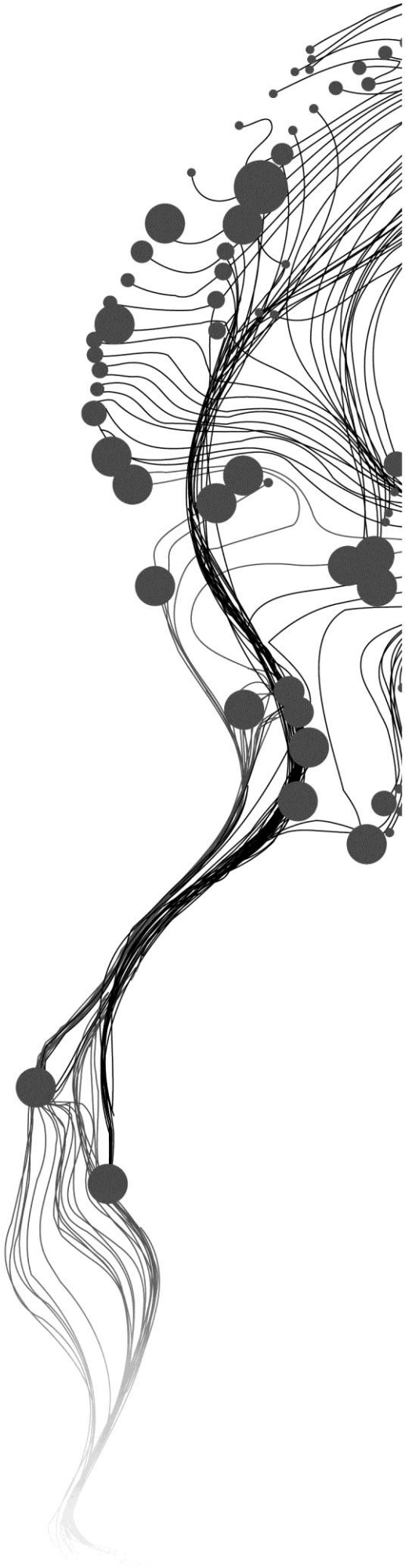
PROCEDURES FOR THE TRANSLATION OF BOUNDARY DESCRIPTION TEXTS INTO GEOGRAPHIC LOCATION FORMAT

MASIDA BENNETTE JEM MBANO

February, 2011

SUPERVISORS:
Prof. Dr. Ir. Martien Molenaar
Prof. Dr. Menno - Jan Kraak



PROCEDURES FOR THE TRANSLATION OF BOUNDARY DESCRIPTION TEXTS INTO GEOGRAPHIC LOCATION FORMAT

MASIDA BENNETTE JEM MBANO
Enschede, The Netherlands, February, 2011

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Geoinformatics

SUPERVISORS:
Prof. Dr. Ir. Martien Molenaar
Prof. Dr. Menno - Jan Kraak

THESIS ASSESSMENT BOARD:
Prof. Dr. Menno - Jan Kraak (Chair)
Ms Dr. J.E. Stoter (External Examiner)

ABSTRACT

In Africa, International boundaries continue to be a major source of conflict between neighbouring countries. Most of these are poorly defined on the ground. These boundaries were agreed upon by colonial states like England, Portugal, France, Belgium and The Netherlands that colonized the African countries. This was done through the signing of International boundary treaty agreements without physically demarcating the boundary on the ground. Interpretation of the texts in the treaties into geographic location format has proved difficult as there are no proper procedures followed and hence lead to different interpretations by parties involved. This research aimed at formulating standardized procedures that can be used to interpret the international boundary description text in the treaties, into a geographic location format. In the formulated procedures, uncertainties that are associated with the interpretation of the text into geographic location format were also formulated and indicated at what scale they occurred. The Anglo-German Treaty [Heligoland-Zanzibar Treaty] (July 1, 1890) was used in this research as an example. The treaty temporarily settled colonial disputes between Germany and Great Britain in Africa. Some uncertainties occurred towards the identification of the boundary line from the old system that is feature based, to the new system that is coordinate based. Flow diagrams that give chaining commands were devised to help in interpreting these uncertainties as they are the ones that usually lead to conflicts. The accuracy of the treaty in identifying the boundary line was at 1:1,000,000 as based on the maps under which the treaties were agreed on. The developed procedures are objective and help identify where decisions have to be made when interpreting the treaty text into geographic location format by following a standardized approach. The procedure will work well in a semi-automated system with the guidance of an experienced land surveyor after acquiring all necessary documentations regarding the boundary line in question.

ACKNOWLEDGEMENTS

First, I would like to thank my Supervisor Professor Dr. Ir. Martien Molenaar for the constant support and guidance throughout my research work. You never stopped reminding me to rest and have fun as well. Your wisdom and great experienced was an encouragement to me as a young scientist for you made me believe that no matter how tough it seems, I could still make it. I also thank Professor Dr. Menno-Jan Kraak, for introducing me to the topic even though it changed from the first one, arranging for the conferences with United Nations headquarters in America and providing me with the map for the status of international boundaries in Africa that you got from United Nations. I also thank Dr. Ir. Rolf de By and Dr. Arbind Tuladhar, for the one shop meeting on which you offered me crucial advice towards my work and also encouraged me on the importance of my research. To Gaurav Singh, the PhD student, I say thank you for sparing your time sometimes to also assist me in my work.

Secondly, I give thanks to Survey Department and the government of Malawi for sponsoring my studies, without which, I could not have been here.

To my wife Anastazia and my son Themba, thank you for patiently waiting for me at home throughout my stay at ITC. To my parents, Zondiwe and Nellie Mbano, thank you so much for installing in me the spirit to always aim high. And to all my friends at ITC, thank you for the moral and spiritual support.

Lastly but not least, I thank the Lord God Almighty for giving me this life I live and sustaining my health throughout my studies at ITC.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

APPENDIX LIST

1. INTRODUCTION

1.1. Motivation and problem statement

In Africa, International boundaries continue to be a major source of conflict between neighbouring countries. Boundary disputes might occur when there are disagreements between neighbouring countries about the location of an agreed boundary or about the actions of one country near the boundary which is perceived to cause detriment to the neighbour (Peterson, 1980). These conflicts sometimes lead to lives being lost and resources being wasted in trying to resolve them.

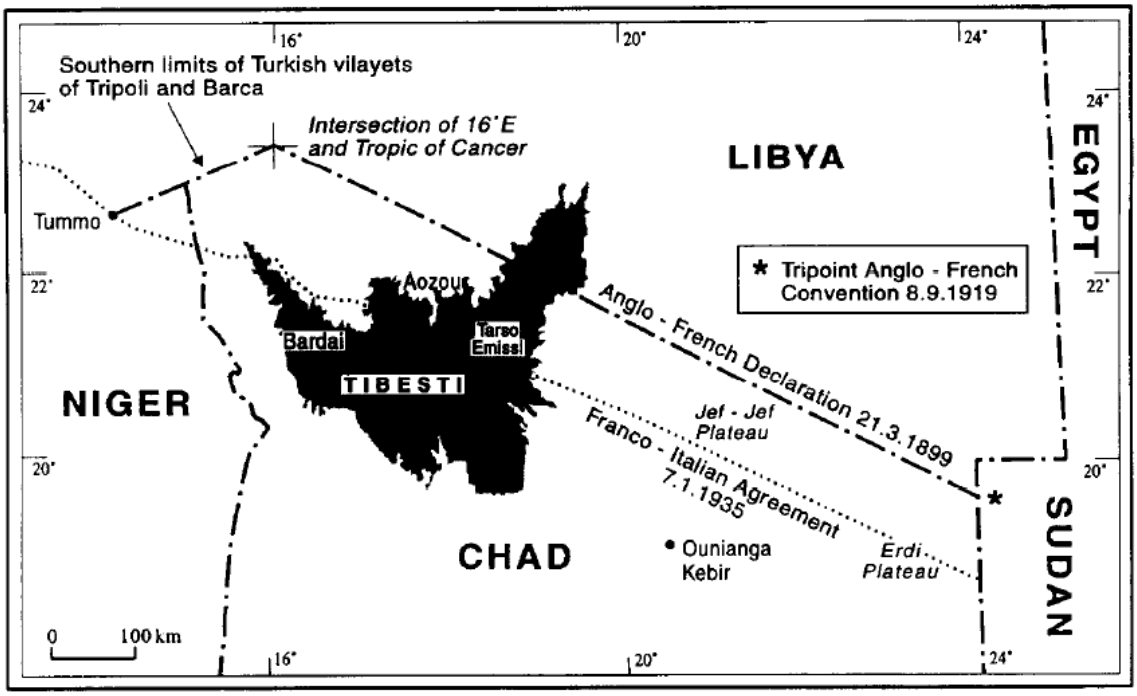
International boundaries of African countries were agreed upon by colonial states that colonized them like Britain, France, Belgium, Portugal and The Netherlands. This was done through the signing of boundary treaties that were in text format and described the boundary based on known features like lakes, rivers, mountains. There was no effort to translate these feature based texts into a geographic location on the ground based on coordinates. Efforts by African nations after attaining their independence to demarcate (field operation whose purpose is to mark the position of the boundary on the ground for all to see) their international boundaries have proved difficult as there are no standards followed in translating these boundary treaty texts into a geographic location on the ground. This leads to different countries interpreting the boundary description texts in different ways. This then leads to disagreements between the involved countries and if not solved, leads to conflicts.

For example, the disagreement between the international boundary of Chad and Libya developed in 1973. Chad relied on the boundary that was agreed in the Franco – Italian Treaty of Friendship and Good Neighbourhood of 10 August 1955, which defined the boundary between Libya and French territories according to various international agreements and treaties that were listed in an annex: the Anglo – French Convention of 14 June 1898 and the Additional Declaration of 21 March 1899, the Franco – Italian Accords of 1 November 1902, the Franco – Turkish Convention of 12 May 1910, the Anglo – French Convention of 8 September 1919 and the Franco – Italian Arrangement of 12 September 1919 (Brownlie & Burns, 1979). On the other hand, Libya relied on a boundary defined in a Franco – Italian Agreement signed on 7 January 1935. This agreement had its origin in the treaty of London devised by British, France, Italy and Russia and signed on 26th April 1915, which was designed to persuade Italy to enter the war on the side of the Allies and included a provision that if Britain and France increased their territories in Africa, at the expense of Germany, then Italy could claim equitable compensation. The boundaries defined in the 1935 Agreement and the 1955 Treaty are showed in Figure 1.

Similarly, the boundary dispute between Burkina Faso and Mali in West Africa started when the two countries attained independence in August and September 1960 respectively. The two governments realised that they held different interpretations of the location of the boundary drawn by the French to separate the previous divisions of French West Africa.

All these conflicts have arisen due to the difficulties encountered in the interpretation of the boundary treaty text into geographic location format. Due to this problem, a lot of International Boundaries in Africa have not yet been demarcated but only delimited (legal process by which two sovereign nations establish and describe in writing the location of their common boundary). Figure 2 shows the status of International boundaries in Africa as of August 2010.

Figure 1: Boundary between Chad and Libya as defined in the 1935 Franco - Italian Agreement and the 1955 Treaty.



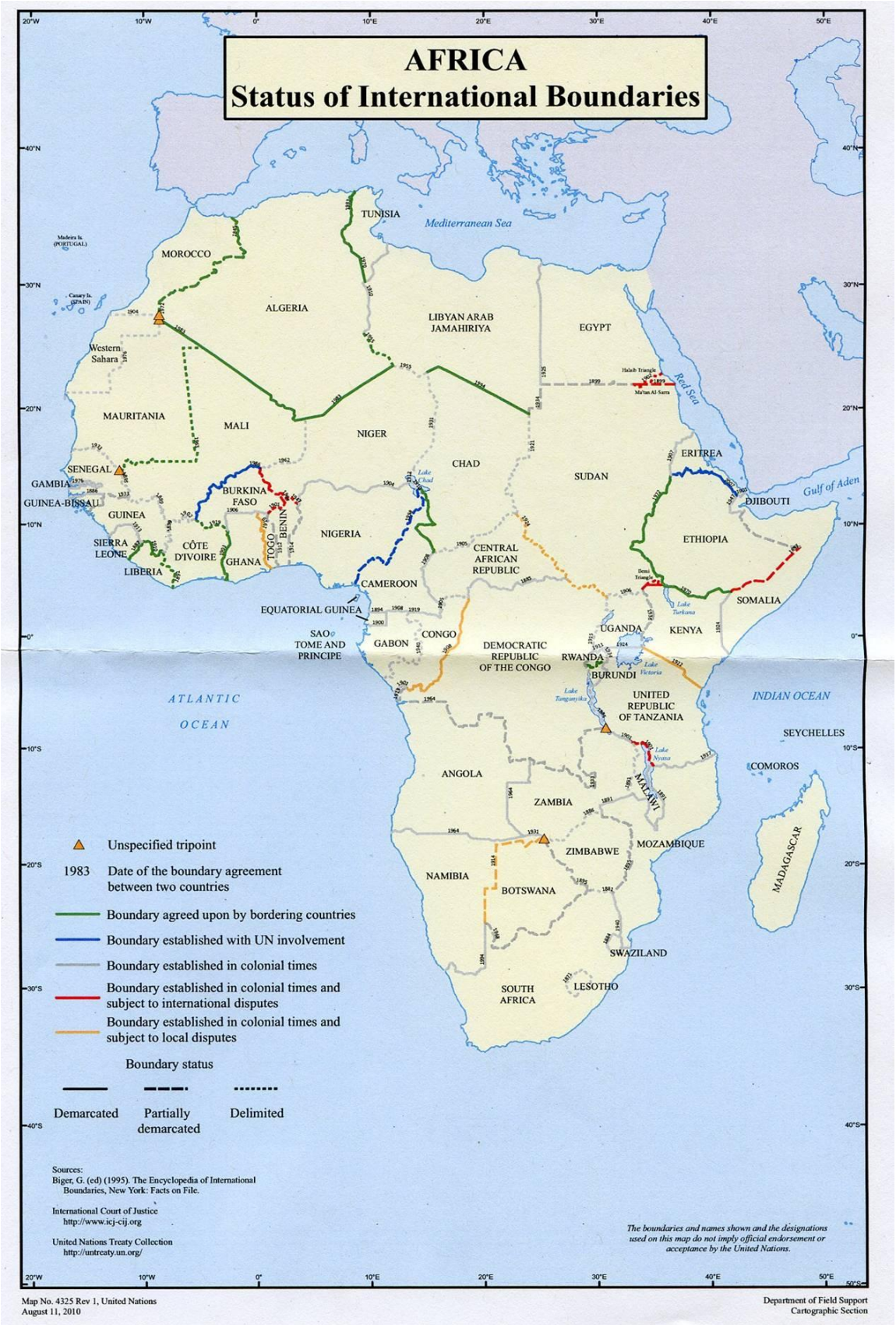
(Source: (Prescott, 1996))

There are some organizations that are working towards minimizing conflicts associated with International boundaries. For example, The International Boundaries Research Unit (IBRU) is a research unit based in United kingdom that works to minimize conflict associated with international boundaries on land and at sea around the world ((IBRU), 2010). It does this by offering training courses and also consultancies in boundary management and resolution making and also helping concerned parties manually identify the boundary line on the map and on the ground whenever conflicts arise.

The United Nations (UN) is also involved in settling many boundary and territorial disputes, the former largely through the International Court of Justice (ICJ) and the latter mainly through Iron-judicial involvement (Prescott, 1996). The United Nations is a large and complex structure. Those parts of the UN that have played the major role in trying to resolve boundary and territorial disputes are the Security Council, the General Assembly, the Committee on the Situation with Regard to the Implementation of the Declaration on the Granting of Independence to Colonial Countries and Peoples (the Special committee on decolonization) and the ICJ (Prescott, 1996).

The ICJ for example, was involved in the Chad and Libya boundary dispute and after careful research, the court found the boundary claim by Chad was the correct boundary in international law and cited the fact that Libya had recognized this by helping to define the eastern terminus of the boundary after 10 August 1955 (Prescott, 1996). Similarly for Burkina Faso and Mali, the ICJ was involved and in December 1986, ICJ delimited a line which the two countries agreed to demarcate. The judgement divided the dispute zone almost equally between the two countries, although Mali secured more of the western sector and Burkina Faso more of the eastern sector. In its unanimous judgement, the Court referred to the principle of *uti possidentis juris*. The primary aim of this principle is to establish respect for the boundaries that existed at the time of independence (Naldi, 1987).

Figure 2: Status of International Boundaries in Africa



(Source: United Nations Cartographic Section – Department of Field Support. (August, 2010)).

Due to the lack of proper procedures, these institutions and individual country mapping agencies have great challenges when involved in the translation of the boundary treaty text into geographic location. Understanding of the boundary treaties and formulating standardised procedures for translating the international boundary text into geographic location is a major challenge and the main objective of this research. Since the developed procedures will be objective, they will help those institutions involved in boundary dispute solving and also individual mapping agencies of different countries to have a unified approach of interpreting the boundary treaties text into geographic location format. Finally, the developed procedures will also help to reduce most of the conflicts associated with international boundaries in Africa.

1.2. Research Identification

1.2.1. Research objectives

The objectives of this research were to:

a. Formulate procedures to be used for the translation of international boundary treaty text into geographic location format.

b. Formulate uncertainty descriptors for the output of these translation procedures.

c. Indicate the scale or resolution levels the location uncertainty of the identified boundaries occurs.

d. Indicate the effects of these uncertainties in terms of disputes or conflicts.

1.2.2. Research Questions

In the context of the research objectives above, the following research questions were worked out. With regards to:

a. the formulation of the procedures,
   - Can the Location and Elevation Interpreter (LEI) method used by (David N. Chin, 1994) on biological specimen be used as a framework to translate boundary description text into a geographic location on the ground?
   - Can the approach used by (Frank, 1992) on qualitative spatial reasoning about distance and direction in geographic space be used to calculate the geographic locations on the ground from qualitative data provided in the boundary description text?
   - Can the formulated procedures be automated?

b. the uncertainty descriptors,
   - What are the uncertainties aspects associated with interpretation of the boundary description text in boundary treaties into geographic location?
   - Can different classes be identified for such uncertainty aspects?

c. the scale and resolution,
   - What are the descriptors per class of uncertainty?
   - What are their uncertainties scales?
   - How can the uncertainties be evaluated?

d. the effect of the uncertainty,
   - Which types of uncertainties are relevant for solving these disputes?

1.3. Research Methodology

The following methodology was used to achieve the research objectives and answer the research questions. The methodology was also used to formulate the standardised procedures for translating the boundary description text into a geographic location format:-

a) Collection of Boundary treaty information from the Office of Legal Affairs (OLA) of UN ((OLA), 2010) for free by the author. The Anglo-German Treaty [Heligoland – Zanzibar Treaty] (July 1, 1980) was used as an example in this research. This is the treaty that settled colonial disputes between German and Great Britain territories in Africa. The treaty is in English.
b) Studying the boundary description text in the treaty and maps on which the boundary treaties were agreed upon. This helped to identify an objective approach to translate the boundary description text it into geographic location format and to also identify the uncertainties aspects associated with the boundary description text.
c) Use of a name gazetteer to identify and correctly classify names of places and other features in the treaty as towns, mountains, rivers, lakes etc.
d) Use of a geodatabase of Africa to identify names of places and features and plot them using their coordinates.
e) Use Distance calculation procedures from qualitative spatial information to geographic space (Frank, 1992).
f) Use of the Location and Elevation Interpreter (LEI) methodology used on biological specimen as a framework for the formulation of the procedures (David N. Chin, 1994). LEI is composed of four main components:
    - The language analyser, PPI. This analyser parses the English location description and produces a collection of spatial relations that relate the actual collection point to geographical objects. It uses knowledge of geographical objects and their associated names from the geographic knowledge base.
    - The geographical reasoned, GR. This translates spatial relations from the language analyser into polygons and performs polygon intersection calculations to obtain the area specified by the spatial relations.
    - The user interface LEIview. This displays maps and allows users to add or modify object locations.
    - The geographic knowledge base, GKB. This contains an object-oriented description of geographical objects such as valleys, streams, and waterfalls with their associated locations and names.

1.4. Thesis Structure

Chapter 1 provides the introduction to the thesis, motivation and problem statement, research objective and questions and research methodology. Chapter 2 provides the literature review on what has been done towards approaches to translate description texts into geographic location format. Chapter 3 provides procedures for analysing and processing the texts in the boundary treaty while Chapter 4 provides the procedures for interpreting the analysed text into geographic location format. Chapter 5 provides the results obtained after analysing the boundary description text using the stipulated procedures. Chapter 6 discusses the uncertainties related with boundary description text when interpreting it into geographic location format and provides an appropriate approach for solving the uncertainty where possible and also the consequences of these uncertainties i.e. how sensitive and important the understanding of the uncertainties is for the boundary issues. Chapter 7 provides the discussion for the findings and lastly, Chapter 8 provide the conclusion and recommendation for further research.

2. LITERATURE REVIEW

Interpretation of boundary description text into geographic location format is mainly a problem that has arisen due to the lack of research in this area. For institutions and even mapping agencies that work in this area, they consistently rely on their vast knowledge and wide consultations in order to make the best decisions. However, there have been some research in other disciplines that helped build a proper framework for the formulation of procedures to translate these boundary description texts into geographic location format.

A research was done by (David N. Chin, 1994) on biological specimens that had historically been labelled with English descriptions of the location of collection. These descriptions included the place description in text e.g. Punaluu Valley, prepositional phrases e.g. from Punaluu to Kaluanui Valley, date of collection of the species, collector's name(s), genus, elevation, the museum's collection number, and collector's accession number. In order to perform spatial or statistical analysis on this historical data, these descriptions were converted into geodetic coordinates, a time consuming process that required eye-straining poring over maps to search for each location. Automating this process, required understanding the natural language descriptions, reasoning about spatial relations described by the natural language, and mapping these into a geographical object base to derive the collection coordinates.

This research on biological specimen had been the benchmark of the research on procedures for the translation of boundary description text into geographic location format. First, it had location descriptions text that needed to be translated to geographic location format and even after converting the location descriptions into the appropriate spatial relation, there were still problems in the correct interpretation of the relations. For example, "along a stream", did not mean the collection site was in the stream, but within some distance of the stream. The problem was what exactly the value of that distance was. Even cardinal directions like "north of" were fuzzy concepts. Even though these biological data were in tabular form, their similarities to the boundary description text, offered a benchmark for this research.

(André, Bosch, Herzog, & Rist, 1987), (Herskovits, 1986) and (Talmy, 1983) like many others, have documented the many problems in interpreting and using spatial prepositions. For example, *in* and *on* have similar but different meanings: 'in the car" means within the car, while "on the car" means on top of the car. However, "on the bus/plane," means within the bus or plane. Also each preposition typically has several different meanings or usage. For example, one says "at home," but "at the bank," and the meaning of "the plane is at Honolulu airport," is within the area of Honolulu airport, but the meaning of "the dog is at the telephone pole" is not within the telephone pole, but near it. These context usages make the interpretation of spatial prepositions problematic. (David N. Chin, 1994), formulated a table that was used for interpreting the prepositions found in the biological specimen text, and their respective spatial relations (Appendix B). Such concepts had been applied in the boundary text, in order to correctly analyse and interpret the prepositions found in the boundary description text into spatial relations.

On the other hand, spatial reasoning is the ability to interpret and make maps, form mental images, and visualize movement or change in those images. This ability is important for generating and conceptualizing solutions to multi-step problems that arise in areas such as architecture, engineering, science, mathematics, art, games, and everyday life. It is extremely important in disciplines such as mathematics and computer

science (Zhan, 2002). But spatial reasoning theories have only been applied in mathematics and computer science and have never been tested on geographic information.

Qualitative spatial reasoning (QSR) abstracts metrical details of the physical world and enables computers to make predictions about spatial relations even when precise quantitative information is unavailable (Moratz & Ragni, 2008). The two main directions in QSR are topological reasoning about regions and reasoning about orientation configurations. Orientations can refer to a global reference system e.g. cardinal directions. Reasoning about relative orientation poses additional difficulties compared to reasoning about orientations in an absolute reference frame. (Davis, 1986), (Kuipers, 2000) and (Frank, 1992) investigated qualitative and/or quantitative reasoning techniques for dealing with spatial relations. (Frank, 1992) gives a set of qualitative deduction rules for a subset of spatial reasoning, namely reasoning with cardinal directions and qualitative distance descriptors, without relying on quantitative calculations (e.g. square roots, trigonometric functions) or analytical geometry. A standard approach to modelling human spatial reasoning is to use Euclidean geometry in the plane or three-dimensional space and represent the task using analytical geometry formulae. Many problems can be expressed as an optimization problem with a set of constraints, such as location of a resource and the shortest path in this framework.
Similarly, the important field of geographic reference frames in natural language has mostly been treated using an analytical geometry approach. Typically, spatial positions are expressed relative to positions of other objects. Examples occur in every day speech in forms such as 'the church is west of the restaurant'. In the past these descriptions were translated into Cartesian coordinate space and the mathematical formulations were analysed. A qualitative approach can deal with imprecise data, and therefore yields less precise results than the quantitative one. This is highly desirable because: precision is not always desirable; and precise, quantitative data is not always available. Also, qualitative reasoning has the advantage that it can deal with imprecise data and need not translate it into a quantitative form (Frank, 1992). (Freeman, 1975), identified commonly used spatial relations. All these approaches above had been used in order to find a standardized approach of translating the boundary description text into geographic location.

(Frank, 1992) also gave algorithms for the analysis of cardinal directions using spatial reasoning while (Li, 2007), analyses algorithms for the four cardinal directional relations between plane regions, viz. *west, east, north, south,* and then defines nine basic relations by using the usual relational operations of intersection and complementation.

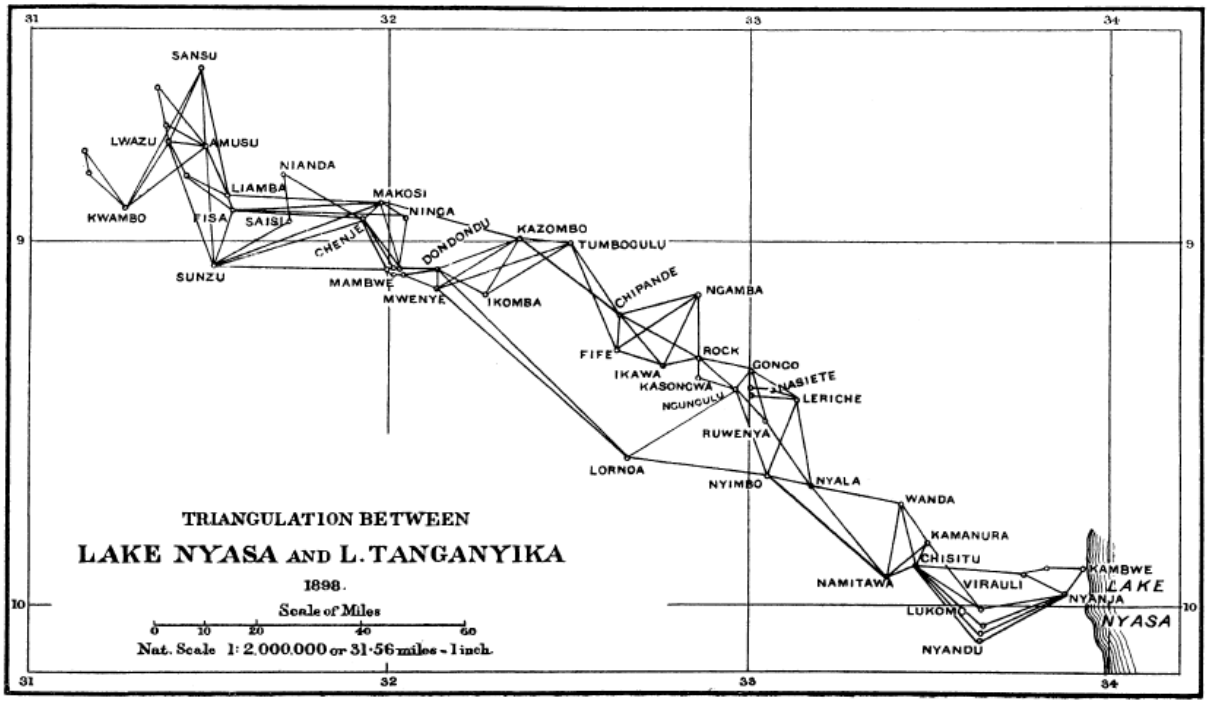
However, not much research has been done on international boundary treaty text especially to analyse it and interpret it into geographic location format. However, there has been some research on how some of the boundary related conflicts were solved. These helped give an insight on the best approaches to be used for interpreting this boundary description text into geographic location format.

(Toset, Gleditsch, & Hegre, 2000), showed how shared water resources have caused or may lead to conflicts or violence between countries that share them. They indicate that there are more than 200 river systems shared by two or more countries. The study was built on newly generated data on boundary – crossing rivers. The result showed that joint rivers do indeed increase the probability of conflicts. (Prescott, 1996), highlights some boundary conflicts that occurs between different nations and how the United Nations helped in solving these conflicts between involved nations. He shows how different countries interpreted the boundary treaty text, how such different interpretations lead to conflicts and how the United Nations helped to solve these conflicts. Even though there were no standardised procedures in the approaches taken by United Nations, they still give an intellectual understanding and insight or how similar problems can be dealt with.

(Brownlie & Burns, 1979), gives details on African boundaries by highlighting the boundary treaties that bind them, which countries signed the treaties, if possible, sketches that were used to roughly indicate how some boundaries would be. He also highlights problems encountered by many African countries in order to translate this boundary description text into geographic location format.

The Anglo – German Treaty [Heligoland – Zanzibar Treaty] (July 1, 1980) is the treaty which has been used as an example in this research. (Boileau & Wallace, 1899), gave a report on which the Anglo – German Treaty was based on. They were part of the members of the Anglo – German Boundary Commission. The report showed the tourist maps produced by Wallace on his voyages in Central Africa and also the triangulation map produced by the surveyor, Captain Boileau. Figure 3 indicates the triangulation map. In their report, they give detailed description on their expeditions, how they got the names of local places and features like rivers. Their report therefore, gives a clear understanding on how the boundary treaties were formulated and also gives an insight on how to solve some uncertainties related to the boundary description text in the Anglo – German treaty.

Figure 3: Triangulation points surveyed by Captain F.F.R. Boileau for the Anglo-German Boundary Commission



The handling of large amounts of information about the natural and built environments, as is necessary in any GIS, is prone to uncertainty in a number of forms. Ignoring that uncertainty can, at best, lead to slightly incorrect predictions or advice and at worst can be completely fatal to the use of the GIS and undermine any trust which might have been put in the work of the system or operator. It is therefore of crucial importance to all users of GIS that awareness of uncertainty and error should be as widespread as possible. A minimal response should be that users of the GIS be aware of the possible complications to their analysis caused by uncertainty, and at best present the user of the analysis with a report of the uncertainty in the final results together with a variety of plausible outcomes. A complete response to uncertainty is to present the results of a full modelling exercise which takes into account all types of uncertainty in the different data themes used in the analysis (Fisher, 1999). Interpretation of boundary description text to a geographic location would not be complete, if the uncertainties associated with this

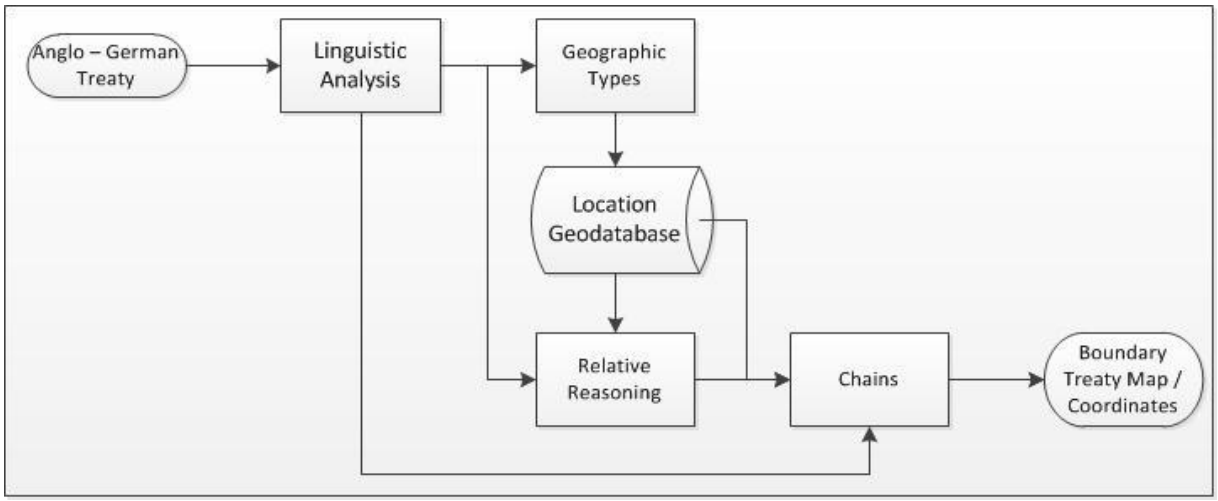
interpretation are not analysed and indicated at which level they occur. Such analysis makes the user to be made aware of what to expect and how to deal with it.

3. LINGUISTIC ANALYSIS

The Anglo – German Treaty, 1980 was used in this research as an example in the translation of the boundary description treaty text into geographic location format. This chapter was as a result of studying the approach used by (David N. Chin, 1994) when translating the text used on biological specimen into geographic location, and analysis of the boundary description text in the Anglo – German Treaty. By comparing the similarities and differences between them, a draft working method was devised.

A study of the report by (Boileau & Wallace, 1899) as members of the Anglo – German Boundary Commission also gave an insight on the basis of the boundary treaty agreements. Wallace made several expeditions to Africa for the sake of sports and had complied several tourist map of Central Africa on the places he had visited. Before the Anglo – German treaty was agreed, the Anglo – German commission was established to verify the maps and places Wallace had visited. The surveyor Boileau, led the surveying team for this commission. He established triangulation points and a list of coordinates for control points he established while confirming the maps done by Wallace. It is from their final report that the Anglo – German Treaty was based on. In their report, they plotted the tourist maps used by Wallace on his voyages in Central Africa and also the triangulation map produced by the surveyor, Captain Boileau. Such maps were thus used by the Anglo – German Commission as a basis for some of the boundaries alignment agreed in the treaty. Studying this report gave a better understanding of the basis of the treaty. By applying all these concepts and using pen and paper, procedures were formulated for the translation of the boundary description text into geographic location format. The formulated procedures highlight all the necessary steps required to be undertaken for analysing the boundary description text in the context it was written and interpreting the text geographically. Figure 4 shows all the devised procedures required to analyse the boundary description text and translate it into a geographic location format.

Figure 4: Procedures for translating the boundary description text into geographic location format

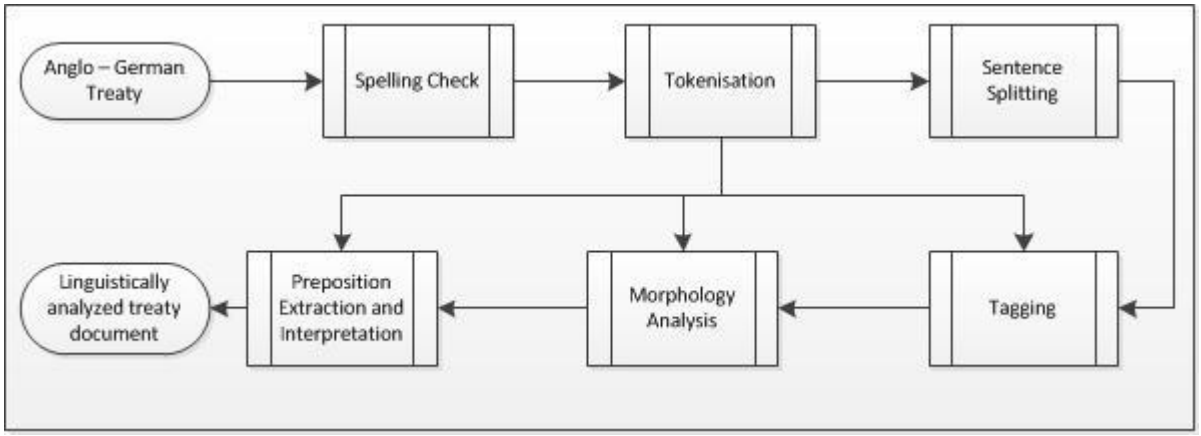


The formulated procedures have been divided into two sections;

- Linguistic analysis and
- Geographic interpretation.

Figure 5 shows all the steps that are involved in the linguistic analysis of the boundary description text.

Figure 5: Procedures involved in the linguistic analysis process of the boundary treaty text



The purpose of the linguistic analysis is to understand and analyse the English natural language used in the boundary treaty, in order to later interpret it geographically in the geographic interpretation process. This section has been divided into six main steps: spelling check, tokenisation, sentence splitting, tagging, morphology analysis and preposition interpretation. Each step has been explained in the following subsections below. Apart from the process of spelling checking, all the other five processes were analysed in GATE software (http://gate.ac.uk/). Gate software is open source free software used for the processing of text documents. GATE is over 15 years old and is in use for all types of computational tasks involving human languages. It is used for analysis of text in different languages like English, German, and French and for all shapes and sizes of documents. That is it can be used to analyse documents from word, portable device format (pdf) etc. of all sizes.

3.1. Spelling Check

This is the first step in the analysis the boundary description text. In this section, the boundary treaty document was analysed for spelling errors. According to (Peterson, 1980), there are two types of spelling programs: spelling checkers and spelling correctors. The spelling checker identifies words that are incorrect from a given input file of a text and the spelling corrector both detects the misspelled words and tries to find the most likely correct word from a reference dictionary. A reference dictionary is any dictionary like Oxford dictionary that can be uploaded into an on-line spelling checker to be used for correction of sentences. This algorithm only solves errors associated with English words. Spellings of local names either if correctly spelled or not, are therefore if not in the dictionary always highlighted. The spelling check therefore is used for two different purposes:

- To check any possible errors in the spelling of English words in the treaty and make appropriate corrections.
- To highlight any local names of places or features like names of rivers, mountains etc. used in the treaty that are not already in the reference dictionary. This acts as a check for the geographic types (look-up) sections, where such names are checked in the local gazetteer to see if they are there and if they are properly spelled and labelled as river, mountains etc. The geographic types sections thus contain a gazetteer of geographic names and their properties. E.g. the name Kilimanjaro has a property of being a range (mountain). As a check, the local names highlighted in the spelling check section, can be verified if in the geographic type section they have been highlighted i.e. if

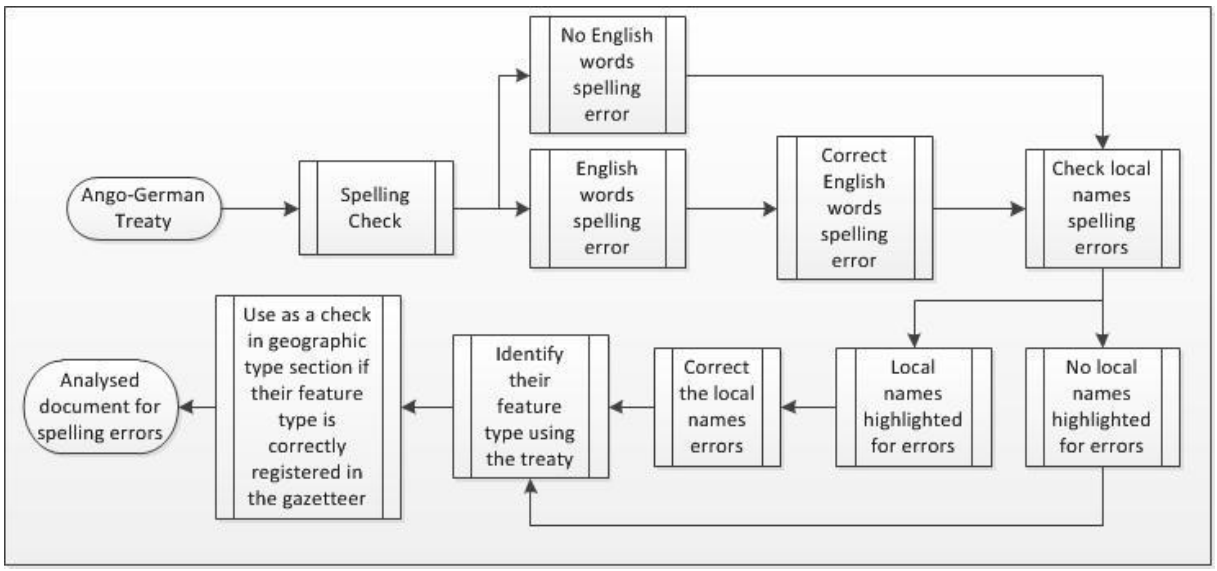
they are present in the gazetteer and if they are properly labelled in the gazetteer as rivers, mountains, etc. The proper labelling can be done by comparing the names in the gazetteer and their features against such names as they appear in the treaty document e.g. lake Victoria, Umba River are the way these names are used in the treaty document.

If no local name is highlighted in the spelling check, it means either;

- There are no local names in the treaty but still the geographic types section needs to be verified if the names in the boundary treaty are contained in the gazetteer or
- The local names in the treaty document already exist in the dictionary. E.g. the name Victoria already exists in most English online dictionaries. Also names of mountains like Kilimanjaro, which is the highest in Africa, also exist in most of the English dictionaries. In this case, still the geographic types section needs to be verified if the names in the boundary treaty are contained in the gazetteer.

Figure 6 shows the summary of the flow chart to be used in the spelling check section for analysing the boundary treaty document.

Figure 6: Procedures involved in the spelling check process of the boundary treaty text



3.2. Tokenisation

In tokenisation, texts are split into very simple tokens such as numbers, punctuations and words of different types. A distinction is also made between words of uppercase and lower case. Each individual word, number or punctuation in a sentence is taken as a separate token and is given its own unique identity. E.g. "To the north," would be split into four tokens, thus "To", "the", "north", and ",". These tokens are classified as to whether they start with an upperInitial letter as in "To" or they are all lowercase letters as in "the" and "north". Lastly, the length of the token is indicated, e.g. the length of "To" is 2 and "north" is 5.

The aim of tokenisation is to obtain maximum efficiency when processing the text as only the unique identity can be used whenever you use the token and also the process acts as a benchmark for all the other processes in the linguistic analysis, that is, the other processes cannot be run if the document has not been

tokenised. The tokeniser also maximises efficiency and enable greater flexibility by placing the burden on the grammar rules, which are more adaptable.
The boundary treaty text was thus tokenized using GATE software. Refer to Figure 19 for a pictorial view of tokenisation process in GATE software and Appendix D for the results obtained after tokenizing the first paragraph of the boundary description text.

3.2.1. Token Type

The following types of token are possible

a. Word: A word is defined as any set of contiguous upper or lower letters, including a hyphen (but no other form of punctuation). A word also has the attribute 'orth' (orthography - the art of writing words with the proper letters, according to accepted usage and correct spelling), for which four values are defined.
   - upperInitial – initial letter is uppercase, the rest are lower case
   - allCaps – all uppercase letters
   - lowerCase – all lowercase letters
   - mixedCaps – any mixture of upper and lowercase letters not included in the above category.

b. Number: A number is defined as a combination of consecutive digits. There is no subdivision of numbers.

c. Symbols: Two types of symbols are defined: currency symbols (e.g. $) and symbols (e.g. '&'). These are represented by any number of consecutive currency or other symbols (respectively).

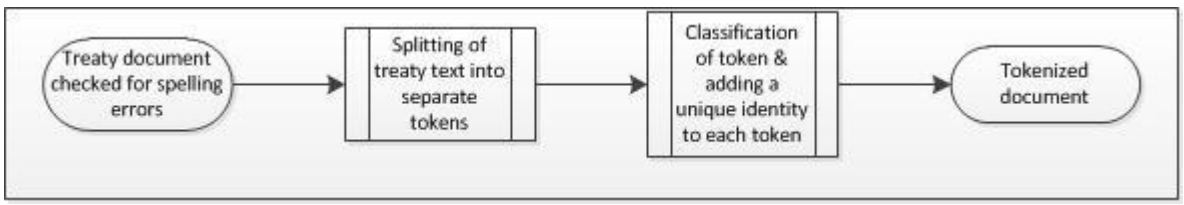
d. Punctuations: Three types of punctuations are defined: start punctuation (e.g. '('), end punctuation (e.g. ')'), and other punctuations (e.g. ':'). Each punctuation is a separate token.

e. Space token: White spaces between words are also taken as token and are divided into two types of space tokens- space and control-according to whether they are pure space characters or control characters

After tokenisation, the tokens have the following three features, length of the token, value or orth of the token and string of the token. For example, from Figure 19 and Appendix D, the length of the tokens "To" is 2 and its value or orth is UpperInitial, meaning "To" starts with an uppercase letter and the rest are lowercase letters. While the length of "north" is 5 and its value or orth is lowercase, meaning "north" is an all lowercase token. And the string of a token is the token itself i.e. the string of "To" is also "To" and the string of "north" is also "north".

Figure 7 shows the procedures involved when tokenising a document

Figure 7: Procedures involved in the tokenisation process of the boundary treaty text



3.3. Sentence Splitter

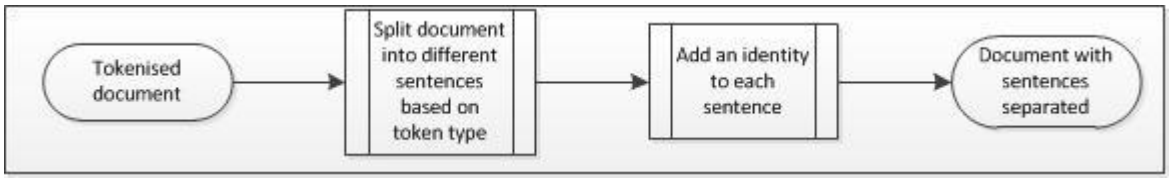
After tokenisation, the sentence splitter segments the boundary treaty text into individual sentences. The splitter uses a gazetteer list of abbreviations to help distinguish sentence-marking full stops from other kinds like commas etc. That is, after the tokenisation process, the sentence splitting section uses punctuation tokens like full stop (“.”), question mark (“?”) and space token to identify where a sentence starts and stops.

The sentence splitter is mainly to do with improving the execution time and robustness, especially when faced with irregular input. Irregular input can be inputs from different languages e.g. French, German etc. and also documents that are in different formats like word, portable devise format (pdf) etc. After the sentences have been spilt, each one is marked with a unique identity, so that it can always be called using its unique identity, thus also improving the execution time unlike if the whole sentence had to be written in full every time it has to be used.

The treaty document text was therefore split into individual sentences using GATE software. Refer to Figure 20 for sentence splitting process in GATE software and Appendix E for the results obtained after splitting the sentences of the first paragraph of the boundary description text.

Figure 8 shows the procedures to be followed when splitting the sentences of the boundary treaty document.

Figure 8: Procedures involved in the sentence splitting process of the boundary treaty text



3.4. Tagger

This process takes as an input a tokenized GATE document and adds a tag to the tokens. The correct tag is determined by using the rules of natural language and also comparing with the inbuilt dictionary that is in the GATE software. These tags label the tokens as nouns, adverbs, numbers etc. There are different tags used to denote different meanings of each token. Appendix A gives a list of all tags used and their interpretation.

The tagging process is an important process in the linguistic analysis of the document for the following reasons:

- In the process to follow of morphology analysis, verbs are split into root and affixes. E.g. “commences” and “runs” are split into roots of “commence” and “run” respectively and affix of “s” and “s” respectively. This increases efficiency in processing the document, by minimising the number of words stored in the dictionary. That is, instead of storing “commences, commence, commenced and commencing” separately as words, the dictionary would only store “commence” as a root and store “s, ed, ing” as affix that can be used to formulate the other words. This also helps use less memory for the processing, hence results are yielded faster. From tokenisation, therefore, all tags for verbs, i.e. VBD, VBG, VBN, VBP, VB, VBZ (meaning past tense, present participle, past participle, singular present, subjective and 3rd person singular present verbs) and

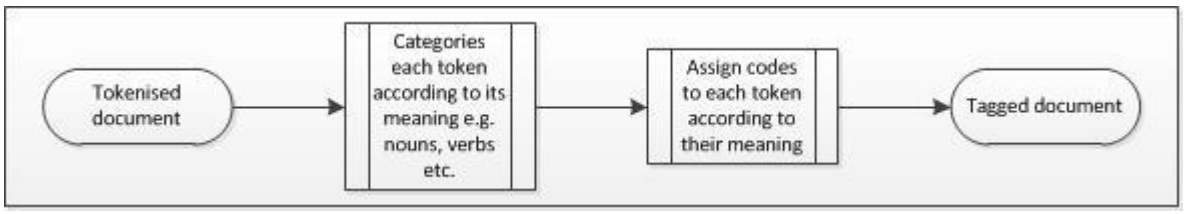
proper nouns (NNP) as from Appendix A, are extracted and used in the morphology analysis process.

- Cardinal directions like south, east, northern and western are also extracted from the treaty using their tags of JJ, RB and NN obtained from tokenisation. Refer to Appendix A for their interpretation. These cardinal directions are then used in the relative reasoning section.

Appendix A gives a summary of all the codes used for tagging in GATE software with their respective interpretations. Refer to Figure 21 for a pictorial view of tagging process in GATE software. After tagging the token, the "category" value is added to the features on the Token annotation. The "category" feature indicates the code that is assigned to each token. The code depends on whether the token is a proper noun, adjective, a number, punctuation, etc. E.g. the token "To" and "by" are assigned the code "TO" and "IN", translating to "literally to" and "preposition" according to Appendix A.

Figure 9 shows the procedures to be followed when tagging the treaty text in the tagging section.

Figure 9: Procedures involved in the tagging process of the boundary treaty text



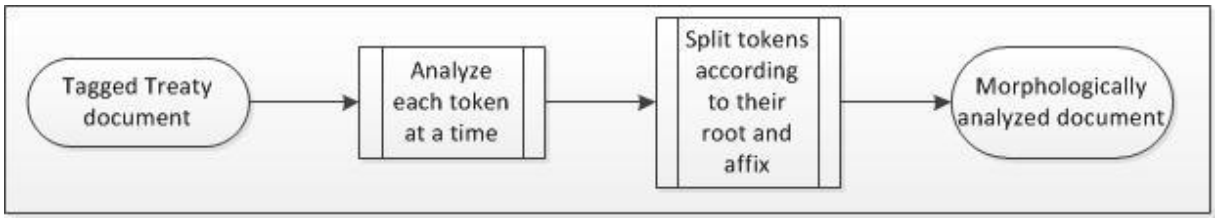
3.5. Morphology Analysis

The morphology analysis takes as an input a tokenized GATE document, considers each token and its tag from the tagging process, one at a time, it identifies its lemma (the structure of the smallest component of word or linguistic unit that has a semantic meaning) and its affix. The affix and root (lemma) values are then added as features on the Token annotation.
Refer to Figure 22 for morphology analysis in GATE software and Appendix D for results obtained after analysis the first paragraph of the boundary treaty text in GATE software.

For example, from Figure 22 it is seen that to the token "To", no affix or root features are added, while the token "commences" has an affix, "s" added and the root "commence" added to the feature category of the token. This process helps improve efficiency in processing the document. As explained in tagging process, the dictionary would only have to store the root of the word like "commence" and store affixes like "s, ed, ing" to form words like commences, commenced, commencing, instead of storing each word separately. Also words with plurals would have to be stored separately and affix like "s" added to make the plural. E.g. "village" would be stored separate and adding affix "s" to make the word "villages".

Figure 10 shows the procedure to be followed when running the text through morphology analysis stage.

Figure 10: Procedures involved in the morphology analysis process of the boundary treaty text



3.6. Preposition Interpretation

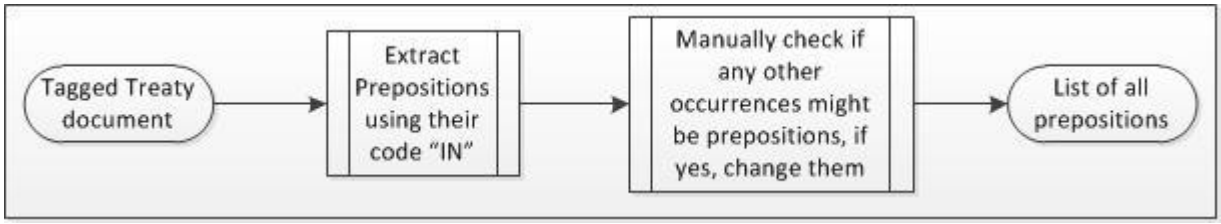
Location descriptions consist almost entirely of actual or implied prepositional phrases, (Futch, Chin, McGranaghan, & Lay, 1992). A typical location description contains several prepositional phrases and thus results in several relations. (Futch, et al., 1992), in their research on biological specimen came up with a table of prepositions, their relation meaning and the property and feature type on which they can be interpreted on. The main goal of the preposition interpretation was therefore to unify each of these prepositional phrases in the boundary treaty with possible relations. This was done by classifying prepositional relationships according to the type and/or form of the ground object that should accompany the preposition, according to whether the spatial reference was relative to an object in the landscape or if the relation implied a fixed point in space (Futch, et al., 1992).

The preposition interpretation is a set of rules that translates English prepositional phrases into relations that can be found and tested geographically. A preposition phrase is a preposition followed by noun-phrase, along with its semantic interpretation, (David N. Chin, 1994). From Appendix A, prepositions are tagged as "IN". However in the example used in this thesis, by using the definition of prepositions by (David N. Chin, 1994) above, some words which are prepositions, were not be tagged as such (refer to section 5.1.6 on results from preposition analysis for an example on this). For example the word "To" was tagged at "TO", meaning "literal to". Since in the example given, not all prepositions were tagged with "IN", which is a code for prepositions from the tagging process, it is therefore important to manually crosscheck if all the prepositions have been tagged as "IN" and if not, changes needs to be made on all the tokens that are prepositions but have not been labelled as such.

Table 2 shows the preposition phrases extracted from the boundary treaty text and their spatial interpretation while Appendix C shows a summarised list of 13 prepositions and their spatial meanings as modelled by (Freeman, 1975).

Figure 11 shows the flow of procedures to be followed when extracting the prepositions from the treaty text

Figure 11: Procedures involved in the prepositions extraction from the boundary treaty text



4. GEOGRAPHIC INTERPRETATION

This section takes as an input a document that has passed through all the stages in linguistic analysis, and provides procedures for interpreting the text in the document into geographic location format. This section has four subsections, thus, geographic types, location geodatabase, relative reasoning and chains. Refer to figure 4 for the flow chart showing how these processes are linked.

4.1. Geographic Types

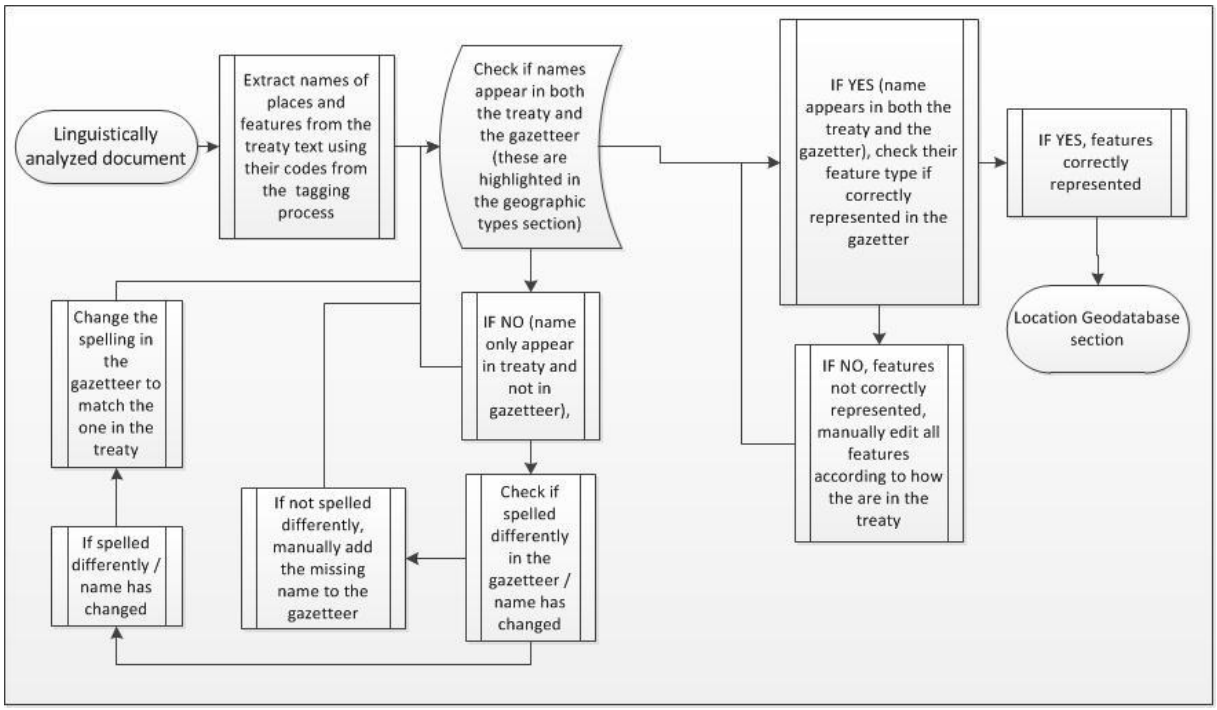
The purpose of the geographic types block was to categorise names of places in the treaty as areas, lines or point objects. This category is very important in the relative reasoning section. For example when analysing prepositions, when the preposition "along" is met, according to Table 2, it can only be associated with a linear feature. Also spatial queries of intersection, joining, touching etc. will yield different results if the features concerned are areas only or area and line, etc. Proper categorising in a tabular form of the feature in the treaty is therefore important.

In this section therefore, first, all names categorised as NNP (proper nouns) are extracted from the treaty document. A look-up gazetteer that contains names of different places and feature is then used to check if the names extracted from the treaty document appear in this gazetteer.
If the names of places and features like lakes in the treaty are also present in the look-up gazetteer, they are then highlighted in the boundary treaty text in GATE software. By clicking on the highlighted names, you can see its feature as provided by the gazetteer. E.g. by clicking on Victoria, shows that it was classified in the gazetteer as a name of a female person, a city and also a province. In the tagging process, names of features like lakes, rivers, mountains, are also categorised as NNP (proper nouns). This therefore helps to compare the feature types as written in the treaty text with the feature types as highlighted in the gazetteer and make appropriate changes in the gazetteer list.
There are a lot of names that in the tagging process are highlighted as NNP (proper nouns) but are not highlighted in the geographic types section since they do not appear in the gazetteer. Such occurrences have to be manually added to the gazetteer.

Refer to section 5.2.1 and Figure 23 for results obtained after running the geographic types section in GATE software and a discussion on geographic types section.

Figure 12 shows the procedures to be followed when analysing the document through the geographic types section.

Figure 12: Procedures involved in the geographic types process of the boundary description text

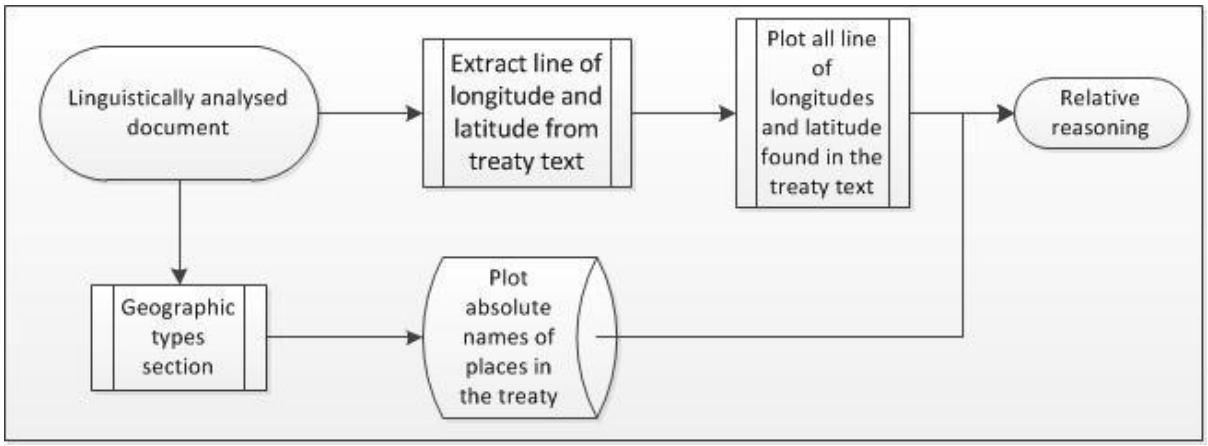


4.2. Location Geodatabase

The location geodatabase is a database that contains names from the gazetteer and their coordinates. In this section therefore, all the absolute places appearing in the treaty after being verified in the geographic types section were plotted using their coordinates from the location geodatabase. This gives the first insight impression of the boundary line. All the remaining locations on the ground, are therefore classifies as relative positions, and are determine in the relative reasoning section.

Locations that followed line of latitude e.g."33rd degree of east latitude" (in the treaty, they are written like this and not "33rd degree latitude east") were also plotted as absolute points. This section therefore categorises the locations as absolute and relative positions and plots only the absolute positions. Figure 13 shows the procedures to be followed when analysing the treaty text in the location geodatabase section.

Figure 13: Procedures involved in the location geodatabase process of the boundary description text



4.3. Relative Reasoning

Relative reasoning contained two steps, understanding of spatial relations and qualitative reasoning about distance and directions in geographic space.

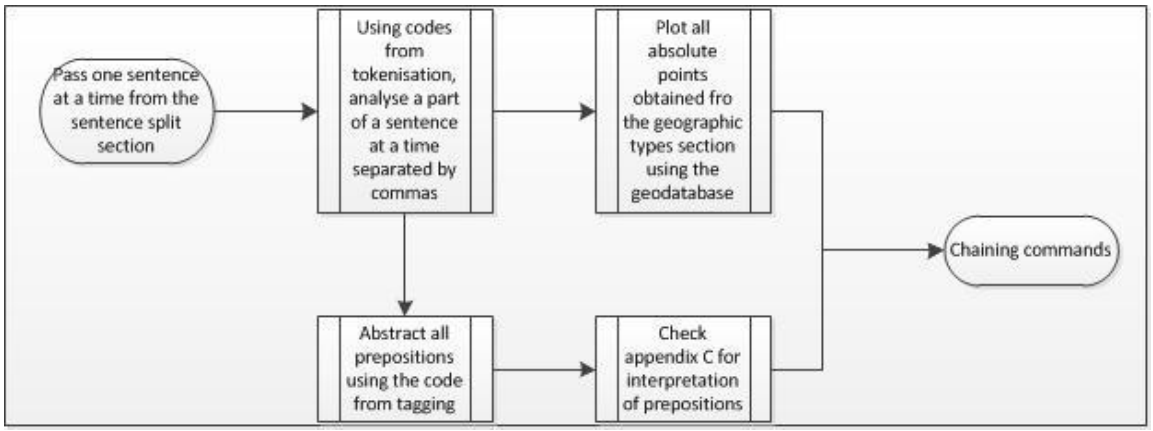
4.3.1. Understanding spatial relations

This step involved understanding the meaning of spatial relationships used in the treaty document. A thorough understanding of prepositions, leads to proper interpretation of the treaty documents. (Freeman, 1975), came up with a summarized list of thirteen spatial relationships along with their corresponding prepositions. Appendix C is therefore a summary of all spatial relations along with the corresponding prepositions.

Before we can have relations between objects, we need to have some representation of the object themselves. This means the relationship depends on whether the object is a point, line or area and this is done in the geographic type section. If we have some representation for the object, we can then decide what relationships apply to those objects.

Relative reasoning gives a flow of chaining commands to follow to make the best decision when interpreting prepositional phrases and analysing spatial relations. Figure 14 shows the chaining commands to follow when interpreting the treaty using relative reasoning.

Figure 14: Procedures involved in the relative reasoning about spatial relationships in the boundary description text



- Each sentence is analysed one at a time following the sentence split process in which the text was split into different sentences.
- Since sentences of the treaty have different sections separated by commas, each section is analysed on its own to interpret its meaning.
- All the locations identified using the codes of proper noun in the geographic type process are plotted first using the location geodatabase.
- Prepositions are also extracted from the treaty using their codes from tagging.
- By reference to Appendix B on prepositions, their spatial relations and their property or type of feature they act on, can be determined.
- Chaining commands as illustrated in Chapter 6, can then be followed in order to come up with the final coordinates.

4.3.2. Understanding cardinal directions in the treaty

Qualitative spatial reasoning (QSR) abstracts metrical details of the physical world. The two main directions in QSR are topological reasoning about regions and reasoning about orientation of point configurations. Orientations can refer to a global reference system, e.g. cardinal directions. (Moratz & Ragni, 2008) use QSR to abstract details of the physical world and enable computers to make predictions about spatial relations even when precise quantitative information is unavailable.

In natural language and many practical applications, topological and directional relations are used together, (Li, 2007). In the boundary treaty text, a lot of cardinal directions like northern, eastern and westwards are used. (Frank, 1992), gives a set of qualitative deduction rules for a subset of spatial reasoning, namely reasoning with cardinal directions and qualitative distance descriptors, without relying on qualitative calculations (e.g. square root, trigonometric functions or analytical geometry). In his approach, it is clear that qualitative approach loses some precision, but simplifies reasoning and allows deductions when precise information is not available.

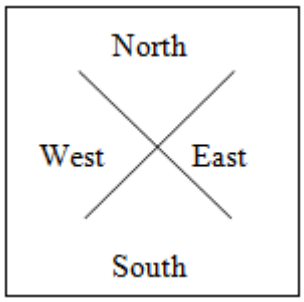
He uses a qualitative distance function which determines from two points on the plane onto a qualitative distance value. Similarly, he uses a qualitative directions function between two points in the plane to determine a symbolic direction or its equivalent, from a path onto a symbolic direction.

In his function, he uses the directions of N, NE, E, SE, S, SW, W, and NW. These techniques were used to understand how to calculate the distance and direction from qualitative information in the boundary description text.

4.3.2.1. Cardinal directions as cones

The most often used prototypical concept of cardinal directions is related to the angular direction between the observer's position and a destination point. This direction is rounded to the next established cardinal direction. This results is a cone-shaped areas for which a symbolic direction is applicable (Frank, 1992). Figure 15 illustrates the cardinal directions used as cones.

Figure 15: Cardinal directions as cones

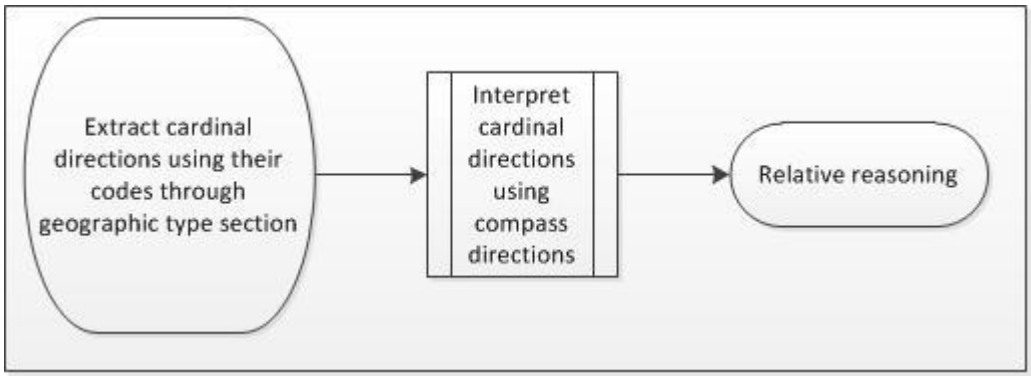


This model of cardinal directions has the property that 'the area of acceptance for any given direction increases with distance' and is sometimes called triangular model. This is the concept that will be used in treating cardinal directions in the boundary treaty.

Cardinal directions in the treaty are to be interpreted as illustrated in Table 6 and Figure 15. They are to follow the directions of the compass and by adding the cone theory as earlier explained in 4.3.2.1, the decision making would be simplified.

Figure 16 shows the chain of commands to be followed to analyse the cardinal directions.

Figure 16: Procedures involved in interpreting cardinal directions in the boundary treaty text

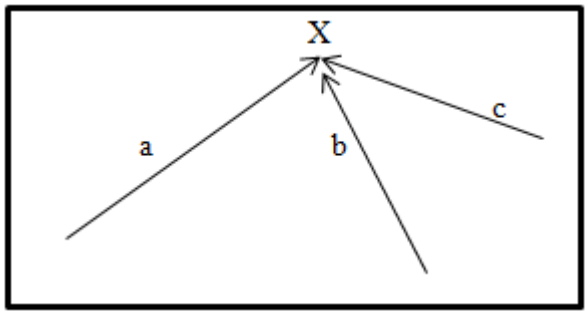


4.4. Chains

A chaining command is a series of steps that are to be followed when analysing the boundary treaty text. These commands give the decision maker the option of when to make which decision and why to make it. They give a set of rules to be followed in order to come to the desired outcome. They are written as flow diagrams. Refer to Chapter 6 for chain of commands flow diagrams. That chapter identifies the types of uncertainties that occur when translating the boundary description text into geographic location. A set of chaining commands are then provided illustrating the paths that are to be taken in order to make the best decisions to interpret the boundary description text into geographic location format.

(Frank, 1992), discusses on the algebraic approach for a qualitative chain of reasoning to deduce a location of a point. If a position is determined by more than one chain of reasoning, then one selects the intersection of all results. This is a widely used convention, assuring that the result of a deduction chain is certainly containing the correct value. Figure 17, illustrates the deduction of a point using more than one chain of commands. Point X has been deduced using executions of sentence a, b and c as indicated, leading to better position of X, than if only one or two sentences were used.

Figure 17: Illustration of deduction of a point using more than one chain of commands



5. RESULTS

For the analysis and results of the treaty, only the first part of the Anglo – German Treaty was used.

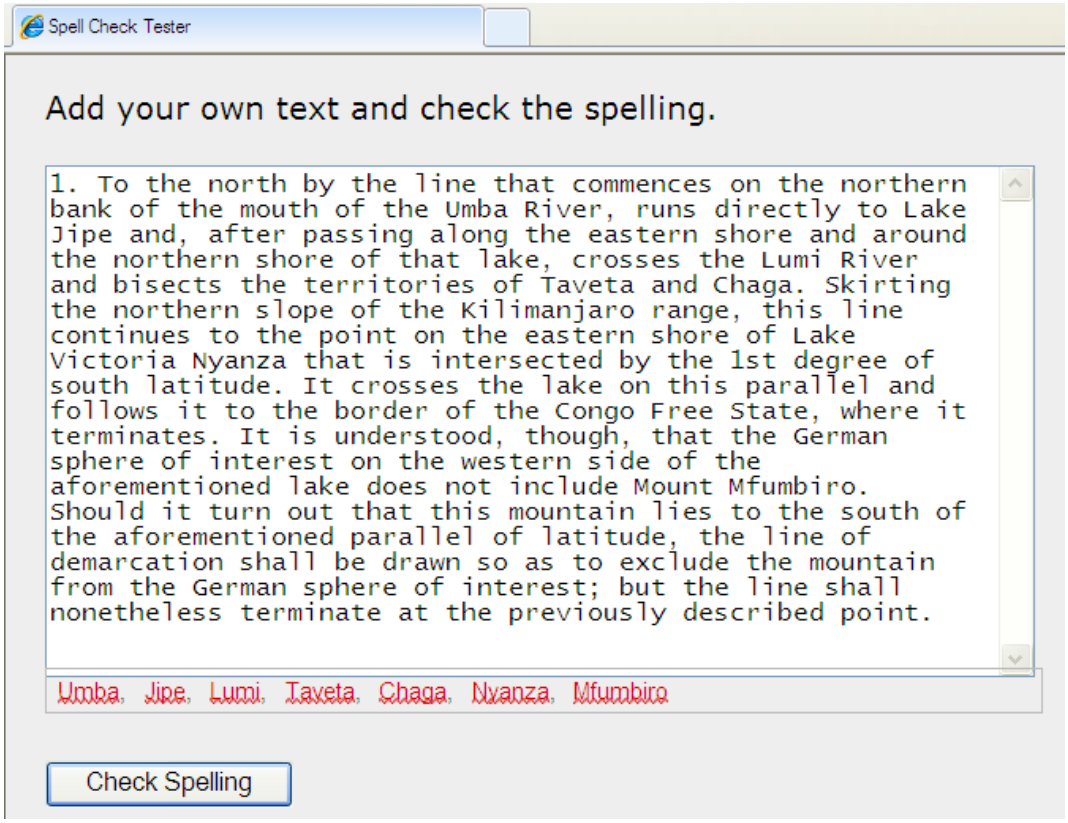
Paragraph_1: *"To the north by the line that commences on the northern bank of the mouth of the Umba River, runs directly to Lake Jipe and, after passing along the eastern shore and around the northern shore of that lake, crosses the Lumi River and bisects the territories of Taveta and Chaga. Skirting the northern slope of the Kilimanjaro range, this line continues to the point on the eastern shore of Lake Victoria Nyanza that is intersected by the 1st degree of south latitude. It crosses the lake on this parallel and follows it to the border of the Congo Free State, where it terminates. It is understood, though, that the German sphere of interest on the western side of the aforementioned lake does not include Mount Mfumbiro. Should it turn out that this mountain lies to the south of the aforementioned parallel of latitude, the line of demarcation shall be drawn so as to exclude the mountain from the German sphere of interest; but the line shall nonetheless terminate at the previously described point"*

5.1. Linguistic Analysis

5.1.1. Spelling Check

After running the paragraph_1 of the treaty document, no English words were found to be misspelled indicating that the English words had been thoroughly checked for errors before the boundary treaties were signed. However, seven local names were highlighted, indicating that they were not in the dictionary used by the online checker. Figure 18 shows the results after running the paragraph in the online spelling checker.

Figure 18: Results (print screen) of spelling checking process of the boundary description text



From Figure 6, on procedures to be followed in the spelling check section, since local names were highlighted, their feature types had to be identified from the treaty to be used in the geographic type section. By looking at the paragraph of the treaty, the feature types of all the seven highlighted names were identified e.g. Umba is referred to as a river in the treaty, so its feature type is a river. Table 1, gives a summary of all the seven highlighted names in the spelling check process and their corresponding type category obtained from the boundary treaty. This table acted as a check in the geographic type section by comparing it to appendix F on results obtained from geographic type section. Refer to section 5.2.1 on geographic type for comparison and analysis of this.

However, there is name confusion by the highlighted name of Nyanza. Nyanza is a common east and southern African way of referring to any lake. The above could be ready and understood by the locals as Lake Victoria Lake. Most names used in the boundary treaty, were acquired by Captain C.F. Boileau and L.A.Wallace as Members of The Anglo - German Boundary Commission. A thorough report is given in The Nyasa-Tanganyika plateau report (Boileau & Wallace, 1899). It is clear in their report that some of the name mismatch, could have come from relying on the locals too much without a second way of verifying

the names as they could not know all names and places. For example, they wrote in their report, "They soon got into the way of finding out and telling me the names of the rivers, mountains, villages, etc., which we passed, and of giving me such other information as they knew I was likely to want, and I am afraid that some were not above occasionally shamelessly inventing such names and information rather than own their ignorance."
A gazetteer of local names could be of importance in such instances to verify the actual name of places.

Table 1: Summarised results of the highlighted names from spelling checker with their feature types

Name	**Type**
Umba	River
Jipe	Lake
Lumi	River
Taveta	Territory
Changa	Territory
Nyanza (Lake Victoria Nyanza	? referring to Lake Victoria
Mfumbilo	Mount

5.1.2. Tokenisation

Running again paragraph_1 of the treaty document in GATE software for Tokenisation, gave the result in Figure 19 as a print screen but refer to appendix D for table with full results:

The shaded text is the paragraph_1 of the treaty text that was used as an input in GATE software for tokenisation. After running the text for tokenisation, according to Figure 7 on steps to be followed when tokenising a document, the following things were observed;

- The text in paragraph_1 was split into different individual tokens as separated by the dashed white lines in Figure 19, e.g. some of the tokens are, "To", "the", and "north".
- Each token was assigned a start node and an end node, the blank spaces were not assigned any nodes. The counting of the nodes started at 0. E.g. the first token "To" started at node 0 and ended at node 2 and the next token "the" started at node number 3 and ended on node number 6. The difference between the end node and the start node is the length of the token.
- Each token was assigned a unique identity as indicated in the "id" section of Figure 19. E.g. the identities for the first token "To" and "the" were 6 and 8 respectively.

The tokens were also split into the following features:

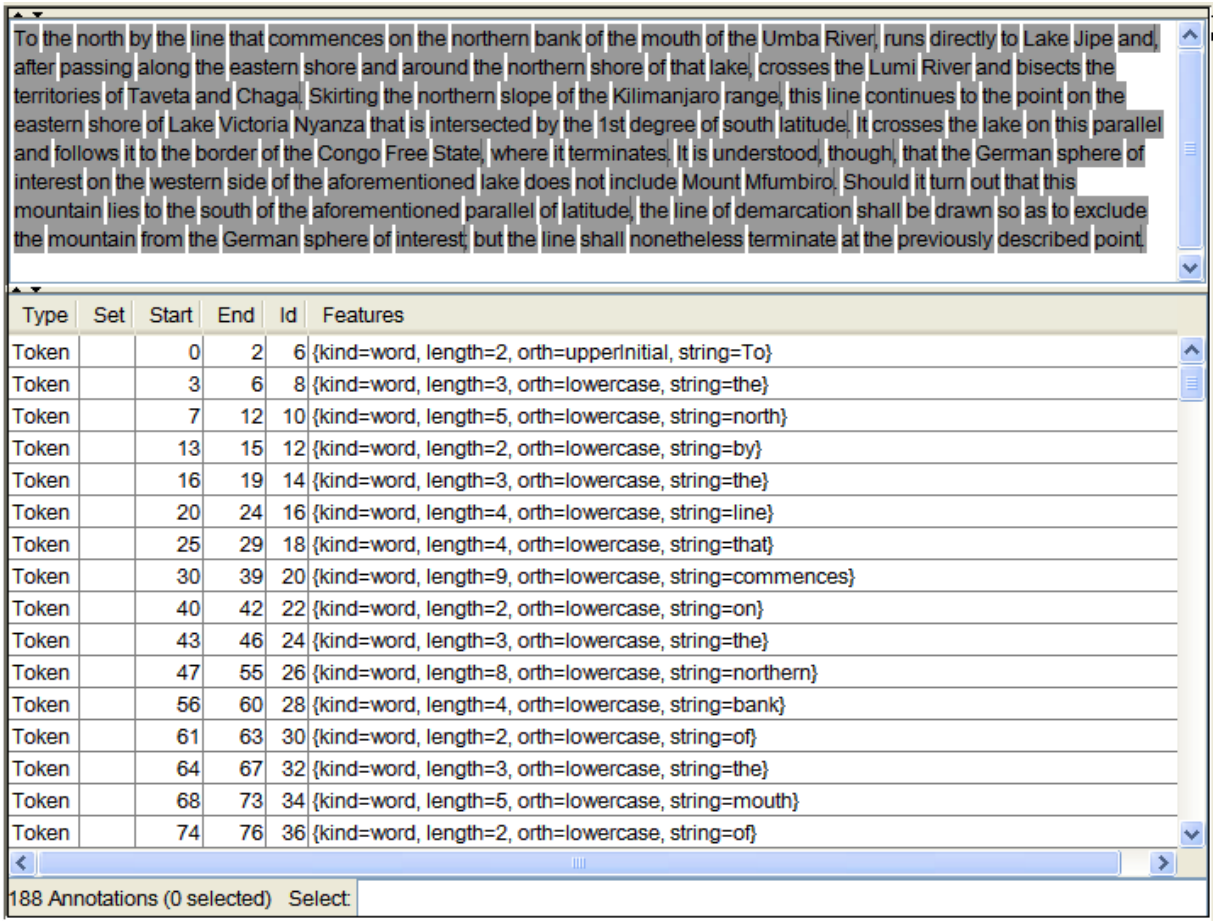
- Kind – This indicated whether the token is a word, number, symbol, punctuation or space token as earlier explained in section 3.2.1. e.g. kind of "To" is word and kind of "," is punctuation.
- Length – This indicated the length of the token which is also the difference between the end node and the start node of the token. E.g. length of "To" is 2 and length of "north" is 5.
- Orth / Value – This as earlier explained in section 3.2.1 indicated whether the tokens initial letter is uppercase and the rest of the letters are lowercase, or a token is an all uppercased word, or is an all lowercased letter word or if has a mixture of upper and lowercase. E.g. the orth of "To" is upperInitial and orth of "north" is lowercase, meaning "To" starts with a

uppercase letter and the rest of the letter are lowercased letters while "north" is an all lowercase lettered word.

- String – The string of the token is the token itself. E.g. the string of "To" is also "To" and the string of "north" is also "north".

188 tokens were annotated from tokenizing the first paragraph of the boundary text. These tokens therefore acted as a basis for the following sections in Linguistic analysis.

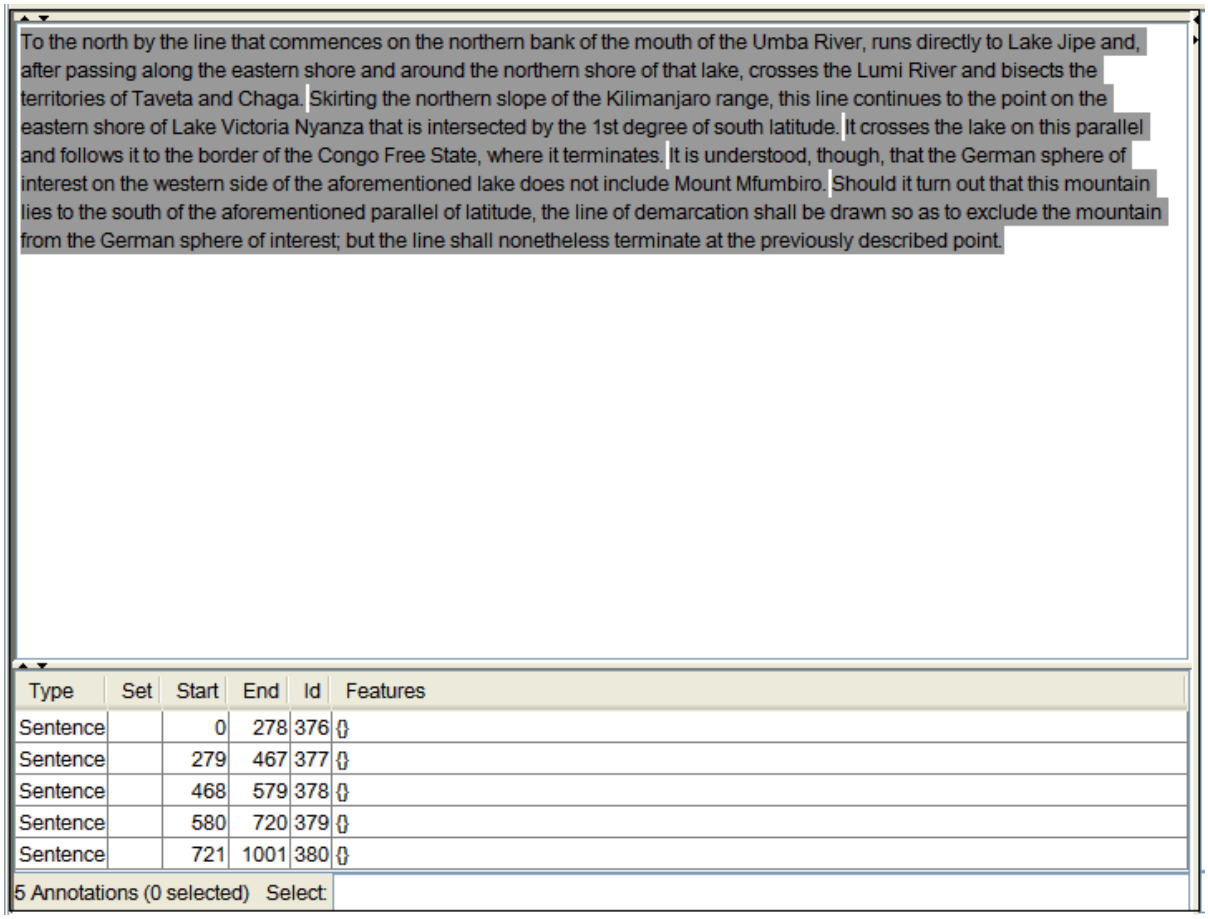
Figure 19: Results (print screen) of tokenisation process of the boundary treaty text in GATE software



5.1.3. Sentence Splitter

Running the same paragraph_1 in GATE software for sentence splitting, gave the results in Figure 20 as a print screen but refer to Appendix E for full results:

Figure 20: Results (print screen) of sentence splitting process of the boundary treaty text in GATE software



The shaded text in Figure 20 is the paragraph_1 of the treaty text that has been used as an input for the sentence splitting section. After running paragraph_1 in GATE software for sentence splitting and according to figure 8 on the procedures to be followed when tokenising a document, the following were observed.

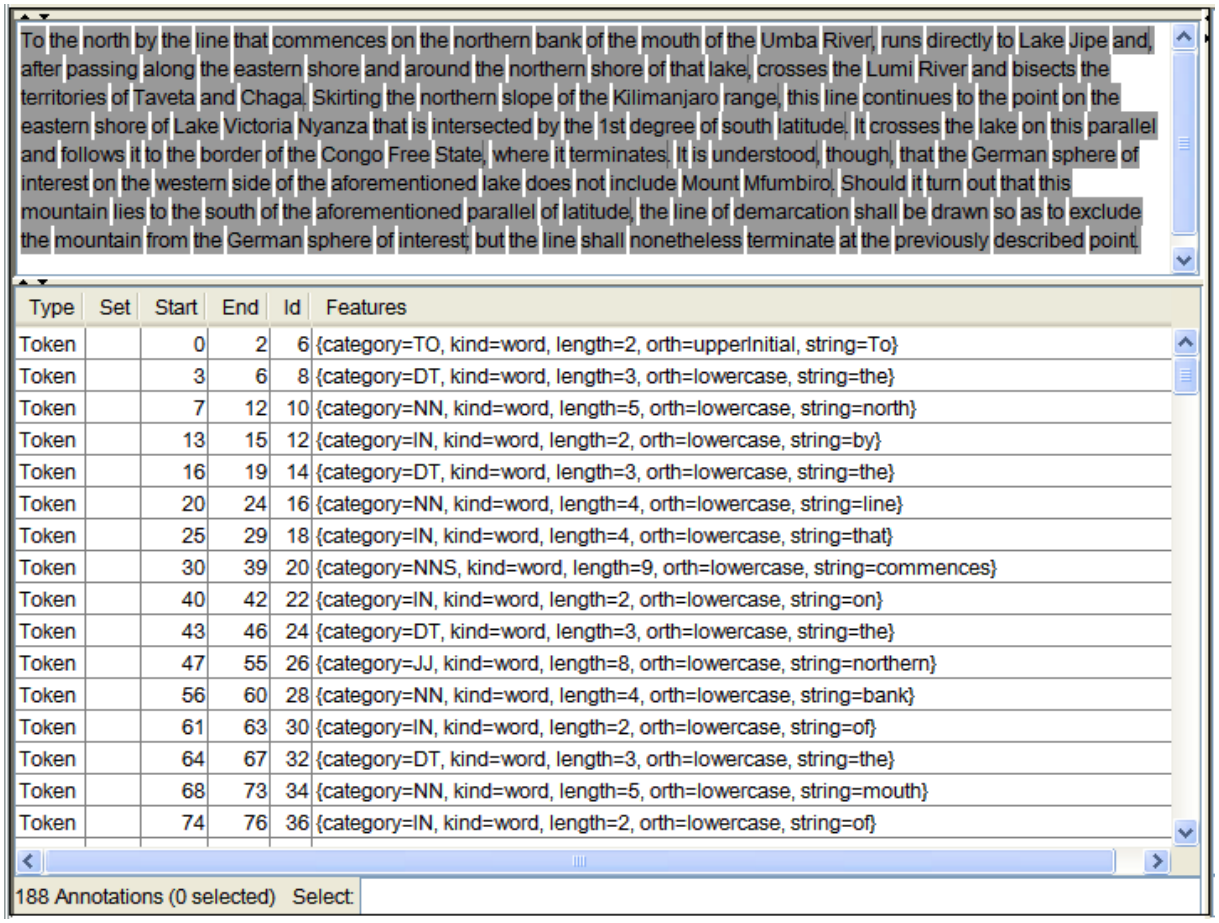
- Paragraph_1 was dividing into five sentences based on sentence ending punctuation as analysed in the tokenisation process i.e. the sentences were split using full stop tokens.
- Each sentence was given a start node and an end node. Each letter in the sentences being counted as a node. E.g. the first sentence starts with node 0 and ends with node 278. The spaces are not counted as nodes.
- Each sentence was given a unique identity. E.g. the identity of the first sentence was 376. The identity is the unique character used to call for individual sentences in any further processing.

Appendix E gives the complete table of results obtained from sentence splitting procedures.

5.1.4. Tagger

Running the same paragraph_1 in GATE software for tagging, gave Figure 21 as a print screen but refer to Appendix D for table with full results:

Figure 21: Results (print screen) of tagging process of the boundary treaty text in GATE software



The shaded text in Figure 21 is the paragraph_1 of the treaty text that has been used as an input for the tagging section. After running paragraph_1 in GATE software for tagging and according to Figure 9 on the procedures to be followed when tagging a document, the following were observed.

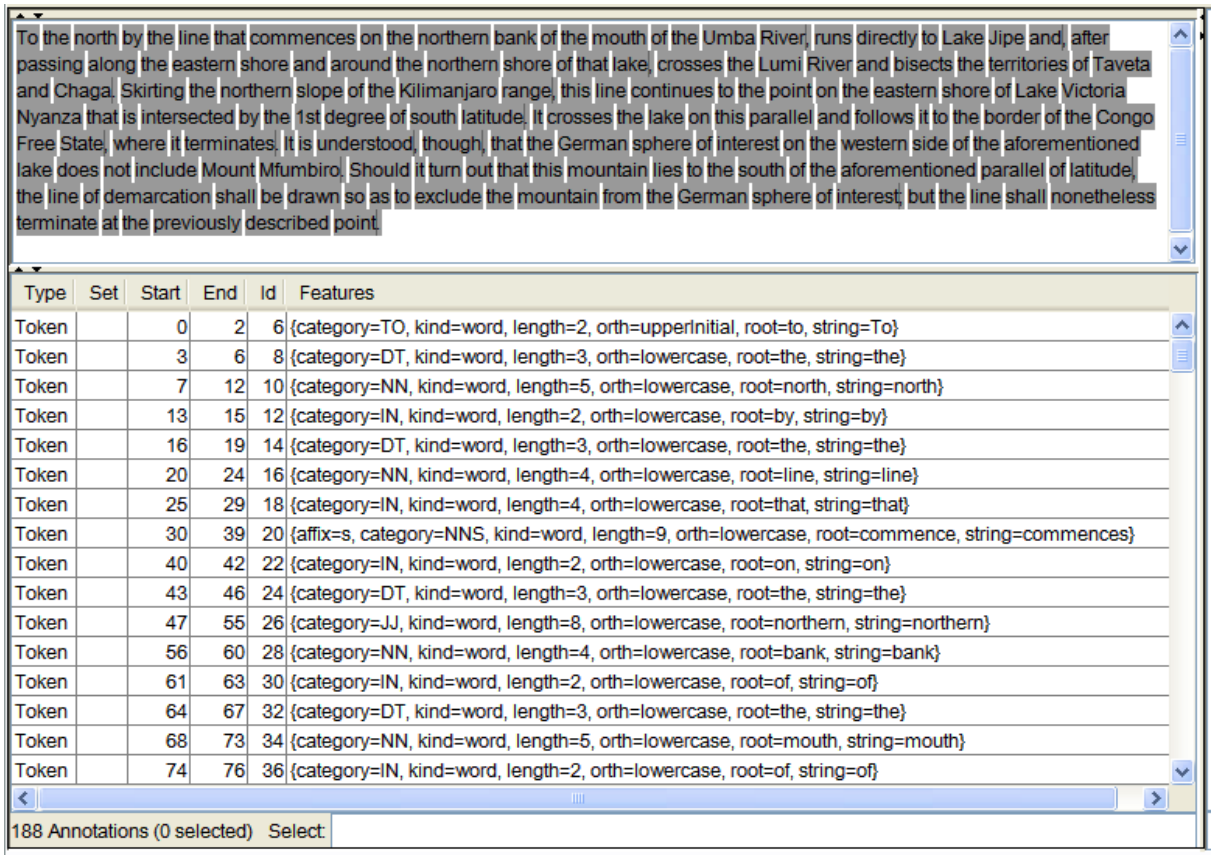
- The number of tokens tagged was the same number of tokens obtained from the tokenisation process, i.e. 188 tokens were tagged.
- The start node and end nodes of the tokens were just as they were obtained from tokenisation process
- The identity of the tokens are as obtained from the tokenisation process
- However, in the feature section, a new feature of "category" was added to each token as a result of the tagging process. The category feature used codes which according to Appendix A, can be interpreted as nouns, adverbs, numbers, punctuation etc. It is from these codes that preposition, cardinal directions etc. are extracted from in the following sections. E.g. the tokens "To" and "the", have been categorised as "TO" and "DT" respectively. From Appendix A, this translates to mean "literal to" and "determiner" respectively.
- The other features of "kind", "length", "orth" and "string", were as they were analysed in the tokenisation process.

For a complete list of the meanings of all tags used in GATE software, refer to Appendix A.

5.1.5. Morphology Analysis

Running the same paragraph_1 in GATE software for morphology analysis, gave Figure 22 as a print screen but refer to Appendix D for table with full results:

Figure 22: Results (print screen) of morphology analysis process of the boundary treaty text in GATE software



To the north by the line that commences on the northern bank of the mouth of the Umba River, runs directly to Lake Jipe and, after passing along the eastern shore and around the northern shore of that lake, crosses the Lumi River and bisects the territories of Taveta and Chaga. Skirting the northern slope of the Kilimanjaro range, this line continues to the point on the eastern shore of Lake Victoria Nyanza that is intersected by the 1st degree of south latitude. It crosses the lake on this parallel and follows it to the border of the Congo Free State, where it terminates. It is understood, though, that the German sphere of interest on the western side of the aforementioned lake does not include Mount Mfumbiro. Should it turn out that this mountain lies to the south of the aforementioned parallel of latitude, the line of demarcation shall be drawn so as to exclude the mountain from the German sphere of interest, but the line shall nonetheless terminate at the previously described point.

Type	Set	Start	End	Id	Features
Token		0	2	6	{category=TO, kind=word, length=2, orth=upperInitial, root=to, string=To}
Token		3	6	8	{category=DT, kind=word, length=3, orth=lowercase, root=the, string=the}
Token		7	12	10	{category=NN, kind=word, length=5, orth=lowercase, root=north, string=north}
Token		13	15	12	{category=IN, kind=word, length=2, orth=lowercase, root=by, string=by}
Token		16	19	14	{category=DT, kind=word, length=3, orth=lowercase, root=the, string=the}
Token		20	24	16	{category=NN, kind=word, length=4, orth=lowercase, root=line, string=line}
Token		25	29	18	{category=IN, kind=word, length=4, orth=lowercase, root=that, string=that}
Token		30	39	20	{affix=s, category=NNS, kind=word, length=9, orth=lowercase, root=commence, string=commences}
Token		40	42	22	{category=IN, kind=word, length=2, orth=lowercase, root=on, string=on}
Token		43	46	24	{category=DT, kind=word, length=3, orth=lowercase, root=the, string=the}
Token		47	55	26	{category=JJ, kind=word, length=8, orth=lowercase, root=northern, string=northern}
Token		56	60	28	{category=NN, kind=word, length=4, orth=lowercase, root=bank, string=bank}
Token		61	63	30	{category=IN, kind=word, length=2, orth=lowercase, root=of, string=of}
Token		64	67	32	{category=DT, kind=word, length=3, orth=lowercase, root=the, string=the}
Token		68	73	34	{category=NN, kind=word, length=5, orth=lowercase, root=mouth, string=mouth}
Token		74	76	36	{category=IN, kind=word, length=2, orth=lowercase, root=of, string=of}

188 Annotations (0 selected) Select:

The coloured text in Figure 22 is the paragraph_1 of the treaty text that has been used as an input for the morphology analysis section. After running paragraph_1 in GATE software for morphology analysis and according to Figure 10 on the procedures to be followed when morphologically analysing a document, the following were observed.

- Start and end nodes were counted in the same way as in the tokenisation process.
- Different identities were used for the token as now they had been split into roots and affixes
- Tokens had been split into roots and affixes.
- A new feature of root and affix was added to the feature column, e.g. the token "commences" has been split into the root "commence" and the affix "s".
- All the rest of the features of category, kind, length, orth and string, were as analysed form the tagging and tokenisation process.

Morphology analysis helped in improving the efficiency of analyzing the documents as it helped keep a smaller dictionary by only keeping the roots of words and adding affixes to form different words, instead of storing all the words separately. E.g. by storing commence as a root and adding "s", "ed" and "ing" as affixes, you can form commences, commenced and commencing, instead of storing them as they are.

5.1.6. Preposition Interpretation

The preposition interpretation is a set of rules that translates English prepositional phrases into relations that can be found and tested geographically. A preposition phrase is a preposition followed by noun-phrase, along with its semantic interpretation, (David N. Chin, 1994). From Appendix A, prepositions are tagged as "IN". However using the definition of prepositions by (David N. Chin, 1994) above, it is important to manually crosscheck if all the prepositions have been tagged as "IN". For example, from Figure 21 on tagging and appendix D, all the tokens used in the example were properly tagged with "IN" except for the token "To". "To" was tagged as "TO", translated according to Appendix A as meaning "literally to". Such occurrences then, need to be manually checked to make sure that all tagging has been done correctly. Table 2 shows the preposition phrases extracted from the boundary treaty text and their spatial interpretation while Appendix C shows a summarised list of 13 prepositions and their spatial meanings as modelled by (Freeman, 1975).

Table 2: Prepositions extracted from the treaty document and their spatial relations

Preposition(In treaty)	Relation	Property-Feature type
To (TO-literal to)	Towards	Oriented object/observer
By	Near	
On	Within	
Of	Within	
After	Beyond	
Around	Near	
From	From	
At	Within	
Along (RB-adverb)	Near	Linear feature

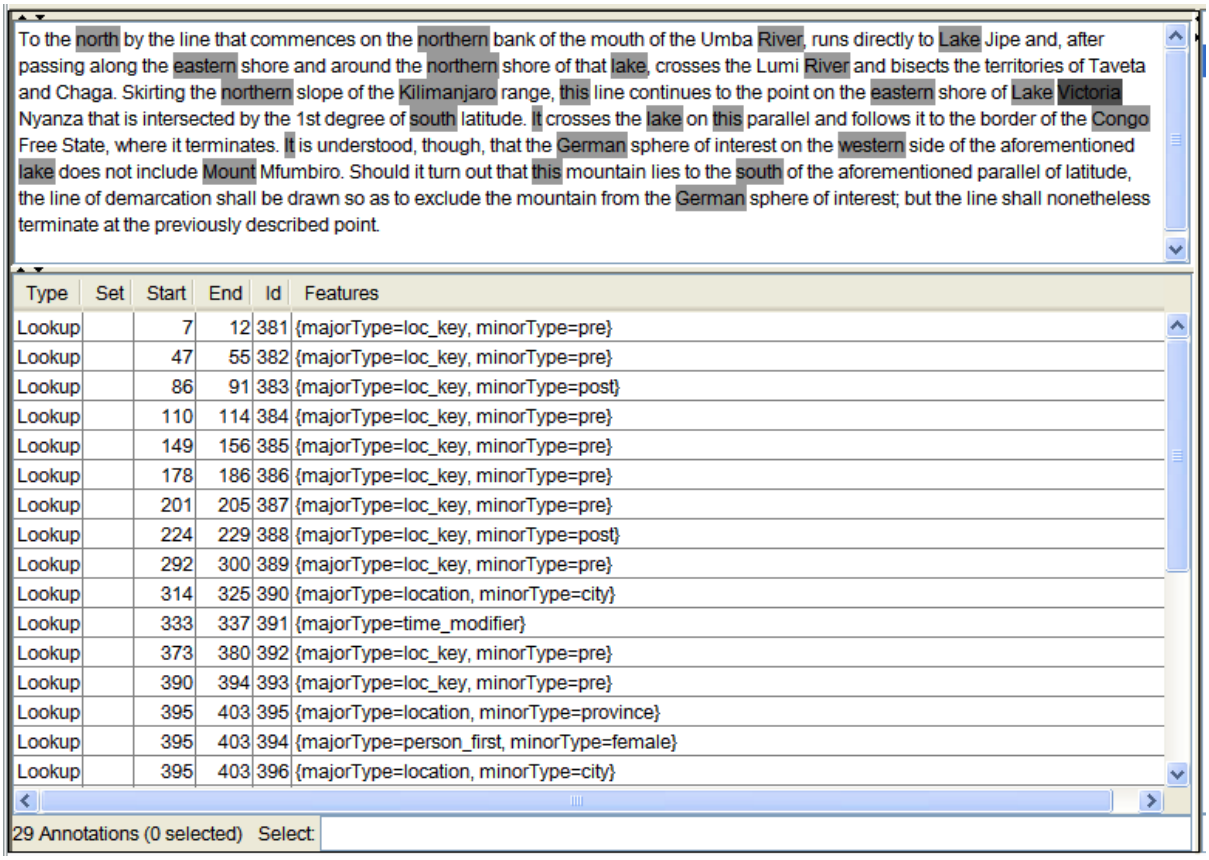
For a complete list of all the prepositions and their relations as modelled by (David N. Chin, 1994), see Appendix B.

5.2. Geographic Interpretation

5.2.1. Geographic Types

Loading the default world-wide gazetteer that comes with GATE software and running the geographic types or lookup table, gave Figure 23. For full results, refer to Appendix F.

Figure 23: Results (print screen) of geographic types process of the boundary treaty text in GATE software



To the north by the line that commences on the northern bank of the mouth of the Umba River, runs directly to Lake Jipe and, after passing along the eastern shore and around the northern shore of that lake, crosses the Lumi River and bisects the territories of Taveta and Chaga. Skirting the northern slope of the Kilimanjaro range, this line continues to the point on the eastern shore of Lake Victoria Nyanza that is intersected by the 1st degree of south latitude. It crosses the lake on this parallel and follows it to the border of the Congo Free State, where it terminates. It is understood, though, that the German sphere of interest on the western side of the aforementioned lake does not include Mount Mfumbiro. Should it turn out that this mountain lies to the south of the aforementioned parallel of latitude, the line of demarcation shall be drawn so as to exclude the mountain from the German sphere of interest; but the line shall nonetheless terminate at the previously described point.

Type	Set	Start	End	Id	Features
Lookup		7	12	381	{majorType=loc_key, minorType=pre}
Lookup		47	55	382	{majorType=loc_key, minorType=pre}
Lookup		86	91	383	{majorType=loc_key, minorType=post}
Lookup		110	114	384	{majorType=loc_key, minorType=pre}
Lookup		149	156	385	{majorType=loc_key, minorType=pre}
Lookup		178	186	386	{majorType=loc_key, minorType=pre}
Lookup		201	205	387	{majorType=loc_key, minorType=pre}
Lookup		224	229	388	{majorType=loc_key, minorType=post}
Lookup		292	300	389	{majorType=loc_key, minorType=pre}
Lookup		314	325	390	{majorType=location, minorType=city}
Lookup		333	337	391	{majorType=time_modifier}
Lookup		373	380	392	{majorType=loc_key, minorType=pre}
Lookup		390	394	393	{majorType=loc_key, minorType=pre}
Lookup		395	403	395	{majorType=location, minorType=province}
Lookup		395	403	394	{majorType=person_first, minorType=female}
Lookup		395	403	396	{majorType=location, minorType=city}

29 Annotations (0 selected) Select:

After running paragraph_1 in GATE software for geographic type section, the following were observed;

- Start and end nodes for the highlighted names were indicated.
- A unique identity was assigned to each highlighted name
- In the feature type, the features major type were shown in which group the feature belonged and minor type which showed the attribute of the major type were also added. E.g. Kilimanjaro was indicated as a location in major type and a city in the minor type.

With reference to Appendix F and Figure 12 on the procedure to be followed when analysing the text in the geographic types section, the geographic section is useful for three reasons;

i. It is used for the extraction of names of places in the treaty and making sure that these names are also present in the loaded gazetteer. Names are extracted from the treaty text through the code of NNP (proper noun) and JJ (adjective) as shown in Appendix A. From Figure 23 above and referring to Appendix D, the name "Congo" was extracted using the code NNP and the name

"German" using the code JJ (German here is an adjective as it is qualifying the noun sphere). These names are only highlighted in the geographic types section if they are also present in the loaded gazetteer. This means that all the other names like Jipe, Taveta, Changa etc. that are still tagged as NNP through tokenisation, have not been highlighted in the geographic types section because they are not present in the loaded gazetteer. Refer to Table 3 for summarised results on name extraction through geographic types.

Table 3: Summarised results on name extraction from geographic type process

Names in treaty	**Code from tagging (Appendix D)**	**Names highlighted in Geographic types (Appendix F)**	**Comment**
Victoria	NNP	Highlighted	Also present in gazetteer
Congo	NNP	Highlighted	Also present in gazetteer
Kilimanjaro	NNP	Highlighted	Also present in gazetteer
German	JJ	Highlighted	Also present in gazetteer
Umba	NNP	Not highlighted	Not present in gazetteer
Jipe	NNP	Not highlighted	Not present in gazetteer
Lumi	NNP	Not highlighted	Not present in gazetteer
Taveta	NNP	Not highlighted	Not present in gazetteer
Changa	NNP	Not highlighted	Not present in gazetteer
Nyanza	NNP	Not highlighted	To be omitted (corrected in spelling check)
Free	NNP	Not highlighted	Not present in gazetteer
State	NNP	Not highlighted	Not present in gazetteer
Mfumbiro	NNP	Not highlighted	Not present in gazetteer

ii. Assigning the correct feature types to names in the treaty. E.g., "Victoria" is referred to as a Lake and "Umba" as a river in the treaty. The name "Victoria" therefore has to be categorising as Lake and the name "Umba" as a river in the feature type from the gazetteer. This is useful as the proper interpretation of spatial relationships can only be done so if the proper type of the feature is known. Features were also extracted using the code NN (nouns), NNP (proper nouns - singular), NNPS (proper nouns – plural), NNS (noun – plural), NP (proper noun – singular), and NPS (proper noun – plural). Table 4 shows the results obtained from extraction of features in geographic types section.

Table 4: Summarised results on feature extraction from geographic types process

Features in treaty	**Code from tagging (Appendix D)**	**Features Highlighted in geographic types**	**Comment**
River	NNP	Highlighted	Defined in gazetteer
Lake	NNP	Highlighted	Defined in gazetteer
Mount	NNP	Highlighted	Defined in gazetteer
Territory	NNS	Not highlighted	Not defined in gazetteer
Range	NN	Not highlighted	Not defined in gazetteer
Mountain	NN	Not highlighted	Not defined in gazetteer

iii. Extraction of cardinal direction in the treaty document. These were extracted using the codes NN (nouns), JJ (adjective) and RB (adverb). This is also an important section as the extracted cardinal directions were used in the reasoning with cardinal directions section. Table 5 below shows the extracted cardinal directions from the treaty.

Table 5: Summarised results on cardinal direction extraction from geographic types process

Cardinal directions in the treaty	**Code from tagging (Appendix D)**	**Cardinal directions highlighted in geographic types**	**Comment**
North	NN	Highlighted	Defined in gazetteer
Northern	JJ	Highlighted	Defined in gazetteer
Eastern	JJ	Highlighted	Defined in gazetteer
South	RB	Highlighted	Defined in gazetteer
Western	JJ	Highlighted	Defined in gazetteer

It should be noted from Table 3 that all the seven local names that were highlighted in the spelling check (Umba, Jipe, Lumi, Taveta, Changa, Nyanza and Mfumbiro) were not highlighted in the geographic type section. This shows the importance of the spelling check when local names are highlighted, as they then act as a check to manually verify if they are in the gazetteer. All the names not highlighted in the geographic types section apart from "Nyanza" (i.e. not present in the gazetteer), need to be manually added to the gazetteer. The name "Nyanza" need to be deleted from the treaty as explained in spelling check. It is a repetition for the word lake in the local language.

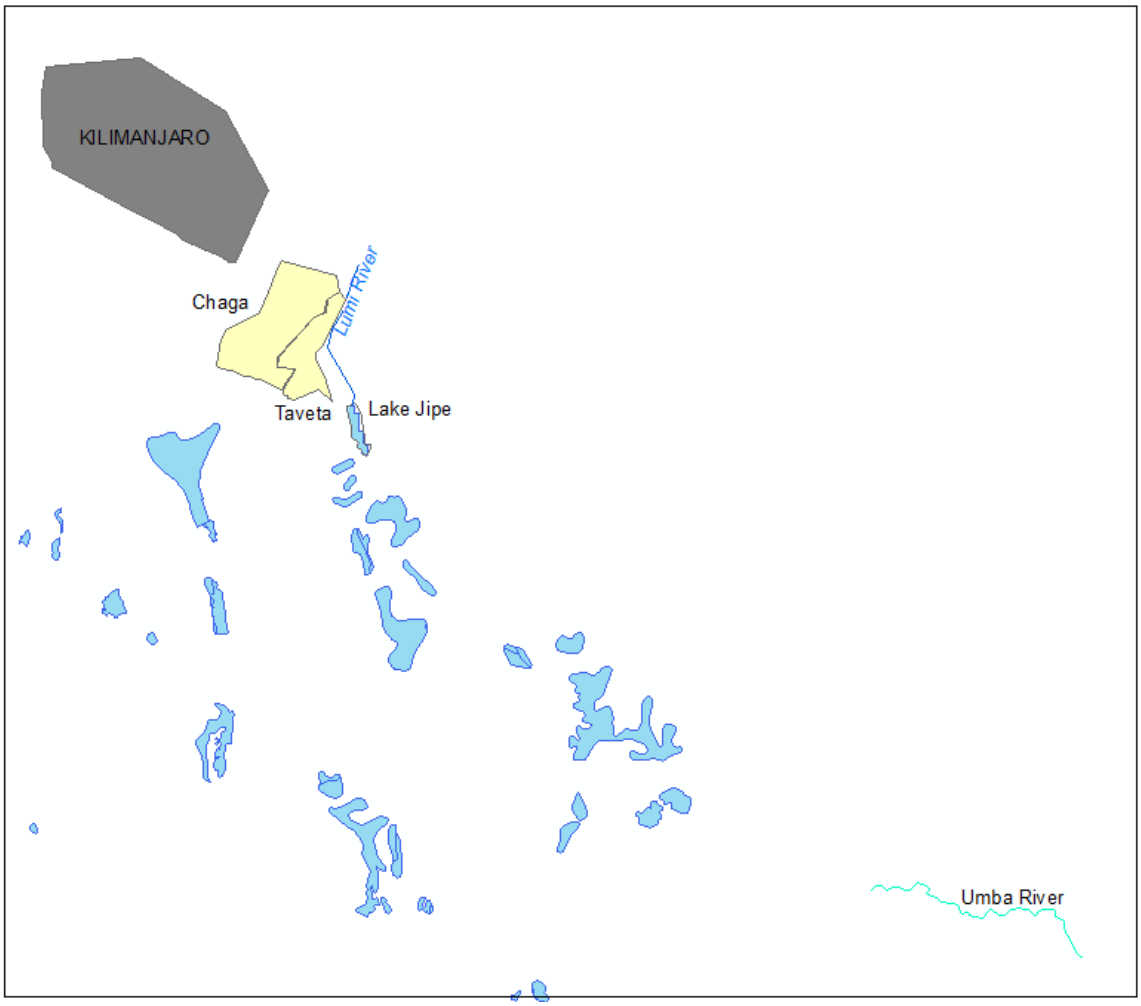
Also form Table 3 and referring to Appendix F on results from geographic types, it is seen that the name "Victoria" in the gazetteer was categorised in the minor type feature section as a city, a province and a name of a female while in the treaty it is referred as a Lake. Such instances need to be checked and corrected. In the gazetteer, therefore, "Victoria" should be changed to mean a lake

From Table 4, all the other features that are not present in the gazetteer i.e. territory, range and mountain, also need to be manually added to the gazetteer. However, form Table 5, it is clear that all cardinal directions were properly extracted from the treaty.

5.2.2. Location Geodatabase

All absolute places from sentence 376 (from Appendix E, using sentence id) were plotted on a map to visualize their position. This made the identification/calculation of the relative points to be easier.
These are names identified as NNP in tagging or proper nouns. These were plotted by using an African location database. Figure 24 below is an example of the plotting of all absolute places of the first sentence of the treaty.

Figure 24: Illustration of plotting of absolute places from the boundary treaty text



5.2.3. Relative reasoning

5.2.3.1. Spatial Reasoning

Chapter 6 gives illustrations on how spatial reasoning was used to make decisions on the translation of the boundary treaty texts into geographic location format.

5.2.3.2. Reasoning with cardinal directions

The following cardinal directions expressions are used in the treaty, north, northern, northwards, east, eastern, south southern, west, westward, western, and southwest as extracted through the geographic types process. Their literal meaning were specified according to the cardinal compass directions of north, south, east and west including their intermediate points of northeast, southeast, southwest and northwest. Table 6 shows how these cardinal directions were interpreted on the ground.

Table 6: Interpretation of Cardinal direction found in the boundary treaty text

Cardinal Direction	Word type	Explanation
North, south, east, west	Nouns	Following compass directions, literally to the north, south, east and west
Northern, southern, eastern, western	Adjectives (a word class that qualifies a noun)	These have to be followed by a noun, but literally meaning to north, south, east and west of the noun they qualify.
Northeast, southeast, southwest, northwest	Nouns	Intermediate positions of the compass directions between the north, east, south and west.

6. INTERPRETATION OF UNCERTAINTY USING RELATIVE REASONING

6.1. Accurate description

Accuracy is the degree of closeness of measurements of a quality to the value taken to be true. For example, analysing sentence 377 (from appendix E, using sentence id), the line of latitude is taken as an accurate description. Figure 25 gives an illustration for a plotted accurate boundary description text.

Figure 25: Illustration of accurate description from the boundary treaty text

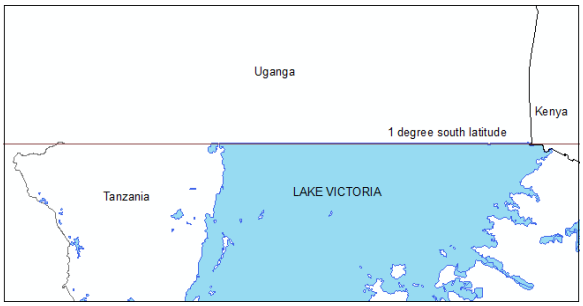
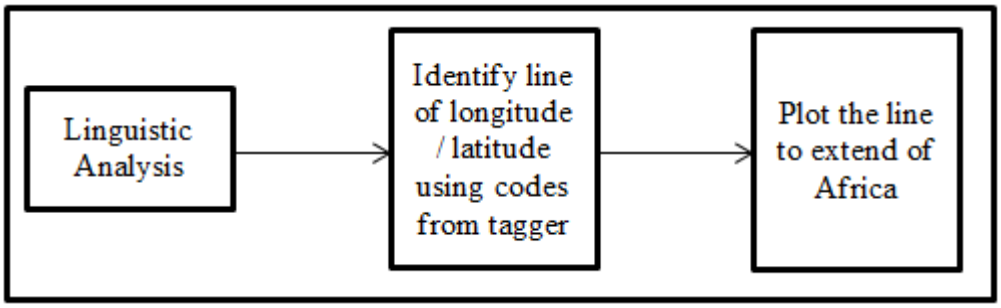


Figure 26 illustrates the chain of commands to be followed to plot accurate descriptions of longitude, latitude.

Figure 26: Procedures involved when plotting accurate boundary treaty text



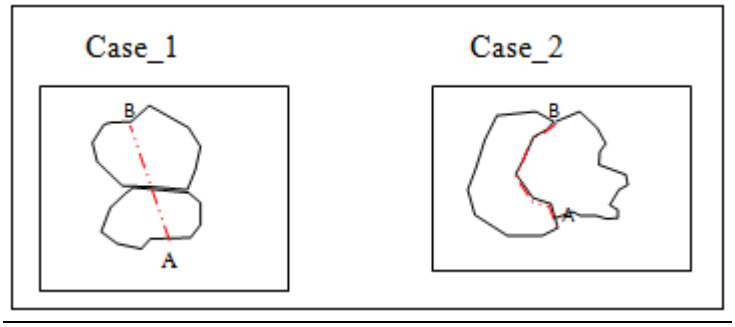
- It is clear that in such cases, comparing with Figure 2 of the status of international boundaries in Africa according to UN, it is seen that there are no international disputes associated with the boundary when it follows such descriptions.
- It would therefore be important after plotting all absolute point positions, to plot the great line of latitude and longitude.
- The graphical representation of the above plotting, could lead to even better analysis of relative point positioning.

- As the break off points of the latitude and longitude lines cannot be calculated yet, it would be necessary to first plot them as a straight line as long as possible to the extent of the African continent.

6.2. Ambiguity

Ambiguity is a term used in writing and math, and under conditions where information can be understood or interpreted in more than one way. For example, analysing sentence 376 (from appendix E, using sentence id), the description of the line bisecting the territories of Taveta and Chaga is an ambiguous concept. It is not clear if the boundary will be a straight line or will follow the boundaries between these two territories, see Figure 27 for an illustration.

Figure 27: Illustration showing ambiguity in the boundary treaty text

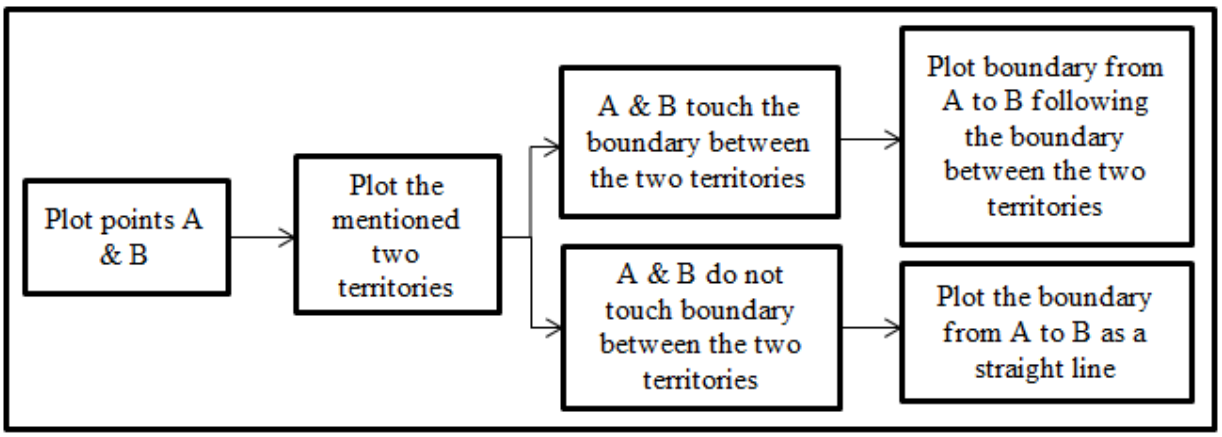


If the A and B are identified points of the boundary, then depending on the boundary outline of the two territories mentioned in the treaty, either the boundary can follow:

- Case_1: If the boundary between the two territories does not collide with points A and B.
- Case_2: If the boundary between the two territories collides with points A and B.

Figure 28 shows procedures of how this problem can be solved.

Figure 28: Procedures involved when solving ambiguity in boundary treaty text



6.3. Vagueness

Vagueness is if there are objects which one cannot say with certainty whether they belong to a group of objects which are identified with this concept or which exhibit characteristics that have this predicate (so-called "border-line cases"). For example, analysing sentence 377 (from appendix E, using sentence id), the concept of the boundary line “Skirting the northern slope of Kilimanjaro”, is vague since the actual spatial extend of the mountain range can be difficult to determine. Where is the northern slope? Where does the mountain itself start since height increase can be gradual? This leads to vagueness in identifying the extent of the northern slope of Kilimanjaro. Also a start of a river is a vague concept as a river can have different streams at its mouth.

Consider Figure 29, as an illustration of solving vague sentences, where A is the last known certain point of the boundary, B cannot be executed since the sentence describing it is vague. Point C is the next certain point described in the treaty text and can be located.

Figure 29: Illustration for solving vague sentences in boundary treaty text

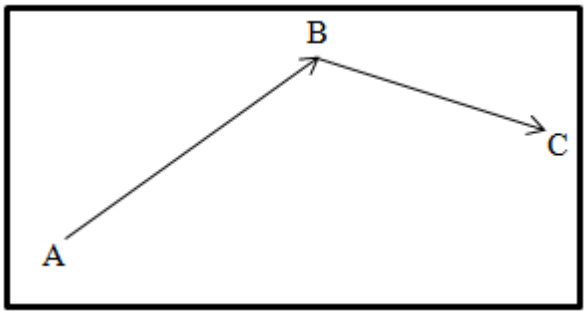
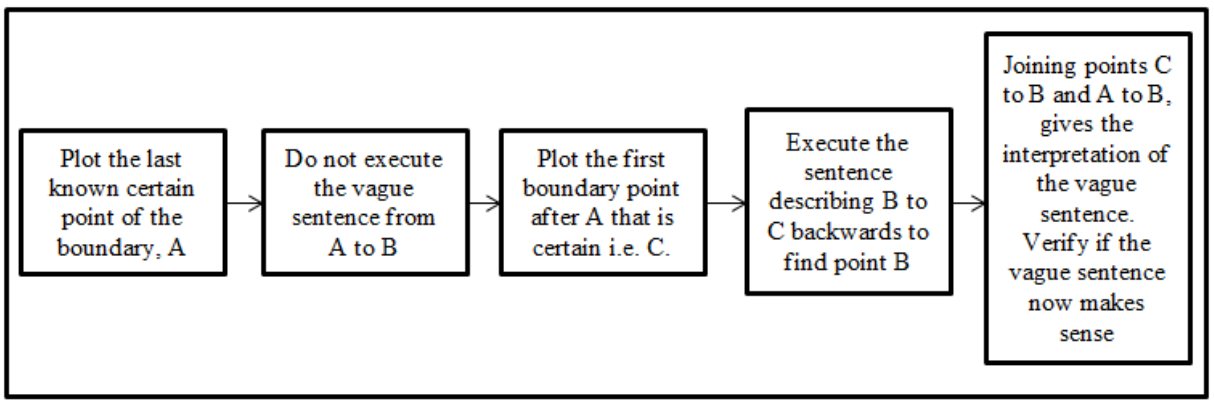


Figure 30 then illustrates chaining processes on how spatial reasoning can be used to deduce the interpretation of the above vagueness.

Figure 30: Procedures involved when solving vagueness in boundary treaty text



If there is more than one vague sentence, the same reasoning shall be applied by executing those parts that are certain and also working backwards and forward on each sentence possible, until all the points have been found.

6.4. Time dependent location

6.4.1. Change of river flows

Some rivers like Songwe that is mentioned in the treaty, changes course every now and then. In the treaty, this river is used to define the boundary between Malawi and Tanzania. This has caused international disputes between Malawi and Tanzania as also shown in Figure 2 of the status of international boundaries in Africa.

Archiving all the necessary data of the course of the river when the treaties were being signed would be the first solution to determine where the boundary ought to be or analysis of the river changes over time can be used to determine the course of the river at a certain period. This however is a challenging job as not all such data can be readily available. Stabilization of such rivers could also be another step of avoiding future conflicts.

6.4.2. Lake diminishing

For example, analysing sentence 376 (from appendix E, using sentence id), the treaty indicates that the boundary runs directly to Lake Jipe and passes along the eastern shore and around the northern shore of that lake.

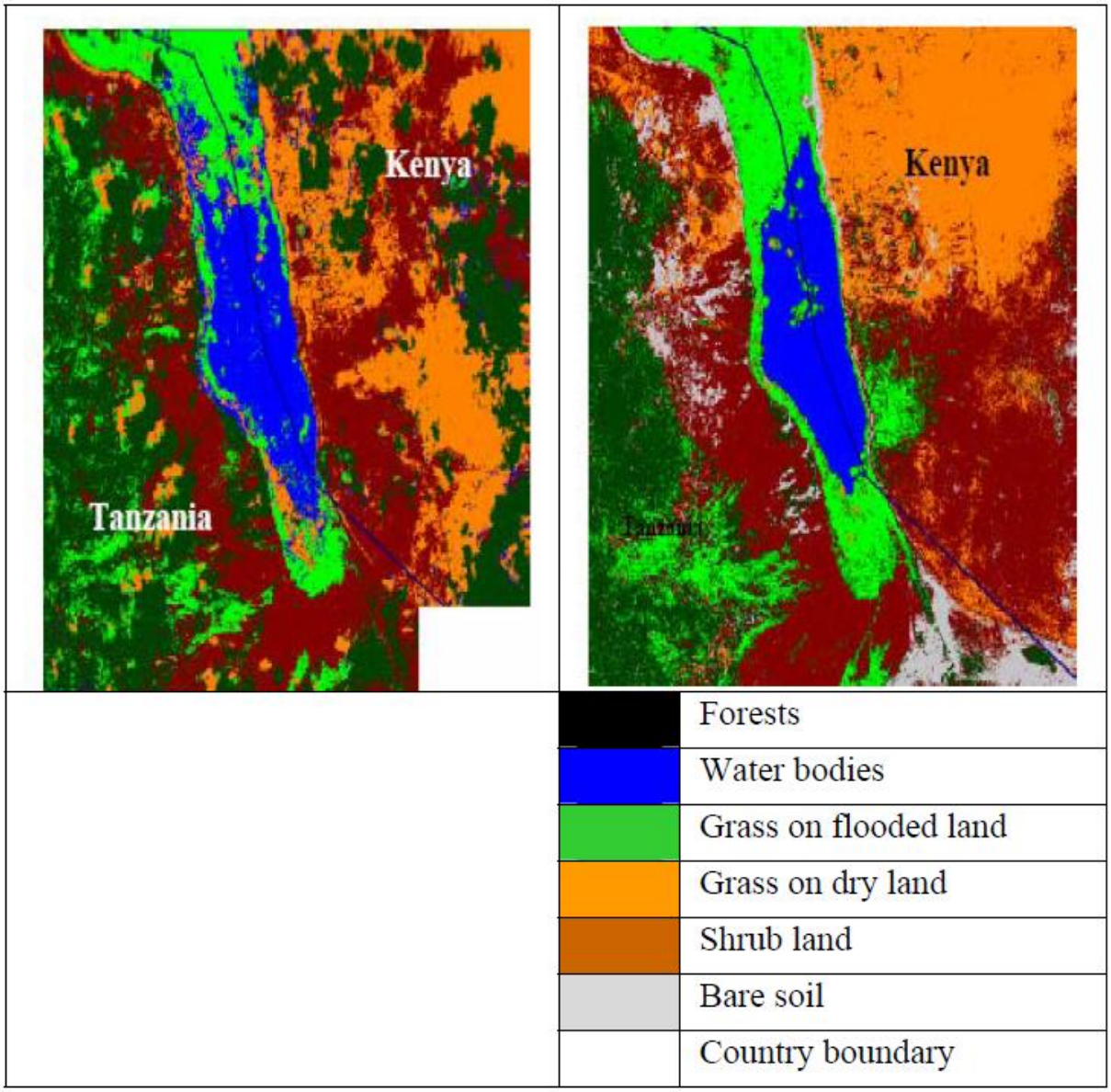
However, (Mtalo, 2005), indicates that Lake Jipe has undergone major changes in Land cover over the last 15 years. He analysed multi-temporal LANDSAT Thematic Map satellite imagery of 1st January 1987 and 4th March 2001 using IDRISI Kilimanjaro digital image processing software to map out the dynamics of the lake changes. The results showed that the surrounding water body has decreased by 32% respectively. Figure 31 shows this diminishing trend in the lake.

The following uncertainties are therefore observed;

- The northern shore was originally identified at a small scale map of 1: 1,000,000. Locating it at larger scales requires additional data from satellite images, etc.
- The northern shore is a vague concept. Refer to Figure 32 for illustration. Where exactly on the shore does this boundary line start from?
- As the Lake size is diminishing, the northern shore also changes with time. To identify the correct northern shores will require analysis of the boundary of the lake when the treaty was being signed.

Therefore, where the boundary is supposed to follow a certain lake, all the necessary data like aerial images of the lake, data on how the lake extend is changing are supposed to be gathered for better analysis.

Figure 31: Temporal and spatial variations of land use and land cover around lake Jipe, left 1987: right 2001 (Mtalo, 2005)

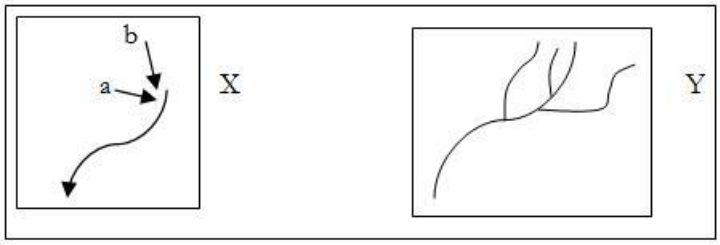


6.5. Combined uncertainities

Each category of uncertainties has been analysed individually, but in practice, they come as a combination of different uncertainties. For example, analysing sentence 376 (From, Appendix E using id), shows that the sentence has different uncertainties associated with it. Figure 32 illustrates this example.
The following are noted;

- As from diagram X, it is ambiguous to determine the point on the northern boundary of the river. Is it "a" or "b". What if the river runs in the west – east direction, where would be the northern bank be?
- Even if the northern bank is established, how far from the river water is the point going to be on the bank?
- From diagram Y, the following questions are asked. If the river has several attributes at its mouth, how do you find the northern bank? Do you use the attribute with most water as the main river?

Figure 32: Illustration of combined uncertainties



Two approaches can be used to analyse these situations;

- Using Figure 30 above as the basis, plot all the absolute positions first and then reason back and forth to calculate the uncertain boundary point in question between two sequential points.
- Acquiring of more data or information on these scenarios can help better understand the uncertainties. These could be by reading reports produced prior to the signing of treaties as they could have definitions of certain terms. E.g. the main stream could have been the one and only one that the explorers were using for navigation at the time the treaties were signed as in the report by (Boileau & Wallace, 1899)

7. DISCUSSION

Conflicts that arise due to misinterpretations or holding different interpretation of the boundary treaty text arise due to the uncertainties in the treaty text. This research provides an approach that is standardised and that can be used as a guide to staff working in the translation of international boundary treaty text, into geographic location format.

7.1. Applicability of procedures

- The developed procedures are to be used as a formal process to instruct staff members working on the translation of the treaty text into geographic location, guiding them how and where to make which decisions.
- The developed procedures have distinct steps for:
    - Specified interpretation options by having distinct steps to follow.
    - Decision making identification by giving chain of commands to be followed and
    - Uncertainties of decision specified by highlighting uncertainties associated with each decision.
- The developed procedures are a first step in the process of translating boundary description text into coordinates.

7.2. Uncertainties

7.2.1. Types of uncertainties

There are a lot of uncertainties concerned with interpretation of boundaries treaty text into geographic location. These were mentioned in Chapter 6 and involve uncertainties on:

- Accuracy – These are descriptions in the treaty document that refer to the boundary line as following the lines of longitude and latitude. These descriptions with the modern technology of Geographical Information Systems (GPS) can be located on the ground to cm accuracy.

- Ambiguity – The treaty specifies some boundaries to follow a river. This ambiguous description has been interpreted differently by individual countries. This is because in some cases, the treaties did not specify if the boundary would be the middle of the water feature or the banks. If it is the banks, then how far away from the bank and which side of the bank would this boundary be.

  In non-navigable rivers or streams, the boundary is usually taken as sinuous line lying along the middle of the water surface and equidistance from the banks so that each country has access to the water. In such situations, it is important to specify the particular stage, such as low water level, along which the surface must be measured, because the rise and fall of the water might otherwise cause a lateral movement of the boundary, especially when the slopes of the opposite banks are unequal. Boundaries in navigable rivers normally coincide with the thalweg, the line of deepest soundings, but diversions may occasionally be necessary to avoid the bisection of mid – stream islands or to adopt a branch of the river that forms the principle shipping channel but does not contain the thalweg (McEwen, 2002).

- Vagueness - It is not surprising that in the treaty, natural features such as watershed, and mountain ranges were used. Such features offered themselves as physical landmarks the geographical identity of which is beyond doubt. Yet the adoption of some of these lead to vague concepts as the treaty did not precisely indicate where on these feature the line must be drawn. Identifying for example where a mountain starts is very vague.

- Time dependent positions – Some of features used in the boundary description are rivers and lakes. Rivers sometimes change courses and the spatial extent of lakes can change with time. Use of satellite images and other technologies to determine the original course and extend of such features, is important.

- Combined uncertainties – In practice uncertainties come as combinations of different uncertainties. The approach given in the procedures is to plot the absolute points first and then reason backwards and forward to locate the uncertain points.

All these uncertainties have been highlighted so as to act as a checklist making sure no item is missed out when translating the boundary text into geographic location.

7.2.2. Effects of uncertainties

- Uncertainties request for additional information like satellite images, volume of water, shape files, maps.
- If no additional information is available, or the additional information is not adequate enough, disputes may arise between parties.
- If these disputes are not solved, it leads to conflicts. Therefore, the resolution of these uncertainties has the highest priority.

7.3. Scale

(Boileau & Wallace, 1899) as members of the Anglo-German Boundary Commission in their report of The Nyasa-Tanganyika plateau showed that the base map on which the Anglo – German boundary treaty was signed on, was at a scale of 1: 1,000,000. The treaty text was therefore agreed upon based on that scale of 1:1,000,000.

At this scale, most uncertainties considering positional accuracy e.g. whether the boundary is in the middle of the river or the bank would make no difference as far as map production is concerned. However, the biggest problem would be to identify the exact location of the point on the ground. The stipulated procedures, therefore provides an approach on choosing the best location of the boundary line.

Also considering that international boundary maps and text were agreed based on a scale of 1:1,000,000, to plot these at larger scales of 1:250,000, 1:100,000, 1:50,000 or 1: 10,000 etc. would require the need for additional information, like more maps, other treaties, more satellite images.

7.4. Language

The Anglo-German Treaty [Heligoland-Zanzibar Treaty] on which this research is based on, was written in English. However there are more treaties in English and also in other languages like French, German and Italian. It would therefore be important to see if these developed procedures would also apply on the treaties written in the other languages.

7.5. Treaties in treaties

Some other treaties are mentioned in other treaties. Such a consideration was not addressed in this research. For example in The Anglo – German Treaty, other treaties of 1885, 1885 are mentioned. This would therefore call for hyperlinks to these treaties.

7.6. Supporting data

It is clear that reading and analysis the treaty on their own could lead to grave mistakes being made as far as interpretation of these texts to geographic location is concerned. The following more data needs to be sourced;

- All other relevant treaties mentioned in the treaty being analysed.
- All old maps and reports on which the boundary treaties were based on.
- Country gazetteers as recognized by relevant mapping agencies.
- Satellite images and other data that help to verify the spatial situation at the time the treaties were being made.

8. CONCLUSION AND RECOMMENDATION

8.1. Conclusion

In reflection on the research objectives and the developed procedures, the following were observed;

- The procedures in this thesis have been developed to provide a non-biased standardized way of interpreting the boundary treaty text in order to translate it into a geographic location format. They only lead to make the best choice when it comes to boundary text interpretation. In this research, they have not been developed to have a boundary map as a final product.
- The developed procedures are to be used as instruction for staff working under International Boundary issues, as a step by step guide towards the interpretation of the boundary treaty text into geographic location format.
- Analysis of the treaty and decisions to follow can be better made with the awareness of the uncertainties associated with this. Analysis without accommodating data uncertainty, can quiet severely limit its usefulness. Yet an appropriate conceptualisation of uncertainty and the application of related analytical methods creates rich analytical environment where decision making based on spatial information is facilitated not only by objective orderings of alternatives but also by giving confidence in those alternatives(Fisher, 1999). The developed procedure therefore highlight the uncertainties associated with the interpretation of the boundary treaty text into geographic location format, thereby creating awareness for the staff on possible challenges and thereby helping them make better decisions.
- The location uncertainties occur at a scale of 1 : 1000,000. This is the scale at which the boundary treaties were based on. Identifying the boundary line at larger scales requires additional information like more maps, satellite images or even treaties.
- The resolution of the uncertainties has the highest priority as it is from the uncertainties that disputes occur from. The procedures should give an approach where by the uncertainties can be resolved.

8.2. Recommendation

The following are the recommendations for further research:

- To completely move out of the manual interpretation of these procedures, coding will need to be done to make a complete semi-automated system for the analysis of the treaty document into geographic location format.
- The developed procedures are only a framework for the interpretation of the boundary treaty text into location format. They are to be used only as a guideline and not as a complete package on their own. User experience and guidance has to be used for a semi-automated system to work.
- Each category of uncertainties has been analysed individually, but in practice, they come as a combination of different uncertainties. Analysis of combined uncertainties has not been done in this research. However, this research recommends that the best reasoning approach would be to plot the absolute point first and reason back and forth to calculate all the other intermediate points that have uncertainties. For future research, it would be important to try this concept.
- The extraction of verbs from the treaty text from the tagging process, will lead to analysis of these verbs and also the formulation of topological relationships. Some of the verbs in the treaty are "commences", "bisect", and "crosses". From these, algorithms for topological relationships like intersect, meet, and touch can be formulated. However, this has not been done in this research. It would therefore be important that during the coding process, an algorithm for developing topological relationships from the treaty verbs be devised in order to find the required boundary point.
- In most instances, international boundaries are usually plotted at larger scales than the 1: 1000,000 as analysed in this research. It would be important for future research to see the practicality of getting the additional information required to plot these maps at these larger scales like 1: 250,000, 1: 50,000, 1: 10,000 etc.
- In this research, only the treaty in English was analysed, it would be appropriate for users familiar and working on the other languages like French and German to apply and see the applicability of these procedures to interpretation of the treaty texts into geographic location format in those languages.
- This research acts as a basis for solving most of the international boundary conflicts that arise in Africa due to disagreements on the location of the boundary line. It can be adopted by any organisation like the United Nations (UN), International Boundary Research Unit (IBRU) and mapping agencies in different countries as a working frame for interpretation of the International boundary treaty text into geographic location format.

DEFINITIONS

Delimitation - Is the legal process by which two sovereign nations establish and describe in writing the location of their common boundary. This is the task of the diplomats and treaty negotiators, and may require more than a single agreement.

Demarcation - Is a field operation; its purpose is to mark the position of the boundary on the ground for all to see. A joint commission, composed of an equal number of members from each country, normally undertakes demarcation.

Delineation - Is the graphical or mathematical representation of the boundary. Quiet often a joint commission undertakes both demarcation and delineation. The commission publishes results consisting of reports, photographs and other illustrations, maps, and tables showing geographic positions of boundary monuments and survey control stations. These documents represent the official record of boundary location.

Thalweg - Is a line drawn to join the lowest points along the entire length of a stream bed Or valley in its downward slope, defining its deepest channel. The thalweg thus marks the natural direction (the profile) of a watercourse. The thalweg is almost always the line of fastest flow in any river.

Orth - (orthography) the art of writing words with the proper letters, according to accepted usage and correct spelling.

APPENDIX

Appendix A: Tags used in GATE Software

Tag	Description
CC	coordinating conjunction: 'and', 'but', 'nor', 'or', 'yet', plus, minus, less, times (multiplication), over (division). Also 'for' (because) and 'so' (i.e., 'so that')
CD	cardinal number
DT	determiner: Articles including 'a', 'an', 'every', 'no', 'the', 'another', 'any', 'some', 'those'
EX	existential there: Unstressed 'there' that triggers inversion of the inflected verb and the logical subject; 'There was a party in progress'
FW	foreign word
IN	preposition or subordinating conjunction
JJ	adjective: Hyphenated compounds that are used as modifiers; happy-go-lucky
JJR	adjective - comparative: Adjectives with the comparative ending '-er' and a comparative meaning. Sometimes 'more' and 'less
JJS	adjective - superlative: Adjectives with the superlative ending '-est' (and 'worst'). Sometimes 'most' and 'least'
JJSS	unknown-, but probably a variant of JJS
LRB	unknown
LS	list item marker: Numbers and letters used as identifiers of items in a list
MD	modal: All verbs that don't take an '-s' ending in the third person singular present: 'can', 'could', 'dare', 'may', 'might', 'must', 'ought', 'shall', 'should', 'will', 'would'
NN	noun - singular or mass
NNP	proper noun - singular: All words in names usually are capitalized but titles might not be
NNPS	proper noun - plural: All words in names usually are capitalized but titles might not be
NNS	noun - plural
NP	proper noun - singular
NPS	proper noun - plural
PDT	predeterminer: Determiner like elements preceding an article or possessive pronoun; 'all/PDT his marbles', 'quite/PDT a mess'
POS	possessive ending: Nouns ending in 's' or '"
PP	personal pronoun
PRPR$	unknown-, but probably possessive pronoun
PRP	unknown-, but probably possessive pronoun
PRP$	unknown, but probably possessive pronoun, such as 'my', 'your', 'his', 'his', 'its', 'one's', 'our', and 'their'

RB	adverb: most words ending in '-ly'. Also 'quite', 'too', 'very', 'enough', 'indeed', 'not', '-n't', and 'never'
RBR	adverb - comparative: adverbs ending with '-er' with a comparative meaning
RBS	adverb - superlative
RP	particle: Mostly monosyllabic words that also double as directional adverbs
STAART	start state marker (used internally)
SYM	symbol: technical symbols or expressions that aren't English words
TO	literal to
UH	interjection: Such as 'my', 'oh', 'please', 'uh', 'well', 'yes'
VBD	verb - past tense: includes conditional form of the verb 'to be'; 'If I were/VBD rich...'
VBG	verb - gerund or present participle
VBN	verb - past participle
VBP	verb - non-3rd person singular present
VB	verb - base form: subsumes imperatives, infinitives and subjunctives
VBZ	verb - 3rd person singular present
WDT	wh-determiner
WP$	possessive wh-pronoun: includes 'whose'
WP	wh-pronoun: includes 'what', 'who', and 'whom'
WRB	wh-adverb: includes 'how', 'where', 'why'. Includes 'when' when used in a temporal sense
::	literal colon
,	literal comma
$	literal dollar sign
-	literal double-dash
(	literal left parenthesis
.	Literal period
#	literal pound sign
)	literal right parenthesis

Appendix B: Table showing preposition, their relations, properties and feature they act on(David N. Chin, 1994).

Preposition	**Relations**	**Properties – Feature_ Types**
Above	At elevation	Greater elevation
Adjacent to	Adjacent to	
Adjacent to	Near	
Along	Near	Linear feature
Among	Within	Between several objects
Around	Near	
At	Within	
At	Near	
At	At elevation	Specified elevation
(In) Back of	Beyond	Oriented object/observer
Behind	Beyond	Oriented object/observer
Below	At elevation	Lower elevation
Beside	Adjacent to	Oriented object/observer
Between	Within	Two ground objects
Beyond	Beyond	Oriented object/observer
By	Near	
Down	Down	Lower elevation and path
East of	Easton	
From	From	Oriented object/observer
(In) Front of	Front of	Oriented object/observer
In	Within	
Inside	Within	
Into	Within	
Left of	Left of	
Near	Near	
Next to	Adjacent to	
North of	North of	
On	Within	
On	Adjacent to	
On top of	Within	
Outside	Without	
Right of	Right of	
South of	South of	

To	Toward	Oriented object/observer
Up	At elevation	Greater elevation and path
West of	West of	

Appendix C: Summary of 13 spatial relations with their corresponding prepositions (Freeman, 1975)

Spatial-Relation	Preposition
Adjacent-to	Adjacent to, besides, next to, on
At-elevation	Above, at, below, down, up
Between	between
Beyond	[in/at] back of, behind, beyond
East-of	East of
From	From
Front-of	Before, [in] front of
Left-of	[to[the]] left [hand side] of
Near	Adjacent-to, along, around, at, by, near, outside [of]
North –of	North of
Right of	[to[the]] right [hand side] of
South-of	South of
Towards	To, towards
West-of	West of
Within	Among, at, in, inside, into, on, on top of

Appendix D: GATE software results from tokenisation, tagging and morphology analysis process

T = Token; St = Start; L = length; Catg = Category (form tagging process); Affix & Root (from morphology process)

(Only first page of results has been appended due to the size of the results – refer to the digital data for complete results of appendix D)

T	**St**	**End**	**Id**	**Feature**						
				Kind	**L**	**Orth**	**String**	**Catg**	**Affix**	**Root**
T	0	2	6	word	2	upperInitial	To	TO		to
T	3	6	8	word	3	lowercase	the	DT		the
T	7	12	10	word	5	lowercase	north	NN		north
T	13	15	12	word	2	lowercase	by	IN		by
T	16	19	14	word	3	lowercase	the	DT		the
T	20	24	16	word	4	lowercase	line	NN		line
T	25	29	18	word	4	lowercase	that	IN		that
T	30	39	20	word	9	lowercase	commences	NNS	affix=s	commence
T	40	42	22	word	2	lowercase	on	IN		on
T	43	46	24	word	3	lowercase	the	DT		the
T	47	55	26	word	8	lowercase	northern	JJ		northern
T	56	60	28	word	4	lowercase	bank	NN		bank
T	61	63	30	word	2	lowercase	of	IN		of
T	64	67	32	word	3	lowercase	the	DT		the
T	68	73	34	word	5	lowercase	mouth	NN		mouth
T	74	76	36	word	2	lowercase	of	IN		of
T	77	80	38	word	3	lowercase	the	DT		the
T	81	85	40	word	4	upperInitial	Umba	NNP		umba
T	86	91	42	word	5	upperInitial	River	NNP		river
T	91	92	43	punctuation	1					
T	93	97	45	word	4	lowercase	runs	VBZ	affix=s	run
T	98	106	47	word	8	lowercase	directly	RB		directly
T	107	109	49	word	2	lowercase	to	TO		to

Appendix E: GATE software results from sentence splitting process

Split sentence	Type	Start	End	Id	Feature
To the north by the line that commences on the northern bank of the mouth of the Umba River, runs directly to Lake Jipe and, after passing along the eastern shore and around the northern shore of that lake, crosses the Lumi River and bisects the territories of Taveta and Chaga.	Sentence	0	278	376	-
Skirting the northern slope of the Kilimanjaro range, this line continues to the point on the eastern shore of Lake Victoria Nyanza that is intersected by the 1st degree of south latitude.	Sentence	279	467	377	-
It crosses the lake on this parallel and follows it to the border of the Congo Free State, where it terminates.	Sentence	468	579	378	-
It is understood, though, that the German sphere of interest on the western side of the aforementioned lake does not include Mount Mfumbiro.	Sentence	580	720	379	-
Should it turn out that this mountain lies to the south of the aforementioned parallel of latitude, the line of demarcation shall be drawn so as to exclude the mountain from the German sphere of interest; but the line shall nonetheless terminate at the previously described point.	Sentence	721	1001	380	-

Appendix F: GATE software results from geographic types process

Token	Type	Start	End	Id	majorType	minorType
north	Lookup	7	12	381	loc_key	pre
northern	Lookup	47	55	382	loc_key	pre
River	Lookup	86	91	383	loc_key	post
Lake	Lookup	110	114	384	loc_key	pre
eastern	Lookup	149	156	385	loc_key	pre
northern	Lookup	178	186	386	loc_key	pre
Lake	Lookup	201	205	387	loc_key	pre
River	Lookup	224	229	388	loc_key	post
northern	Lookup	292	300	389	loc_key	pre
Kilimanjaro	Lookup	314	325	390	location	city
this	Lookup	333	337	391	time_modifier	
eastern	Lookup	373	380	392	loc_key	pre
Lake	Lookup	390	394	393	loc_key	pre
Victoria	Lookup	395	403	394	location	city
Victoria	Lookup	395	403	395	location	province
Victoria	Lookup	395	403	396	person_first	female
south	Lookup	452	457	397	loc_key	pre
It	Lookup	468	470	398	stop	
lake	Lookup	483	487	399	loc_key	pre
this	Lookup	491	495	400	time_modifier	
Congo	Lookup	541	546	401	location	country
It	Lookup	580	582	402	stop	
German	Lookup	615	621	403	country_adj	
western	Lookup	648	655	404	loc_key	pre
lake	Lookup	683	687	405	loc_key	pre
mount	Lookup	705	710	406	loc_key	pre
this	Lookup	745	749	407	time_modifier	
south	Lookup	771	776	408	loc_key	pre
German	Lookup	899	905	409	country_adj	

LIST OF REFERENCES

(IBRU), I. B. R. U. (2010). International Boundaries Research Unit : Boundary news - Durham University. Retrieved 12 - 08, 2010, from http://www.dur.ac.uk/ibru/news/boundary_news/?itemno=10181&rehref=%2Fibru%2F&resubj=Boundary+news%20Headlines

(OLA), U. N. U. O. o. L. A. (2010). United Nations Treaty Collection. Retrieved 16 - 08, 2010

André, E., Bosch, G., Herzog, G., & Rist, T. (1987). Coping with the intrinsic and deictic uses of spatial prepositions. *Artificial intelligence II: Methodology, systems, applications*, 375-382.

Boileau, F. F. R., & Wallace, L. A. (1899). The Nyasa-Tanganyika Plateau. *The Geographical Journal, 13*(6), 577-621.

Brownlie, I., & Burns, I. (1979). *African boundaries: A legal and diplomatic encyclopaedia*: C. Hurst & Co. Publishers.

David N. Chin, M. M., Tung-Tse Chen. (1994). Understanding Location Descriptions in the LEI Systempp. 138--143).

Davis, E. (1986). *Representing and acquiring geographic knowledge*: Courant Inst. of Mathematical Sciences, New York Univ., NY.

Fisher, P. F. (1999). Models of uncertainty in spatial data. *Geographical information systems, 1*, 191-205.

Frank, A. U. (1992). Qualitative spatial reasoning about distances and directions in geographic space. [doi: DOI: 10.1016/1045-926X(92)90007-9]. *Journal of Visual Languages & Computing, 3*(4), 343-371.

Freeman, J. (1975). The modelling of spatial relations. *Computer Graphics and Image Processing, 4*(2), 156-171.

Futch, S., Chin, D., McGranaghan, M., & Lay, J. (1992). Spatial-linguistic reasoning in LEI. In A. Frank, I. Campari & U. Formentini (Eds.), *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space* (Vol. 639, pp. 318-327): Springer Berlin / Heidelberg.

Herskovits, A. (1986). Language and spatial cognition: An interdisciplinary study of the prepositions in English.

Kuipers, B. (2000). The Spatial Semantic Hierarchy* 1. *Artificial Intelligence, 119*(1-2), 191-233.

Li, S. (2007). *Combining topological and directional information for spatial reasoning*. Paper presented at the Proceedings of the 20th international joint conference on Artifical intelligence.

McEwen, A. (2002). The demarcation and maintenance of International Boundaries.

Moratz, R., & Ragni, M. (2008). Qualitative spatial reasoning about relative point position. [10.1016/j.jvlc.2006.11.001]. *Journal of Visual Languages & Computing, 19*(1), 75-98.

Mtalo, F. (2005). Water Resources Management Issues and conflict
resolutions at a catchment Level. A Case Study of Pangani
River Basin, Tanzania.

Naldi, G. (1987). The case concerning the Frontier dispute (Burkina Faso/Republic of Mali): Uti possidetis in an African perspective. *International and Comparative Law Quarterly, 36*(04), 893-903.

Peterson, J. L. (1980). Computer programs for detecting and correcting spelling errors. *Commun. ACM, 23*(12), 676-687.

Prescott, V. (1996). *Contributions of the United Nations to solving boundary and territorial disputes, 1945-1995*.Unpublished manuscript.

Talmy, L. (1983). *How language structures space*: Cognitive Science Program, Institute of Cognitive Studies, University of California at Berkeley.

Toset, H. P. W., Gleditsch, N. P., & Hegre, H. (2000). Shared rivers and interstate conflict. [Proceedings Paper]. *Political Geography, 19*(8), 971-996.

Zhan, C. (2002). Picture Smart:Spatial Reasoning and Its Role in Cognition. *Serendip*. Retrieved from http://serendip.brynmawr.edu/bb/neuro/neuro02/web3/czhan.html