua kwa mzunguko wa damu
1 "13vt:103" kuenea kwa ugonjwa wa upofu
1 "13vt:103" kuenea kwa ugonjwa wa malaria

HC-V + GEN-CON + N + GEN-CON + N "13vt:104"

1 "13vt:104" kuumbuka kwa kichwa cha mtoto
1 "13vt:104" kuambukiza kwa njia ya mtu

HC-V + GEN-CON + HC-N + GEN-CON + N "13vt:105"

1 "13vt:105" kuziba kwa mishipa ya limfatiki

HC-V + GEN-CON + N + GEN-CON + HC-N "13vt:106"

6 "13vt:106" kuzuiwa kwa maendeleo ya uzazi
3 "13vt:106" kupasuka kwa mfuko wa maji
2 "13vt:106" kupasuka kwa mji wa mimba
1 "13vt:106" utunzaji wa vyanzo vya maji
1 "13vt:106" usambazaji wa matokeo ya utafiti
1 "13vt:106" kuziba kwa virija vya mayai
1 "13vt:106" kuziba kwa njia ya mkojo
1 "13vt:106" kuvimba kwa mifuko ya mbegu
1 "13vt:106" kukua kwa mazoea ya chakula
1 "13vt:106" kuenea kwa matumizi ya dawa
1 "13vt:106" kuenea kwa magonjwa ya zinaa
1 "13vt:106" kuenea kwa magonjwa ya uasherati

HC-V + GEN-CON + N + GEN-CON + INF "13vt:107"

1 "13vt:107" kueneza kwa njia ya kugusana

HC-V + PREP + N + GEN-CON + HC-N "13vt:108"

3 "13vt:108" kupasuka kwa mfuko wa maji
2 "13vt:108" kuenea kwa magonjwa ya uasherati
1 "13vt:108" kuzuiwa kwa maendeleo ya uzazi
1 "13vt:108" kuzuia kwa njia ya chanjo
1 "13vt:108" kuziba kwa njia ya mkojo
1 "13vt:108" kutunga nje_ya nyumba ya uzazi
1 "13vt:108" kusaidia katika programu za chanjo

1 "13vt:108" kusaidia katika huduma za afya
 1 "13vt:108" kupungua kwa mji wa mimba
 1 "13vt:108" kuishi katika hali ya afya
 1 "13vt:108" kuenea kwa magonjwa ya zinaa
 1 "13vt:108" kuenea katika pango la nyonga
 HC-V + PREP + HC-N + GEN-CON + N "13vt:109"
 1 "13vt:109" kuvimba kwa chanjo ya ndui
 HC-V + PREP + N + GEN-CON + N "13vt:111"
 1 "13vt:111" kutibiwa kwa madawa ya kisasa
 1 "13vt:111" kukua kwa akili ya mtoto
 1 "13vt:111" kuenea kwa magonjwa ya uashebatii
 HC-V + N + GEN-CON + HC-N "13vt:115"
 2 "13vt:115" kuugua magonjwa ya zinaa
 HC-V + HC-N + GEN-CON + HC-N "13vt:116"
 1 "13vt:116" kudoea damu ya viumbe
 INF + N + GEN-CON + HC-N "13vt:118"
 1 "13vt:118" kutatua tatizo la maji
 1 "13vt:118" kutatua matatizo ya lishe
 1 "13vt:118" kutatua matatizo ya kiafya
 1 "13vt:118" kulishwa mchanganyiko wa maziwa
 1 "13vt:118" kula mabaki ya vyakula
 1 "13vt:118" kujadili vipindi vya afya
 1 "13vt:118" kufuata kanuni za usafi
 1 "13vt:118" kufikia kiwango cha afya
 1 "13vt:118" kuelewa matumizi ya vyakula
 1 "13vt:118" kuelewa hali ya mgonjwa
 1 "13vt:118" kuchangia vifaa vya upasuaji
 1 "13vt:118" kuchangia huduma ya afya
 1 "13vt:118" kuchafua vyanzo vya maji
 INF + HC-N + GEN-CON + HC-N "13vt:119"
 1 "13vt:119" kupewa dawa za usingizi
 1 "13vt:119" kupewa dawa ya minyoo
 1 "13vt:119" kuepuka mimba za usichana
 INF + HC-N + GEN-CON + N "13vt:120"
 1 "13vt:120" kupelekwa kliniki ya watoto
 1 "13vt:120" kupelekea maambukizo ya hatari
 1 "13vt:120" kula mboga za kijani
 1 "13vt:120" kubeba vijidudu vya magonjwa
 INF + N + GEN-CON + HC-V "13vt:122"
 1 "13vt:122" kujua tarehe ya kuzaliwa
 HC-V + N + GEN-CON + INF "13vt:126"
 1 "13vt:126" kuota ndoto za kutisha
 INF + HC-N + LOC "13vt:128"
 5 "13vt:128" kutoka uchafu ukeni
 5 "13vt:128" kutoka damu ukeni

2 "13vt:128" kukosa maji mwilini
 1 "13vt:128" kutoka usaha ukeni
 1 "13vt:128" kutoka uchafu puani
 1 "13vt:128" kutoka maji ukeni
 1 "13vt:128" kuingiza vijidudu majini
 HC-V + HC-N + LOC "13vt:129"
 1 "13vt:129" kuota kuyu kinywani
 HC-N + A-INFL + GEN-CON + HC-N + GEN-CON + HC-N "14nt:1"
 1 "14nt:1" dalili kubwa za ugonjwa wa homa
 1 "14nt:1" dalili kubwa za homa ya matumbo
 INF + ADV + PREP + N + GEN-CON + HC-N "14vt:127"
 1 "14vt:127" kutoka nje kwa mji wa mimba
 INF + GEN-CON + HC-N + GEN-CON + N + GEN-CON + HC-N "14vt:129"
 1 "14vt:129" kuzuka kwa kansa ya shingo ya uzazi
 1 "14vt:129" kuongezeka kwa nguvu ya mwendo wa damu
 INF + HC-N + PREP + HC-N + GEN-CON + HC-N "14vt:148"
 1 "14vt:148" kusababisha vijidudu katika kibofu cha mkojo
 HC-N + GEN-CON + HC-N + PREP + HC-N + GEN-CON + HC-N "14nt:16"
 1 "14nt:16" vimelea vya tauni katika mwili wa viroboto
 N + GEN-CON + HC-N + GEN-CON + HC-N + GEN-CON + N "14nt:168"
 1 "14nt:168" kanuni za usafi wa chakula cha mtoto
 HC-N + GEN-CON + HC-N + PREP + N + GEN-CON + HC-N "14nt:17"
 3 "14nt:17" viotea vya ngozi katika sehemu za siri
 1 "14nt:17" viotea vya ngozi kwenye sehemu za siri
 1 "14nt:17" mzunguko wa damu katika mji wa mimba
 1 "14nt:17" msongo wa pumzi kutokana na upungufu wa oksijeni
 1 "14nt:17" mfumo wa afya katika wizara ya afya
 N + GEN-CON + HC-N + GEN-CON + N + GEN-CON + HC-N "14nt:170"
 1 "14nt:170" rasi ya mfupa wa pingili za nyonga
 N + GEN-CON + HC-N + GEN-CON + N + GEN-CON + N "14nt:171"
 1 "14nt:171" uvimbe wa ngozi ya kichwa cha mtoto
 N + GEN-CON + HC-N + PREP + HC-N + GEN-CON + HC-N "14nt:178"
 3 "14nt:178" kitengo cha utafiti kuhusu mfumo wa afya
 2 "14nt:178" mafunzo ya utafiti kuhusu mfumo wa afya
 1 "14nt:178" sera za utafiti kuhusu mfumo wa afya
 1 "14nt:178" mpango wa utafiti kuhusu mfumo wa afya
 N + GEN-CON + N + GEN-CON + HC-N + N "14nt:192"
 1 "14nt:192" uamuzi wa shirika la afya duniani
 1 "14nt:192" mshikamano wa shirika la afya duniani
 1 "14nt:192" msaada wa shirika la afya ulimwenguni
 1 "14nt:192" mratibu wa mradi wa lishe mkoani

1 "14nt:192" mkutano wa shirika la afya ulimwenguni
N + GEN-CON + N + GEN-CON + N + HC-N "14nt:193"
1 "14nt:193" kifupisho cha ukosefu wa kinga mwilini
N + GEN-CON + N + PREP + HC-N + GEN-CON + N "14vt:199"
1 "14nt:199" maisha ya udowezi ndani_ya mwili wa binadamu
1 "14nt:199" jukumu la wazazi katika afya ya mtoto
N + GEN-CON + N + PREP + N + GEN-CON + HC-N "14vt:200"
2 "14nt:200" kichwa cha mtoto katika pango la nyonga
1 "14nt:200" mahudhurio ya watu katika taasisi za afya
1 "14nt:200" dhima ya jumuiya katika programu za chanjo
N + INF + PREP + HC-N + GEN-CON + N "14vt:206"
1 "14nt:206" oxygen kutoka katika damu ya mama
HC-N + A-INFL + GEN-CON + N + GEN-CON + HC-N "14nt:3"
1 "14nt:3" mikazo dhaifu ya mji wa mimba
1 "14nt:3" dalili kubwa ya athari za ugonjwa
HC-N + GEN-CON + N + PREP + N + GEN-CON + HC-N "14nt:48"
1 "14nt:48" mishipa ya fahamu katika pango la nyonga
HC-N + GEN-CON + N + PREP + N + GEN-CON + N "14nt:49"
1 "14nt:49" afya ya msingi katika ngazi ya wilaya
N + GEN-CON + HC-N + GEN-CON + N + GEN-CON + N "14vt:171"
2 "14vt:71" kupasuka mapema kwa mfuko wa maji
1 "14vt:74" kulegea kwa misuli ya mji wa mimba
HC-V + GEN-CON + N + GEN-CON + N + GEN-CON + HC-N "14vt:80"
5 "14vt:80" kufunguka kwa shingo ya mji wa mimba
1 "14vt:80" kuchanika kwa shingo ya mji wa mimba
HC-V + HC-N + PREP + HC-N + GEN-CON + N "14vt:98"
1 "14vt:98" kuoga mwili kwa maji ya vuguvugu

APPENDIX 4

Literature terms recovered in the extraction

Length 1 terms

LT-N	"llnt:"		31	"llnt:1"	ubunifu
436	"llnt:1"	fasihi	31	"llnt:1"	tamaduni
395	"llnt:1"	lugha	30	"llnt:1"	utenzi
359	"llnt:1"	riwaya	29	"llnt:1"	utafiti
286	"llnt:1"	mwandishi	29	"llnt:1"	mtazamo
269	"llnt:1"	tamthilia	29	"llnt:1"	kilele
228	"llnt:1"	maudhui	28	"llnt:1"	uhakiki
224	"llnt:1"	ushairi	28	"llnt:1"	tajiriba
212	"llnt:1"	hadithi	28	"llnt:1"	mandhari
196	"llnt:1"	shairi	27	"llnt:1"	usemi
160	"llnt:1"	mhusika	27	"llnt:1"	kisa
141	"llnt:1"	tamthiliya	24	"llnt:1"	urari
141	"llnt:1"	sanaa	23	"llnt:1"	urasimi
140	"llnt:1"	nadharia	23	"llnt:1"	mkondo
126	"llnt:1"	mshororo	23	"llnt:1"	mdundo
106	"llnt:1"	ujumbe	20	"llnt:1"	mswada
106	"llnt:1"	fani	19	"llnt:1"	mada
102	"llnt:1"	mizani	19	"llnt:1"	lahaja
90	"llnt:1"	utamaduni	18	"llnt:1"	tamathali
90	"llnt:1"	mtindo	18	"llnt:1"	itikadi
82	"llnt:1"	ubeti	17	"llnt:1"	kipawa
81	"llnt:1"	methali	17	"llnt:1"	kaida
75	"llnt:1"	ngano	16	"llnt:1"	takriri
74	"llnt:1"	falsafa	16	"llnt:1"	mhakiki
73	"llnt:1"	taswira	15	"llnt:1"	ubwege
70	"llnt:1"	vichekesho	15	"llnt:1"	tashihisi
67	"llnt:1"	nyimbo	15	"llnt:1"	taharuki
66	"llnt:1"	maandishi	15	"llnt:1"	jadi
65	"llnt:1"	utunzi	15	"llnt:1"	diwani
59	"llnt:1"	fasihi_simulizi	15	"llnt:1"	bahari
57	"llnt:1"	hisia	14	"llnt:1"	upeo
57	"llnt:1"	dhamira	13	"llnt:1"	malenga
56	"llnt:1"	wimbo	13	"llnt:1"	hulka
55	"llnt:1"	ngoma	13	"llnt:1"	arudhi
54	"llnt:1"	dhana	12	"llnt:1"	utungaji
52	"llnt:1"	mshairi	12	"llnt:1"	ufafanuzi
52	"llnt:1"	mchezo	12	"llnt:1"	simulizi
51	"llnt:1"	jazanda	11	"llnt:1"	upelelezi
48	"llnt:1"	muktadha	11	"llnt:1"	umenke
47	"llnt:1"	mtunzi	11	"llnt:1"	uchambuzi
47	"llnt:1"	drama	11	"llnt:1"	hekaya
44	"llnt:1"	muundo	11	"llnt:1"	fumbo
44	"llnt:1"	msimulizi	10	"llnt:1"	uwasilishaji
43	"llnt:1"	taashira	10	"llnt:1"	onyesho
42	"llnt:1"	msimamo	10	"llnt:1"	kichekesho
40	"llnt:1"	sadfa	10	"llnt:1"	fasihi_andishi
38	"llnt:1"	usimulizi	10	"llnt:1"	changamano
36	"llnt:1"	mtiririko	9	"llnt:1"	wasikilizaji
35	"llnt:1"	uandishi	9	"llnt:1"	uyakinifu
34	"llnt:1"	mgogoro	9	"llnt:1"	ufasaha
34	"llnt:1"	hadhira	9	"llnt:1"	nadhari
33	"llnt:1"	utanzu	9	"llnt:1"	mgongano
33	"llnt:1"	uhusika	9	"llnt:1"	chimbuko
33	"llnt:1"	msanii	8	"llnt:1"	tahakiki
33	"llnt:1"	masimulizi	8	"llnt:1"	shujaa
32	"llnt:1"	utungo	8	"llnt:1"	nadhari

8	"llnt:1"	mila	3	"llnt:1"	amali
7	"llnt:1"	uigizaji	2	"llnt:1"	wasifu
7	"llnt:1"	tanakali	2	"llnt:1"	uundaji
7	"llnt:1"	mwanatamthilia	2	"llnt:1"	uumbaji
7	"llnt:1"	mtazamaji	2	"llnt:1"	uthakili
7	"llnt:1"	maono	2	"llnt:1"	utendi
7	"llnt:1"	mantiki	2	"llnt:1"	ushujaa
7	"llnt:1"	kibwagizo	2	"llnt:1"	ujumi
6	"llnt:1"	usanifu	2	"llnt:1"	tamthilia-teuliwa
6	"llnt:1"	tarbia	2	"llnt:1"	tamrini
6	"llnt:1"	ruwaza	2	"llnt:1"	ridhimu
6	"llnt:1"	mwanafasihi	2	"llnt:1"	ramli
6	"llnt:1"	mawaidha	2	"llnt:1"	ngomani
6	"llnt:1"	kunga	2	"llnt:1"	mviga
6	"llnt:1"	kitendawili	2	"llnt:1"	mtagusano
6	"llnt:1"	kigezo	2	"llnt:1"	mori
6	"llnt:1"	kasumba	2	"llnt:1"	mlingano
6	"llnt:1"	karama	2	"llnt:1"	miigo
6	"llnt:1"	iktisadi	2	"llnt:1"	mchezoni
5	"llnt:1"	waigizaji	2	"llnt:1"	matini
5	"llnt:1"	visasili	2	"llnt:1"	mashabiki
5	"llnt:1"	uwanda	2	"llnt:1"	mapigo
5	"llnt:1"	takhmisa	2	"llnt:1"	makala
5	"llnt:1"	taarabu	2	"llnt:1"	makafara
5	"llnt:1"	ngonjera	2	"llnt:1"	mahubiri
5	"llnt:1"	muwala	2	"llnt:1"	lafudhi
5	"llnt:1"	mtambaji	2	"llnt:1"	kileleni
5	"llnt:1"	mtafaruku	2	"llnt:1"	kikoa
5	"llnt:1"	matapo	2	"llnt:1"	kiimbo
5	"llnt:1"	mahadhi	2	"llnt:1"	balagha
5	"llnt:1"	lakabu	1	"llnt:1"	zeze
5	"llnt:1"	kejeli	1	"llnt:1"	wanachuoni
5	"llnt:1"	desturi	1	"llnt:1"	wachoraji
4	"llnt:1"	ukasuku	1	"llnt:1"	vioja
4	"llnt:1"	ukale	1	"llnt:1"	uumbaji
4	"llnt:1"	uchanganuzi	1	"llnt:1"	utumbuizo
4	"llnt:1"	uasili	1	"llnt:1"	ususi
4	"llnt:1"	tumbuizo	1	"llnt:1"	usuli
4	"llnt:1"	tasfida	1	"llnt:1"	upembuzi
4	"llnt:1"	taathira	1	"llnt:1"	unyago
4	"llnt:1"	nahau	1	"llnt:1"	umalenga
4	"llnt:1"	mwimbaji	1	"llnt:1"	ulinganifu
4	"llnt:1"	mvuto	1	"llnt:1"	ukomo
4	"llnt:1"	muziki	1	"llnt:1"	uhafidhina
4	"llnt:1"	mtazamo-kike	1	"llnt:1"	ufasihi
4	"llnt:1"	mawanda	1	"llnt:1"	uchapwa
4	"llnt:1"	mappingiti	1	"llnt:1"	uangavu
4	"llnt:1"	malumbano	1	"llnt:1"	tendi
4	"llnt:1"	malimwengu	1	"llnt:1"	tawasifu
4	"llnt:1"	majigambo	1	"llnt:1"	tashwishi
4	"llnt:1"	kafara	1	"llnt:1"	tapo
4	"llnt:1"	dhima	1	"llnt:1"	tamthilia-mwavuli
4	"llnt:1"	ala	1	"llnt:1"	sitiari
3	"llnt:1"	vumba	1	"llnt:1"	shairi-huru
3	"llnt:1"	utopia	1	"llnt:1"	pungwa
3	"llnt:1"	utambaji	1	"llnt:1"	pigo
3	"llnt:1"	ulimbwende	1	"llnt:1"	mwanatamthiliya
3	"llnt:1"	tanzia	1	"llnt:1"	mwanasanaa
3	"llnt:1"	tafakuri	1	"llnt:1"	mwanariwaya
3	"llnt:1"	mwanafalsafa	1	"llnt:1"	mwanamapoeko
3	"llnt:1"	msemu	1	"llnt:1"	mwanagenzi
3	"llnt:1"	mguso	1	"llnt:1"	muundo-msingi
3	"llnt:1"	majibizano	1	"llnt:1"	muhula
3	"llnt:1"	maigizo	1	"llnt:1"	mtungaji
3	"llnt:1"	dhahania	1	"llnt:1"	mtizamo

1 "11nt:1"	mtabiri	1 "11nt:1"	fanani
1 "11nt:1"	msuluhishi	1 "11nt:1"	dibaji
1 "11nt:1"	msimuliaji	1 "11nt:1"	changizi
1 "11nt:1"	mshindo	1 "11nt:1"	chakacha
1 "11nt:1"	mpambaji	1 "11nt:1"	andiko
1 "11nt:1"	mirindimo		
1 "11nt:1"	masivina	LT-V "11vt"	
1 "11nt:1"	mashairi-andishi		
1 "11nt:1"	masaibu	46 "11vt:2"	kutunga
1 "11nt:1"	mapitio	32 "11vt:2"	kubuni
1 "11nt:1"	manju	10 "11vt:2"	kusimulia
1 "11nt:1"	lelemama	7 "11vt:2"	kuimba
1 "11nt:1"	kutaharuki	5 "11vt:2"	kuiga
1 "11nt:1"	kivumba	5 "11vt:2"	kuhakiki
1 "11nt:1"	kitambo	5 "11vt:1"	kukejeli
1 "11nt:1"	kishujaa	2 "11vt:2"	kutamba
1 "11nt:1"	kimila	2 "11vt:2"	kusanii
1 "11nt:1"	kimantiki	2 "11vt:2"	kusanifu
1 "11nt:1"	kimahadhi	2 "11vt:2"	kusadifu
1 "11nt:1"	kimafumbo	2 "11vt:2"	kunukuu
1 "11nt:1"	kilinge	2 "11vt:2"	kukariri
1 "11nt:1"	kijazanda	1 "11vt:2"	kusadifu
1 "11nt:1"	kihulka	1 "11vt:2"	kutanabahi
1 "11nt:1"	kidokezo	1 "11vt:2"	kutabiri
1 "11nt:1"	kiamu	1 "11vt:2"	kusaliti
1 "11nt:1"	juzuu	1 "11vt:2"	kufunda
1 "11nt:1"	jando	1 "11vt:2"	kufinyanga
1 "11nt:1"	hadithi-pendezi	1 "11vt:2"	kufariji
1 "11nt:1"	guni	1 "11vt:2"	kusimulia
1 "11nt:1"	filosofia		

Length 2 terms

LT-N + LT-N "12nt:1"

4 "12nt:1"	nadharia mtazamo-kike	1 "12nt:1"	taharuki nyimbo
4 "12nt:1"	fasihi simulizi	1 "12nt:1"	shairi guni
3 "12nt:1"	mashairi simulizi	1 "12nt:1"	sanaa andishi
1 "12nt:1"	utambaji hadithi	1 "12nt:1"	riwaya changamano
1 "12nt:1"	atafiti nyimbo	1 "12nt:1"	mhusika msimulizi
1 "12nt:1"	atafiti mafumbo	1 "12nt:1"	maudhui changizi
		1 "12nt:1"	fasihi changamano

LT-N + N "12nt:2"

4 "12nt:2"	tamthilia toto	1 "12nt:2"	taswira mfululizo
3 "12nt:2"	mhusika mkuu	1 "12nt:2"	sanaa halali
2 "12nt:2"	watunzi chipukizi	1 "12nt:2"	ngano chapwa
2 "12nt:2"	utamaduni mseto	1 "12nt:2"	mtunzi stadi
2 "12nt:2"	msimamo thabiti	1 "12nt:2"	mtazamo tofauti
2 "12nt:2"	mhusika msaidizi	1 "12nt:2"	mshairi msifika
1 "12nt:2"	wanafasihi	1 "12nt:2"	mizani nanenane
	nyanjani	1 "12nt:2"	mitazamo tofauti
1 "12nt:2"	wanachuoni mufti	1 "12nt:2"	mikondo tofauti
1 "12nt:2"	wahusika duara	1 "12nt:2"	mhusika mkwezwa
1 "12nt:2"	wahusika bapa-sugu	1 "12nt:2"	mhusika jumui
1 "12nt:2"	waandishi	1 "12nt:2"	mhusika jina
	wanamapinduzi	1 "12nt:2"	mhusika duara
1 "12nt:2"	waandishi wakongwe	1 "12nt:2"	mhusika bwana
1 "12nt:2"	waandishi	1 "12nt:2"	mhusika buruto
	chipukizi	1 "12nt:2"	mhusika binadamu
1 "12nt:2"	utamaduni asili	1 "12nt:2"	mhusika abunuasi
1 "12nt:2"	atafiti nyanjani	1 "12nt:2"	maudhui haba
1 "12nt:2"	ujumbe jumla	1 "12nt:2"	mashairi
1 "12nt:2"	uigizaji jukwaani		wanamuziki

1 "12nt:2"	lugha njema	1 "12nt:2"	kisa mchangamano
1 "12nt:2"	lugha asili	1 "12nt:2"	bahari tesi
1 "12nt:2"	kisa mwisio	1 "12nt:2"	bahari tesi
N + LT-N "12nt:4"			
2 "12nt:4"	uchezaaji ngoma	1 "12nt:4"	umuhimu kifalsafa
2 "12nt:4"	mtu shujaa	1 "12nt:4"	umma kifasihi
2 "12nt:4"	mpiga ramli	1 "12nt:4"	ukwasi riwaya
2 "12nt:4"	elimu lahaja	1 "12nt:4"	ufundishaji
1 "12nt:4"	wasomi wanagenzi		ushairi
1 "12nt:4"	wapelelezi	1 "12nt:4"	ploti changamano
	mashujaa	1 "12nt:4"	nguvu lugha
1 "12nt:4"	wanafunzi	1 "12nt:4"	muhtasari ngano
	wasimulizi	1 "12nt:4"	mtetea utamaduni
1 "12nt:4"	wachapishaji	1 "12nt:4"	msamiati
	watunzi		kimuktadha
1 "12nt:4"	vipengele	1 "12nt:4"	mcheza ngoma
	changamano	1 "12nt:4"	matukio hadithini
1 "12nt:4"	upigaji muziki	1 "12nt:4"	kazi changamano
1 "12nt:4"	uonyeshaji		
	tamthiliya		
LT-V + LT-N "12vt:5"			
4 "12vt:5"	kutunga tamthiliya	1 "12vt:5"	kutabiri migogoro
4 "12vt:5"	kusimulia hadithi	1 "12vt:5"	kuimba nyimbo
3 "12vt:5"	kutunga riwaya	1 "12vt:5"	kuimba ngano
3 "12vt:5"	kutunga mashairi	1 "12vt:5"	kuiga mitindo
2 "12vt:5"	kutunga vichekesho	1 "12vt:5"	kuhakiki
1 "12vt:5"	kutunga utenzi		tamthiliya
1 "12vt:5"	kutunga shairi	1 "12vt:5"	kuhakiki mchezo
1 "12vt:5"	kutunga sanaa	1 "12vt:5"	kubuni wahusika
1 "12vt:5"	kutunga kichekesho		
LT-V + N "12vt:6"			
1 "12vt:6"	kutunga istilahi	1 "12vt:6"	kuiga uzungu
1 "12vt:6"	kusimulia tabia	1 "12vt:6"	kuiga tungo
1 "12vt:6"	kusimulia maisha	1 "12vt:6"	kuiga tabia
1 "12vt:6"	kusanii kazi	1 "12vt:6"	kuhakiki kazi
1 "12vt:6"	kusanifu kazi	1 "12vt:6"	kufinyanga
1 "12vt:6"	kunukuu kauli		mazingira
1 "12vt:6"	kunukuu kalima	1 "12vt:6"	kubuni tatizo
1 "12vt:6"	kukariri maadili		
INF + LT-N "12vt:7"			
9 "12vt:7"	kuwasilisha ujumbe	3 "12vt:7"	kutafsiri
8 "12vt:7"	kufundisha		maandishi
	mashairi	3 "12vt:7"	kusoma riwaya
8 "12vt:7"	kuandika riwaya	3 "12vt:7"	kuelewa maudhui
6 "12vt:7"	kufundisha	2 "12vt:7"	kuzingatia
	tamthilia		nadharia
6 "12vt:7"	kufundisha fasihi	2 "12vt:7"	kutoa ujumbe
6 "12vt:7"	kuendeleza maudhui	2 "12vt:7"	kuteka hisia
5 "12vt:7"	kukuza fasihi	2 "12vt:7"	kutafsiri fasihi
5 "12vt:7"	kufundisha methali	2 "12vt:7"	kupitia taswira
5 "12vt:7"	kuandika	2 "12vt:7"	kukuza lugha
	tamthiliya	2 "12vt:7"	kujenga fasihi
4 "12vt:7"	kuzingatia fani	2 "12vt:7"	kufurahia hadithi
4 "12vt:7"	kuwasilisha ujumbe	2 "12vt:7"	kufundisha ushairi
4 "12vt:7"	kutumia methali	2 "12vt:7"	kufundisha riwaya
4 "12vt:7"	kusoma shairi	2 "12vt:7"	kufundisha
4 "12vt:7"	kuandika hadithi		mshororo

2	"12vt:7"	kufundisha maudhui	1	"12vt:7"	kusherehekea
2	"12vt:7"	kufundisha fani			kilele
2	"12vt:7"	kufufua kaida	1	"12vt:7"	kusheheneza
2	"12vt:7"	kufanya utafiti			taswira
2	"12vt:7"	kuelezea dhana	1	"12vt:7"	kushadidia ujumbe
2	"12vt:7"	kuelewa nadharia	1	"12vt:7"	kusambaza mashairi
2	"12vt:7"	kuchambua shairi	1	"12vt:7"	kuruhusu ubunifu
2	"12vt:7"	kuandika tamthilia	1	"12vt:7"	kurejelea shairi
2	"12vt:7"	kuandika mashairi	1	"12vt:7"	kurejelea muktadha
1	"12vt:7"	kuzungumzia	1	"12vt:7"	kupuuza muundo
		tamthilia	1	"12vt:7"	kupotosha fasihi
1	"12vt:7"	kuzungumzia	1	"12vt:7"	kupinga ushairi
		nadharia	1	"12vt:7"	kupendekeza
					tasfida
1	"12vt:7"	kuzingatia ubunifu	1	"12vt:7"	kupenda tamaduni
1	"12vt:7"	kuzingatia	1	"12vt:7"	kupata maudhui
		tamathali	1	"12vt:7"	kupata kilele
1	"12vt:7"	kuzingatia mhusika	1	"12vt:7"	kupata dhamira
1	"12vt:7"	kuzingatia mdundo	1	"12vt:7"	kupanga ushairi
1	"12vt:7"	kuzingatia maudhui	1	"12vt:7"	kuorodhesha
1	"12vt:7"	kuzingatia	1	"12vt:7"	vidokezo
		majibizano			kuonyesha vipawa
1	"12vt:7"	kuzingatia mada	1	"12vt:7"	kuonyesha
1	"12vt:7"	kuzingatia hadithi	1	"12vt:7"	vichekesho
1	"12vt:7"	kuweka vidokezo			kuonyesha usanifu
1	"12vt:7"	kuweka migongano	1	"12vt:7"	kuonyesha
1	"12vt:7"	kuwasilisha ngano	1	"12vt:7"	tamthiliya
1	"12vt:7"	kuvuta tafakuri			kuonyesha
1	"12vt:7"	kuua utamaduni	1	"12vt:7"	migongano
1	"12vt:7"	kuua fasihi			kuondoa kasumba
1	"12vt:7"	kutwaa utamaduni	1	"12vt:7"	kunajisi ushairi
1	"12vt:7"	kutumika kijazanda	1	"12vt:7"	kumwaya falsafa
1	"12vt:7"	kutumia vipawa	1	"12vt:7"	kumulika maudhui
1	"12vt:7"	kutumia mizani	1	"12vt:7"	kukuza vichekesho
1	"12vt:7"	kutumia mandhari	1	"12vt:7"	kukuza tamaduni
1	"12vt:7"	kutumia fani	1	"12vt:7"	kukuza maudhui
1	"12vt:7"	kutoa tamthiliya	1	"12vt:7"	kukuza kipawa
1	"12vt:7"	kutoa midundo	1	"12vt:7"	kukuza hisia
1	"12vt:7"	kutoa hisia	1	"12vt:7"	kukusanya ngano
1	"12vt:7"	kutiana mori	1	"12vt:7"	kukosa muwala
1	"12vt:7"	kuteua maudhui	1	"12vt:7"	kukosa msimamo
1	"12vt:7"	kuteua lugha	1	"12vt:7"	kukosa mantiki
1	"12vt:7"	kutenganisha visa	1	"12vt:7"	kukopa arudhi
1	"12vt:7"	kutegemea ubunifu	1	"12vt:7"	kukamilisha
1	"12vt:7"	kutegemea muktadha			ubunifu
1	"12vt:7"	kutegemea itikadi			kukamilisha
1	"12vt:7"	kutegemea dhamira	1	"12vt:7"	tamthiliya
1	"12vt:7"	kutazama uigizaji			kukamilisha
1	"12vt:7"	kutazama	1	"12vt:7"	dhamira
		tamthiliya			kukadiria taathira
1	"12vt:7"	kutayarisha	1	"12vt:7"	kujengea maudhui
		wasanii	1	"12vt:7"	kujenga taswira
1	"12vt:7"	kutanzua	1	"12vt:7"	kujenga taharuki
		kitendawili	1	"12vt:7"	kujenga shairi
1	"12vt:7"	kutalii fasihi	1	"12vt:7"	kujenga riwaya
1	"12vt:7"	kutahini	1	"12vt:7"	kujenga kaida
		vitendawili	1	"12vt:7"	kujadili
1	"12vt:7"	kutafiti fumbo	1	"12vt:7"	tamthiliya
1	"12vt:7"	kustawisha			kujadili sanaa
		utamaduni	1	"12vt:7"	kuingiza drama
1	"12vt:7"	kusoma tamthilia	1	"12vt:7"	kuingilia
1	"12vt:7"	kusoma mashairi	1	"12vt:7"	tamthiliya
1	"12vt:7"	kusoma lugha			kuibuka maandishi
1	"12vt:7"	kusoma fasihi	1	"12vt:7"	kuhozi methali
1	"12vt:7"	kusikiliza hadithi	1	"12vt:7"	

1 "l2vt:7"	kuhifadhi	1 "l2vt:7"	kuendeleza tanzu
1 "l2vt:7"	fasihi simulizi	1 "l2vt:7"	kuendeleza sanaa
1 "l2vt:7"	kuhesabu takriri	1 "l2vt:7"	kuendeleza nadharia
1 "l2vt:7"	kugusa hisia	1 "l2vt:7"	kuendeleza lugha
1 "l2vt:7"	kugawa vipawa	1 "l2vt:7"	kuendeleza hisia
1 "l2vt:7"	kufurahia lugha	1 "l2vt:7"	kuendeleza fasihi
1 "l2vt:7"	kufupisha ujumbe	1 "l2vt:7"	kuendeleza fani
1 "l2vt:7"	kufupisha matini	1 "l2vt:7"	kuendeleza dhana
1 "l2vt:7"	kufunza tamthilia	1 "l2vt:7"	kueleza tajiriba
1 "l2vt:7"	kufunza lugha	1 "l2vt:7"	kueleza hisia
1 "l2vt:7"	kufunza fani	1 "l2vt:7"	kueleza falsafa
1 "l2vt:7"	kufundisha	1 "l2vt:7"	kueleza dhana
1 "l2vt:7"	visasili	1 "l2vt:7"	kuelewa utunzi
1 "l2vt:7"	kufundisha taswira	1 "l2vt:7"	kuelewa utamaduni
1 "l2vt:7"	kufundisha tanzu	1 "l2vt:7"	kuelewa ushairi
1 "l2vt:7"	kufundisha tamathali	1 "l2vt:7"	kuelewa shairi
1 "l2vt:7"	kufundisha takriri	1 "l2vt:7"	kuelewa muundo
1 "l2vt:7"	kufundisha nyimbo	1 "l2vt:7"	kuelewa mashairi
1 "l2vt:7"	kufundisha ngano	1 "l2vt:7"	kuelewa hisia
1 "l2vt:7"	kufundisha nathari	1 "l2vt:7"	kudhihirisha
1 "l2vt:7"	kufundisha mishororo	1 "l2vt:7"	maudhui
1 "l2vt:7"	kufumbua mafumbo	1 "l2vt:7"	kudharau uasili
1 "l2vt:7"	kufumbua fumbo	1 "l2vt:7"	kuchukua msimamo
1 "l2vt:7"	kufufua utamaduni	1 "l2vt:7"	kucheza muziki
1 "l2vt:7"	kufuata mkondo	1 "l2vt:7"	kucheza ala
1 "l2vt:7"	kufua visa	1 "l2vt:7"	kuchanganya lahaja
1 "l2vt:7"	kufisha fasihi	1 "l2vt:7"	kuchagua migongano
1 "l2vt:7"	kufikia kilele	1 "l2vt:7"	kuchagua methali
1 "l2vt:7"	kuficha hisia	1 "l2vt:7"	kuchafua itikadi
1 "l2vt:7"	kufanya upelelezi	1 "l2vt:7"	kubeba maudhui
1 "l2vt:7"	kufanikisha utunzi	1 "l2vt:7"	kubainisha mikondo
1 "l2vt:7"	kufanikisha sanaa	1 "l2vt:7"	kubainisha maudhui
1 "l2vt:7"	kufahamu mandhari	1 "l2vt:7"	kubainisha falsafa
1 "l2vt:7"	kufafanua dhana	1 "l2vt:7"	kubadilisha mtindo
1 "l2vt:7"	kufafanua dhamira	1 "l2vt:7"	kuathiri hisia
1 "l2vt:7"	kueneza nadharia	1 "l2vt:7"	kuanzisha drama
1 "l2vt:7"	kuendesha drama	1 "l2vt:7"	kuandika drama
		1 "l2vt:7"	kuandika beti
LT-N + GEN-CON + LT-N "l2nt:8"			
16 "l2nt:8"	utunzi wa	3 "l2nt:8"	usimulizi wa
	tamthiliya		hadithi
8 "l2nt:8"	tamathali za usemi	3 "l2nt:8"	ujumbe wa
7 "l2nt:8"	utanzu wa fasihi		mwandishi
7 "l2nt:8"	mwandishi wa	3 "l2nt:8"	uandishi wa
	riwaya		tamthiliya
7 "l2nt:8"	falsafa ya	3 "l2nt:8"	nadharia ya fasihi
	mwandishi	3 "l2nt:8"	maudhui ya
6 "l2nt:8"	urari wa mizani		tamthilia
6 "l2nt:8"	sanaa za maonyesho	3 "l2nt:8"	lugha ya fasihi
6 "l2nt:8"	lugha ya kishairi	3 "l2nt:8"	dhamira ya
5 "l2nt:8"	uandishi wa riwaya		mwandishi
5 "l2nt:8"	lugha ya kifasihi	2 "l2nt:8"	wimbo wa kejeli
4 "l2nt:8"	ujumbe wa shairi	2 "l2nt:8"	watunzi wa
4 "l2nt:8"	tamthiliya za		tamthiliya
	drama	2 "l2nt:8"	waandishi wa
4 "l2nt:8"	sanaa ya lugha		riwaya
4 "l2nt:8"	msimamo wa	2 "l2nt:8"	utunzi wa
	mwandishi		vichekesho
4 "l2nt:8"	maudhui ya hadithi	2 "l2nt:8"	utunzi wa
3 "l2nt:8"	utunzi wa riwaya		kichekesho
3 "l2nt:8"	utanzu wa riwaya	2 "l2nt:8"	utungaji wa riwaya
		2 "l2nt:8"	utanzu wa ushairi

2	"12nt:8"	usimulizi wa ngano	1	"12nt:8"	utungaji wa
2	"12nt:8"	upeo wa fasihi			mashairi
2	"12nt:8"	uhakiki wa ushairi	1	"12nt:8"	utungaji wa fasihi
2	"12nt:8"	tamthiliya za	1	"12nt:8"	utambaji wa fasihi
		vichekesho	1	"12nt:8"	utamaduni wa
2	"12nt:8"	nadharia za fasihi			mwandishi
2	"12nt:8"	nadharia ya	1	"12nt:8"	utafiti wa fasihi
		uyakinifu	1	"12nt:8"	usimulizi wa
2	"12nt:8"	nadharia ya			mwandishi
		ushairi	1	"12nt:8"	usimulizi wa
2	"12nt:8"	nadharia ya ubwege			kinathari
2	"12nt:8"	mwandishi wa	1	"12nt:8"	urasimi wa fasihi
		mchezo	1	"12nt:8"	upeo wa ushairi
2	"12nt:8"	muundo wa ushairi	1	"12nt:8"	upeo wa riwaya
2	"12nt:8"	muundo wa	1	"12nt:8"	ulimbwende wa
		tamthilia			tamthilia
2	"12nt:8"	muundo wa shairi	1	"12nt:8"	ujumbe wa ngano
2	"12nt:8"	muktadha wa	1	"12nt:8"	ujumbe wa mshairi
		tamthilia	1	"12nt:8"	ujumbe wa hadithi
2	"12nt:8"	mtindo wa sadfa	1	"12nt:8"	uhusika wa
2	"12nt:8"	mtindo wa ngano			wahusika
2	"12nt:8"	mtindo wa methali	1	"12nt:8"	uhusika wa
2	"12nt:8"	mtazamo wa mtunzi			tamthilia
2	"12nt:8"	mshororo wa shairi	1	"12nt:8"	uhusika wa mhusika
2	"12nt:8"	mkondo wa hadithi	1	"12nt:8"	uhakiki wa
2	"12nt:8"	maudhui ya shairi			tamthilia
2	"12nt:8"	maudhui ya	1	"12nt:8"	uhakiki wa riwaya
		mwandishi	1	"12nt:8"	uhakiki wa
2	"12nt:8"	mashairi ya kijadi			mashairi
2	"12nt:8"	lugha ya ushairi	1	"12nt:8"	uhakiki wa
2	"12nt:8"	lugha ya nathari			kimuundo
2	"12nt:8"	fasihi ya lugha	1	"12nt:8"	uhakiki wa
2	"12nt:8"	dhana ya fasihi			kimtindo
2	"12nt:8"	dhamira ya shairi	1	"12nt:8"	uhakiki wa
1	"12nt:8"	wimbo wa hadithini			kifasihi
1	"12nt:8"	watunzi wa	1	"12nt:8"	uhakiki wa fasihi
		vichekesho	1	"12nt:8"	ufasihi wa methali
1	"12nt:8"	watunzi wa fasihi	1	"12nt:8"	ufasaha wa lugha
1	"12nt:8"	wasimulizi wa	1	"12nt:8"	ufafanuzi wa
		hadithi			wanagenzi
1	"12nt:8"	wasanii wa fasihi	1	"12nt:8"	ufafanuzi wa lugha
1	"12nt:8"	wahusika wa	1	"12nt:8"	uchanganuzi wa
		tamthilia			shairi
1	"12nt:8"	wahusika wa riwaya	1	"12nt:8"	ubunifu wa lugha
1	"12nt:8"	wahusika wa	1	"12nt:8"	ubeti wa utenzi
		kidhania	1	"12nt:8"	uandishi wa mswada
1	"12nt:8"	wahusika wa	1	"12nt:8"	uandishi wa
		hadithi			mashairi
1	"12nt:8"	wahusika wa	1	"12nt:8"	tanzu za ushairi
		fasihi_simulizi	1	"12nt:8"	tanzia ya lugha
1	"12nt:8"	vidokezo vya	1	"12nt:8"	tamthiliya ya
		kitamthiliya			drama
1	"12nt:8"	uwasilishaji wa	1	"12nt:8"	tamathali za semi
		ujumbe	1	"12nt:8"	tamathali za lugha
1	"12nt:8"	uwasilishaji wa	1	"12nt:8"	tajiriba ya
		tajiriba			ushairi
1	"12nt:8"	uwasilishaji wa	1	"12nt:8"	taashira ya
		sanaa			utamaduni
1	"12nt:8"	uwanda wa utafiti	1	"12nt:8"	taashira ya ngoma
1	"12nt:8"	utunzi wa ushairi	1	"12nt:8"	taashira ya
1	"12nt:8"	utunzi wa sanaa			migongano
1	"12nt:8"	utungo wa	1	"12nt:8"	taashira ya
		mishororo			migogoro
1	"12nt:8"	utungo wa kishairi	1	"12nt:8"	shairi la tarbia
1	"12nt:8"	utungo wa drama			

1 "12nt:8"	sanaa ya	1 "12nt:8"	mikondo ya ubunifu
1 "12nt:8"	vitendawili	1 "12nt:8"	midundo ya ngoma
1 "12nt:8"	sanaa ya ushairi	1 "12nt:8"	michezo ya
1 "12nt:8"	sanaa ya tanzu	1 "12nt:8"	washairi
1 "12nt:8"	sanaa ya nathari	1 "12nt:8"	mhusika wa
1 "12nt:8"	sanaa ya maonyesho	1 "12nt:8"	mchezoni
1 "12nt:8"	sanaa ya fasihi	1 "12nt:8"	mhakiki wa ushairi
1 "12nt:8"	riwaya za	1 "12nt:8"	mhakiki wa
	upelelezi	1 "12nt:8"	tamthilia
1 "12nt:8"	riwaya ya shujaa	1 "12nt:8"	mdundo wa ushairi
1 "12nt:8"	riwaya ya mtunzi	1 "12nt:8"	mdundo wa shairi
1 "12nt:8"	nyimbo za taarabu	1 "12nt:8"	mdundo wa muziki
1 "12nt:8"	nyimbo za ngomani	1 "12nt:8"	mawaidha ya
1 "12nt:8"	ngoma za kijadi	1 "12nt:8"	msimulizi
1 "12nt:8"	ngano za	1 "12nt:8"	maudhui ya
	kisimulizi	1 "12nt:8"	vichekesho
1 "12nt:8"	ngano za	1 "12nt:8"	maudhui ya ushairi
	fasihi_simulizi	1 "12nt:8"	maudhui ya tanzu
1 "12nt:8"	nadharia ya utanzu	1 "12nt:8"	maudhui ya riwaya
1 "12nt:8"	nadharia ya	1 "12nt:8"	maudhui ya ngano
	urasimi	1 "12nt:8"	maudhui ya
1 "12nt:8"	nadharia ya		kichekesho
	uhakiki	1 "12nt:8"	maudhui ya fasihi
1 "12nt:8"	mwimbaji wa	1 "12nt:8"	masimulizi ya
	taarabu		ngoma
1 "12nt:8"	mwandishi wa	1 "12nt:8"	masimulizi ya
	tamthiliya		mwandishi
1 "12nt:8"	mwandishi wa	1 "12nt:8"	masimulizi ya
	tamthilia		mhusika
1 "12nt:8"	muundo wa	1 "12nt:8"	masimulizi ya
	usimulizi		hadithi
1 "12nt:8"	muktadha wa utunzi	1 "12nt:8"	mashairi ya
1 "12nt:8"	muktadha wa utungo		masimulizi
1 "12nt:8"	muktadha wa shairi	1 "12nt:8"	mashairi ya jadi
1 "12nt:8"	muktadha wa	1 "12nt:8"	majibizano ya
	kitamaduni		kishairi
1 "12nt:8"	muktadha wa	1 "12nt:8"	mahadhi ya wimbo
	iktisadi	1 "12nt:8"	maandishi ya sanaa
1 "12nt:8"	muktadha wa	1 "12nt:8"	maandishi ya
	hadithi		nathari
1 "12nt:8"	mtunzi wa	1 "12nt:8"	maandishi ya
	vichekesho		mwandishi
1 "12nt:8"	mtiririko wa	1 "12nt:8"	maandishi ya
	hadithi		kinathari
1 "12nt:8"	mtindo wa mashairi	1 "12nt:8"	lugha ya
1 "12nt:8"	mtindo wa		tamthiliya
	kimafumbo	1 "12nt:8"	lugha ya sanaa
1 "12nt:8"	mtindo wa jadi	1 "12nt:8"	lugha ya riwaya
1 "12nt:8"	mtazamo wa watunzi	1 "12nt:8"	lugha ya mhusika
1 "12nt:8"	mtazamo wa	1 "12nt:8"	lahaja ya ushairi
	mwandishi	1 "12nt:8"	kunga za utungaji
1 "12nt:8"	msimulizi wa kisa	1 "12nt:8"	kunga za ushairi
1 "12nt:8"	msimulizi wa	1 "12nt:8"	kilele cha ubunifu
	hadithi	1 "12nt:8"	kilele cha
1 "12nt:8"	msimamo wa		tamthilia
	kinadharia	1 "12nt:8"	kilele cha mchezo
1 "12nt:8"	mkondo wa ngano	1 "12nt:8"	kilele cha hadithi
1 "12nt:8"	mkondo wa mizani	1 "12nt:8"	kilele cha falsafa
1 "12nt:8"	mizani ya fasihi	1 "12nt:8"	kiimbo cha lugha
1 "12nt:8"	mitindo ya	1 "12nt:8"	kigezo cha uhusika
	uandishi	1 "12nt:8"	jazanda ya ubwege
1 "12nt:8"	misimamo ya	1 "12nt:8"	jazanda ya
	wahusika		iktisadi
1 "12nt:8"	mishororo ya	1 "12nt:8"	itikadi za
	mashairi		kinadharia

1 "l2nt:8" hulka za mhusika
 1 "l2nt:8" hulka ya ushairi
 1 "l2nt:8" hulka ya mhusika
 1 "l2nt:8" hisia za hadhira
 1 "l2nt:8" hadhira ya
 mwandishi
 1 "l2nt:8" hadhira ya
 fasihi_simulizi
 1 "l2nt:8" fasihi ya
 kinathari
 1 "l2nt:8" fani za sanaa
 1 "l2nt:8" fani ya ushairi
 1 "l2nt:8" fani ya sanaa
 1 "l2nt:8" fani ya fasihi
 1 "l2nt:8" falsafa ya mhakiki
 1 "l2nt:8" dhana ya tamthilia
 1 "l2nt:8" dhana ya shairi
 1 "l2nt:8" dhamira za
 mwandishi
 1 "l2nt:8" dhamira ya
 tamthiliya
 1 "l2nt:8" dhamira ya riwaya
 1 "l2nt:8" dhamira ya hekaya
 1 "l2nt:8" chimbuko la
 vitendawili
 1 "l2nt:8" chimbuko la riwaya
 1 "l2nt:8" chimbuko la fasihi

LT-N + GEN-CON + N "l2nt:9"

21 "l2nt:9" fasihi ya kiswahili
17 "l2nt:9" ushairi wa kiswahili
13 "l2nt:9" riwaya ya kiswahili
7 "l2nt:9" mashairi ya kiswahili
5 "l2nt:9" tanakali za sauti
5 "l2nt:9" shairi la kiswahili
5 "l2nt:9" ngoma ya ng'wanamalundi
5 "l2nt:9" mshororo wa mwisho
5 "l2nt:9" mashairi ya kisasa
5 "l2nt:9" lugha ya mkato
4 "l2nt:9" wimbo wa mtoto
4 "l2nt:9" tamthilia ya kibwege
4 "l2nt:9" mtiririko wa vituko
3 "l2nt:9" wasanii wa kibwege
3 "l2nt:9" utamaduni wa jamii
3 "l2nt:9" ufasaha wa maneno
3 "l2nt:9" tamthilia ya kilio
3 "l2nt:9" sanaa ya kibwege
3 "l2nt:9" nadharia ya utabia
3 "l2nt:9" nadharia ya kihafidhina
3 "l2nt:9" muktadha wa matumizi
3 "l2nt:9" muktadha wa jamii
3 "l2nt:9" mashairi ya kimapokeo
3 "l2nt:9" falsafa ya kijumla
2 "l2nt:9" wimbo wa taifa
2 "l2nt:9" utungaji wa kazi
2 "l2nt:9" utamaduni wa watu
2 "l2nt:9" utamaduni wa tanzania
2 "l2nt:9" utamaduni wa binadamu
2 "l2nt:9" ubeti wa mwisho
2 "l2nt:9" uandishi wa kiswahili
2 "l2nt:9" taswira ya mtu
2 "l2nt:9" tamthilia ya kiswahili
2 "l2nt:9" tamthilia ya kiafrika
2 "l2nt:9" tajiriba za jamii
2 "l2nt:9" tajiriba za binadamu
2 "l2nt:9" taashira ya kifo
2 "l2nt:9" riwaya za kiswahili
2 "l2nt:9" riwaya ya kisasa
2 "l2nt:9" nadharia ya udhanaishi
2 "l2nt:9" nadharia ya kimaendeleo
2 "l2nt:9" mwandishi wa kiswahili
2 "l2nt:9" muundo wa jamii
2 "l2nt:9" muktadha wa kijamii
2 "l2nt:9" mtiririko wa vitendo
2 "l2nt:9" mtindo wa kitendawili
2 "l2nt:9" mtazamo wa kimagharibi
2 "l2nt:9" mshairi wa kiswahili
2 "l2nt:9" mhusika wa miraba
2 "l2nt:9" lugha za uzunguni
2 "l2nt:9" lugha za kiasili
2 "l2nt:9" lugha ya taifa
2 "l2nt:9" kaida za tungo
2 "l2nt:9" fasihi ya nchi
2 "l2nt:9" fasihi ya kizungu
2 "l2nt:9" falsafa ya udhanifu
2 "l2nt:9" falsafa ya kiufundi
2 "l2nt:9" drama ya wageni
2 "l2nt:9" dhana ya maisha
1 "l2nt:9" wimbo wa ushindi
1 "l2nt:9" wimbo wa uhusiano

1 "l2nt:9" wimbo wa tamati
1 "l2nt:9" wimbo wa ndege
1 "l2nt:9" wimbo wa matumaini
1 "l2nt:9" wimbo wa kimapinduzi
1 "l2nt:9" wimbo wa kifo.
1 "l2nt:9" wimbo wa fahari
1 "l2nt:9" wasifu wa mtu
1 "l2nt:9" washairi wa kiswahili
1 "l2nt:9" washairi wa kisasa
1 "l2nt:9" washairi wa kimapokeo
1 "l2nt:9" washairi wa jamii
1 "l2nt:9" wanafasihi wa mataifa
1 "l2nt:9" wanafasihi wa kizungu
1 "l2nt:9" wahusika wa tabaka
1 "l2nt:9" wahusika wa mraba
1 "l2nt:9" wahusika wa mjini
1 "l2nt:9" wahusika wa kihalisia
1 "l2nt:9" wahusika wa kibwege
1 "l2nt:9" wahusika wa historia
1 "l2nt:9" wahakiki wa kiswahili
1 "l2nt:9" waandishi wa kiswahili
1 "l2nt:9" waandishi wa kibwege
1 "l2nt:9" waandishi wa kiafrika
1 "l2nt:9" vitendawili vya mtihani
1 "l2nt:9" vitendawili vya jamii
1 "l2nt:9" visa vya kihistoria
1 "l2nt:9" uyakunifu wa mambo
1 "l2nt:9" uwasilishaji wa kauli
1 "l2nt:9" uundaji wa matabaka
1 "l2nt:9" uundaji wa jamii
1 "l2nt:9" utunzi wa kiswahili
1 "l2nt:9" utungaji wa kiswahili
1 "l2nt:9" uthakili wa mfumo
1 "l2nt:9" utenzi wa kiarabu
1 "l2nt:9" utamaduni wa uzungu-ubepari
1 "l2nt:9" utamaduni wa nchi
1 "l2nt:9" utamaduni wa mjini
1 "l2nt:9" utamaduni wa mabwana
1 "l2nt:9" utamaduni wa kizungu
1 "l2nt:9" utamaduni wa kimwambao
1 "l2nt:9" utamaduni wa kikatikati
1 "l2nt:9" utamaduni wa kiasili
1 "l2nt:9" utamaduni wa binaadamu
1 "l2nt:9" usimulizi wa matukio
1 "l2nt:9" ushairi wa mapokeo
1 "l2nt:9" ushairi wa kimapokezi
1 "l2nt:9" ushairi wa kihistoria
1 "l2nt:9" ushairi wa jamii
1 "l2nt:9" usanifu wa tungo
1 "l2nt:9" urasimi wa kiswahili
1 "l2nt:9" urasimi wa jamii
1 "l2nt:9" upeo wa matamano
1 "l2nt:9" upeo wa maendeleo
1 "l2nt:9" ulimbwende wa vitushi
1 "l2nt:9" ukomo wa uwezo
1 "l2nt:9" ujumi wa kazi
1 "l2nt:9" ujumbe wa jumla
1 "l2nt:9" uhakiki wa jumuiya
1 "l2nt:9" ufafanuzi wa wataalamu
1 "l2nt:9" ufafanuzi wa matini
1 "l2nt:9" uchambuzi wa maisha
1 "l2nt:9" uchambuzi wa kitarakimu
1 "l2nt:9" uchambuzi wa kimaielezo

1 "l2nt:9" ubunifu wa mwanafunzi
 1 "l2nt:9" uandishi wa vitabu
 1 "l2nt:9" uandishi wa tanzania
 1 "l2nt:9" uandishi wa magazeti
 1 "l2nt:9" tenzi za kiarabu
 1 "l2nt:9" taswira ya mkombozi
 1 "l2nt:9" taswira ya maisha
 1 "l2nt:9" taswira ya kitu
 1 "l2nt:9" tapo la wanamabadiliko
 1 "l2nt:9" tanzu za mwanzo
 1 "l2nt:9" tamthiliya za redio
 1 "l2nt:9" tamthiliya za kiswahili
 1 "l2nt:9" tamthiliya za kiingereza
 1 "l2nt:9" tamthiliya ya redio
 1 "l2nt:9" tamthilia za kiswahili
 1 "l2nt:9" tamthilia za kibwege
 1 "l2nt:9" tamthilia za kiafrika
 1 "l2nt:9" tamthilia ya kimajaribio
 1 "l2nt:9" tamthilia ya kigiriki
 1 "l2nt:9" tamaduni za watu
 1 "l2nt:9" tamaduni za nchi
 1 "l2nt:9" tamaduni za kizungu
 1 "l2nt:9" tamaduni za binadamu
 1 "l2nt:9" tahakiki ya kauli
 1 "l2nt:9" tafakuri ya hisi
 1 "l2nt:9" taashira za kimazingira
 1 "l2nt:9" taashira za kikati
 1 "l2nt:9" taashira ya takilifu
 1 "l2nt:9" taashira ya safari
 1 "l2nt:9" taashira ya nchi
 1 "l2nt:9" taashira ya mzinga
 1 "l2nt:9" taashira ya mizoga
 1 "l2nt:9" taashira ya matumizi
 1 "l2nt:9" taashira ya maangamizi
 1 "l2nt:9" taashira ya jipu
 1 "l2nt:9" taashira ya chombo
 1 "l2nt:9" simulizi za kiafrika
 1 "l2nt:9" shairi la kimapokeo
 1 "l2nt:9" sanaa za watu
 1 "l2nt:9" sadfa ya mwisho
 1 "l2nt:9" riwaya za kiingereza
 1 "l2nt:9" riwaya za dunia
 1 "l2nt:9" nyimbo za makabila
 1 "l2nt:9" ngoma za kiafrika
 1 "l2nt:9" ngoma ya utawala
 1 "l2nt:9" ngoma ya misituni
 1 "l2nt:9" nadharia za maisha
 1 "l2nt:9" nadharia za elimu
 1 "l2nt:9" nadharia ya udenguuzi
 1 "l2nt:9" nadharia ya kimapinduzi
 1 "l2nt:9" nadharia ya kijamii
 1 "l2nt:9" nadharia ya kibwege
 1 "l2nt:9" nadharia ya kibinafsi
 1 "l2nt:9" nadharia ya jumuiya
 1 "l2nt:9" nadharia ya elimu
 1 "l2nt:9" nadhari ya msomaji
 1 "l2nt:9" mwandishi wa kimapinduzi
 1 "l2nt:9" mwanafalsafa wa kifaransa
 1 "l2nt:9" muundo wa tungo
 1 "l2nt:9" muundo wa sentensi
 1 "l2nt:9" muundo wa maana
 1 "l2nt:9" muundo wa kisémantiki
 1 "l2nt:9" muundo wa kimapokezi

1 "l2nt:9" muktadha wa vijijini
 1 "l2nt:9" muktadha wa usomaji
 1 "l2nt:9" muktadha wa kazi
 1 "l2nt:9" mtiririko wa mawazo
 1 "l2nt:9" mtiririko wa maneno
 1 "l2nt:9" mtiririko wa maelezo
 1 "l2nt:9" mtindo wa kizungu
 1 "l2nt:9" mtindo wa kingano
 1 "l2nt:9" mtindo wa kimapokezi
 1 "l2nt:9" mtazamo wa wasomi
 1 "l2nt:9" mtazamo wa udhanaishi
 1 "l2nt:9" mtazamo wa kirasimi
 1 "l2nt:9" msimamo wa mawazo
 1 "l2nt:9" msemu wa kiingereza
 1 "l2nt:9" mpambaji wa jukwaa
 1 "l2nt:9" mori wa kimapinduzi
 1 "l2nt:9" mlingano wa sauti
 1 "l2nt:9" mkondo wa uzuri
 1 "l2nt:9" mkondo wa ubaya
 1 "l2nt:9" miziki ya kizungu
 1 "l2nt:9" mizani za kiswahili
 1 "l2nt:9" mizani ya uelewa
 1 "l2nt:9" mizani ya sauti
 1 "l2nt:9" mizani ya mwisho
 1 "l2nt:9" mizani ya kisayansi
 1 "l2nt:9" mizani ya kihalisia
 1 "l2nt:9" mitindo ya kikale
 1 "l2nt:9" mitindo ya binadamu
 1 "l2nt:9" mila ya kiafrika
 1 "l2nt:9" mikondo ya mikosi
 1 "l2nt:9" mhusika wa kisaikolojia
 1 "l2nt:9" mhusika wa kimapinduzi
 1 "l2nt:9" mhakiki wa kirusi
 1 "l2nt:9" methali za kikati
 1 "l2nt:9" methali ya kiswahili
 1 "l2nt:9" mdundo wa makelele
 1 "l2nt:9" mchezo wa redio
 1 "l2nt:9" mawanda ya elimu
 1 "l2nt:9" mawaidha ya kiuongozi
 1 "l2nt:9" maudhui za ugandamizaji
 1 "l2nt:9" maudhui ya ukombozi
 1 "l2nt:9" maudhui ya kisiasa
 1 "l2nt:9" maudhui ya dini
 1 "l2nt:9" masimulizi ya vijana
 1 "l2nt:9" masimulizi ya mdomo
 1 "l2nt:9" mashujaa wa kitaifa
 1 "l2nt:9" mashujaa wa kijamii
 1 "l2nt:9" mashairi ya kirasmi
 1 "l2nt:9" mashairi ya kimapokezi
 1 "l2nt:9" mashairi ya kimapokea
 1 "l2nt:9" mashairi ya jamii
 1 "l2nt:9" mapigo ya zana
 1 "l2nt:9" mandhari ya kitajiri
 1 "l2nt:9" mandhari ya kimaskini
 1 "l2nt:9" mandhari ya dunia
 1 "l2nt:9" maigizo ya maisha
 1 "l2nt:9" mada ya ubaguzi
 1 "l2nt:9" mada ya mapenzi
 1 "l2nt:9" maandishi ya taaluma
 1 "l2nt:9" maandishi ya kimagharibi
 1 "l2nt:9" lugha za kimagharibi
 1 "l2nt:9" lugha za kikoloni
 1 "l2nt:9" lugha ya mkoloni

1 "l2nt:9" lugha ya kizungu
 1 "l2nt:9" lugha ya kisasa
 1 "l2nt:9" lugha ya hadhi
 1 "l2nt:9" lahaja ya kimvita
 1 "l2nt:9" lahaja ya kikae
 1 "l2nt:9" kunga za uandikaji
 1 "l2nt:9" kisa cha omari
 1 "l2nt:9" kisa cha nyongo
 1 "l2nt:9" kisa cha mwana
 1 "l2nt:9" kisa cha chuki
 1 "l2nt:9" kipawa cha mtu
 1 "l2nt:9" kilele cha mapinduzi
 1 "l2nt:9" kilele cha maisha
 1 "l2nt:9" kilele cha maasi
 1 "l2nt:9" kilele cha anasa
 1 "l2nt:9" kasumba ya tabia
 1 "l2nt:9" kaida za kimapokezi
 1 "l2nt:9" kaida ya mtu
 1 "l2nt:9" kaida ya maumbile
 1 "l2nt:9" jazanda ya ukanganyifu
 1 "l2nt:9" jazanda ya uchawi
 1 "l2nt:9" jazanda ya mvua
 1 "l2nt:9" jazanda ya mbinu
 1 "l2nt:9" jazanda ya matumizi
 1 "l2nt:9" jazanda ya matumaini
 1 "l2nt:9" jadi ya kimaskini
 1 "l2nt:9" itikadi za dini
 1 "l2nt:9" itikadi ya mtu
 1 "l2nt:9" hulka za binadamu
 1 "l2nt:9" hisia za wanyama
 1 "l2nt:9" hisia za wakoloni
 1 "l2nt:9" hisia za msomaji
 1 "l2nt:9" hisia za kibinadamu
 1 "l2nt:9" hisia za huruma
 1 "l2nt:9" hisia ya utumwa
 1 "l2nt:9" fasihi ya tanzania
 1 "l2nt:9" fasihi ya mababu
 1 "l2nt:9" fasihi ya kisasa
 1 "l2nt:9" fasihi ya kirasimi
 1 "l2nt:9" fasihi ya kimagharibi
 1 "l2nt:9" fasihi ya kikati
 1 "l2nt:9" fasihi ya kiingereza
 1 "l2nt:9" fasihi ya kidhanaishi
 1 "l2nt:9" fasihi ya kiarabu
 1 "l2nt:9" fasihi ya kiafrika
 1 "l2nt:9" fasihi ya awali
 1 "l2nt:9" falsafa ya utu
 1 "l2nt:9" falsafa ya plato
 1 "l2nt:9" dhana za maisha
 1 "l2nt:9" dhana za maana
 1 "l2nt:9" dhana ya utenganishaji
 1 "l2nt:9" arudhi ya kiswahili
 1 "l2nt:9" arudhi ya kiarabu
 1 "l2nt:9" amali ya jamii

LT-N + GEN-CON + ADV "l2nt:10"

2 "l2nt:10" ushairi wa leo
 1 "l2nt:10" ushairi wa sasa
 1 "l2nt:10" ujumbe wa ndani
 1 "l2nt:10" uandishi wa moja_kwa_moja
 1 "l2nt:10" tajiriba za hali
 1 "l2nt:10" riwaya za awali

1 "l2nt:10" nyimbo za kila_siku
 1 "l2nt:10" mtiririko wa moja_kwa_moja
 1 "l2nt:10" mtindo wa siku_hizi
 1 "l2nt:10" msimamo wa karibu
 1 "l2nt:10" mhusika wa kike
 1 "l2nt:10" mashairi ya zamani
 1 "l2nt:10" lugha ya kila_siku
 1 "l2nt:10" fasihi ya kike
 1 "l2nt:10" fasihi ya kesho

N + GEN-CON + LT-N "l2nt:11"

37 "l2nt:11" kazi ya fasihi
 26 "l2nt:11" matumizi ya lugha
 16 "l2nt:11" kazi ya sanaa
 12 "l2nt:11" vipengele vya fani
 8 "l2nt:11" mwisho wa hadithi
 7 "l2nt:11" ufundishaji wa fasihi
 7 "l2nt:11" maana ya shairi
 7 "l2nt:11" jamii ya riwayani
 4 "l2nt:11" uonyeshaji wa tamthiliya
 4 "l2nt:11" somo la fasihi
 4 "l2nt:11" mbinu za uandishi
 4 "l2nt:11" kazi za fasihi
 3 "l2nt:11" vipengele vya fasihi
 3 "l2nt:11" umuhimu wa fasihi
 3 "l2nt:11" ukuzaji wa fasihi
 3 "l2nt:11" ujuzi wa utunzi
 3 "l2nt:11" ufundishaji wa ushairi
 3 "l2nt:11" tafsiri za fasihi
 3 "l2nt:11" mwisho wa tamthilia
 3 "l2nt:11" mwisho wa riwaya
 3 "l2nt:11" mwalimu wa fasihi
 3 "l2nt:11" msingi wa tamthilia
 3 "l2nt:11" msingi wa nadharia
 3 "l2nt:11" msingi wa mashairi
 3 "l2nt:11" maoni ya mwandishi
 3 "l2nt:11" maelezo ya mwandishi
 3 "l2nt:11" kipindi cha urasimi
 3 "l2nt:11" historia ya riwaya
 3 "l2nt:11" akili ya mhusika
 2 "l2nt:11" vitushi vya hadithi
 2 "l2nt:11" vitabu vya fasihi
 2 "l2nt:11" vipengele vya kifasihi
 2 "l2nt:11" vipande vya mishororo
 2 "l2nt:11" uzuri wa nadharia
 2 "l2nt:11" uzuri wa lugha
 2 "l2nt:11" uwezo wa mwandishi
 2 "l2nt:11" utumiaji wa lugha
 2 "l2nt:11" umbo la shairi
 2 "l2nt:11" umbo la riwaya
 2 "l2nt:11" ulimwengu wa tamthilia
 2 "l2nt:11" ukweli wa methali
 2 "l2nt:11" ukuzaji wa lugha
 2 "l2nt:11" ukosefu wa mantiki
 2 "l2nt:11" ufundishaji wa tamthilia
 2 "l2nt:11" ufundishaji wa riwaya
 2 "l2nt:11" uendelezaji wa maudhui
 2 "l2nt:11" uelewaji wa shairi
 2 "l2nt:11" uelewaji wa fasihi
 2 "l2nt:11" ubingwa wa lugha
 2 "l2nt:11" tatizo la mwandishi
 2 "l2nt:11" tabia za wahusika

2 "l2nt:11" somo la fasihi_simulizi
 2 "l2nt:11" mwanzo wa hadithi
 2 "l2nt:11" mwalimu wa ushairi
 2 "l2nt:11" msamiati wa lugha
 2 "l2nt:11" mgongo wa ushairi
 2 "l2nt:11" mfano wa shairi
 2 "l2nt:11" mfano wa mhusika
 2 "l2nt:11" mawazo ya hadhira
 2 "l2nt:11" matumizi ya methali
 2 "l2nt:11" matendo ya wahusika
 2 "l2nt:11" matatizo ya utunzi
 2 "l2nt:11" maoni ya wanafalsafa
 2 "l2nt:11" lengo la shairi
 2 "l2nt:11" kikundi cha drama
 2 "l2nt:11" kijejezi cha ushairi
 2 "l2nt:11" kijejezi cha dhana
 2 "l2nt:11" kiini cha tamthiliya
 2 "l2nt:11" jukumu la mhakiki
 2 "l2nt:11" jukumu la fasihi
 2 "l2nt:11" jina la utungo
 2 "l2nt:11" ishara ya utamaduni
 2 "l2nt:11" imani za jadi
 2 "l2nt:11" funzo la hadithi
 2 "l2nt:11" fikira ya kishairi
 2 "l2nt:11" dosari ya nadharia
 2 "l2nt:11" chombo cha fasihi
 2 "l2nt:11" asili ya ushairi
 1 "l2nt:11" wizara ya utamaduni
 1 "l2nt:11" weledi wa mhusika
 1 "l2nt:11" watetezi wa mashairi
 1 "l2nt:11" watazamaji wa uigizaji
 1 "l2nt:11" wasomaji wa hadithi
 1 "l2nt:11" wapenzi wa fasihi
 1 "l2nt:11" wanamapokezi wa ushairi
 1 "l2nt:11" wakongwe wa fasihi
 1 "l2nt:11" waigaji wa watunzi
 1 "l2nt:11" waasisi wa nadharia
 1 "l2nt:11" vizingiti vya kimila
 1 "l2nt:11" viwango vya maudhui
 1 "l2nt:11" viumbe wa sanaa
 1 "l2nt:11" viumbe wa hadithini
 1 "l2nt:11" viumbe vya hadithini
 1 "l2nt:11" vitushi vya riwaya
 1 "l2nt:11" vitushi vya kisa
 1 "l2nt:11" vitendo vya wahusika
 1 "l2nt:11" vitendo vya mhusika
 1 "l2nt:11" vipengele vya tamthiliya
 1 "l2nt:11" vipengele vya kisanaa
 1 "l2nt:11" vipengele vya kidhahania
 1 "l2nt:11" vipengele vya hadithi
 1 "l2nt:11" vipande vya lugha
 1 "l2nt:11" uzoefu wa masimulizi
 1 "l2nt:11" uzito wa sadfa
 1 "l2nt:11" uzito wa amali
 1 "l2nt:11" uzingatifu wa lugha
 1 "l2nt:11" uwezo wa utunzi
 1 "l2nt:11" uwezo wa uhakiki
 1 "l2nt:11" uwezo wa uchambuzi
 1 "l2nt:11" uwezo wa sanaa
 1 "l2nt:11" uwezo wa fasihi
 1 "l2nt:11" uwanja wa riwaya
 1 "l2nt:11" uwanja wa masimulizi
 1 "l2nt:11" uvunjaji wa kaida

1 "l2nt:11" utumizi wa taswira
 1 "l2nt:11" utumizi wa mizani
 1 "l2nt:11" utumizi wa methali
 1 "l2nt:11" utumizi wa lahaja
 1 "l2nt:11" uthabiti wa tamaduni
 1 "l2nt:11" uteuzi wa lugha
 1 "l2nt:11" uteule wa dhamira
 1 "l2nt:11" uteteaji wa ushairi
 1 "l2nt:11" utendaji wa utungo
 1 "l2nt:11" utendaji wa mashairi
 1 "l2nt:11" utelezi wa nadharia
 1 "l2nt:11" utekwaji wa hisia
 1 "l2nt:11" utasa wa fasihi
 1 "l2nt:11" utaratibu wa wahusika
 1 "l2nt:11" utaratibu wa kimaandishi
 1 "l2nt:11" utaratibu wa hadithi
 1 "l2nt:11" utamu wa ushairi
 1 "l2nt:11" utambuzi wa vipawa
 1 "l2nt:11" usuhubiano wa visa
 1 "l2nt:11" usomaji wa mashairi
 1 "l2nt:11" usahili wa lugha
 1 "l2nt:11" urefu wa mshororo
 1 "l2nt:11" uonyeshaji wa drama
 1 "l2nt:11" unakidi wa lugha
 1 "l2nt:11" umuhimu wa waandishi
 1 "l2nt:11" umuhimu wa taswira
 1 "l2nt:11" umuhimu wa nadharia
 1 "l2nt:11" umuhimu wa mtindo
 1 "l2nt:11" umuhimu wa maudhui
 1 "l2nt:11" umuhimu wa drama
 1 "l2nt:11" umri wa wahusika
 1 "l2nt:11" umbuji wa msanii
 1 "l2nt:11" umbo la utungo
 1 "l2nt:11" umbo la kisanaa
 1 "l2nt:11" umbo la hadithi
 1 "l2nt:11" umalizaji wa tamthilia
 1 "l2nt:11" ulimwengu wa riwaya
 1 "l2nt:11" ulimwengu wa fasihi_simulizi
 1 "l2nt:11" ulimi wa mashairi
 1 "l2nt:11" ulengaji wa wahusika
 1 "l2nt:11" ulazimishaji wa lugha
 1 "l2nt:11" ukweli wa usemi
 1 "l2nt:11" ukwasi wa ushairi
 1 "l2nt:11" ukwasi wa maudhui
 1 "l2nt:11" ukwasi wa lugha
 1 "l2nt:11" ukuzaji wa fasihi_andishi
 1 "l2nt:11" ukuta wa ukale
 1 "l2nt:11" ukumbukaji wa mashairi
 1 "l2nt:11" ukiukaji wa kaida
 1 "l2nt:11" ukinzani wa lugha
 1 "l2nt:11" ukinzani wa fasihi
 1 "l2nt:11" ukati wa hadithi
 1 "l2nt:11" ukanganyifu wa lugha
 1 "l2nt:11" ukamilifu wa mchezo
 1 "l2nt:11" ujuzi wa taswira
 1 "l2nt:11" ujuzi wa mtunzi
 1 "l2nt:11" ujuzi wa lugha
 1 "l2nt:11" ujenzi wa wahusika
 1 "l2nt:11" ujenzi wa utamaduni
 1 "l2nt:11" ujenzi wa uhusika
 1 "l2nt:11" ujenzi wa dhamira
 1 "l2nt:11" ujenzaji wa maudhui
 1 "l2nt:11" ujelezi wa dhana

1 "l2nt:11" ujanja wa mhusika
1 "l2nt:11" uhodari wa masimulizi
1 "l2nt:11" ufupi wa maonyesho
1 "l2nt:11" ufundishaji wa mashairi
1 "l2nt:11" ufundi wa lugha
1 "l2nt:11" ufundi wa kisanana
1 "l2nt:11" ufumbuzi wa ujumbe
1 "l2nt:11" ufumbuzi wa taswira
1 "l2nt:11" ufumbuzi wa mafumbo
1 "l2nt:11" ufinyanzi wa lugha
1 "l2nt:11" ufichuaji wa ujumi
1 "l2nt:11" ufanano wa mashairi
1 "l2nt:11" uenezaji wa falsafa
1 "l2nt:11" uendelezi wa maudhui
1 "l2nt:11" uendelezaji wa wahusika
1 "l2nt:11" uendelezaji wa tanzu
1 "l2nt:11" uelimishaji wa fasihi
1 "l2nt:11" uelezaji wa dhana
1 "l2nt:11" uelewekaji wa maudhui
1 "l2nt:11" uelewa wa lugha
1 "l2nt:11" uchawi wa mwandishi
1 "l2nt:11" uchaguzi wa maudhui
1 "l2nt:11" ubingwa wa utunzi
1 "l2nt:11" ubingwa wa mwandishi
1 "l2nt:11" uandisbi wa mswada
1 "l2nt:11" uafaka wa waigizaji
1 "l2nt:11" tungo za washairi
1 "l2nt:11" tungo za mwandishi
1 "l2nt:11" tungo za mizani
1 "l2nt:11" tungo za kijadi
1 "l2nt:11" tungo za fasihi
1 "l2nt:11" tukio la tajiriba
1 "l2nt:11" tokeo la mgongano
1 "l2nt:11" tofauti ya mashairi
1 "l2nt:11" tatizo la nahau
1 "l2nt:11" tatizo la lugha
1 "l2nt:11" tasnifu ya tamthilia
1 "l2nt:11" taratibu za muundo
1 "l2nt:11" tafsiri za lugha
1 "l2nt:11" tafsiri ya utungo
1 "l2nt:11" tafsiri ya ubeti
1 "l2nt:11" tafsiri ya shairi
1 "l2nt:11" tafsiri ya msimamo
1 "l2nt:11" tafsiri ya kisa
1 "l2nt:11" tafrija ya hadithi-pendezi
1 "l2nt:11" tabia za kisanana
1 "l2nt:11" tabia ya utungaji
1 "l2nt:11" tabia ya upembuzi
1 "l2nt:11" swala la utamaduni
1 "l2nt:11" swala la mgogoro
1 "l2nt:11" swala la fani
1 "l2nt:11" sura ya vichekesho
1 "l2nt:11" silisili za mizani
1 "l2nt:11" silisili za mila
1 "l2nt:11" sifa za wahusika
1 "l2nt:11" sifa za utenzi
1 "l2nt:11" sifa za urasimi
1 "l2nt:11" shina la utamaduni
1 "l2nt:11" sheria za utunzi
1 "l2nt:11" sheria za mishororo
1 "l2nt:11" sheria ya iktisadi
1 "l2nt:11" sherehe za kitamaduni
1 "l2nt:11" sherehe za jando

1 "l2nt:11" shabaha ya shairi
 1 "l2nt:11" shabaha ya riwaya
 1 "l2nt:11" sehemu ya utamaduni
 1 "l2nt:11" sehemu ya tamthilia
 1 "l2nt:11" sehemu ya muundo
 1 "l2nt:11" sehemu ya mchezo
 1 "l2nt:11" sehemu ya maudhui
 1 "l2nt:11" sehemu ya lugha
 1 "l2nt:11" sehemu ya fani
 1 "l2nt:11" sauti ya uwasilishaji
 1 "l2nt:11" sauti ya shairi
 1 "l2nt:11" pingu za utamaduni
 1 "l2nt:11" nyenzo ya uandishi
 1 "l2nt:11" nyayo za tamthiliya
 1 "l2nt:11" nusu ya riwaya
 1 "l2nt:11" njia za usimulizi
 1 "l2nt:11" njia ya ushairi
 1 "l2nt:11" njia ya usanifu
 1 "l2nt:11" njia ya nyimbo
 1 "l2nt:11" njia ya mvuto
 1 "l2nt:11" njia ya mashairi
 1 "l2nt:11" njia ya mafumbo
 1 "l2nt:11" njia ya maandishi
 1 "l2nt:11" njia ya kishairi
 1 "l2nt:11" nguzo za riwaya
 1 "l2nt:11" nguzo za muundo
 1 "l2nt:11" nguvu za taswira
 1 "l2nt:11" nchi ya riwayani
 1 "l2nt:11" nafsi za hadhira
 1 "l2nt:11" nafsi ya mwandishi
 1 "l2nt:11" nafsi ya hadhira
 1 "l2nt:11" nafasi ya wasanii
 1 "l2nt:11" nafasi ya msimuliaji
 1 "l2nt:11" nafasi ya mhusika
 1 "l2nt:11" nafasi ya fasihi
 1 "l2nt:11" mwisho wa ubeti
 1 "l2nt:11" mwisho wa shairi
 1 "l2nt:11" mwisho wa mshairi
 1 "l2nt:11" mwisho wa kisa
 1 "l2nt:11" mwisho la mshororo
 1 "l2nt:11" mwisho cha fani
 1 "l2nt:11" mweledi wa lugha
 1 "l2nt:11" mweledi wa fasihi
 1 "l2nt:11" mwavuli wa tamthilia
 1 "l2nt:11" mwasilishaji wa hisia
 1 "l2nt:11" mwanzo wa tajiriba
 1 "l2nt:11" mwanzo wa riwaya
 1 "l2nt:11" mwanzo wa mishororo
 1 "l2nt:11" mwanzo wa maudhui
 1 "l2nt:11" mwanafunzi wa fasihi
 1 "l2nt:11" mwalimu wa tamthilia
 1 "l2nt:11" mwalimu wa lugha
 1 "l2nt:11" muhtasari wa ushairi
 1 "l2nt:11" muhtasari wa masimulizi
 1 "l2nt:11" muhtasari wa falsafa
 1 "l2nt:11" mtirikio wa hadithi
 1 "l2nt:11" mtawala wa hadithini
 1 "l2nt:11" mstari wa maandishi
 1 "l2nt:11" msomaji wa riwaya
 1 "l2nt:11" msingi wa urasimi
 1 "l2nt:11" msingi wa riwaya
 1 "l2nt:11" msingi wa itikadi
 1 "l2nt:11" mseto wa ngano

1 "l2nt:11" mseto wa mila
1 "l2nt:11" msamiati wa shairi
1 "l2nt:11" mpangilio wa tamthilia
1 "l2nt:11" mpangilio wa lugha
1 "l2nt:11" mkururo wa vichekesho
1 "l2nt:11" mkurugenzi wa maigizo
1 "l2nt:11" mkabala wa uyakinifu
1 "l2nt:11" mjadala wa wahusika
1 "l2nt:11" mizizi ya lugha
1 "l2nt:11" mivutano ya kiiktisadi
1 "l2nt:11" miundo ya tamthilia
1 "l2nt:11" mitihani ya vitendawili
1 "l2nt:11" misukumo ya migongano
1 "l2nt:11" mistari ya ubeti
1 "l2nt:11" misingi-teule ya kishairi
1 "l2nt:11" misingi ya utamaduni
1 "l2nt:11" misingi ya ushairi
1 "l2nt:11" misingi ya ukale
1 "l2nt:11" misingi ya sanaa
1 "l2nt:11" misingi ya mwandishi
1 "l2nt:11" misingi ya fasihi simulizi
1 "l2nt:11" mipaka ya utamaduni
1 "l2nt:11" mipaka ya ushairi
1 "l2nt:11" mipaka ya tamaduni
1 "l2nt:11" mipaka ya tajiriba
1 "l2nt:11" mipaka ya riwaya
1 "l2nt:11" mipaka wa mashairi
1 "l2nt:11" mifupa ya hadithi
1 "l2nt:11" mifano ya uhakiki
1 "l2nt:11" mifano ya tahakiki
1 "l2nt:11" mifano ya ngano
1 "l2nt:11" mienendo ya wahusika
1 "l2nt:11" mhimili wa maudhui
1 "l2nt:11" mfumo wa iktisadi
1 "l2nt:11" mfumko wa dhana
1 "l2nt:11" mfuasi wa lugha
1 "l2nt:11" mfano wa utenzi
1 "l2nt:11" mfano wa sanaa
1 "l2nt:11" mfano wa riwaya
1 "l2nt:11" mfano wa ngonjera
1 "l2nt:11" mfano wa fasihi
1 "l2nt:11" mbinu za utambaji
1 "l2nt:11" mbinu za kifasihi
1 "l2nt:11" mbinu ya vichekesho
1 "l2nt:11" mbinu ya utunzi
1 "l2nt:11" mbinu ya kisanaa
1 "l2nt:11" mbadiliko wa hadithi
1 "l2nt:11" mazungumzo ya kitamthilia
1 "l2nt:11" mazingira ya ubunifu
1 "l2nt:11" mazingira ya msanii
1 "l2nt:11" mawazo ya mwandishi
1 "l2nt:11" mawazo ya mtunzi
1 "l2nt:11" mawazo ya kishairi
1 "l2nt:11" maumivu ya kifalsafa
1 "l2nt:11" matumizi ya usanifu
1 "l2nt:11" matumizi ya tamathali
1 "l2nt:11" matumizi ya taashira
1 "l2nt:11" matumizi ya kaida
1 "l2nt:11" matumizi ya jazanda
1 "l2nt:11" matatizo ya watunzi
1 "l2nt:11" masomo ya fasihi
1 "l2nt:11" mashiko ya falsafa
1 "l2nt:11" mashaka ya lugha

1 "l2nt:11" mapokezi ya hadhira
1 "l2nt:11" mapato ya wasimulizi
1 "l2nt:11" maoni ya wanamapokeo
1 "l2nt:11" maneno ya nyimbo
1 "l2nt:11" maneno ya mashairi
1 "l2nt:11" maneno ya lugha
1 "l2nt:11" maneno ya kilahaja
1 "l2nt:11" mambo ya mila
1 "l2nt:11" mambo ya kijadi
1 "l2nt:11" mama wa nadharia
1 "l2nt:11" malengo ya miviga
1 "l2nt:11" malengo ya hadithi
1 "l2nt:11" majitapo ya wahusika
1 "l2nt:11" majibu ya vitendawili
1 "l2nt:11" maisha ya wahusika
1 "l2nt:11" mahitimisho ya maono
1 "l2nt:11" mahitaji ya ushairi
1 "l2nt:11" mafunzo ya ngano
1 "l2nt:11" mafunzo ya hadithi
1 "l2nt:11" mafundi wa nadharia
1 "l2nt:11" maendeleo ya mchezo
1 "l2nt:11" maelezo ya taashira
1 "l2nt:11" maelezo ya mtindo
1 "l2nt:11" maelezo ya msanii
1 "l2nt:11" maelezo ya kidhahania
1 "l2nt:11" maelezo ya jazanda
1 "l2nt:11" madhumuni ya vichekesho
1 "l2nt:11" madhumuni ya msanii
1 "l2nt:11" madhumuni ya drama
1 "l2nt:11" mabongo ya wasikilizaji
1 "l2nt:11" mabadiliko ya sanaa
1 "l2nt:11" mabadiliko ya mhusika
1 "l2nt:11" maana ya ushairi
1 "l2nt:11" maana ya usanifu
1 "l2nt:11" maana ya tamthiliya
1 "l2nt:11" maana ya riwaya
1 "l2nt:11" maana ya methali
1 "l2nt:11" maana ya maudhui
1 "l2nt:11" maana ya fasihi
1 "l2nt:11" maana ya falsafa
1 "l2nt:11" maadui wa lugha
1 "l2nt:11" lengo la waandishi
1 "l2nt:11" lengo la vichekesho
1 "l2nt:11" lengo la sanaa
1 "l2nt:11" lengo la mizani
1 "l2nt:11" lengo la miviga
1 "l2nt:11" lengo la mapitio
1 "l2nt:11" lengo la fasihi
1 "l2nt:11" kurunzi ya wahusika
1 "l2nt:11" kurunzi ya msimulizi
1 "l2nt:11" kongamano la ubunifu
1 "l2nt:11" kongamano la fani
1 "l2nt:11" kizingiti cha mhusika
1 "l2nt:11" kiwango cha usanifu
1 "l2nt:11" kiwango cha ubunifu
1 "l2nt:11" kiwango cha kisananaa
1 "l2nt:11" kiwango cha fasihi
1 "l2nt:11" kiwango cha fani
1 "l2nt:11" kitabu cha fasihi
1 "l2nt:11" kipengele cha hadithi
1 "l2nt:11" kipengele cha fasihi
1 "l2nt:11" kipengele cha fani
1 "l2nt:11" kijejezi cha utanzu

1 "l2nt:11" kijejezi cha tashihisi
 1 "l2nt:11" kifungu cha utungo
 1 "l2nt:11" kielelezo cha mtunzi
 1 "l2nt:11" kianzio cha mshororo
 1 "l2nt:11" kazi za wasanii
 1 "l2nt:11" kazi za mshairi
 1 "l2nt:11" kazi za kisanaa
 1 "l2nt:11" kazi ya ubunifu
 1 "l2nt:11" kazi ya uandishi
 1 "l2nt:11" kazi ya tamthilia
 1 "l2nt:11" kazi ya mwandishi
 1 "l2nt:11" kazi ya kisanaa
 1 "l2nt:11" kazi ya kifasihi
 1 "l2nt:11" kauli za lahaja
 1 "l2nt:11" kauli ya mwandishi
 1 "l2nt:11" kasoro ya msemu
 1 "l2nt:11" kanuni za ushairi
 1 "l2nt:11" kanuni za kishairi
 1 "l2nt:11" kanuni za kijadi
 1 "l2nt:11" jumbe za riwaya
 1 "l2nt:11" jumbe za ngano
 1 "l2nt:11" jukwaa la tamthilia
 1 "l2nt:11" jukwaa la michezo
 1 "l2nt:11" jukwaa la mguso
 1 "l2nt:11" jukumu la mshairi
 1 "l2nt:11" juhudi ya mwandishi
 1 "l2nt:11" jina la lakabu
 1 "l2nt:11" jamii ya hadithini
 1 "l2nt:11" ithibati ya mtafaruku
 1 "l2nt:11" imani za kitamaduni
 1 "l2nt:11" imani za kijadi
 1 "l2nt:11" imani ya uyaikinifu
 1 "l2nt:11" imani ya kifalsafa
 1 "l2nt:11" idara ya sanaa
 1 "l2nt:11" historia za ushairi
 1 "l2nt:11" historia za lugha
 1 "l2nt:11" historia ya tamthiliya
 1 "l2nt:11" historia ya fasihi
 1 "l2nt:11" historia ya drama
 1 "l2nt:11" heshima ya lugha
 1 "l2nt:11" heba ya mwandishi
 1 "l2nt:11" hatua za uchambuzi
 1 "l2nt:11" hatua za uandishi
 1 "l2nt:11" hatima ya kilele
 1 "l2nt:11" hali ya ubwege
 1 "l2nt:11" hali ya migogoro
 1 "l2nt:11" hali ya fasihi
 1 "l2nt:11" haja ya mwandishi
 1 "l2nt:11" hadhari ya lugha
 1 "l2nt:11" gogo la hadithi
 1 "l2nt:11" funzo la hekaya
 1 "l2nt:11" fikra ya kishairi
 1 "l2nt:11" fikra ya kidhahania
 1 "l2nt:11" eneo la usimulizi
 1 "l2nt:11" eneo la masimulizi
 1 "l2nt:11" elimu ya kijadi
 1 "l2nt:11" dunia ya lugha
 1 "l2nt:11" dira ya mwandishi
 1 "l2nt:11" darasa la riwaya
 1 "l2nt:11" darasa la miviga
 1 "l2nt:11" darasa la mashairi
 1 "l2nt:11" darasa la lugha
 1 "l2nt:11" darasa la fasihi_simulizi

1 "l2nt:11" darasa la fasihi
1 "l2nt:11" dai la washairi
1 "l2nt:11" chemichemi ya fasihi
1 "l2nt:11" chemichemi ya dhana
1 "l2nt:11" chembe ya fasihi
1 "l2nt:11" chanzo cha fasihi
1 "l2nt:11" bwawa la fasihi simulizi
1 "l2nt:11" busara ya methali
1 "l2nt:11" bingwa wa lugha
1 "l2nt:11" azma ya washairi
1 "l2nt:11" athari za utamaduni
1 "l2nt:11" athari ya nadharia
1 "l2nt:11" athari ya lugha
1 "l2nt:11" athari ya fasihi
1 "l2nt:11" asilimia ya ubunifu
1 "l2nt:11" asili ya utumbuizo
1 "l2nt:11" asili ya mashairi
1 "l2nt:11" akilini mwa wasikilizaji
1 "l2nt:11" akili za wasikilizaji
1 "l2nt:11" akili ya mwandishi
1 "l2nt:11" ainaaina za uhusika
1 "l2nt:11" ainaaina ya lugha
1 "l2nt:11" adhama ya lugha
1 "l2nt:11" adabu ya lugha

LT-V + GEN-CON + LT-N "l2vt:12"

1 "l2vt:12" kuimba kwa mdundo

INF + GEN-CON + LT-N "l2vt:14"

2 "l2vt:14" kueleweka kwa maudhui
1 "l2vt:14" kutiririka kwa muwala
1 "l2vt:14" kutamalaki kwa utamaduni
1 "l2vt:14" kupitia kwa msimulizi
1 "l2vt:14" kupitia kwa mhusika
1 "l2vt:14" kunena kwa lugha
1 "l2vt:14" kukua kwa fasihi
1 "l2vt:14" kukomaa kwa washairi
1 "l2vt:14" kuimbika kwa shairi
1 "l2vt:14" kuibuka kwa ushairi
1 "l2vt:14" kuendelea kwa sanaa
1 "l2vt:14" kuelezea kwa uyaakinifu
1 "l2vt:14" kubuniwa kwa lugha
1 "l2vt:14" kubadilikabadilika kwa lugha
1 "l2vt:14" kubadilika kwa ushairi
1 "l2vt:14" kubadilika kwa mhusika

LT-N + PREP + LT-N "l2nt:15"

4 "l2nt:15" mshororo baada_ya mshororo
3 "l2nt:15" tamthilia katika tamthilia
2 "l2nt:15" ubeti kwa ubeti
2 "l2nt:15" tamthilia ndani_ya tamthilia
2 "l2nt:15" msimulizi katika riwaya
2 "l2nt:15" lugha bila fasihi
2 "l2nt:15" fani katika ngano
1 "l2nt:15" utunzi katika lugha
1 "l2nt:15" usimulizi katika ngano
1 "l2nt:15" ushairi ndani_ya tamthilia
1 "l2nt:15" upelelezi katika fasihi
1 "l2nt:15" umalenga katika ushairi
1 "l2nt:15" ujumbe katika ngano

1 "l2nt:15" uhusika katika riwaya
1 "l2nt:15" shairi katika hadithi
1 "l2nt:15" semi katika shairi
1 "l2nt:15" riwayani katika mafumbo
1 "l2nt:15" mwandishi hadi mwandishi
1 "l2nt:15" mtiririko katika tamthilia
1 "l2nt:15" mtindo katika ushairi
1 "l2nt:15" msimulizi pamoja_na wasikilizaji
1 "l2nt:15" msimulizi katika ngano
1 "l2nt:15" mshororo kwa mshororo
1 "l2nt:15" mizani katika mashairi
1 "l2nt:15" mishororo katika mashairi
1 "l2nt:15" mhusika katika tamthilia
1 "l2nt:15" mhusika katika riwaya
1 "l2nt:15" mhusika katika hadithi
1 "l2nt:15" mgogoro katika ushairi
1 "l2nt:15" methali katika uandishi
1 "l2nt:15" mchezo ndani_ya mchezo
1 "l2nt:15" majigambo katika utunzi
1 "l2nt:15" lugha katika shairi
1 "l2nt:15" fasihi katika utamaduni
1 "l2nt:15" fasihi katika usemi
1 "l2nt:15" fani katika tamthilia
1 "l2nt:15" fani katika fasihi_simulizi

N + PREP + LT-N "l2nt:16"

2 "l2nt:16" kitendo katika tamthiliya
1 "l2nt:16" wanafunzi katika vitendawili
1 "l2nt:16" walibora katika riwaya
1 "l2nt:16" wakwasi katika riwaya
1 "l2nt:16" vimelea katika tamthilia
1 "l2nt:16" vijitamthilia ndani_ya tamthilia-mwavuli
1 "l2nt:16" urogi katika tamthilia
1 "l2nt:16" upya katika fani
1 "l2nt:16" ugeni katika ushairi
1 "l2nt:16" udhaifu kwa_ajili_ya umenke
1 "l2nt:16" uchunguzi katika ushairi
1 "l2nt:16" uchawi katika tamthilia
1 "l2nt:16" ubarakala katika riwaya
1 "l2nt:16" tofauti katika shairi
1 "l2nt:16" tofauti kati_ya vichekesho
1 "l2nt:16" tofauti baina_ya riwaya
1 "l2nt:16" neno katika shairi
1 "l2nt:16" mwelekeo kuhusu maudhui
1 "l2nt:16" mwangaza kuhusu hulka
1 "l2nt:16" mwanga kuhusu dhamira
1 "l2nt:16" mvutano baina_ya utamaduni
1 "l2nt:16" mushkili katika urari
1 "l2nt:16" msikilizaji katika masimulizi
1 "l2nt:16" mshiriki katika kichekesho
1 "l2nt:16" mpango katika ubeti
1 "l2nt:16" mpaka baina_ya visa
1 "l2nt:16" mitihani katika vitendawili
1 "l2nt:16" mazungumzo katika uandishi
1 "l2nt:16" mawasiliano katika sanaa
1 "l2nt:16" matukio katika hadithi
1 "l2nt:16" maneno nje_ya muktadha
1 "l2nt:16" maneno katika mishororo
1 "l2nt:16" mambo katika tamthilia
1 "l2nt:16" majivuno katika riwaya
1 "l2nt:16" majadiliano baada_ya uwasilishaji
1 "l2nt:16" maingiliano baina_ya nadharia

1 "l2nt:16" maingiliano baina_ya wahusika
 1 "l2nt:16" maingiliano baina_ya mada
 1 "l2nt:16" maelezo katika uandishi
 1 "l2nt:16" maelezo katika fasihi
 1 "l2nt:16" mabadiliko katika maudhui
 1 "l2nt:16" maana kutokana_na muktadha
 1 "l2nt:16" ukale katika tamthilia
 1 "l2nt:16" kasi katika masimulizi
 1 "l2nt:16" chuku katika masimulizi

LT-N + PREP + N "l2nt:17"

2 "l2nt:17" muundo katika kazi
 1 "l2nt:17" ushairi katika mataifa
 1 "l2nt:17" ujumbe kuhusu matukio
 1 "l2nt:17" ujumbe katika kazi
 1 "l2nt:17" uandishi katika shule
 1 "l2nt:17" uandishi bila vitendo
 1 "l2nt:17" tamthiliya katika nchi
 1 "l2nt:17" tamthiliya katika mapana
 1 "l2nt:17" tamthilia kuhusu mivutano
 1 "l2nt:17" taathira kwa wanajamii
 1 "l2nt:17" shairi katika upana
 1 "l2nt:17" sanaa kwa ujumla
 1 "l2nt:17" riwaya katika vitabu
 1 "l2nt:17" nadharia kwa jumla
 1 "l2nt:17" mwandishi katika ubunaji
 1 "l2nt:17" mtafaruku katika jamii
 1 "l2nt:17" mtafaruku baina_ya vijana
 1 "l2nt:17" miviga katika jamii
 1 "l2nt:17" migongano katika matendo
 1 "l2nt:17" mguso kwa jamii
 1 "l2nt:17" mgogoro baina_ya binadamu
 1 "l2nt:17" mchezo baina_ya binadamu
 1 "l2nt:17" mashairi katika mataifa
 1 "l2nt:17" majibizano katika mazungumzo
 1 "l2nt:17" itikadi katika uelezaji

INF + PREP + LT-N "l2vt:20"

1 "l2vt:20" kuzungumza kuhusu ubwege
1 "l2vt:20" kutofautiana katika lafudhi
1 "l2vt:20" kutendeka mbele_ya hadhira
1 "l2vt:20" kutambulikana kwa_sababu_ya ukale
1 "l2vt:20" kusema katika riwaya
1 "l2vt:20" kurudi katika uasili
1 "l2vt:20" kuonyesha katika onyesho
1 "l2vt:20" kukua kwa fasihi
1 "l2vt:20" kukamilika kwa maudhui
1 "l2vt:20" kujulikana kwa lakabu
1 "l2vt:20" kujibizana kuhusu mashairi
1 "l2vt:20" kujadili kuhusu jazanda

LT-N + GEN-CON + INF "l2nt:22"

5 "l2nt:22" mchezo wa kuigiza
2 "l2nt:22" hadithi ya kupendeza
1 "l2nt:22" wimbo wa kusifu
1 "l2nt:22" wahusika wa kuaminika
1 "l2nt:22" utungo wa kushajiishana
1 "l2nt:22" uhusika wa kuaminika
1 "l2nt:22" taswira ya kuogofya
1 "l2nt:22" sanaa ya kuchora
1 "l2nt:22" muktadha wa kuchongwa
1 "l2nt:22" michezo ya kuigiza
1 "l2nt:22" mahadhi ya kutilika
1 "l2nt:22" hadithi ya kusikitisha
1 "l2nt:22" dhana ya kustaladhi
1 "l2nt:22" dhamira za kupotoka
1 "l2nt:22" desturi ya kubadilika

N + GEN-CON + LT-V "l2nt:23"

7 "l2nt:23" sheria za kutunga
3 "l2nt:23" uhuru wa kutunga
2 "l2nt:23" mbinu za kutunga
2 "l2nt:23" kazi ya kubuni
1 "l2nt:23" ujuzi wa kutunga
1 "l2nt:23" mighani ya kubuni
1 "l2nt:23" mbinu ya kutunga
1 "l2nt:23" matendo ya kutunga
1 "l2nt:23" kiumbe wa kubuni
1 "l2nt:23" kipaji cha kubuni

LT-N + POSS + LT-N "l2nt:27"

1 "l2nt:27" utungo wenye maudhui
1 "l2nt:27" ushairi wenye mizani
1 "l2nt:27" nadharia yenye uyakinifu
1 "l2nt:27" mhusika mwenye upeo
1 "l2nt:27" mchezo wenye mtiririko
1 "l2nt:27" maandishi yenye maudhui
1 "l2nt:27" lugha yenye taswira
1 "l2nt:27" lugha yenye mdundo
1 "l2nt:27" hadithi yenye mtiririko

LT-N + POSS + N "l2nt:28"

1 "l2nt:28" wahusika wenye udhaifu
1 "l2nt:28" wahusika wenye nafsi
1 "l2nt:28" wahusika wake maneno

1 "l2nt:28" ushairi wenye kanuni
 1 "l2nt:28" nyimbo zenye mistari
 1 "l2nt:28" mtiririko wenye mzuano
 1 "l2nt:28" msimamo wenye mashiko
 1 "l2nt:28" mhusika mwenye akili
 1 "l2nt:28" maudhui yenye umuhimu
 1 "l2nt:28" mashairi yenye vipokeo
 1 "l2nt:28" lugha zenye majukumu
 1 "l2nt:28" hadithi yenye mchangamano
 1 "l2nt:28" fumbo lenye maana
 1 "l2nt:28" fasihi yenye maana

N + POSS + LT-N "l2nt:29"

1 "l2nt:29" wanawake wenye itikadi
 1 "l2nt:29" wanaume wenye itikadi
 1 "l2nt:29" umbo lenye mizani
 1 "l2nt:29" mtu mwenye msimamo
 1 "l2nt:29" mbinu zenye mantiki
 1 "l2nt:29" kazi yenye taharuki
 1 "l2nt:29" kawafi yenye mizani
 1 "l2nt:29" binadamu mwenye ubunifu

LT-V + ADV "l2nt:30"

1 "l2nt:30" kutunga hapo_kwa_hapo
 1 "l2nt:30" kuimba pamoja

Length 3 terms

LT-N + A-INFL + GEN-CON + N "l3nt:1"

1 "l3nt:1" visa virefu vya kihistoria
 1 "l3nt:1" muundo mzima wa matini
 1 "l3nt:1" muktadha mpana wa uimbaji
 1 "l3nt:1" kigezo kikuu cha uenyeji
 1 "l3nt:1" fasihi mpya ya kiswahili

LT-N + A-INFL + GEN-CON + LT-N "l3nt:2"

1 "l3nt:2" utanzu mmojawapo wa fasihi
 1 "l3nt:2" ushairi teule wa nyimbo
 1 "l3nt:2" taswira pana ya shairi
 1 "l3nt:2" riwaya nzuri za waandishi
 1 "l3nt:2" muundo mzima wa utunzi
 1 "l3nt:2" falsafa pana ya mwandishi
 1 "l3nt:2" dhana kuu ya fasihi
 1 "l3nt:2" dhamira kuu ya ngano
 1 "l3nt:2" dhamira kuu ya mwandishi

N + A-INFL + GEN-CON + LT-N "l3nt:3"

1 "l3nt:3" vipengele vikuu vya fani
 1 "l3nt:3" umilisi mkubwa wa lugha
 1 "l3nt:3" umbo zima la riwaya
 1 "l3nt:3" ufundishaji mwema wa fasihi
 1 "l3nt:3" tasnifu kuu ya sanaa
 1 "l3nt:3" sifa kuu za fasihi
 1 "l3nt:3" sifa kubwa ya ushairi
 1 "l3nt:3" nyenzo kuu za usimulizi
 1 "l3nt:3" misingi mipya ya upelelezi
 1 "l3nt:3" mihimili mikuu ya tamthilia
 1 "l3nt:3" matumizi mapevu ya lugha

1 "l3nt:3" lengo kuu la mwandishi
 1 "l3nt:3" kazi nzuri ya sanaa
 1 "l3nt:3" kazi nzima ya fasihi
 1 "l3nt:3" historia fupi ya riwaya

LT-N + GEN-CON + LT-N + GEN-CON + ADV "l3nt:7"

1 "l3nt:7" utunzi wa tamthiliya wa hapo_kwa_hapo

LT-N + GEN-CON + N + GEN-CON + ADV "l3nt:8"

2 "l3nt:8" taharuki ya hali ya juu
 1 "l3nt:8" ubunifu wa hali ya juu
 1 "l3nt:8" mizani za kiswahili cha leo

N + GEN-CON + LT-N + GEN-CON + ADV "l3nt:9"

1 "l3nt:9" waigaji wa watunzi wa zamani
 1 "l3nt:9" hali ya migogoro ya leo

N + GEN-CON + LT-N + GEN-CON + N "l3nt:10"

2 "l3nt:10" ufundishaji wa ushairi wa kiswahili
 2 "l3nt:10" historia ya riwaya ya kiswahili
 1 "l3nt:10" wizara ya utamaduni wa taifa
 1 "l3nt:10" wanamapokezi wa ushairi wa kiswahili
 1 "l3nt:10" uzito wa amali za jamii
 1 "l3nt:10" uthabiti wa tamaduni za kiafrika
 1 "l3nt:10" uteteaji wa ushairi wa kimapinduzi
 1 "l3nt:10" ulazimishaji wa lugha za kimagharibi
 1 "l3nt:10" ukuzaji wa lugha ya kiswahili
 1 "l3nt:10" ukuzaji wa fasihi ya kiswahili
 1 "l3nt:10" ufundishaji wa riwaya ya kiswahili
 1 "l3nt:10" ufundi wa lugha ya mkato
 1 "l3nt:10" ufichuaji wa ujumi wa kazi
 1 "l3nt:10" uenezaji wa falsafa ya ujamaa
 1 "l3nt:10" uendelezaji wa tanzu za fasihi-andishi
 1 "l3nt:10" uchunguzi wa ushairi wa kiswahili
 1 "l3nt:10" ubingwa wa lugha ya kiswahili
 1 "l3nt:10" tafsiri ya utungo wa kiarabu
 1 "l3nt:10" shina la utamaduni wa kiafrika
 1 "l3nt:10" sherehe za kitamaduni za jamii
 1 "l3nt:10" sehemu ya utamaduni wa jamii
 1 "l3nt:10" ploti ya tamthilia ya kibwege
 1 "l3nt:10" mweledi wa lugha ya kiswahili
 1 "l3nt:10" mwalimu wa lugha ya kiswahili
 1 "l3nt:10" msingi wa riwaya ya kisasa
 1 "l3nt:10" msamiati wa lugha ya kiswahili
 1 "l3nt:10" mpangilio wa lugha ya mkato
 1 "l3nt:10" mkururo wa vichekesho vya maskani
 1 "l3nt:10" miundo ya tamthilia za mataifa
 1 "l3nt:10" misingi ya utamaduni wa kizungu
 1 "l3nt:10" mhimili wa maudhui ya elimu
 1 "l3nt:10" maelezo ya taashira ya mzinga
 1 "l3nt:10" maana ya usanifu wa kazi
 1 "l3nt:10" maana ya shairi la kiswahili
 1 "l3nt:10" maana ya falsafa ya kijumla
 1 "l3nt:10" lengo la waandishi wa kibwege
 1 "l3nt:10" kitabu cha fasihi ya kiswahili
 1 "l3nt:10" kielelezo cha mtunzi wa kisasa
 1 "l3nt:10" kazi ya tamthilia ya kigiriki
 1 "l3nt:10" kazi ya sanaa ya kirasimi
 1 "l3nt:10" kanuni za ushairi wa kiswahili

1 "13nt:10" jukumu la mshairi wa kiswahili
1 "13nt:10" ishara ya utamaduni wa kizungu
1 "13nt:10" imani ya uyakinifu wa kijamii
1 "13nt:10" historia ya fasihi ya kiswahili
1 "13nt:10" dosari ya nadharia ya utabia
1 "13nt:10" dosari ya nadharia ya kihafidhina
1 "13nt:10" bingwa wa lugha ya kiswahili
1 "13nt:10" athari ya lugha ya kiingereza
1 "13nt:10" asili ya ushairi wa kiswahili
1 "13nt:10" asili ya mashairi wa kiswahili

N + GEN-CON + N + GEN-CON + LT-N "13nt:11"

2 "13nt:11" uti wa mgongo wa ushairi
1 "13nt:11" watungaji wa kazi za sanaa
1 "13nt:11" wakusanyaji wa kazi za mshairi
1 "13nt:11" uti wa mgongo wa uhakiki
1 "13nt:11" uti wa mgongo wa tamthilia
1 "13nt:11" uti wa mgongo wa fasihi_simulizi
1 "13nt:11" uti wa mgongo wa fasihi
1 "13nt:11" usambazaji wa ujuzi wa utunzi
1 "13nt:11" upevu wa tungo za kijadi
1 "13nt:11" upana wa kazi ya fasihi
1 "13nt:11" unakidi wa kazi za sanaa
1 "13nt:11" ukuzaji wa kazi ya sanaa
1 "13nt:11" ujuzi wa mbinu za uandishi
1 "13nt:11" ubaya wa tabia za wahusika
1 "13nt:11" sura ya mwanzo ya riwaya
1 "13nt:11" sifa za ajabu za sanaa
1 "13nt:11" shabaha ya kazi ya fasihi
1 "13nt:11" picha ya jumla ya shairi
1 "13nt:11" nguvu za mbinu za kifasihi
1 "13nt:11" mwisho wa kazi ya fasihi
1 "13nt:11" muhtasari wa tasnifu ya tamthilia
1 "13nt:11" msomaji wa kazi ya fasihi
1 "13nt:11" msingi wa ujenzi wa utamaduni
1 "13nt:11" msikilizaji wa kazi ya fasihi
1 "13nt:11" mshehenezo wa vipengele vya kifasihi
1 "13nt:11" mpangilio wa mistari ya ubeti
1 "13nt:11" misingi ya kazi za sanaa
1 "13nt:11" mchango wa kazi za fasihi
1 "13nt:11" mazingira ya mseto wa mila
1 "13nt:11" malengo ya imani za jadi
1 "13nt:11" malengo ya elimu ya kijadi
1 "13nt:11" mabadiliko ya umbo la shairi
1 "13nt:11" maana ya matumizi ya lugha
1 "13nt:11" maadili ya wakongwe wa fasihi
1 "13nt:11" kiwakilishi cha imani za kijadi
1 "13nt:11" kivuli cha mwavuli wa tamthilia
1 "13nt:11" kitovu cha kazi ya sanaa
1 "13nt:11" kipindi cha utasa wa fasihi
1 "13nt:11" anwani ya kazi ya fasihi

N + GEN-CON + LT-N + GEN-CON + LT-N "13nt:12"

1 "13nt:12" uvunjaji wa kaida za lugha
1 "13nt:12" ujuzi wa utunzi wa tamthiliya
1 "13nt:12" ubingwa wa lugha wa msimulizi
1 "13nt:12" sheria za utunzi wa mashairi
1 "13nt:12" mwanzo wa mishororo ya ubeti
1 "13nt:12" msingi wa urasimi wa ushairi
1 "13nt:12" mjadala wa wahusika wa riwaya
1 "13nt:12" mifano ya uhakiki wa tamthilia

1 "l3nt:12" mbinu za uandishi wa tamthiliya
 1 "l3nt:12" mbinu za uandishi wa riwaya
 1 "l3nt:12" kipindi cha urasimi wa fasihi
 1 "l3nt:12" kijelezi cha utanzu wa sanaa
 1 "l3nt:12" kijelezi cha dhana ya ushairi
 1 "l3nt:12" kijelezi cha dhana ya ubwege
 1 "l3nt:12" kazi ya uandishi wa riwaya
 1 "l3nt:12" kanuni za kijadi za utunzi
 1 "l3nt:12" hatua za uandishi wa riwaya

LT-N + GEN-CON + LT-N + GEN-CON + N "l3nt:13"

2 "l3nt:13" nadharia ya uyakinifu wa kijamii
 1 "l3nt:13" wahusika wa tamthilia za kibwege
 1 "l3nt:13" utunzi wa ushairi wa kiswahili
 1 "l3nt:13" utunzi wa mashairi ya kiswahili
 1 "l3nt:13" utungaji wa fasihi ya kiswahili
 1 "l3nt:13" urasimi wa fasihi ya kiswahili
 1 "l3nt:13" uhakiki wa ushairi wa kiswahili
 1 "l3nt:13" taashira ya utamaduni wa kizungu
 1 "l3nt:13" taashira ya migogoro ya kisiasa
 1 "l3nt:13" sanaa ya ushairi wa kiswahili
 1 "l3nt:13" muundo wa ushairi wa kiswahili
 1 "l3nt:13" muundo wa tamthilia ya kigiriki
 1 "l3nt:13" mhakiki wa tamthilia ya kiswahili
 1 "l3nt:13" masimulizi ya ngoma za kiasili
 1 "l3nt:13" lugha ya sanaa ya kibwege
 1 "l3nt:13" fasihi_andishi ya lugha ya kiswahili
 1 "l3nt:13" fasihi ya lugha ya taifa
 1 "l3nt:13" chimbuko la riwaya ya kiswahili

LT-N + GEN-CON + N + GEN-CON + LT-N "l3nt:14"

1 "l3nt:14" wahusika wa kazi ya fasihi
 1 "l3nt:14" waandishi wa kazi ya sanaa
 1 "l3nt:14" utunzi wa tungo za vichekesho
 1 "l3nt:14" utungaji wa kazi ya sanaa
 1 "l3nt:14" utungaji wa kazi ya fasihi
 1 "l3nt:14" uandishi wa vitabu vya fasihi
 1 "l3nt:14" mtindo wa kazi ya sanaa
 1 "l3nt:14" kaida za kimapokezi za utunzi
 1 "l3nt:14" dhamira ya kazi ya fasihi

LT-N + GEN-CON + LT-N + GEN-CON + LT-N "l3nt:15"

2 "l3nt:15" utunzi wa tamthiliya za drama
 1 "l3nt:15" taashira ya migongano ya kitamaduni
 1 "l3nt:15" mtazamo wa watunzi wa vichekesho
 1 "l3nt:15" mshororo wa shairi la tarbia
 1 "l3nt:15" mkondo wa ngano za fasihi_simulizi
 1 "l3nt:15" maandishi ya sanaa za maonyesho
 1 "l3nt:15" fani za sanaa za maonyesho

LT-N + GEN-CON + N + GEN-CON + N "l3nt:16"

2 "l3nt:16" utenzi wa vita vya maumau
 1 "l3nt:16" wahusika wa tabaka la kibepari
 1 "l3nt:16" waandishi wa nchi za ulaya
 1 "l3nt:16" uwasilishaji wa kauli za kikoloni
 1 "l3nt:16" uthakili wa mfumo wa uongozi
 1 "l3nt:16" utamaduni wa nchi za kimagharibi
 1 "l3nt:16" utamaduni wa kiasili wa afrika
 1 "l3nt:16" usanifu wa tungo za kiswahili

1 "l3nt:16" urasimi wa jamii za kigiriki
 1 "l3nt:16" upeo wa matamano ya maisha
 1 "l3nt:16" uhakiki wa jumuiya ya madola
 1 "l3nt:16" uchambuzi wa maisha ya jamii
 1 "l3nt:16" ubwege wa maisha ya binadamu
 1 "l3nt:16" ubwege wa bara la afrika
 1 "l3nt:16" tamthilia ya kiswahili ya kisasa
 1 "l3nt:16" tahakiki ya kauli ya kikoloni
 1 "l3nt:16" taashira ya mzinga wa nyuki
 1 "l3nt:16" riwaya ya kiswahili ya kisasa
 1 "l3nt:16" riwaya ya kiswahili ya awali
 1 "l3nt:16" riwaya ya kisasa ya kiswahili
 1 "l3nt:16" nadharia za kimsingi za elimu
 1 "l3nt:16" mlingano wa sauti za herufi
 1 "l3nt:16" mitindo ya maisha ya binadamu
 1 "l3nt:16" mitindo ya binadamu ya ubwanyenye
 1 "l3nt:16" mada ya mapenzi ya kidungu
 1 "l3nt:16" maandishi ya kimagharibi ya udhanaishi
 1 "l3nt:16" kisa cha mwana wa mfalme
 1 "l3nt:16" jazanda ya ukanganyifu wa jamii
 1 "l3nt:16" jazanda ya mbinu za kidhalimu
 1 "l3nt:16" jazanda ya matumizi ya nguvu
 1 "l3nt:16" hadithi ya kaizari ya kihistoria
 1 "l3nt:16" dhana za maisha ya binadamu

LT-N + GEN-CON + LT-N + PREP + LT-N "l3nt:17"

1 "l3nt:17" uhusika wa wahusika katika riwaya
 1 "l3nt:17" taashira ya ngoma katika tamthilia
 1 "l3nt:17" riwaya za upelelezi katika fasihi
 1 "l3nt:17" nadharia ya ubwege katika sanaa
 1 "l3nt:17" mtazamo wa mtunzi kuhusu dhamira
 1 "l3nt:17" msimamo wa mwandishi kuhusu dhamira
 1 "l3nt:17" falsafa ya mwandishi katika riwaya

LT-N + GEN-CON + N + PREP + LT-N "l3nt:18"

1 "l3nt:18" uyakini wa mambo katika muktadha
 1 "l3nt:18" taashira za kikati katika riwaya
 1 "l3nt:18" sadfa ya mwisho katika hadithi
 1 "l3nt:18" riwaya za kiswahili katika fasihi
 1 "l3nt:18" fasihi ya kiswahili kwa maandishi

LT-N + GEN-CON + LT-N + PREP + N "l3nt:19"

1 "l3nt:19" uwasilishaji wa ujumbe kwa maneno
 1 "l3nt:19" utunzi wa tamthiliya kwa jumla
 1 "l3nt:19" sanaa za maonyesho kwa jumla
 1 "l3nt:19" mtazamo wa mwandishi kuhusu migaragazo
 1 "l3nt:19" falsafa ya mwandishi kwa jumla

N + GEN-CON + LT-N + PREP + LT-N "l3nt:21"

1 "l3nt:21" vipengele vya fani katika riwaya
 1 "l3nt:21" umuhimu wa taswira katika fasihi
 1 "l3nt:21" umuhimu wa fasihi katika utamaduni
 1 "l3nt:21" ufumbuzi wa ujumbe katika mashairi
 1 "l3nt:21" ubingwa wa mwandishi katika lugha
 1 "l3nt:21" tafsiri ya ubeti kwa ubeti
 1 "l3nt:21" matumizi ya lugha kwa ufasaha
 1 "l3nt:21" maana ya riwaya katika fasihi

N + GEN-CON + N + PREP + LT-N "13nt:22"

1 "13nt:22" visimamizi vya nafsi katika shairi
1 "13nt:22" ugumu wa mawasiliano katika sanaa
1 "13nt:22" swali la kimsingi katika nadharia
1 "13nt:22" sura ya kitanzania kwa upande wa maudhui
1 "13nt:22" mpangilio wa maneno katika mishororo
1 "13nt:22" mbinu za kiisimu katika uhakiki
1 "13nt:22" matokeo ya mitihani katika vitendawili
1 "13nt:22" masharti ya lazima kwa ushairi
1 "13nt:22" maana ya maneno nje ya muktadha
1 "13nt:22" imani za kikale katika tamthilia

N + GEN-CON + LT-N + PREP + N "13nt:23"

2 "13nt:23" maana ya shairi kwa jumla
1 "13nt:23" ufundishaji wa fasihi katika shule
1 "13nt:23" nafasi ya fasihi katika jamii
1 "13nt:23" maumivu ya kifalsafa kuhusu maisha
1 "13nt:23" matumizi ya lugha kwa mchujo
1 "13nt:23" jukumu la fasihi kwa jamii
1 "13nt:23" jukumu la fasihi katika maisha
1 "13nt:23" historia ya drama kwa ufupi

LT-N + GEN-CON + INF + GEN-CON + N "13nt:27"

1 "13nt:27" taswira ya kuogofya ya damu
1 "13nt:27" dhamira za kupotoka kwa vijana

N + GEN-CON + INF + GEN-CON + LT-N "13nt:30"

1 "13nt:30" sifa ya kuimbika kwa shairi
1 "13nt:30" lengo la kueleweka kwa maudhui
1 "13nt:30" jukwaa la kuelezea kwa uyakinifu

LT-N + GEN-CON + INF + LT-N "13nt:31"

1 "13nt:31" utunzi wa kutumia vichekesho
1 "13nt:31" utamaduni wa kuandikia tamthilia
1 "13nt:31" mtindo wa kuwasilishia hadithi
1 "13nt:31" kipawa cha kuchuja maono

LT-N + GEN-CON + INF + N "13nt:32"

1 "13nt:32" wimbo wa kusifu ulafi
1 "13nt:32" mtindo wa kutumia picha
1 "13nt:32" desturi ya kubadilika kitabia

LT-N + GEN-CON + LT-V + LT-N "13nt:33"

1 "13nt:33" mtindo wa kusimulia hadithi

LT-N + GEN-CON + LT-V + N "13nt:34"

1 "13nt:34" mandhari ya kubuni tatizo

N + GEN-CON + LT-V + LT-N "13nt:35"

6 "13nt:35" sheria za kutunga mashairi
1 "13nt:35" ujuzi wa kutunga tamthiliya
1 "13nt:35" mbinu za kutunga tamthiliya
1 "13nt:35" mbinu za kutunga mashairi

N + GEN-CON + LT-V + N "13nt:36"

1 "13nt:36" njia ya kuiga wajuzi

N + GEN-CON + INF + LT-N "13nt:37"

1 "13nt:37" wajibu wa kukuza fasihi
1 "13nt:37" uwezo wa kufumbua fumbo
1 "13nt:37" uwezo wa kufua visa
1 "13nt:37" ustadi wa kucheza ala
1 "13nt:37" ujasiri wa kujaribu maudhui
1 "13nt:37" ubingwa wa kuandika mashairi
1 "13nt:37" njia ya kuwasilishia ujumbe
1 "13nt:37" njia ya kufundisha methali
1 "13nt:37" mwito wa kukuza lugha
1 "13nt:37" mwito wa kufufua tamaduni
1 "13nt:37" mbinu za kufundishia tamthilia
1 "13nt:37" mbinu ya kufafanua dhamira
1 "13nt:37" matayarisho ya kufundisha mashairi
1 "13nt:37" majazi ya kuendeleza maudhui
1 "13nt:37" mahitaji ya kutafsiri maandishi
1 "13nt:37" maelezo ya kufundisha mshororo
1 "13nt:37" lengo la kushadidia ujumbe
1 "13nt:37" lengo la kukamilisha dhamira
1 "13nt:37" lengo la kufundisha fasihi
1 "13nt:37" kusudi la kumulikia maudhui
1 "13nt:37" kipindi cha kuondoa kasumba
1 "13nt:37" kazi ya kutafiti fumbo
1 "13nt:37" juhudi ya kuelewa muundo
1 "13nt:37" hoja ya kupinga ushairi

LT-N + GEN-CON + N + CC + N "13nt:44"

2 "13nt:44" utenzi wa adamu na hawa
1 "13nt:44" uchambuzi wa kimaelezo na kitarakimu
1 "13nt:44" tamthiliya za redio na televisheni
1 "13nt:44" riwaya za kiingereza na kiswahili
1 "13nt:44" mizani ya kur'ani na biblia
1 "13nt:44" mgongano wa mnyanyasaji na mnyanyaswaji
1 "13nt:44" maudhui ya kisiasa na kijiitimai
1 "13nt:44" lugha ya kiarabu na kur'ani
1 "13nt:44" falsafa ya plato na aristotle
1 "13nt:44" dhana ya utenganishaji na upembuaji
1 "13nt:44" chimbuko la ubaguzi na unyonyaji

LT-N + GEN-CON + N + CC + LT-N "13nt:46"

2 "13nt:46" lugha ya kiingereza na fasihi
1 "13nt:46" nadharia za elimu na fasihi
1 "13nt:46" nadharia ya kihafidhina na fasihi
1 "13nt:46" lugha za kiasili na fasihi

LT-N + GEN-CON + LT-N + CC + LT-N "13nt:47"

2 "13nt:47" utunzi wa riwaya na hadithi
1 "13nt:47" ujumbe wa mwandishi na mhusika
1 "13nt:47" mwandishi wa riwaya na hadithi

N + GEN-CON + LT-N + CC + LT-N "13nt:48"

1 "13nt:48" vizingiti vya kimila na kitamaduni
1 "13nt:48" viwango vya maudhui na fani
1 "13nt:48" vipengele vya kisanaa na kimuundo

1 "13nt:48" umuhimu wa drama au vichekesho
 1 "13nt:48" ujengaji wa maudhui na dhamira
 1 "13nt:48" uendelezi wa maudhui na dhamira
 1 "13nt:48" ubingwa wa utunzi na uandishi
 1 "13nt:48" sherehe za jando na unyago
 1 "13nt:48" njia ya nyimbo na mashairi
 1 "13nt:48" historia ya riwaya na hadithi
 1 "13nt:48" heshima ya lugha na fasihi

LT-N + GEN-CON + N + LT-V "13nt:49"

1 "13nt:49" lugha ya kiswahili sanifu

LT-N + GEN-CON + N + PREP + N "13nt:54"

1 "13nt:54" visa vya chaka katika historia
 1 "13nt:54" ujumbe wa jumla kuhusu matukio
 1 "13nt:54" tenzi za kiarabu kwa kiswahili
 1 "13nt:54" taswira ya maisha baada_ya udhalimu
 1 "13nt:54" nadharia ya utabia katika elimu
 1 "13nt:54" hisia za huruma kwa wasomaji
 1 "13nt:54" dhana ya moto kwa moto
 1 "13nt:54" dhana ya maskani katika taifa
 1 "13nt:54" dhana ya maisha baada_ya kifo

LT-N + GEN-CON + PREP + N "13nt:61"

1 "13nt:61" vichekesho vya baada_ya azimio
 1 "13nt:61" fasihi ya kabla_ya uhuru

N + GEN-CON + PREP + LT-N "13nt:63"

1 "13nt:63" fikira ya ndani_ya mchezo

LT-N + N + GEN-CON + N "13nt:64"

2 "13nt:64" wanafalsafa waanzilishi wa siasa
 1 "13nt:64" utamaduni asili wa kiafrika
 1 "13nt:64" lugha asili za kiafrika

LT-N + LT-N + GEN-CON + N "13nt:66"

1 "13nt:66" shairi guni la kiswahili

N + N + GEN-CON + LT-N "13nt:69"

1 "13nt:69" vitabu mwafaka vya fasihi
 1 "13nt:69" uwasilishi habari wa lugha
 1 "13nt:69" utoaji muhtasari wa masimulizi
 1 "13nt:69" mwalimu mweledi wa fasihi
 1 "13nt:69" mshika kioo wa hadithi
 1 "13nt:69" matumizi ainaaina ya lugha
 1 "13nt:69" maelezo mafupimafupi ya mshairi
 1 "13nt:69" jicho zowefu la mwandishi

N + LT-N + GEN-CON + LT-N "13nt:70"

1 "13nt:70" vipengele changamano vya fani

N + POSS + LT-N + CARD "13nt:72"

1 "13nt:72" kawafi zenye mizani sita

LT-N + POSS + N + CARD "l3nt:73"

1 "l3nt:73" wahusika wenye nafsi moja
1 "l3nt:73" utungo wenye vipande vitatu

LT-N + PREP + N + GEN-CON + LT-N "l3nt:74"

1 "l3nt:74" ujumbe kwa njia ya methali
1 "l3nt:74" muundo katika kazi ya fasihi

LT-N + PREP + LT-N + GEN-CON + LT-N "l3nt:75"

1 "l3nt:75" umalenga katika ushairi wa taarabu
1 "l3nt:75" majigambo katika utunzi wa tamthiliya

LT-N + PREP + LT-N + GEN-CON + N "l3nt:76"

1 "l3nt:76" utunzi katika lugha ya kiswahili
1 "l3nt:76" fasihi katika utamaduni wa tanzania

N + PREP + N + GEN-CON + LT-N "l3nt:78"

1 "l3nt:78" tungo katika kiwango cha usanifu
1 "l3nt:78" picha katika akili za wasikilizaji
1 "l3nt:78" mweledi katika ufumbuzi wa mafumbo
1 "l3nt:78" mwanzo hadi mwisho wa tamthilia
1 "l3nt:78" mwanzo hadi mwisho wa riwaya
1 "l3nt:78" mwanzo hadi mwisho wa hadithi
1 "l3nt:78" mwalimu katika darasa la mashairi
1 "l3nt:78" mabishano kati_ya watetezi wa mashairi
1 "l3nt:78" maana katika umbo la hadithi
1 "l3nt:78" athari katika ujenzi wa uhusika

N + PREP + LT-N + GEN-CON + LT-N "l3nt:79"

1 "l3nt:79" upya katika fani ya tamthiliya
1 "l3nt:79" msikilizaji katika masimulizi ya ngano
1 "l3nt:79" mabadiliko katika maudhui ya tamthiliya
1 "l3nt:79" athari katika utunzi wa tamthiliya

N + PREP + LT-N + GEN-CON + N "l3nt:80"

1 "l3nt:80" ugeni katika ushairi wa kiswahili
1 "l3nt:80" uchunguzi katika ushairi wa kiswahili
1 "l3nt:80" mazungumzo katika uandishi wa kiswahili
1 "l3nt:80" maelezo katika uandishi wa kiswahili
1 "l3nt:80" maelezo katika fasihi ya kiswahili

N + GEN-CON + ADV + GEN-CON + LT-N "l3nt:81"

1 "l3nt:81" maana ya juujuu ya shairi
1 "l3nt:81" kiwango cha juu cha lugha

LT-N + GEN-CON + ADV + GEN-CON + LT-N "l3nt:82"

1 "l3nt:82" ujumbe wa ndani wa mwandishi

N + POSS + LT-N + GEN-CON + LT-N "l3nt:84"

1 "l3nt:84" wanawake wenye itikadi za kijadi

N + POSS + LT-N + GEN-CON + N "13nt:86"
 1 "13nt:86" wanaume wenye itikadi ya kimaendeleo

LT-N + POSS + LT-N + GEN-CON + LT-N "13nt:87"
 1 "13nt:87" maandishi yenye maudhui ya upelelezi

LT-N + POSS + N + GEN-CON + N "13nt:89"
 1 "13nt:89" maonyesho yenye asili ya nchi

LT-N + POSS + LT-N + GEN-CON + N "13nt:90"
 1 "13nt:90" mhusika mwenye upeo wa mawazo

N + GEN-CON + ORD + GEN-CON + LT-N "13nt:91"
 1 "13nt:91" sura ya kwanza ya riwaya
 1 "13nt:91" sehemu ya kwanza ya mashairi
 1 "13nt:91" herufi za kwanza za mishororo

LT-N + GEN-CON + ORD + GEN-CON + LT-N "13nt:92"
 1 "13nt:92" waandishi wa kwanza wa mashairi

LT-N + GEN-CON + ORD + GEN-CON + N "13nt:93"
 1 "13nt:93" riwaya ya kwanza ya kiswahili
 1 "13nt:93" lugha ya kwanza ya mfasiri

INF + GEN-CON + LT-N + GEN-CON + LT-N "13vt:100"
 1 "13vt:100" kupitia kwa uhusika wa mhusika

INF + GEN-CON + LT-N + GEN-CON + N "13vt:102"
 1 "13vt:102" kutamalaki kwa utamaduni wa kimagharibi
 1 "13vt:102" kuibuka kwa ushairi wa kisasa
 1 "13vt:102" kubadilika kwa ushairi wa kiswahili

LT-V + N + PREP + N "13vt:117"
 1 "13vt:117" kusimulia maisha katika mapana

LT-V + LT-N + PREP + N "13vt:118"
 1 "13vt:118" kusimulia hadithi kwa mpango

INF + LT-N + PREP + LT-N "13vt:119"
 1 "13vt:119" kupeleka tamthilia kwenye kilele
 1 "13vt:119" kuandika tamthiliya kutokana_na kichekesho

INF + N + PREP + LT-N "13vt:120"
 1 "13vt:120" kuwasilisha wazo kwa hadhira
 1 "13vt:120" kutahini wanafunzi katika vitendawili
 1 "13vt:120" kupiga chuku katika masimulizi
 1 "13vt:120" kupata mafunzo kutokana_na tajiriba
 1 "13vt:120" kupata maana kutokana_na muktadha
 1 "13vt:120" kupanga vitushi kwa mtiririko
 1 "13vt:120" kuleta mushkili katika urari

1 "13vt:120" kudhihirika bayana katika riwaya

INF + LT-N + PREP + N "13vt:121"

1 "13vt:121" kuwasilishia ujumbe katika kazi
1 "13vt:121" kuwasilisha ngano katika darasa
1 "13vt:121" kutazama shairi katika ujumla
1 "13vt:121" kuhifadhi fasihi simulizi katika vitabu
1 "13vt:121" kuanzisha drama katika shule
1 "13vt:121" kuandika riwaya kwa ajili ya watoto

LT-V + N + GEN-CON + LT-N "13vt:122"

1 "13vt:122" kutunga kazi ya sanaa
1 "13vt:122" kuhakiki kazi ya fasihi

LT-V + N + GEN-CON + N "13vt:124"

1 "13vt:124" kuiga tungo za watangulizi

INF + N + GEN-CON + LT-N "13vt:125"

1 "13vt:125" kuzingatia matumizi ya lugha
1 "13vt:125" kuzingatia maana ya maudhui
1 "13vt:125" kuvuka mipaka ya tamaduni
1 "13vt:125" kutumia nyenzo ya uandishi
1 "13vt:125" kutumia mbinu ya vichekesho
1 "13vt:125" kutoa mawazo ya kishairi
1 "13vt:125" kutegemea vitendo vya wahusika
1 "13vt:125" kutegemea nguvu za taswira
1 "13vt:125" kutazama matatizo ya watunzi
1 "13vt:125" kutambua umuhimu wa fasihi
1 "13vt:125" kutafuta umbo la riwaya
1 "13vt:125" kusikiliza kazi ya fasihi
1 "13vt:125" kusaili ukosefu wa mantiki
1 "13vt:125" kupuuza umuhimu wa maudhui
1 "13vt:125" kupotosha lengo la sanaa
1 "13vt:125" kupanua kiwango cha fasihi
1 "13vt:125" kuonyesha ukweli wa methali
1 "13vt:125" kuonyesha ukosefu wa mantiki
1 "13vt:125" kuongelea matatizo ya utunzi
1 "13vt:125" kukuza msamiati wa lugha
1 "13vt:125" kukidhisha misingi-teule ya kishairi
1 "13vt:125" kukabili kazi ya fasihi
1 "13vt:125" kujua mawazo ya hadhira
1 "13vt:125" kuhukumu kazi ya fasihi
1 "13vt:125" kufupisha kazi ya fasihi
1 "13vt:125" kufundisha mbinu ya utunzi
1 "13vt:125" kufundisha matumizi ya lugha
1 "13vt:125" kufundisha maneno ya kilahaja
1 "13vt:125" kufuata taratibu za muundo
1 "13vt:125" kufuata fikira ya kishairi
1 "13vt:125" kufaulisha kazi ya fasihi
1 "13vt:125" kufahamu kijelezi cha tashihisi
1 "13vt:125" kufafanua nguzo za muundo
1 "13vt:125" kuepuka tabia za wahusika
1 "13vt:125" kueneza utumizi wa methali
1 "13vt:125" kueleza nafasi ya wasanii
1 "13vt:125" kuelewa maana ya methali
1 "13vt:125" kuchunguza matumizi ya lugha
1 "13vt:125" kuchukua nafasi ya msimuliaji
1 "13vt:125" kuchukua mkabala wa uyaqinifu
1 "13vt:125" kuchambua matumizi ya tamathali

1 "13vt:125" kuchambua kazi ya fasihi
 1 "13vt:125" kubadili umbo la riwaya
 1 "13vt:125" kubadili sura ya vichekesho
 1 "13vt:125" kuashiria athari za utamaduni
 1 "13vt:125" kuandika kazi za kisanaa

INF + LT-N + GEN-CON + LT-N "13vt:126"

2 "13vt:126" kuendeleza maudhui ya shairi
 1 "13nt:126" kuandiki tamthiliya ya drama

INF + LT-N + GEN-CON + N "13vt:127"

2 "13vt:127" kukuza fasihi ya kiswahili
 1 "13vt:127" kuzingatia nadharia ya utabia
 1 "13vt:127" kuweka vidokezo vya jukwaa
 1 "13vt:127" kuwa mwandishi wa kiswahili
 1 "13vt:127" kutumia lugha ya muhtasari
 1 "13vt:127" kutumia lugha ya mchongaji
 1 "13vt:127" kutumia lugha ya kikoloni
 1 "13vt:127" kutiana mori wa kimapinduzi
 1 "13vt:127" kutegemea muktadha wa matumizi
 1 "13vt:127" kupitia taswira ya kitu
 1 "13vt:127" kupima ushairi wa kiswahili
 1 "13vt:127" kupata maudhui ya kazi
 1 "13vt:127" kuonyesha usanifu wa tungo
 1 "13vt:127" kukuza lugha za mataifa
 1 "13vt:127" kukadiria taathira ya maneno
 1 "13vt:127" kuhesabu takriri za vokali
 1 "13vt:127" kugusa hisia za msomaji
 1 "13vt:127" kufundisha shairi la kimapokeo
 1 "13vt:127" kuendeleza hisia za umoja
 1 "13vt:127" kuelezea tamthilia ya kigiriki
 1 "13vt:127" kuelezea tajiriba za binadamu
 1 "13vt:127" kueleza tajiriba za binadamu
 1 "13vt:127" kucheza muziki wa waltz
 1 "13vt:127" kucheza mchezo wa shetani
 1 "13vt:127" kuandika riwaya ya maana

INF + LT-N + GEN-CON + INF "13vt:128"

1 "13vt:128" kutumia mtindo wa kufundisha
 1 "13vt:128" kutumia mtindo wa kueleza

INF + LT-N + GEN-CON + LT-V "13vt:130"

1 "13vt:130" kuzingatia hadithi za kubuni
 1 "13vt:130" kutafsiri maandishi ya kutunga

INF + LT-N + LOC "13vt:135"

1 "13vt:135" kutumbukiza shairi hapa na pale
 1 "13vt:135" kufundisha ushairi darasani
 1 "13vt:135" kufundisha tamthilia darasani
 1 "13vt:135" kufundisha riwaya darasani
 1 "13vt:135" kufundisha fasihi darasani

INF + LT-N + N "13vt:138"

1 "13vt:138" kuwasilisha ujumbe kiubingwa
 1 "13vt:138" kutungia lugha istilahi
 1 "13vt:138" kukosa msimamo thabiti
 1 "13vt:138" kuchukua msimamo thabiti

INF + LT-N + LT-N "13vt:139"

- 1 "13vt:139" kufuata nadharia mtazamo-kike
- 1 "13vt:139" kuchambua shairi mshororo

INF + N + LT-N "13vt:140"

- 2 "13vt:140" kuelezea neno fasihi

LT-V + LT-N + N "13vt:143"

- 1 "13vt:143" kutunga vichekesho kiutendaji

N + INF + PREP + N + GEN-CON + LT-N "14nt:207"

- 1 "14vt:207" watazamaji kuingizwa katika ulimwengu wa tamthilia

Length 4 terms

N + GEN-CON + N + PREP + N + GEN-CON + LT-N "14nt:200"

- 1 "14vt:200" uteuzi wa maneno kwa njia ya usanifu
- 1 "14vt:200" udondoshaji wa silabi kwa lengo la mizani

N + GEN-CON + N + PREP + LT-N + GEN-CON + N "14nt:199"

- 1 "14vt:199" maeneo ya afrika kabla_ya usuli wa ukoloni

N + GEN-CON + N + PREP + LT-N + GEN-CON + LT-N "14nt:198"

- 1 "14vt:198" mbinu za kiisimu katika uhakiki wa fasihi

N + GEN-CON + N + N + GEN-CON + LT-N "14nt:197"

- 1 "14vt:197" uteuzi wa vitabu mwafaka vya fasihi

N + GEN-CON + N + LT-N + GEN-CON + LT-N "14nt:194"

- 1 "14vt:194" nia ya kuandiki tamthiliya ya drama

N + GEN-CON + N + GEN-CON + LT-N + N "14nt:192"

- 1 "14vt:192" uti wa mgongo wa uhakiki baada-ukoloni

N + GEN-CON + LT-N + PREP + N + GEN-CON + N "14nt:180"

- 1 "14vt:180" umuhimu wa fasihi katika elimu ya jamii
- 1 "14vt:180" ufundishaji wa fasihi katika shule za mataifa
- 1 "14vt:180" maumivu ya kifalsafa kuhusu maisha ya binadamu
- 1 "14vt:180" jukumu la fasihi katika maisha ya binadamu

N + GEN-CON + LT-N + PREP + N + GEN-CON + LT-N "14nt:179"

- 1 "14vt:179" ujuzi wa utunzi kutokana_na somo la fasihi

N + GEN-CON + LT-N + PREP + LT-N + GEN-CON + LT-N "14nt:178"

- 1 "14vt:178" mbinu ya vichekesho kwa utunzi wa tamthiliya

N + GEN-CON + LT-N + GEN-CON + N + N "14nt:172"

- 1 "14vt:172" kikundi cha drama cha tanzania film

1 "l4vt:172" kikundi cha drama cha tanganyika textile

N + GEN-CON + LT-N + GEN-CON + N + GEN-CON + N "l4nt:171"

1 "l4vt:171" taratibu za muundo wa tungo za kiswahili
1 "l4vt:171" miundo ya tamthilia za mataifa ya kimagharibi
1 "l4vt:171" maelezo ya taashira ya mzinga wa nyuki

N + GEN-CON + LT-N + GEN-CON + N + GEN-CON + LT-N "l4nt:170"

1 "l4vt:170" ufichuaji wa ujumi wa kazi ya sanaa
1 "l4vt:170" mbinu ya utunzi wa uonyeshaji wa tamthiliya
1 "l4vt:170" maana ya usanifu wa kazi ya fasihi

N + GEN-CON + LT-N + GEN-CON + LT-N + GEN-CON + N "l4nt:168"

1 "l4vt:168" tofauti ya mashairi ya washairi wa mapote
1 "l4vt:168" sheria za utunzi wa mashairi ya kiswahili
1 "l4vt:168" nguzo za muundo wa ushairi wa kiswahili
1 "l4vt:168" msingi wa urasimi wa ushairi wa kiswahili
1 "l4vt:168" kipindi cha urasimi wa fasihi ya kiswahili
1 "l4vt:168" idara ya sanaa za maonyesho ya chuo

INF + N + PREP + LT-N + GEN-CON + N "l4vt:164"

1 "l4vt:164" kupata maana kutokana_na muktadha wa matumizi

INF + LT-N + PREP + N + GEN-CON + N "l4vt:154"

1 "l4vt:154" kuandika riwaya kwa_ajili_ya watoto wa shule

INF + LT-N + PREP + N + GEN-CON + LT-N "l4vt:152"

1 "l4vt:152" kuandika tamthilia kwa njia ya ushairi

INF + N + PREP + N + GEN-CON + LT-N "l4vt:120"

1 "l4vt:120" kuelimisha umma juu_ya umuhimu wa fasihi

LT-N + GEN-CON + N + PREP + LT-N + GEN-CON + N "l4nt:47"

1 "l4nt:47" uyakinifu wa mambo katika muktadha wa jamii
1 "l4nt:47" riwaya za kiswahili katika fasihi ya kiswahili

LT-N + GEN-CON + N + PREP + LT-N + GEN-CON + LT-N "l4nt:46"

1 "l4nt:46" lugha ya kiswahili katika upeo wa riwaya

LT-N + GEN-CON + N + GEN-CON + N + N "l4nt:38"

1 "l4nt:38" mada ya mapenzi ya kidungu utu
1 "l4nt:38" dhana pevu ya uwezo wa sanaa

LT-N + A-INFL + GEN-CON + N + GEN-CON + LT-N "l4nt:3"

1 "l4nt:10" kilele cha falsafa ya shaaban robert

LT-N + A-INFL + GEN-CON + LT-N + GEN-CON + LT-N "l4nt:1"

1 "l4nt:1" muundo mzima wa utunzi wa kichekesho