(Multilingual) Knowledge Representation for Epistemological Access

By

Mary-Jane Antia

Supervised by

Professor Tommie Meyer

A dissertation presented for the degree of
Master of Science

Department of Computer Science
University of Cape Town
December 2017

Dedicated to

My Family

(Multilingual) Knowledge Representation for Epistemological Access

Mary-Jane Antia

Submitted for the degree of Master of Science
December 2017

Abstract

In South Africa, the inability of many students to acquire knowledge and to successfully demonstrate acquired knowledge in assessments has, in part, been attributed to their low levels of proficiency in the language of learning and teaching (which is also the language of textbooks). This dissertation investigates the possibility of using knowledge representation to enhance epistemological access for learners, specifically, enhancing understanding of school science by learners with limited proficiency in academic English. In the dissertation, sample texts from a life science textbook were modeled into entities and relations, using the conceptual graphs formalism. While the labels for the entities were in English, the links between the entities were provided in both English and in the variety of Afrikaans called Kaaps. A knowledge-based application, using Jupyter notebook and a graph database, was subsequently developed on the basis of the modeled texts.

To test the impact of this graph-based resource, an experimental study was designed involving grade 10 learners in a Cape Town high school. One group was exposed to the graphically modeled content and another group was limited only to the text content. The null hypothesis formulated was as follows: there will be no difference in the performance scores of those learners who are exposed to the knowledge modeled in the application and scores of those learners exposed only to knowledge in text without the model. On six of the seven questions, the experimental group performed better than the control group. The performance of the experimental study was further verified using inferential statistics, which showed that the results were statistically significant. Given that the experimental group performed better than the control group, the null hypothesis was rejected.

During interviews, participants' subjective experiences indicated that the graphically modeled knowledge allowed for a better understanding of the text. Although findings from larger studies are clearly required, the current study indicates that the implementation of graph-based knowledge systems is a promising means for intervening to enhance the understanding of English-language science textbooks by learners who may not be proficient in English.

Declaration

The work in this dissertation is based on the research carried out in the department of computer science at the University of Cape Town as part of the completion of my Master's degree. I declare that no part of this work has been submitted elsewhere for any purpose. I declare that these are my own words except where references are given of texts or diagrams. I declare that I have submitted an original work written in my own words. With the signature, I declare that I have been informed about normal academic citation rules and I conform to citation convention customary to the sciences. This written work may be tested electronically for plagiarism.

Signature: Signed by candidate Date: December 2017

Acknowledgement

I am grateful to the NRF and the DST-CSIR for scholarships that enabled me to carry out this research. I am grateful to my supervisor, Professor Tommie Meyer, for his guidance, understanding and encouragement throughout the process of completing this dissertation.

I am thankful to Mr. Nkosi, a science teacher, without whose support the study would have been extremely difficult to carry out. I also thank the study participants as well as the authorities at the school where data was collected. I am also grateful to Chloe Stiglingh, Geraldine Hartman and Rolene Martins for advice on schools and for translating test material.

I am also very thankful to Claude Kakoko for assistance with the inferential statistical analysis. Approval by the Western Cape Education Department to carry out the study is appreciated.

I am forever thankful to my family, especially my husband, for all the support I received during my studies.

Table of Contents

List of Figures and Tables

Glossary

API	Application Programming Interface
CBN	Causal Bayesian Network
CG	Conceptual Graph
CQL	Cypher Query Language
CSIR	Center for science and Research
DL	Description Logic
FOL	First Order Logic
KR	Knowledge Representation
KRR	Knowledge Representation and Reasoning
RDFS	Resource Description Frameworks
NRF	National Research Foundation
SNePS	Semantic networks Processing Systems
SQL	Standard Query Language
WCED	Western Cape Education Department

Chapter One: Introduction

It has been claimed that the "language used for learning and teaching is crucial for learners' acquisition of knowledge and understanding and the development of their skills, and for learners' ability to demonstrate their acquired knowledge effectively in assignments and examinations" (1). This makes language an issue of epistemological access, which has been defined in the basic education context as "access to knowledge that could lead to successful schooling outcomes" (2). In South Africa, the dominant language of learning and teaching is English, and the performance of learners in many content subjects in the matriculation examinations in the Western Cape has been shown to be stratified according to whether the home language of learners is English or not. English home language learners were found to be the top achievers (3). Concerns with performance in school subjects, especially science, is however not limited to students with limited proficiency in English (4). The reason why understanding of content and demonstration of understanding are problems for many learners is because "Scientific texts are … difficult to read; and this is said to be because they are written in 'scientific language', a 'jargon' which has the effect of making the learner feel excluded and alienated from the subject-matter" (4). A number of the difficulties which students have with 'scientific texts' have been identified (4). There is, for instance, syntactic ambiguity, where the structure of a sentence can lead to different interpretations of the sentence. An example can be seen in the sentence "lung cancer death rates are clearly associated with increased smoking" (4); the verb group 'are clearly associated with' can be understood as 'is caused by' or 'causes/leads to', but the meaning intended is the former.

It is features of scientific English writing such as in the sentence above that contribute to learners' difficulties with understanding science. The computer science challenge that this problem raises may be stated as follows: how can the knowledge in school textbooks in English be represented in a computer system in a way that prevents scientific English from being an obstacle to understanding? In other words: can the knowledge on a topic in an English textbook on a subject such as life science be graphically modeled so that the learners can have access to the main entities and relations on the topic, and in such a way that their level of English knowledge is less a barrier than is the case currently?

The problem the dissertation seeks to address can be resolved using knowledge representation (KR), also referred to as knowledge representation and reasoning (KRR). Knowledge Representation (KR) involves analysing knowledge (related to a real world issue) and making it possible for this knowledge to be processed (i.e. reasoned with) by the computer (5). Information from a given domain, when extracted, can be used to solve problems in that domain with the help of a computer (6). For the computer to do this a knowledge base of the domain is created and a KR scheme is used to determine what an appropriate form of representation will be. These KR schemes, include logical representation schemes (used in translating natural language into a language defined by logic), procedural representation schemes, which make use of production rules (i.e. ensuring a condition is satisfied before proceeding), and create a series of "if then-else" instances; network representation schemes, which presents graphical representation of concepts and relationships; and structured representation schemes, which can present a combination of different forms of representations (6, 7).

The main advantage of a KR lies in its ability to "make things explicit" and to "resolve ambiguities" (8). Since a KR is ultimately about processing by machine, the representation of reality that is modeled has to be clear content-wise. Ambiguities that humans can perhaps deal with also have to be removed. KRs are known to be well researched and easily verifiable by whatever method of cross-referencing applied to it (9, 10). Some known KR applications are expert systems and knowledge-based systems (23).

Aim, objectives and research hypothesis

Aim of the study

This dissertation explores how knowledge in life-sciences textbooks meant for learners in high schools can be represented in a graphic knowledge based computer system as a strategy for enhancing epistemological access, that is, helping learners to overcome some of the complexities of scientific English.

Objectives of the study

The following are the objectives of the study:

a) To identify texts from student textbooks and model them using entities and relations.

b) To create an application with an underlying database in which the entities and relations will be stored (relations will be displayed in two languages, creating a multilingual format).
c) To create a user interface for the application with a visualization property
d) To experimentally determine the effect on a group of learners of knowledge modeled on the application.

Research Hypothesis:

In experimentally determining the effect of the knowledge modeled on the application, the following null hypothesis is formulated:

> In the test to be conducted, there will be no difference in the performance scores of those learners who are exposed to the knowledge modeled in the application and scores of those learners exposed only to knowledge in text without the model.

Ethics clearance

Ethical clearance for this project was obtained from the Faculty of Science Ethics committee of the University of Cape Town (UCT) (see Appendix A). Approval was also obtained from the Western Cape Education Department (see Appendix B).

Dissertation Outline

In Chapter two, we provide a background to the study and review literature from both computer science and linguistics. The computer science literature focuses on different knowledge representation formalisms. Although the focus is on Conceptual graphs, the formalism employed in this study, the review also covers other graph-based knowledge representation formalisms such as Concept graphs and Semantic networks. The linguistics literature examines the challenges in learning science, and presents certain frameworks (multimodality and translanguaging) as being easily combined with insights from KR to provide a response to the learning challenges of South African learners.

In chapter three, we describe the methods employed in this research. We describe the extraction of knowledge from life sciences textbooks. We describe the assumptions that guided the modeling. We also describe the tools and the programming languages used in developing the application database and user interface.

In chapter four, we present and discuss the findings of the study. In reporting on the quantitative data, we use both descriptive and inferential statistics. The quantitative data are supplemented by qualitative data derived from interviews conducted with participants in order to elicit their subjective experiences of the graphically modeled knowledge.

In chapter five, we present an overview of the entire research and suggest future lines of research. The references and appendixes are provided afterwards.

Chapter Two: Review of Related Literature

In this chapter, we review the literature on knowledge representation as well as on select knowledge representation formalisms. This review provides a context for processing the challenges experienced by students in their encounter with scientific English. After a review of these challenges, the chapter provides an overview of research in linguistics on multisemiocity through which it is obvious, for instance, that a knowledge representation formalism, being a graphical mode, is one of several legitimate modes through which texts may be understood or meaning extracted.

Knowledge Representation (KR): Overview and Principles

Knowledge Representation (KR), also known as knowledge representation and reasoning (KRR), involves the analysis of knowledge (related to a real world issue) in order for this knowledge to be processed (i.e. reasoned with) by the computer (5). Shapiro (11), while reviewing Sowa (10), defines Knowledge Representation (KR) as “the application of logic and ontology to the task of constructing computable models for some domain” (11). Also, Shapiro (11) describes several fundamental principles underlying all KR. We will now review these principles in order to get clearer insights into the above definition.

Principle 1 is that all KR is a surrogate for the real world knowledge, useful and necessary for reasoning about the real knowledge. One implication of principle 1 is that knowledge models are simplifications of the real knowledge, and therefore not a perfect match to the real knowledge.

Principle 2 is a consequence of Principle 1, and it considers KR as a set of ontological commitments, that is, perspectives on what entities exist. A particular analysis of knowledge underpins all KR. KR is, therefore, a particular conceptualization of the real knowledge. In other words, different KR formalisms impose different views of knowledge, attending to certain dimensions while ignoring others. Attention may be on entities and the relations between them, on prototypical objects and on attributes-objects-values, or on rules of inference which are applied to create inferencing abilities for a would-be KR system.

Principle 3 is that all KR derives from an idea of how reasoning (intelligent reasoning) should take place, but KR is ultimately a partial picture of this idea because no KR is able to account for all of the complexities associated with how (intelligent) reasoning takes place. The reasoning theory of a KR may be determined by examining its handling of inferences, for example, in terms of indicating what can be inferred, or prescribing what should be inferred.

Principle 4 is that since KR is ultimately about getting the machine to behave intelligently, it is important that its design or language or algorithm be such that the machine can efficiently behave in the way intended, e.g. quickly draw inferences, etc.

Principle 5 is that all KR is a "medium of expression and communication in which we tell the machine (and perhaps one another) about the world" (11). The implication is that the KR must not only be adequate in expressing the representation, its expression, must also be understood by the interpreter (machine or human). In other words, its medium must be adequate to think or reason with.

The above account shows KR as a means of simplifying complexity and as guiding and allowing reasoning over a model of knowledge. Although processing by the machine is essential to the very conceptualization of KR, principle 5 allows for a view of KRs as languages in which humans communicate with other humans, and whose effectiveness need to be determined. It is therefore clear why KR is important for "making things explicit" and "resolving ambiguities" (12). Since a KR is ultimately about processing by machine, the representation of reality that is modeled has to be clear content-wise. Ambiguities that humans can perhaps deal with also have to be removed.

Knowledge Representation (KR) Schemes

Information from a given domain, when extracted, can be used to solve problems in that domain with the help of a computer. For the computer to do this, a knowledge base (which is a method of storing information in a manner usable by a computer) of the domain is created and a KR scheme is applied. A KR scheme is used to determine how the knowledge of the user (widely referred to as agent) will be represented (6). Several KR schemes used in creating KRs include:

a) Logical representation schemes: They have very powerful expressivity with the help of inference rules, and are used in translating natural language into a language defined by logic. An example is first-order predicate logic (6, 7).
b) Procedural representation schemes: These representations are based on satisfying a specified condition, after which an action is then carried out. They make use of production rules and create a series of "if then-else" instances (6, 7).
c) Network representation schemes: information is represented as graphs; nodes and arc are used to display concepts and relationships respectively. Examples include semantic networks and conceptual graphs (6, 7).
d) Structured representation schemes: these are complex networks consisting of slots that contain information in various representation forms such as logical sentences and procedural forms. An example is frames (6, 7).

Knowledge Representation (KR) formalisms

There are several KR formalisms in use today. Some are graph-based while others are not (such as frames, ontologies, rules, descriptive logic, first-order logic). These KR formalisms are what KR schemes described above utilize in creating knowledge representations. Graph based KR formalism are mostly used in network representation and structured representation schemes. The focus in this review will be on a graph-based KR formalism.

Graph-based KR formalisms include conceptual graphs, Concept graphs (which is very similar to conceptual graphs), Semantic networks, Inference graphs and Causal Bayesian networks. Though communicating information effectively is the key goal of KR, being able to present information visually has long been known to provide effective communication to humans (34).

Representing information as graphs helps to reduce the size/volume of the information being represented because it allows for visualizations, gives clear description of the knowledge (that is, its concepts, relations and relationships between them), and also forms a good basis for formal description of knowledge processing (34).

My study seeks to create graphical visualization of text content and graph-based KR formalisms will suit this purpose. As it is the formalism to be used in this study, the Conceptual graph formalism will be reviewed in detail, while some of the other graph-based KR formalisms (semantic networks and concept graphs) will only be reviewed briefly. The reason Conceptual graphs are used in this research is because it is close to natural language expressions, allows for unambiguous expression, and makes fewer processing demands on the human interpreter.

Conceptual graphs

Definition

Conceptual graphs (CG) has been defined as a system of logic based on the existential graphs of Charles Sanders Peirce and the semantic networks of artificial intelligence which communicates meaning in a form that is logically precise, humanly readable and computationally tractable (8, 9, 10, 23, 24). Existential graphs formalism basically makes the case of a graph nested within another graph. This simply means if ‘graph B’ is nested in ‘graph A’, then graph B only exists if Graph A exists (9, 24); Semantic networks will be explained later on. CGs are bipartite graphs that consist of nodes called concepts and relations that are linked by an arc that is usually directional.

Architecture (Linear and Graphical forms)

CG can be presented in two forms: either as a linear form or as a display form. The linear form is when a CG is drawn using linear text with square brackets and parentheses; an example is:

[Concept_1]→ (relation)→[Concept_2]

FIGURE 1: CG LINEAR FORM

The square brackets are used to represent the concepts, while the parentheses are used to represent the relations. The text above is read following the directions of the arrowheads; in this case it is from left to right. Therefore, concept 1 is (name of relation) of concept 2. The display or graphical form is when the CG is drawn using rectangles to represent the concepts and circles/ovals to represent the relations. An example of the display form is shown in figure 2 below:

FIGURE 2: A SIMPLE GRAPH WITH GENERIC CONCEPTS (24)

It also reads as follows: “The relation of Concept_1 is Concept_2”; if the directions of the arrows are reversed, it then reads: “The relation of Concept_2 is Concept_1”. The translation of the two CG forms provides the same information without any changes. A concept can stand alone on a graph, i.e. it can be the only part of a given CG (e.g. [Person], which means ‘there is a person”). It is important to note that two concepts cannot be connected to each other directly without a relation between them (9, 25).

A relation consists of only a relation type (this is the name given to the relation; it tells us what kind of relation we are working with) and cannot stand on its own without an arc on a CG. A relation has certain properties: the valence (which indicates the number of arcs that belongs to a given relation) and the signature (which is the number of concept types involved in a given relation). Valence and signature can only be determined by the relation type (25). The arc is an integral part of a CG with a very important role. Concepts are linked to relations with an arc that is usually directional. The pointing of the arc away from a relation or towards a relation has a direct effect on how a CG is read and interpreted (9, 24, 26).

It is important to note that the reading of the graph in many cases may not be grammatically correct but will be very clear to the reader. It is interesting to see how the conceptual graph formalism avoids some of the complexities of natural language. A sentence such as the following can easily confuse learners in schools who are not comfortable with (academic) English or are unable to analyze carefully. For example:

> “Sensory neurons: along whose fibres messages are passed into the central nervous system from sensory cells or receptors”

Learners may have difficulty answering a question like: Where are messages from sensory cells sent to? How are messages sent to the central nervous system? But a conceptual graph representation could make the sentence much easier to understand, or even the process of building the graph can compel learners to get all of the information needed to understand it well. Below is a graphical representation of the above example:

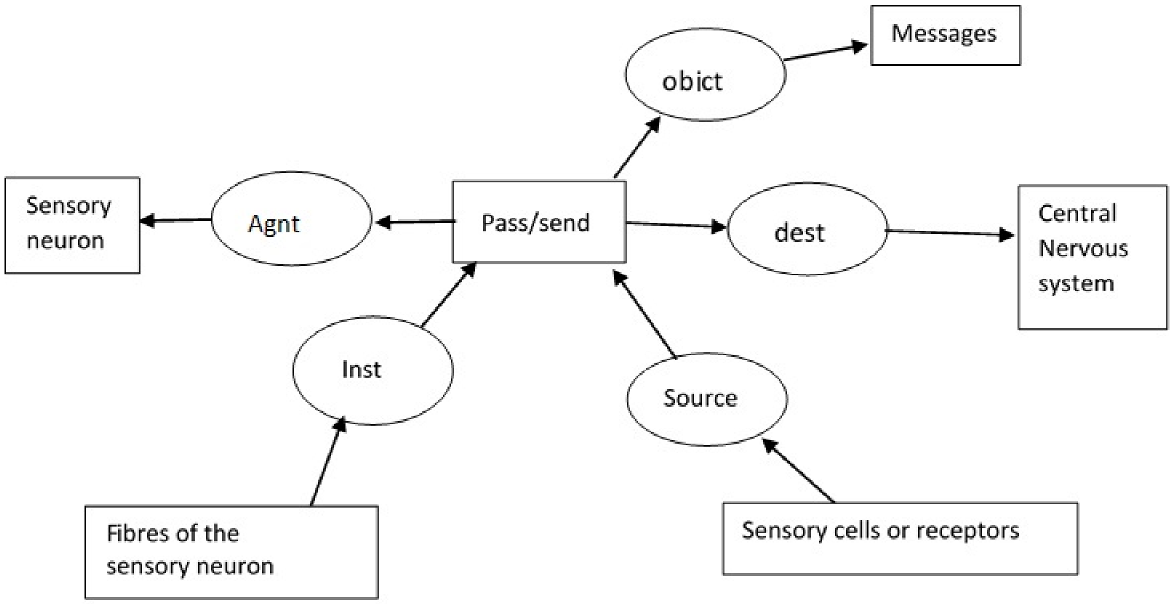


FIGURE 3: A CONCEPTUAL GRAPH REPRESENTATION OF "SENSORY NEURONS: ALONG WHOSE FIBRES MESSAGES ARE PASSED INTO THE CENTRAL NERVOUS SYSTEM FROM SENSORY CELLS OR RECEPTORS"

The names of the relations in the circles can of course be simplified, e.g. 'doer' in place of 'agnt:agent', 'something' in place of 'objct:object', 'to' in place of 'dest:destination', 'through what means' or 'how' in place of 'inst:instrument', and 'from (where)' in place of 'source'. Also, the names of the relations can also be provided in languages other than English. However, in its current form, the graph reads as follows:

- The agent of pass/send which is sensory neurons
- The object of pass/send is messages
- The destination of (what is) passed/sent is central nervous system
- The source for (what is) passed/sent is sensory cells or receptors
- The instrument for passing/sending is fibres of the sensory neurons.

With this representation, answering the kinds of questions posed earlier (Where are messages from sensory cells sent to? How are messages sent to the central nervous system?) is presumably easier.

Logic /reasoning

The CG formalism has a number of mechanisms that makes it a powerful means for representing knowledge and adding value to data (through inferring new knowledge). Some of these features or mechanisms are as follows:

a) *Type Hierarchy*

A concept type is usually a name assigned to a group of entities with similar traits, while a referent is usually an instance of the concept type (e.g. Person: John). A concept type can have a "supertype" and/or a "subtype" (where the subtype inherits all the attributes of the supertype, as would happen in an object-orientation environment). This is known as a type hierarchy. An example is "Manager < Employee" where manager is a subtype of a more general entity, Employee. A type may be defined:

1) By extension: here, a list of references to all members are provided.
2) By intension: it defines rules/properties that clearly states how members are identified (from the example above, all managers are employees)
3) By reference to other types: here a clear description of what sets a particular type apart from the others. Consider this example:

   Animal < Animate, Mobile_Entity,

   Physical_Object. Animate < Entity.

   Mobile_Entity < Entity.

   Physical_Object < Entity.

   Animal in this example will have the characteristics of all its supertypes being animate, a mobile entity and a physical object (24, 25).

b) *Canonical formation rules*:

In Conceptual graphs, canonical formation rules affect the logical relationship between an original graph *u* and its resultant graph *v* when applied (8). According to Harmelen et al, (8) the six canonical rules can be grouped into three:

1. Equivalence rules: which are "Copy and simplify" and are responsible for the generation of the equivalent of the original graph. This means with "Copy rule", the original graph is identical to the resultant graph. Therefore, *graph u* and its resultant *graph v* are exact copy of each other. With the "simplify rule", if two relations x and y in original graph u are duplicates of each other, then one of them can be removed completely from the resultant graph v along with all its arcs. Thus, the equivalent rules thus show that $u \supset v$ and $v \supset u$) (8, 45).

2. Specialization rules: which are "Join and restrict" which generate a graph *v* that implies the original. This is how the "Join rule" works: suppose there was another graph w, which was the same as original graph u; if *concept b* in graph u is identical to *concept c* in graph w, then let in the resultant graph v, *concept c* be deleted and the arcs of its conceptual relations be linked to *concept b* (8, 45). With the "restrict rule": for a given *concept c* in the original graph, type (c) can be substituted with a subtype (c). The specialization rules show that the resultant graph v is partially identical to the original graph u due to some existing restrictions (8, 45).
3. Generalization rules: which are "Detach and unrestrict". The detach rule refers to being able to erase a given conceptual relation along with all arcs belonging to it from a graph. We could say that detach is an inverse of the Join rule; while unrestrict is the inverse of the restrict rule. These generalization rules are responsible for the generation of a graph as insinuated by the original graph (meaning, that there are no restrictions to the extent in which original graph can imply the resultant graph) (8, 45).

c) Inferences:

The inference in CG is based on Pierce's existential graph logic ("if- then" logic). It basically says that since B is found in A (i.e. they are nested in together), Graph B can only be created if Graph A is present. To illustrate this, consider figure 4 below: Graph B is contained in Graph A and Graph A is contained in the Mother Graph. What this shows is that the content of Graph B can be found in Graph A, and that of Graph A can be found in the Mother Graph. Pierce's logic here will be read as "**if** Graph A **then** Graph B". Figure 4 also shows that the Mother graph dominates graph A and graph A dominates graph B (24).

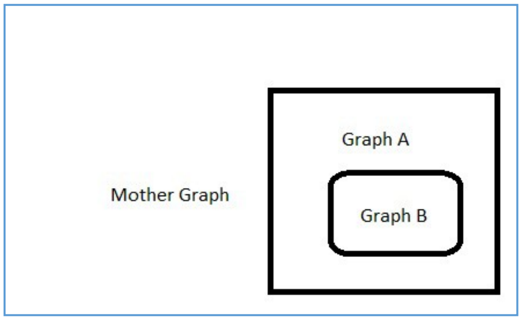


FIGURE 4: PIERCE'S "IF-THEN" LOGIC (ADAPTED FROM 24)

In CG, the possibility of generating new knowledge, which was not originally explicit in the graph, becomes very easy and doable (9, 24).

Applications and Tools

Many researchers have documented several advantages of using graphs. Graphs have been noted to present better efficiency in that a lot of information can be gathered at a glance compared to having to read through linear texts. According to Sowa (9), the structural nature of graphs makes it possible for them to "simplify many algorithms for reasoning, searching, indexing, and pattern matching". Sowa (9) went on to describe CG as having "direct mapping ability to language, and serving as an intermediate language for translating computer-oriented formalisms to and from natural languages" (9, 26). With their graphic representation, they serve as a readable alternative language.

CGs have been implemented in a variety of projects for information retrieval, database design, expert systems, and natural language processing (12, 23, 24, 26, 27, 28). Conceptual graphs have been deployed for knowledge representation to solve problems in a variety of areas: Health (9, 29), Computer Science (9, 27, 29), Geographical Sciences (29, 30), Biological Sciences (21). These applications show how versatile CGs are.

At the University of Surrey (Guildford, UK), Ahmad (12) and his team developed a conceptual graph editor to support the visualization of terms in a terminology database (12). Some of the assumptions of Ahmad's work include the following:

a) Specialist terms encode much of the knowledge (that is, concepts) in a discipline (e.g. neuron, central nervous system, etc.).
b) These specialist concepts and their terms are structured in the sense that they exhibit different types of relations or links (e.g. sensory neuron is a type of neuron, motor neurons is a type of neuron).
c) Unfortunately, when terms are entered into a database, there is no means of visualizing the relations which these concepts enter.
d) Also, when terms are stored in a conventional database, there is no way of navigating them in customized ways and inferring new knowledge from them.

Given these problems, Ahmad and his team developed a conceptual graph editor for representing terminological knowledge (concepts and the relationships between them). From a database of pre-stored simple graphs linked to a terminology database (for example on genes), the conceptual graph editor makes it possible for the user to build and visualize graphs. When a user performs a search on a concept (e.g. genome), the user can click the editor's build function, which will apply the formation rules automatically to display the graphs for the selected concept. A filter option allows the user to set a constraint on the types of relations to be displayed (e.g. is_a (type of), part_of). There is also an option that activates the inheritance mechanism.

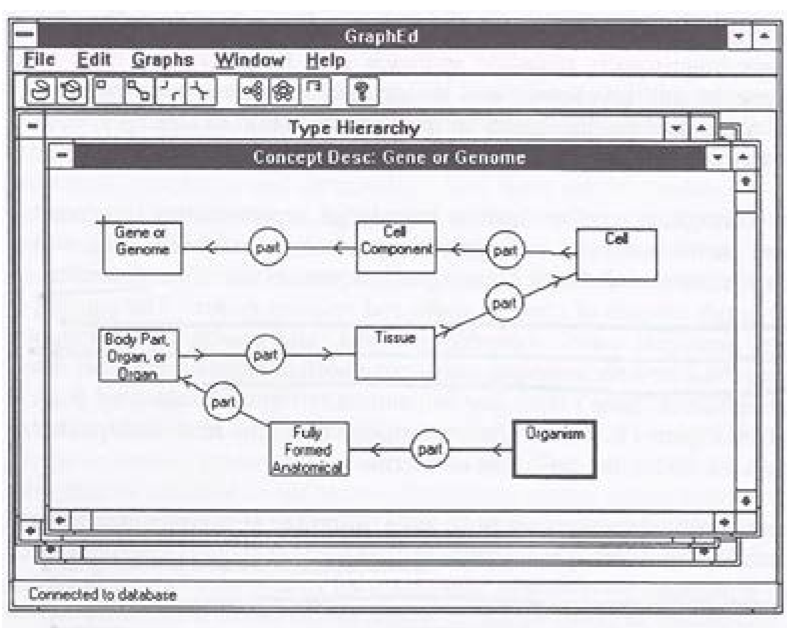


FIGURE 5: CONCEPTUAL GRAPH EDITOR DEVELOPED BY AHMAD AND HIS TEAM (12)

In Figure 5 above, the constraint, set on the search is ‘part_of’; that is why the results presented only display this relation type (part_of) associated with the search concept: gene/genome.

Within the context of the classroom specifically, one widely used graph application inspired by conceptual graph is the concept map (Cmap). According to Robinson, “A concept map is a diagram consisting of nodes that represent concepts and labelled lines that indicate the relationship between those concepts” (13). Several studies have been conducted to show the different ways the Cmap has been used to enhance understanding especially amongst science students (15, 16, 17, 18, 19, 20). These various ways have been classified into four groups including: acting as a learning strategy, as an instructional strategy, as a tool for instructional design process, and as an assessment tool (13, 14).

Most relevant to my study is the use of Cmap as a learning strategy. In one Brazilian study, undergraduate students having challenges with reading academic texts in English in their respective fields were taught to create Cmap and were asked to model their academic texts with Cmap. According to the researchers, the results showed that the use of Cmap enhanced their comprehension of texts in English. Also according to the researchers, “students’ own narratives revealed the positive impact the process of map drawing had on their awareness of the reading process and how they managed to have more control over reading comprehension in English by visually representing what was conveyed in the texts they read”. Also from the findings of the study, the researchers clamed that “the creation of concept maps for enhancing comprehension in English as second language (L2) can empower students in two important ways. First, in the development of their autonomy concerning ways to organize knowledge acquired from texts and, secondly, in their awareness that they can read well in English for academic purposes once they use adequate strategies”(20).

A second example of the application of conceptual graph comes from the representation of biological knowledge in the context of teaching and learning. The researchers wanted to improve students’ ability to learn and retain the vast knowledge required of them during their study. In order to assist the teachers in disseminating this knowledge, technology assisted tools were applied. The researchers developed a concept mapping model which was used to map and analyze the conceptual parts of the work. (21).

Computerized CG based tools have been developed and applied across different fields. The official site for conceptual graphs (26) lists the following:

a. “Amine- a multi-layer platform dedicated to the development of Intelligent Systems and Multi-agent Systems.

b. CharGer - a prototype conceptual graph editor developed at the University of Alabama in Huntsville, free for noncommercial use, and runs under Java.

c. CG Mars Lander - fast conceptual graph retrieval and question answering tool, available for joint development and industrial funding.

d. CoGITANT - several useful utilities: a set of library routines in C++ for conceptual modeling, some knowledge bases in conceptual graphs, and an XML specification for CGXML.

e. CPE - a modular environment that provides modules to give functionality to a user without having to take the whole environment. Currently, there is a CGIF editor, ARC Edit and other CG Operations (Projection and Maximal Join) should be available in the future.

f. GoGui - a free graph-based visual tool, developed in Java, for building Conceptual Graph knowledge bases represented in COGXML format, compatible with CoGITaNT (see above).

g. Prolog+CG - an object-oriented extension of PROLOG, based on CG. CG (both simple and compound) is a basic data structure, like term. PROLOG+CG is implemented with Java.

h. WebKB - tools for information retrieval and knowledge representation” (26)

In concluding this section, it is important to note the comparisons and contrasts that have been made between Conceptual graphs and Concept graphs. Conceptual graphs and Concept graphs are two very similar KR formalisms with a few differences. They are similar in their structure, semantics, mathematical methods as well as inference mechanisms (23, 31, 32). While Conceptual graphs were developed by Sowa (9, 31, 32), Concept graphs was first developed by Willie from the combination of conceptual graphs and formal concept analysis (31, 32). Quite notable is that Concept graphs are represented graphically as multi-hypergraphs and bipartite graphs similar to conceptual graphs but use multi-hypergraphs only in its mathematical representations (31, 32, 33).

Unlike conceptual graphs, concept graphs allow references with two or more individual markers (34). In a recent study, Agrawal et al (35) made use of both conceptual graphs and concept graph for a similarity search where preferred web contents are gathered from several web documents.

While several conceptual graph tools and applications are more in circulation than those of concept graphs, Aubert et al (31) explained that choosing to use concept graphs or conceptual graphs "was a matter of taste, since their mathematical objects are very similar". There are some known concept graphs tools, including cTakes 3.1.2 with YTEX graph drawing software, graph databases such as Neo4j and Ondex software, which is an API platform (34).

Semantic Networks

Definition:

Semantic networks can be defined as "a graph structure for representing knowledge in patterns of interconnected nodes and arcs" (36). Another way of looking at semantic networks is as an association between abstractions, concepts and ideas, which may turn out to be vague. While having been successfully used in developing expert systems in biomedical studies and other areas, critics have pointed to its ambiguous nature (12, 37).

Semantic networks were originally used in areas such Philosophy, Psychology and Linguistics before its application in artificial intelligence and machine translations (36). Semantic networks have evolved over the years; they were used to represent knowledge in defined areas (domains), then expanded to represent knowledge on the web and subsequently used for decision making in machine processing (12, 36, 37).

Architecture, Logic /reasoning:

For reasoning, semantic networks use a process called Inheritance, which ensures that the subtype concept inherits all the characteristics of the super-type concept (except otherwise stated). In addition, subtypes can have many super-types allowing for multiple inheritance to occur. For inference, semantic networks allow for implicit knowledge inferencing by applying a process called inverse links (36, 38).

Graphically, Semantic networks can be represented as a labelled directed graph made up of nodes (concepts) and arcs (relationships) used to connect the concepts. Concepts in semantic networks are of two types: the concept in domain (that is, the first concept in a relationship between A and B) and the concept in range (that is, the second concept in a relationship between A and B). A predicate logic formula (which is the arc) can in some cases, be used to represent a semantic network. For example, F(x, y): F is the predicate logic formula, X is the concept in domain and Y is the concept in range (34).

Figure 6 presents an example of a semantic network graph:

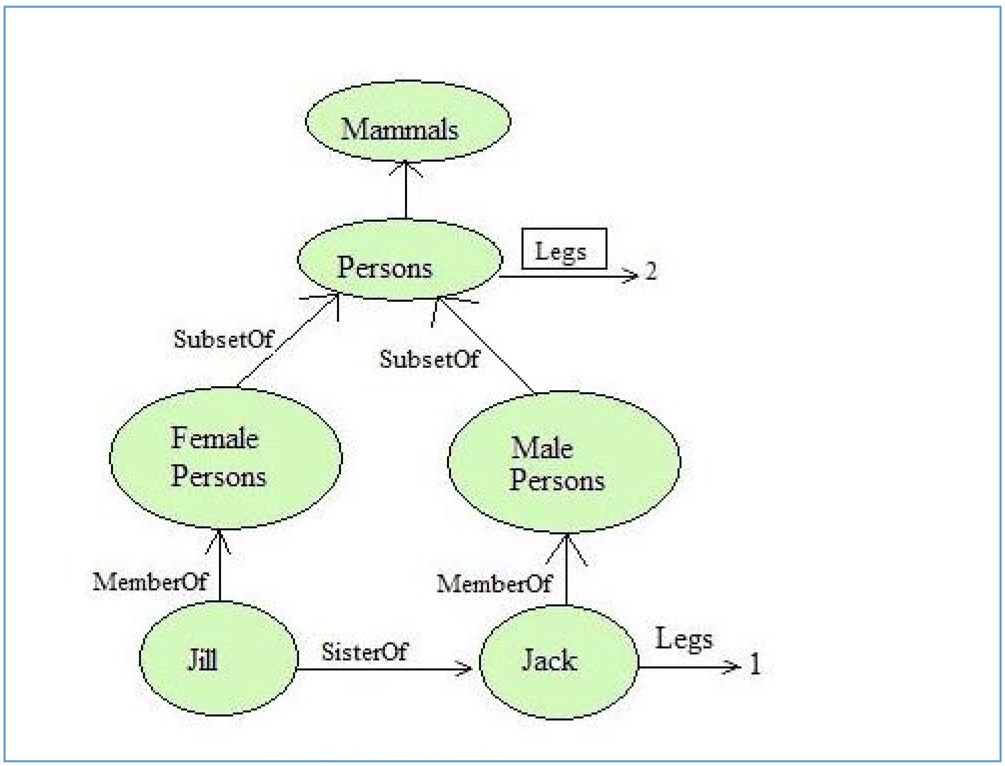


FIGURE 6: EXAMPLE OF SEMANTIC NETWORKS (38)

In figure 6 above, arcs connect the concepts (presented as ovals). The subtypes FemalePersons" and "MalePersons" will inherit all the characteristics of the supertype "Persons" and are linked by the arc "SubsetOf". The subtypes themselves are connected to other subtypes by the arcs "MemberOf". Thus "Jill" and "Jack" will belong to the category with which they are linked and will inherit the characteristics of this catergory (38).

Types of Semantic networks:

Sowa (36) describes different ways in which semantic networks can be classified, including: definitional networks, assertional networks, implicational networks and hybrid networks

1) ***Definitional networks:***
   These were first used in the 3rd century AD by a Greek philosopher, Porphyry, to describe a method of category definitions of types and to distinguish between subtypes and super-type as shown in the tree of Porphyry.

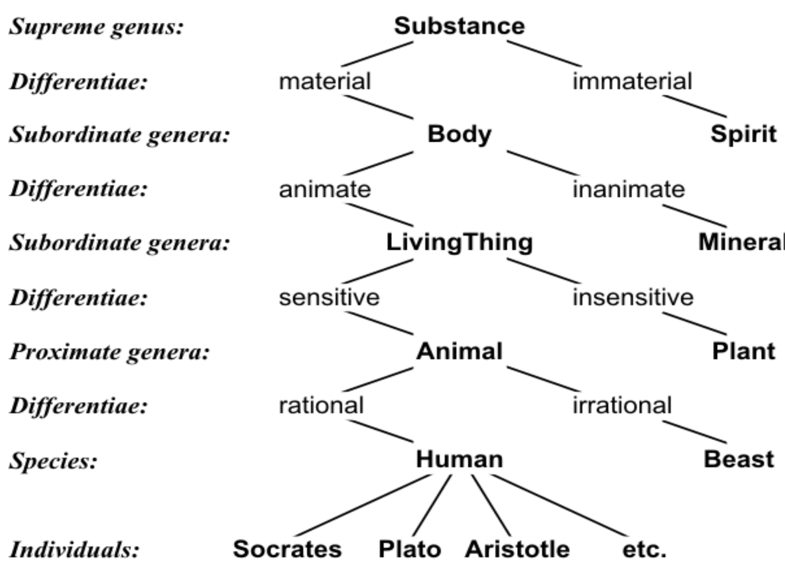


FIGURE 7: TREE OF PORPHYRY (36)

In figure 7 above, Porphyry describes the breakdown and inheritance from **Supreme genus** to **Individuals.** Substance is divided into "material" and "immaterial", which then creates subordinate genera: **Body** and **Spirit**. The subsequent breakdown and inheritance can be seen all the way down to **Humans.** Figure 7 has been referred to as the origin of the modern day method of hierarchical definitions used to create the rule of inheritance, which guides how a sub-type inherits from a super-type (36). Knowledge represented by this type of network is assumed to be true since most definitions are true (36).

2) Assertional networks:

This type of network is the reason why semantic networks are referred to as a propositional network because of their capacity to assert propositions. They are known to produce more truthful information in comparison to the Definitional networks. Formalisms such as existential graphs and conceptual graphs draw their semantic aspects from the Assertional networks.

Some of the early uses of propositional networks were in Linguistics where Lucien Tesniere applied it his system of dependency grammar which later became the basis for many other theories in this field including the Valency theory and the Meaning-Text theory (36, 38).

The Mind system was the first known application of assertional or propositional networks in the area of Artificial intelligence. This system later became what is now known as Semantic Network Processing System (SNePS)

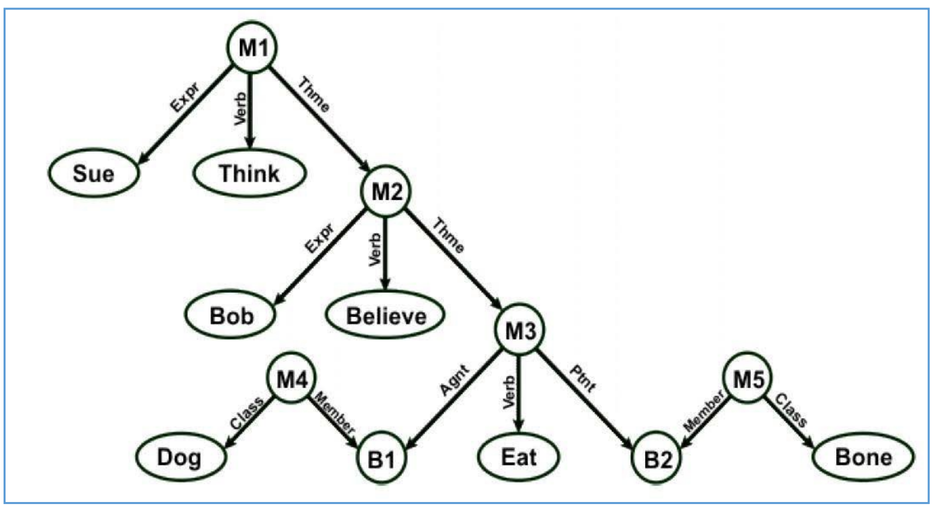


FIGURE 8: MIND SYSTEM (36)

As Sowa (35) explains, in figure 8 above, the "proposition M1 states that Sue is the experiencer (Expr) of the verb think, whose theme (Thme) is another proposition M2.

For M2, the experiencer is Bob, the verb is believe, and the theme is a proposition M3. For M3, the agent (Agnt) is some entity B1, which is a member of the class Dog, the verb is eat, and the patient (Ptnt) is an entity B2, which is a member of the class Bone" (36).

3) Implicational networks:

This is also a type of propositional network; here, the first relation is consequential. That is, it is the reason for the direct adjoining node. The example below (figure 9) shows how related or unrelated the rainy season is to the wetness of grass and its subsequent slipperiness. If it is a rainy season, then T would be true showing that a recent rain is the reason for the wet grass and it being slippery; however, if it is not a rainy season (making T false), F would be true, showing that the sprinkler in use is the reason for the wet grass and it being slippery (36).

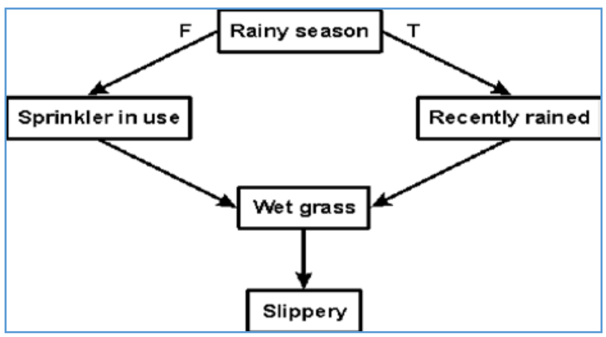


FIGURE 9: IMPLICATIONAL NETWORK (36)

The reasoning system of implicational network could be logic-based or probabilistic depending on the approach used (36).

applied in one application process. An example is when a web-based application is created using HTML, an object oriented programming language (Python) and a database. Unified Modeling Language (UML) is a widely used hybrid notation.

4) **Learning networks**:

These are networks that are able to change its constitution as a result of new knowledge/information received. It does this either by removing previous nodes and arcs as result of the more recent knowledge acquired (called “Restructuring”) or by including the new knowledge to the one already present (called “Rote memory”); or by removing the strength carried by numbers associated with certain nodes and arc called “weight” (in a process is called “changing weights”) (36).

5) **Hybrid networks:**

This is the combining of different types of semantic networks together to use in one process (36). This means there could be a combination of the types of semantic networks applied in one application process. An example is when a web-based application is created using HTML, an object oriented programming language (Python) and a database. Unified Modeling Language (UML) is a widely used hybrid notation.

Tools and applications:

Several tools and applications have been developed in recent times to implement semantic networks formalism. Semantic networks processing systems (SNePS), WordNet and MultiNets are some of the well-known ones (36).

Linguistic Challenges in Learning Science

Academic English is a register. In linguistics, a register is defined as a “cluster of associated features having a greater-than-random … tendency to co-occur” (4). In many scientific texts (e.g. journal articles), the features that may co-occur (that is, appear together) more frequently than could be explained by chance might include the use of the passive, lengthy sentences, nominal (noun-based) constructions, technical terms, and so on. With registers of science as of other fields, the relevant cluster of features is functional, as it allows experts to formulate their thoughts and to make their arguments in a way which would otherwise have been difficult. However, while this register is suited to the expert’s needs, it causes “difficulty to the novice” (4). It causes difficulty both to the mother tongue user of English and the user of English as a second language. There are at least six sources of difficulty of academic English as outlined in (4).

The first is **interlocking definitions**. Definitions may sometimes contain so many other concepts that are used to define each other and that first have to be simultaneously understood before the main concept being defined can be understood. Consider the following example:

> "A circle is a plane curve with the special property that every point on it is at the same distance from a particular point called the centre. This distance is called the radius of the circle. The diameter of a circle is twice the radius. The length of the circle is called its circumference" (4)

In the example above, understanding circle requires that at the same time definitions of other concepts (such as plain curve, radius) be understood. It is taken for granted that plane curve and distance are understood. Diameter is defined in relation to radius, circumference in relation to circle. It can be challenging for a novice to process so many definitions at the same time.

The second challenge, related to the first, has to do with **technical taxonomies**, that is, the fact that scientific concepts occur in clusters within which they are organized according to specific criteria of classification, e.g. relationship of superordination (e.g. x is a y) or relationship of composition (e.g. x is part of y). Problems which these taxonomies may pose include the following:

- they can be complex and have several layers of organization: w is a type of x, and x is a type of y, which is a type of z,
- they are at times not made explicit or clear to the learner,
- the criterion or parameter used in establishing them may not always be clear to the learner. In description of a group of organisms, some of them may be classified according to habitat while others are classified according to some other parameter.

A third challenge is **lexical density,** that is, the number of words in a sentence that are technical terms or parts of terms. In academic English, many of these words may occur within a noun group. See, especially, the initial noun group in bold in the following example:

> "**The conical space rendering of cosmic strips gravitational properties** applies only to straight strings" (4)

There is so much information condensed into the noun group functioning as the subject of the sentence that unpacking it can be challenging for novices.

A fourth difficulty has to do with **syntactic ambiguity**. Consider the following sentence:

> “lung cancer death rates are clearly associated with increased smoking” (4)

The verb group ‘are clearly associated with’ can be ambiguous. The expert may have enough background knowledge to determine that it means ‘caused by’. A novice could process it as meaning ‘causes’ or ‘leads to’ (people/smokers upset by all the talk about lung cancer deciding to smoke/more). Equally ambiguous is the noun group that functions as the subject of the sentence. ‘Lung cancer death rates’ could mean:

- how many people die of lung cancer
- how quickly people die of lung cancer
- how quickly people’s lungs die of lung cancer

The second noun group ‘increased smoking’ could mean:

- more people smoking
- people smoking more
- more people smoking more

Novices may not always have enough background knowledge initially to resolve these ambiguities.

A fifth challenge has to do with **grammatical metaphor**, that is, the shift or change from one grammatical class (e.g. verb) to another (e.g. noun). In ordinary language use, reality tends to be expressed as Participants (animals, people, things, abstract objects), as Processes (verbs), and as Circumstances (how, adverbs). However, academic English tends to convert many of these categories into nouns. Such conversion leads to loss of certain kinds of information and therefore to the kinds of ambiguity seen above. Instead of the underlying idea of ‘how quickly people die from lung cancer’, which comprises Circumstance (how quickly), Participants (people, lung cancer), Process (die), Preposition (from), academic English says ‘lung cancer death rate’. In this noun group, the original adverb (how quickly) has become the noun ‘rate’; the original verb ‘die’ has become the noun ‘death’; the original idea of ‘people’ has been lost and the preposition ‘from’ has also been lost. Such grammatical metaphor represents a view of the world in which virtually everything has turned into a noun (4). Thinking in this way may not always be easy for novices.

A sixth challenge, **semantic discontinuity**, arises when writers of scientific texts “sometimes make semantic leaps, across which the reader is expected to follow them in order to reach a required conclusion” (4). Consider the following example:

> “A stem also transports water and mineral salts from the roots to the leaves and manufactured food from the leaves to all parts of the plants. It therefore has well developed xylem vessels, xylem trachieds and phloem” (39).

The information in the above example does not clearly state that xylem vessels, xylem trachieds and phloem are part of the stem. It also does not state explicitly what functions they perform.

In conclusion, these and possibly many other characteristics of written academic English may explain why students would find this English difficult, irrespective of whether they are mother tongue users or second language users. The section of this chapter on knowledge representation formalism highlights opportunities for addressing the above challenges of scientific English.

Multisemiotic representation

In this section, we review literature from linguistics that provides a context for seeing graph-based knowledge representation formalisms as highlighting the different modalities through which people understand knowledge

Multimodality

Meaning is usually not obtained from one information channel or source. It has been claimed that “Writing is not always the central meaning making resource in applications for use in school English and Science” (40). As an example of the multiple ways of making meaning or multisemiocity, there is the concept of multimodality which “deals with all the means we have for making meanings – the modes of representation - and considers their specific way of configuring the world” (41). These modes of representation are varied. We can represent meaning in specific ways in writing (even through particular words, letters), in speaking (also through a variety of strategies), in images (drawings, pictures, diagrams, etc.), in colour, in dressing, etc.

The aspects of meaning contributed by one mode may not be the same as those contributed by another mode. In other words, “different modes offer specific resources for meaning making, and the ways in which modes contribute to people’s meaning making vary” (40). This also means that different modes may compel certain decisions to be taken and encourage certain interpretations. When we write or say ‘every cell has a nucleus’ or ‘there is a nucleus in every cell’, we are establishing a logical relationship between two entities. This relationship is not spatially defined. In drawing, however, the nucleus has to be placed somewhere in the cell. Wherever it is placed, a reader may assume that the location is the ideal, exact or only location. See figure 10 below

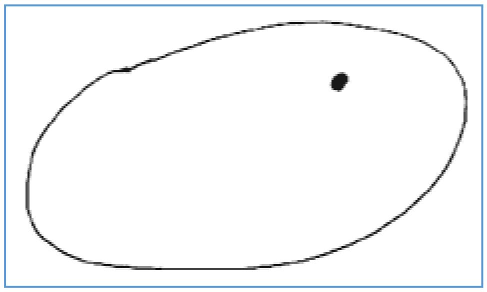


FIGURE 10: A CELL AND ITS NUCLEUS (41)

It is, therefore, clear that a task in multimodality research is to describe the potentials and limitations for meaning which inhere in different modes” (41).

The following text makes several potential meanings.

> “Although most angiosperms are terrestrial, living in areas with adequate water supply (mesophytes), some angiosperms are hydrophytes, living in and around water, with still others being xerophytes, living in very dry areas” (39).

Firstly, taken from a textbook written by experts, recommended by government, used by novices, this written text comes with some in-built authority, which has several effects. One potential effect is that “it forces the reader (and the listener) to stick to its order: the elements have to be read in the sequence in which they occur” (41). Of course, writing requires a sequence, but the authoritative nature of this text places more pressure on the novice reader to follow the sequence.

Secondly, the written text has a complex structure because it does two things simultaneously and may compel (because of its authoritative context) the novice to want to also process these two things simultaneously. The two things are: stating facts, then contrasting and comparing these facts.

It is even possible that apart from the original context of authority, a text such as this seeks to get even more authority by using a complex construction. As a result, even though the facts can be stated more simply in writing, this may not frequently happen.

As has been observed, the task of the reader "is to attempt to follow the pre-given ordering of the written text, embodying the authority of the author, working assiduously to reproduce the meaning which the author had intended for the reader" (41).

A mixed mode (writing and graph) has different potentials as seen in figure 11 below.

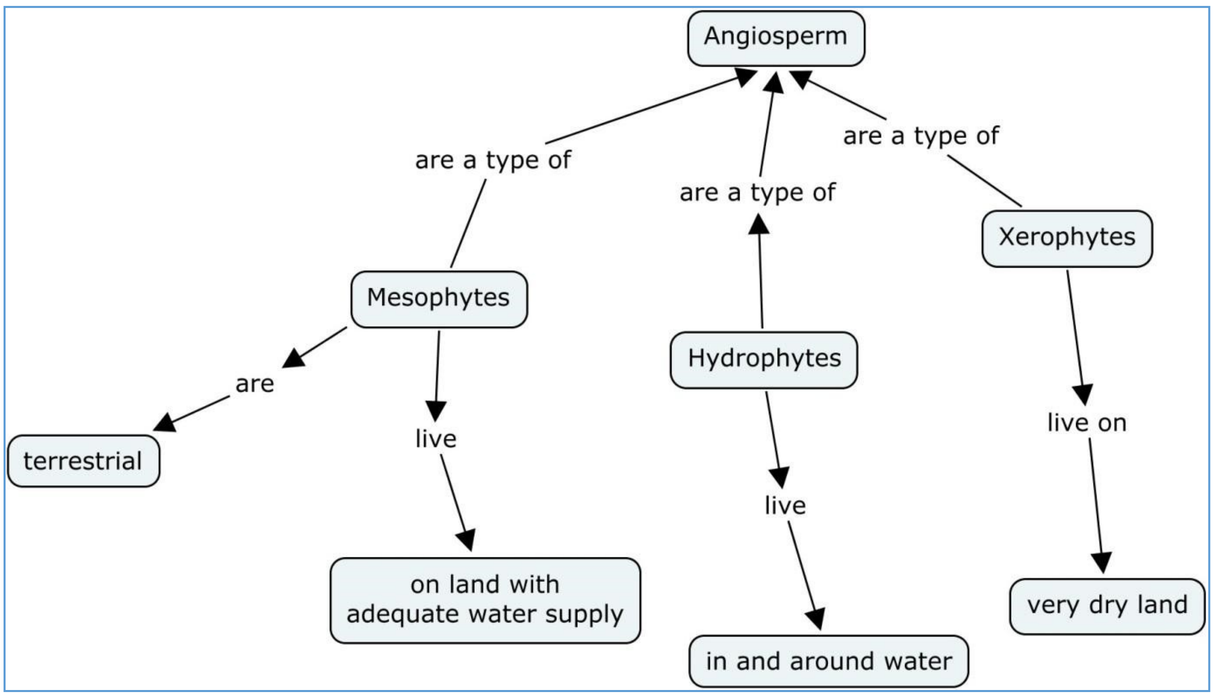


FIGURE 11: MIXED MODE WRITING AND GRAPH (CONCEPT GRAPH REPRESENTATION)

Although different graphic interpretations of the text are possible, the one in figure 11 above shows several differences between writing and graphic visualization as modes. Firstly, the impression of authority (at least some of it) is lost. Graphics are often seen as supplements to written texts and do not always have the kind of authority which writing has. Secondly, the reader does not have to stick to any reading path. Reading can easily begin from xerophytes. So the prescribed or fixed sequence of writing/reading is broken. Thirdly, much of the complexity of the written text is no longer there. The facts are broken down (e.g. hydrophytes are a type of angiosperm; hydrophytes live in and around water). Fourthly, a possible source of confusion is removed (that is, the reference to territorial, which occurs in the same structural position as hydrophytes and xerophytes and can affect an answer to how many types of angiosperms are mentioned). Fifthly, the comparative perspective imposed on the reader by the written text is no longer there. The responsibility for comparison is transferred to the reader. Sixthly, the graphic representation perhaps makes it easier to focus on or ask questions of the several habitats; that is, how they differ.

The above account of the potentials of graphics underscores some of the functions of knowledge representation in artificial intelligence. As seen earlier, in artificial intelligence, representation is "about making things explicit, is about resolving ambiguities" (12). The major point about multimodal representation is that it allows for taking advantage of the potentials of the several modes in such a way that the limitations of each are reduced.

Multilingualism

Just as humans sometimes infer meaning in different modes, we also infer meaning across different languages. The potentials for inference associated with the mix of written text and graphic as seen in figure 11 can be improved by introducing another dimension: multilingualism. Although the graphical visualization already removes the sentence complexity of the text on angiosperms, by using such connectors as 'type of' and 'lives in', these connectors can be expressed in the home languages of the learners, while the core concepts remain in English (because equivalents in the home language may not be known).

Of course, if equivalents are known or definitions can be provided in the home language, these can also be included. The outcome would be knowledge that is both multimodally and multilingually represented. In other words, such representation would allow for meaning to be made across modes and across languages. A relevant notion here is translanguaging. Translanguaging is defined as ‘‘the process of making meaning, shaping experiences, gaining understanding and knowledge through the use of two languages’’ (41).

According to Lewis et al (43), “In the classroom, translanguaging tries to draw on all the linguistic resources of the child to maximize understanding and achievement. Thus, both languages are used in a dynamic and functionally integrated manner to organize and mediate mental processes in understanding, speaking, literacy, and, not least, learning”.

There are a number of advantages of translanguaging. These include the following: “It may promote a deeper and fuller understanding of the subject matter. It may help the development of the weaker language” (43).

Possible implications of multisemiotic representation for my study

In summary, this section provides a basis from language research for the view that exposure to knowledge that is represented multimodally and multilingually will reduce the challenges, which academic English poses to learners. Currently in South Africa, many local languages are not currently well developed in terms of their terminology. If because of problems of terminology, the entities in the graphic model cannot be translated into the first language of the learner, it may at least be possible to translate the relations (e.g. is a type of, causes, used for, part of, etc.) in order to further improve understanding.

Chapter Three: Methodology

The purpose of this study is to investigate if using a graph-based KR formalism (particularly Conceptual graph) to represent knowledge from a science textbook would enhance epistemological access, specifically the understanding of science by high school learners. This requires that an application be developed, populated with obtained data and then tested (in an experimental research). This chapter will describe each of these activities.

Data gathering

The process here was to create a collection of texts that covers part of the selected topic in Life science, and to model the text for use by the learners. To achieve this, the following steps were executed:

Step1: *Text identification*

A high school teacher of life science was requested to identify topics taught, and to identify from the curriculum guide what the learning outcomes of the identified topics are. The teacher also helped to identify the (English) textbooks used in teaching.

Step2: *Extracting of entities and relations and Modeling*

After the identification of the topics, a text on "Angiosperms" from one textbook was subsequently selected. I read and identified the entities (i.e. how they are described) and the relations (i.e. the links or interconnectors) which connect the entities using the methods developed in the conceptual graph model. The texts were then modeled into entities and relations using the principles of Conceptual graphs such as:

1) All concepts are modeled as entities (nodes) while relations (i.e. interconnector) are links to one or more entities to each other.
2) The use of hierarchy on both the concepts and the relations were also applied such that if a node $A \leq B$ and $B \leq C$, then $A \leq C$.
3) The use of a marker (sometimes referred to as individual markers) for specific entities. That is, if a marker conforms to a type A then it must also conform to a super-type of A (9).

Step3: *Verification of the model and translation*

After identifying the entities and relations, we took the developed Conceptual graph model back to the teacher for verification. The teacher cross- checked the entities and relations to ensure they were correctly linked.

After verification by the teacher, the next step was to have the relations translated into Kaaps (a version of the Afrikaans language spoken widely among the local community) which is the language spoken in the Elsie's River area of Cape Town, the site of the high school where learners were recruited to participate in the study.

Development of Application

We developed an application with a graphical database, and a user interface with a visualization component integrated into it. For the graphical database, the Neo4j graphical database management system was used. Since our study design required the production of graphical representations of data, we found the Neo4j graphical database suitable. It stores data as nodes and relations. Neo4j also stores the relationships between data. It produces the graphical representations of the entities, relations and relationships associated with the data, unlike relational databases which would store data as lists in different tables. The Neo4j database management system like other relational databases is manipulated by queries. It however uses the Cypher query language (CQL) instead of SQL. CQL is an "SQL-inspired" query language that is capable of performing create, read, update and delete (CRUD) operations similar to those performed using SQL. In graphs, it helps to show patterns visually with help of an ascii-art syntax (47).

Part of the guiding principles of the Neo4j database comes from the Concept graphs formalism. As stated in chapter 2, the Conceptual graph and Concept graph formalism are very similar and thus using either one of them is a matter of preference (31, 34). The graphical representations created in Neo4j are concept graphs.

For the user interface, we used the Jupyter notebook application. The Jupyter notebook, an open source web application, allows for the creation and sharing of notebook documents containing codes, visualizations, equations and different types of graph representations (44). Although several programming languages can be used on the Jupyter notebook, we chose Python because we are comfortable working with it. The user interface was intended as a reading-support for learners, hence the choice of a notebook design.

We created a connection between Jupyter notebook and the Neo4j database through the use of "Py2neo client library" and then programmed the entities (nodes) and relations using the Python language in Jupyter. The database was subsequently reflected in the database management system as a result of the connection.

The entities and relations that were previously modeled were the data used to populate the graphical database. The translated relations were used as the links to connect nodes to each other while the relations in English were used as the "type" attribute on each relation. This allowed the learners to view both the local language relation as well as the English language relation on mouse over. The database presents the data as graphs, tables and individual nodes and relations. A graphical representation of a section of the text is obtained through the use of CQL queries which extracts the required set of nodes and relations and presents them as a sub-graph.

The visual representation of the content of the database was created with a combination of a Javascript library called VIS.js (designed to handle and manipulate large amounts of data with a capability to display them as a network of nodes and edges) and python scripts containing Cypher queries. With the help of the existing algorithm, we created visualizations of sub-graphs corresponding to different portions of the text; the data was extracted from the database and displayed on the notebook.

Developing questions

The life sciences texts used for the study were analysed against the background of the problems of scientific writing identified in chapter two. This was done to find out what types of potential problems could be found in the different texts. Below we present the results of the analysis along with the questions posed on the basis of perceived difficulties in the text (difficulties that would not be present in the graphical model).

Ambiguity (possible ambiguity) of the text is the basis for questions 1 and 2.

1. How many types of angiosperms are there in the text?

2. Mention the types of angiosperms found in the text

The text reads: "Angiosperm plants are the most varied and successful of all plants. Although most angiosperms are terrestrial, living in areas with adequate water supply (mesophytes), some angiosperms are hydrophytes, living in and around water, with still others being xerophytes, living in very dry areas."

The pattern for hydrophytes and xerophytes (term followed by habitat) is not used for mesophytes. Mesophytes is placed after the explanation. What is placed in front of the explanation instead is the term "terrestrial", which may mislead some readers into thinking terrestrial is a type of angiosperm. This potential for confusion could determine how many types are identified and how they are named.

Also based on a potential ambiguity of the text is question 3:

3. Based on the text, explain to your friend how he or she can distinguish between the types of angiosperms that require water?

As a result of the position of territorial as discussed above, land is not mentioned as habitat in the same way as water is mentioned in hydrophytes. The question is whether the students will be able to retrieve this habitat from "terrestrial" and not consider "terrestrial" as a type of angiosperm.

Questions 4 and 5 are based on a form of organization that could easily lead to semantic discontinuity.

4. What makes it possible for the stem to transport minerals salts, water and manufactured food to different parts of the plant?

5. The xylem vessels, xylem trachieds and phloem can be found in the stem. True or False?

The relevant text reads: “A stem also transports water and mineral salts from the roots to the leaves and manufactured food from the leaves to all parts of the plants. It therefore has well developed xylem vessels, xylem trachieds and phloem”.

The text first states what the stem does, before stating what in the stem makes this function possible. It is the adverb (therefore) that has the responsibility of communicating the inference or conclusion. The reader has to infer that if the stem is able to do X, it must have Y. For a learner who misses the meaning of this adverb, both sentences will be semantically discontinuous.

Questions 6 and 7 are all based on a potential difficulty arising from a technical taxonomy and of identifying whether a string is a term or expresses a taxonomy.

The text reads: The leaf’s flattened form presents a large surface area for the absorption of light. It has mesophyll cells with chloroplasts containing chlorophyll to trap light energy. It has a cuticle to reduce water loss by transpiration.
The questions read:

6. Which of the following traps light energy? (Circle the right answer)

a. Mesophyll cells
b. Chloroplast
c. Chlorophyll

7. Mesophyll cells contain chloroplast. True or False? (Write out the correct answer in the space below)

Given a taxonomy such as ‘*The leaf has mesophyll cells with chloroplast, which has chlorophyll,*’ the challenge is whether it is clear which of these components performs the function of trapping light. There is also (with respect to question 7) the challenge of determining whether *mesophyll cells with chloroplast* is primarily expressing a term or a taxonomy (chloroplast is contained in mesophyll cells).

In summary, questions 1, 2 and 3 arise from the potential ambiguity in text; questions 4 and 5 from potential semantic discontinuity, and questions 6 and 7 from taxonomy.

Usability testing by Grade 10 learners in a resource-poor school

After obtaining approval from the provincial education department and a public high school (the high school is attended mainly by learners from the local community who speak Kaaps at home), the application was tested on grade10 learners of the school. Due to the lack of facilities (in this case, working computers), the learners could not have online access to the application. The visualization of knowledge modeled as concept graphs was then printed out on paper for the learners.

Experimental design

The learners were randomly selected and placed into two groups. One group was assigned to the experimental group while the other group was assigned to the control group. The control group (made up of 17 learners) was given only the English texts with the questions to answer while the experimental group (24 learners) was given texts (the same texts as the control group) and the visualization of the knowledge modeled, answering the same questions as the control group.

After the written test, an interview was conducted. This was done to ascertain from the learners what their experiences were during the test.

Chapter Four: Results and Discussion

Description of application developed

The concept graph application in which knowledge was modeled was created as a Jupyter notebook document, and can be viewed or shared through the "Jupyter nbviewer" at (http://nbviewer.jupyter.org/). Going online with a computer or a cell phone, a user navigates to the above web address; a web page like the one seen in figure 12 below opens up. Next, the user enters the relevant URL for the document in the position indicated by the arrow in figure 12.

FIGURE 12: JUPYTER NBVIEWER

Afterwards, the relevant document containing the visualized graphs opens on the website. The entire text is presented in a graph representation like figure 13 below generated from the application:

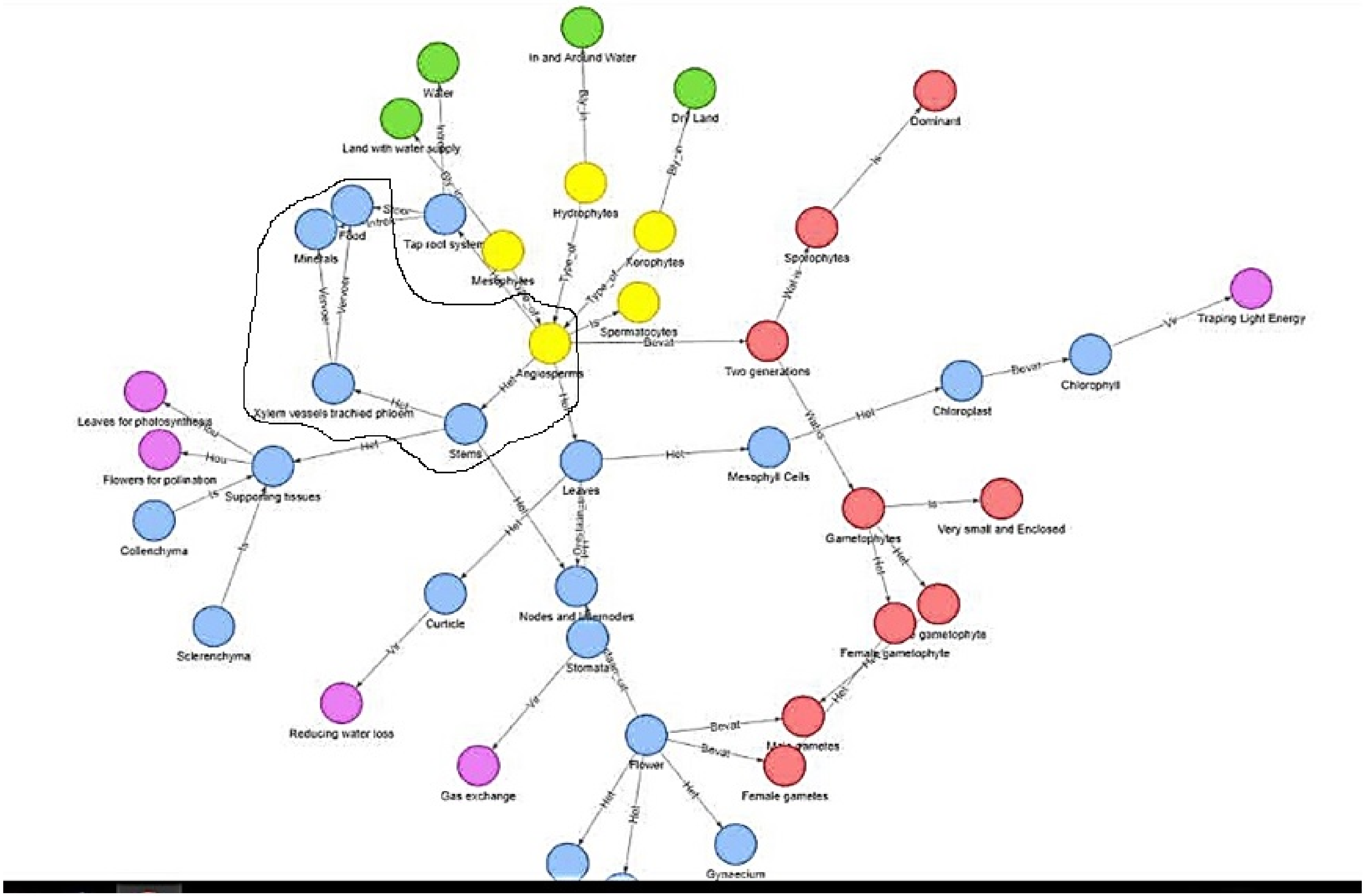


FIGURE 13: GRAPHICAL REPRESENTATION OF TEXT ON ANGIOSPERM

Although in a static representation outside of Jupyter the graph (in figure 13) is in parts difficult to read, because of the amount of information, in the Jupyter Notebook environment the user can extend the graphs in those areas that are cluttered to enhance legibility. For instance, the area that is demarcated (encircled) on the graph above can be seen clearly in the next figure (14) after being expanded.

Subgraphs are created to represent different paragraphs from the text on Angiosperms. Here is an example paragraph from the text in this study:

> **Text 1**: "A stem also transports water and mineral salts from the roots to the leaves and manufactured food from the leaves to all parts of the plants. It therefore has well developed xylem vessels, xylem trachieds and phloem."

As was seen in the analysis of the questions, the claim was that some learners would have difficulty when asked questions regarding the presence of xylem and phloem in the stem; and what is used for the transport of mineral from/to the stem (see table 1 on results on question 4). The graphical representation of the above text, shown below (figure 14), probably helps learners to understand that "xylem and phloem" are located in the stem and are used to transport food and minerals.

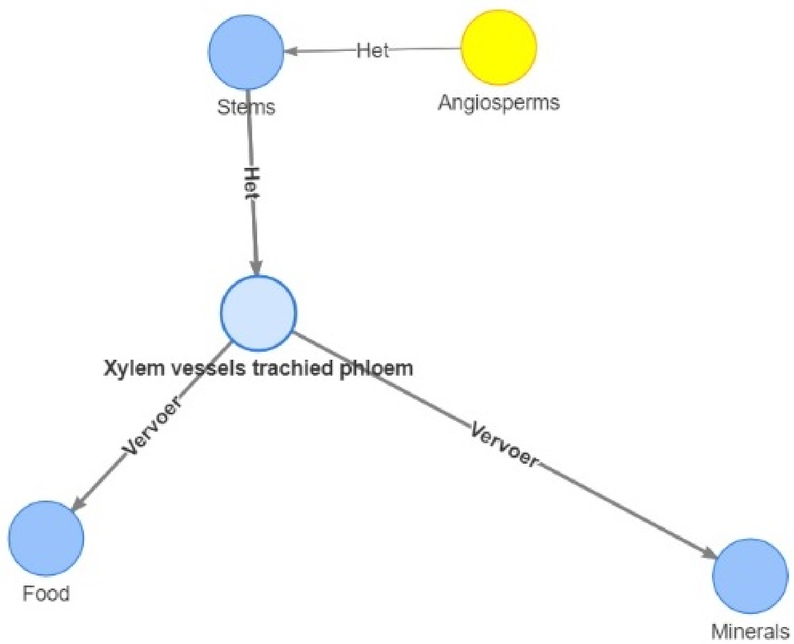


FIGURE 14: SUB-GRAPH FOR TEXT 1

Findings of experimental study

It may be recalled that, in the experimental study, the learners were randomly divided into two groups. One group was assigned to the experimental group while the other group was assigned to the control group. The control group was given only the English texts with the questions to answer while the experimental group was given the same texts and a visualization of the modeled knowledge.

The hypothesis of the study was as follows: “In the test to be conducted, there will be no difference in the performance scores of those learners who are exposed to the knowledge modelled in the application and scores of those learners exposed only to knowledge in text without the model.”

Below I present the statistical results obtained in this study, both descriptive and inferential statistical results.

The results of the raw scores and their corresponding percentages (table 1) show that the experimental group performed much better than the control group overall. This can be seen in table 1 below:

Questions	Control group (n=17) **correct answer**	Control group **wrong answer**	Experimental group (n=24) **correct answer**	Experimental group **wrong answer**
1	15 (88%)	2 (12%)	17 (71%)	7 (29%)
2	10 (59%)	7 (41%)	18 (75%)	6 (25%)
3	8(47%)	9(53%)	14 (58%)	10 (42%)
4	6 (35%)	11 (65%)	18 (75%)	6 (25%)
5	7 (41%)	10 (59%)	23 (96%)	1 (4%)
6	13 (76%)	4 (24%)	24 (100%)	0 (0%)
7	6 (35%)	11 (65%)	22 (92%)	2(28%)

TABLE 1: RAW SCORES OF CONTROL AND EXPERIMENTAL GROUPS

Both groups performed very poorly on question 3. Only on question 1 do we see the control group perform better than the experimental group especially when looking at percentages.

In order to answer the hypothesis, inferential statistical analysis was conducted by a statistician. To find out if there is a difference in the performance of the control group and the experimental group, the hypothesis testing is usually carried out using normal distribution when the population variance is known. When the population variance is unknown, a pool of sample variances are used and a t-distribution is used to carry out the hypothesis testing (37). In this study, the population variance is unknown, therefore the sample variances is used. To find the sample variances, the "F-Test: Two-Sample for Variances" is conducted.

The F-Test: Two-Sample for Variances is performed to determine if the sample variances (i.e. the measurement of the spread of the numbers in a dataset) of the two groups are equal or not (in this case, the control and experimental groups). The one-tail or two-tail tests to determine the statistical significance are performed in the process. If the value of F is less than the value of "F critical" (F<F critical), then automatically the value of P>0.05, thus the difference between the pool variance of the two groups is significant. However, if the value of F is more than or equal to the value of "F critical" (F≥F critical), then automatically P≤0.05. When the value of P is less or equal to 0.05, it shows that the difference between the variance of the two group is not significant. We then go ahead to conduct a t-test assuming equal variance on our data.

For the *T-Test: Two-Sample Assuming Equal Variances*, This test is performed to determine if the null hypothesis (assuming the means of two population groups are equal) can be accepted or rejected. If the value of t < 0, then P (T <= t) one-tail is the probability that a value of the t-statistic would be observed that is more negative than t. In the same way, if the value of t >0, then P (T<=t) one-tail is the probability that a value of the t-statistic would be observed that is more positive than t. The P (T <=t) two-tail is the probability that a value of the t-Statistic would be observed that is larger in absolute value than t. The details and results of the tests for both correct and wrong answers are given below:

Results for Correct Scores for both experimental and control groups

F-Test: Two-Sample for Variances

	Control group correct answer	*Experimental group Correct answer*
Mean	9.285714286	19.42857143
Variance	12.57142857	13.28571429
Observations	7	7
Df	6	6
F	0.946236559	
P(F<=f) one-tail	0.474115382	
F Critical one-tail	0.233434021	

TABLE 2: SHOWING RESULTS OF F-TEST FOR CORRECT SCORES

When we performed the F-test for variance between the control and experimental correct scores, we found that there was no significant difference between the two variances (see table 1). That means we can use “T-test two-sample assuming equal variance” between the control group and experiment group scores. The results of the test will determine if we reject or accept the null hypothesis. The results are presented below as well as the graph representation of the results:

t-Test: Two-Sample Assuming Equal Variances		
	Control group	*Experimental group*
Mean	9.285714286	19.42857143
Variance	12.57142857	13.28571429
Observations	7	7
Pooled Variance	12.92857143	
Hypothesized Mean Difference	0	
Df	12	
t Stat	-5.277388438	
P(T<=t) one-tail	9.7716E-05	
t Critical one-tail	1.782287556	
P(T<=t) two-tail	0.000195432	
t Critical two-tail	2.17881283	

TABLE 3: SHOWING RESULTS OF T-TEST FOR CORRECT SCORES

From the results of the t-Test (see table 3), the P (T<=t) two-tail is 0.00 thus less than 0.05. This shows that there is significant difference between the mean of control group correct scores and experimental group correct scores. Thus, we reject the null hypothesis. Figure 15 below shows the graph of the results described.

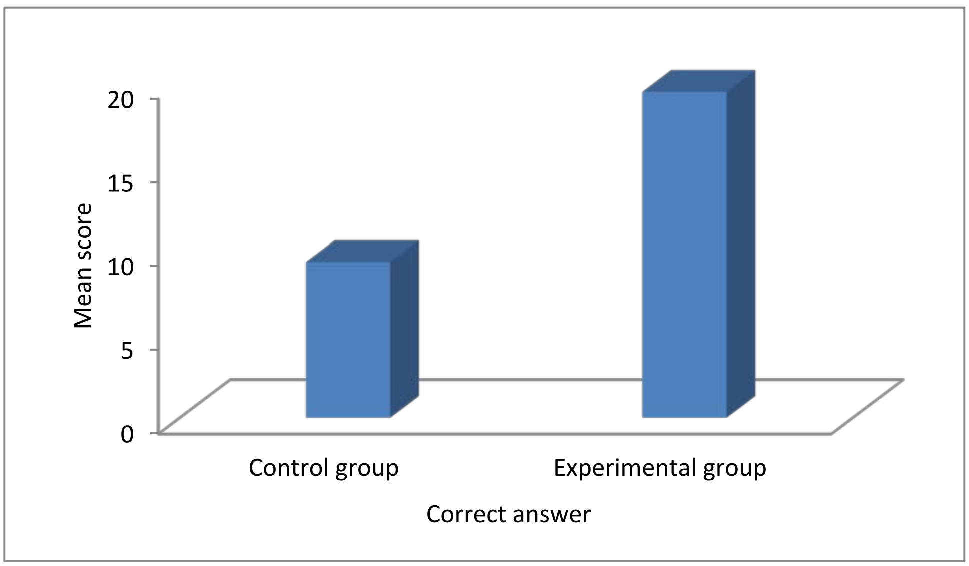


FIGURE 15: MEAN OF CORRECT SCORES FROM BOTH GROUPS

Table 4 below presents the mean of the correct score percentages of all questions for both the control and experimental groups.

Control group	*Experimental group*
54.43%	81%

TABLE 4: SHOWING THE MEAN OF THE PERCENTAGES OF CORRECT SCORES

From table 4, we see that the experimental group had a higher mean percentage score than the control group for the correct answers given.

Below in table 5, we present the F-test for variance between the control and experimental score percentages for the correct answers to determine if there is significant difference between the two variance.

F-Test Two-Sample for Variances

	Control group correct answer	*Experimental group Correct answer*
Mean	54.42857143	81
Variance	433.952381	234.6666667
Observations	7	7
Df	6	6
F	1.849228896	
P(F<=f) one-tail	0.236680756	
F Critical one-tail	4.283865714	

TABLE 5: SHOWING THE RESULTS OF F-TEST FOR PERCENTAGES OF CORRECT SCORES

From Table 5 we find that there is no significant difference between the two variances. Thus, we use "T-test two-sample assuming equal variance" between the control group and experiment group scores. The results of the t-test will determine if we reject or accept the null hypothesis. The results are presented below (table 6) as well as the graph (figure15) representation of the results:

t-Test: Two-Sample Assuming Equal Variances

	Control group correct answer	*Experimental group Correct answer*
Mean	54.42857143	81
Variance	433.952381	234.6666667
Observations	7	7
Pooled Variance	334.3095238	
Hypothesized Mean Difference	0	
Df	12	
t Stat	-2.718783039	
P(T<=t) one-tail	0.009324883	
t Critical one-tail	1.782287556	
P(T<=t) two-tail	0.018649766	
t Critical two-tail	2.17881283	

TABLE 6: SHOWING THE RESULTS OF T-TEST FOR PERCENTAGES OF CORRECT SCORES

From the results of the t-Test (see table 6), the P(T<=t) two-tail is 0.02 thus less than 0.05; this shows that there is significant difference between the mean of control group percentage score for correct answers and experimental group percentage score for correct answers. Thus we reject the null hypothesis. Figure 16 below shows the graph of the results described.

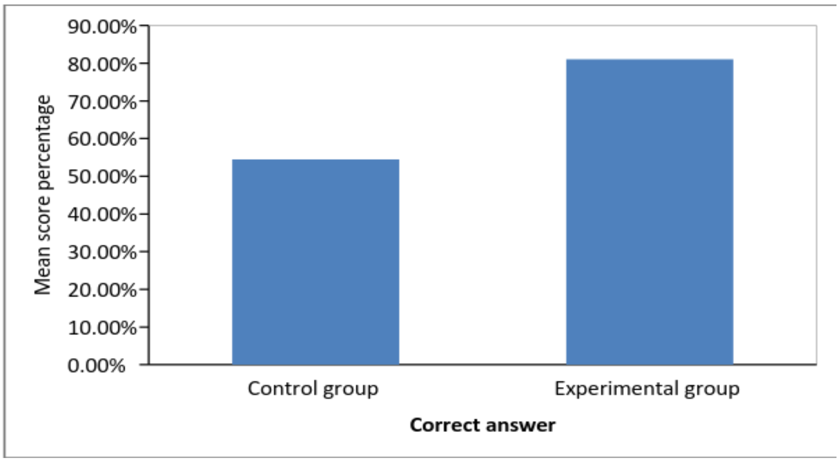


FIGURE 16: MEAN OF PERCENTAGE CORRECT SCORES FOR BOTH GROUPS

Let us turn to an analysis of wrong answer scores. Table 7 below presents the mean of the wrong scores for both the control and experimental groups.

Control group	*Experimental group*
6.285714286	3.571428571

TABLE 7: SHOWING THE MEAN OF WRONG SCORES

Table 7 shows that the experimental group had a much lower mean of wrong answers than the control group. On the next table (table 8), we present the results of the F-Test for the wrong answers for both control and experimental groups.

F-Test Two-Sample for Variances

	Control group wrong answer	*Experimental group Wrong answer*
Mean	6.285714286	3.571428571
Variance	15.9047619	14.95238095
Observations	7	7
Df	6	6
F	1.063694268	
P(F<=f) one-tail	0.471083185	
F Critical one-tail	4.283865714	

TABLE 8: SHOWING THE RESULTS OF F-TEST FOR WRONG SCORES

In table 8, when we performed the F-test for variance between the control and experimental wrong scores, we found that there was no significant difference between the two variances (see table 8).

With this result, we can use “T-test two-sample assuming equal variance” between the control group and experiment group scores. The results of the t-test will determine if we reject or accept the null hypothesis. The results of the t-test are presented below in table 9 as well as in figure 17.

t-Test: Two-Sample Assuming Equal Variances

	Control group wrong answer	*Experimental group Wrong answer*
Mean	6.285714286	3.571428571
Variance	15.9047619	14.95238095
Observations	7	7
Pooled Variance	15.42857143	
Hypothesized Mean Difference	0	
Df	12	
t Stat	1.292786253	
P(T<=t) one-tail	0.11021091	
t Critical one-tail	1.782287556	
P(T<=t) two-tail	0.22042182	
t Critical two-tail	2.17881283	

TABLE 9: SHOWING THE RESULTS OF T-TEST FOR WRONG SCORES

From the results of the t-Test (see table 9), the P(T<=t) two-tail is 0.2, thus greater than 0.05; this shows that there is no significant difference between the mean of control group wrong answers and experimental group wrong answers. Thus, we do not reject the null hypothesis. Figure 17 below shows the graph of the results described.

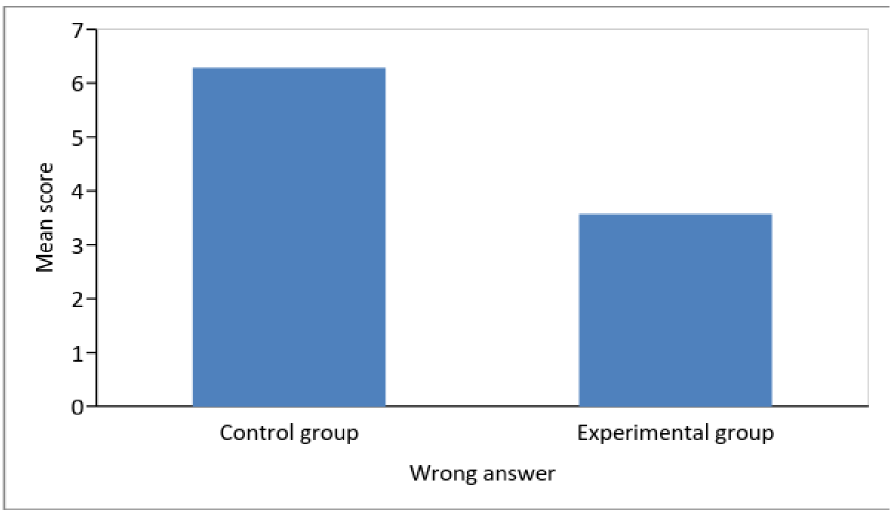


FIGURE 17: MEAN OF WRONG SCORES FROM BOTH GROUPS

Table 10 below presents the mean of the wrong scores percentages of all questions for both the control and experimental groups.

Control group	*Experimental group*
37.14%	17.71%

TABLE 10: SHOWING THE MEAN OF THE PERCENTAGES OF WRONG SCORES

In table 11 below, we present the result of the F-test for the percentages of wrong scores for both the control and experimental groups.

F-Test Two-Sample for Variances

	Control group wrong answer	*Experimental group Wrong answer*
Mean	37.14285714	17.71428571
Variance	549.1428571	274.9047619
Observations	7	7
Df	6	6
F	1.997574918	
P(F<=f) one-tail	0.210276218	
F Critical one-tail	4.283865714	

TABLE 11: SHOWING THE RESULTS OF F-TEST FOR PERCENTAGES OF WRONG SCORES

Table 10 presents the mean of the percentages of wrong scores for both the control and experimental groups. When we performed the F-test for variance between the control and experimental wrong scores, we found that there was no significant difference between the two variances (see table 11). With this result, we can use “T-test two-sample assuming equal variance” between the control group and experiment group percentage scores. The results of the test will determine if we reject or accept the null hypothesis. The results of the t-test are presented below in table 12 as well as in figure 18:

t-Test: Two-Sample Assuming Equal Variances

	Control group wrong answer	*Experimental group Wrong answer*
Mean	37.14285714	17.71428571
Variance	549.1428571	274.9047619
Observations	7	7
Pooled Variance	412.0238095	
Hypothesized Mean Difference	0	
Df	12	
t Stat	1.79066248	
P(T<=t) one-tail	0.049293469	
t Critical one-tail	1.782287556	
P(T<=t) two-tail	0.098586939	
t Critical two-tail	2.17881283	

TABLE 12: SHOWING THE RESULTS OF T-TEST FOR PERCENTAGES OF WRONG SCORES

From the results of the t-Test (see table 13), the P(T<=t) two-tail is 0.09 thus greater than 0.05; this shows that there is no significant difference between the mean of control group wrong answers and experimental group wrong answers. Thus we do not reject the null hypothesis.

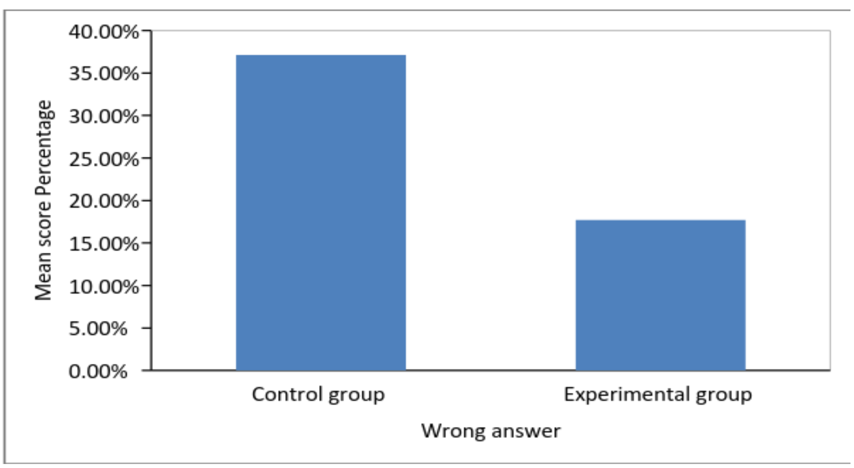


FIGURE 18: MEAN OF PERCENTAGES FOR WRONG SCORES FROM BOTH GROUPS

Discussion

The statistical analysis in this chapter aims to determine the performance of the learners in the two groups, the experimental group and the control group. The experimental group was exposed to an

intervention (i.e. the graphical representations of the texts) and the research aims to determine whether the experimental group performed better than the control group after this exposure.
The results of the study showed that the learners from the experimental group performed better than the control group. This clearly indicates that the exposure to graphical representation as well as the text may have enabled the learners to understand and process the knowledge in the text better than their counterparts who only had the text. With the above findings, we reject the hypothesis of the study.

In respect of question 1, some of the learners were not able to identify the number of angiosperms that were mentioned in the first text. Many only identified 2 while others identified 4 instead of 3. The students who identified 4 types of angiosperms were from the control group. They thought that there were four types (i.e. terrestrial, mesophytes, hydrophytes and xerophytes) given the way the text was written. The students who identified just two types of angiosperm were from both the control and experimental groups. It is unclear why this occurred.

With respect to question 2 on naming the angiosperms identified, some of the learners in the control group confirmed the view that the text was ambiguous when they included “terrestrial” as a type of angiosperm.

Though both groups performed poorly on question 3, the task seemed more difficult for the learners of the control group Most of the learners in the control group, although they knew that hydrophytes were water angiosperms, could not tell that mesophytes were land angiosperms but only needed water to survive. The graphical representation of the text made this distinction between them clear, and the learners of the experimental group were able to answer the question correctly.

With respect to question 4, the learners of the control group found it difficult to make the connection between the stem’s ability to transport food and minerals to other parts of the plant and the presence of xylem vessels, xylem trachieds and phoenix in the stem. This shows clearly the problem of semantic discontinuity.
With respect to question 5, although most learners knew that chlorophyll was used to trap light energy, there was more difficulty in realizing that chloroplast was housed in the mesophyll cells.

Analysis of the interview

The interview conducted with the learners showed that the learners found the graphical representation of the text very useful in answering the questions. As may be recalled, in chapter three, questions were analyzed for challenges they were expected to pose to the learners. Below we provide learners' accounts of their experiences while answering the questions.

Question 1: Our envisaged challenge was that the text was ambiguous and would create confusion as to how many types of angiosperms there really was in the text. It was interesting to see from the interview data that some of learners felt that the text was not clear and they could easily find the number of types an angiosperms had from the graph. Other learners said the graph helped them to confirm the number of types that they found already in the text.

Below, I quote some of the answers:

1) "the graph help me find the answer because I understand easily"
2) "the text and graph give all the answer and make easy to find "

Question 2: Our envisaged challenge was also that the text was ambiguous and would create confusion as to what the types of angiosperms were in the text. From the interview data, learners shared the same view of being confused as to what the types of angiosperms were from the text alone. Most learners were not able to correctly give the 3 types of angiosperms from the text alone. Where they were able to give 3, they named 2 correctly and the third one was wrong. Learners believed that the graph gave them the answer to the question clearly. They are quoted as follows:

1) "I found 3 types in the graph"
2) "3 in graph are mesophytes, hydrophytes and xerophytes"

Question 3: Our envisaged challenge was that the text was ambiguous and would create confusion as to knowing the habitat of the types of angiosperms from the text. Many of our learners from the interview data were sure that it was clear in the text where the habitat of different angiosperms were. However when asked further to differentiate the habitat of mesophytes and hydrophytes, they found it difficult to do and had to go to the graphs to distinguish between the two habitats.

Question 4 and 5: Our envisaged challenge here was on the text organization, which we referred to as semantic discontinuity. From our interview data, learners found that the graph provided the guide that the xylem vessels and phloem were located within the stem. Other learners felt the text was clear. Learners said the following and I quote:

1) “the graph tell me that xylem vessels and phloem are in the stem”
2) “the text is written plain and simple”

Question 6 and 7: Our envisaged challenge here was on the difficulty of technical taxonomy. From the interview data, learners felt that the graph was easier to read than the text. One learner said “the graph shows you were everything comes from clearly”. Some learners attested to knowing the answers from previously taught topics rather than from the text presented.

Though no specific language related questions were asked, it was very interesting to hear the learners point to the relations, which were in the local language, as something that made it very easy for them to understand the graphs.

Although designed differently and using different tools, our study has led to results that confirm findings of related research reported earlier. For instance, it would be recalled from the Brazilian Cmap study that students who did not have English as their home language found that using Cmap enhanced their experience of reading academic texts in English (20). While in that study the students had to actually develop the Cmaps, in our study the learners had to process previously modeled texts. It is interesting that the learners in our study still reported benefits even though using pre-constructed graphs meant that they had spent less time engaging with the knowledge representation models. As in the Brazilian study, it would seem that graphical knowledge models in our study also made it possible for features of academic/scientific English such as reviewed earlier to be less of a problem for learners.

It is acknowledged that the study was conducted using a small sample size, and that such a size limits the generalizability of the findings. In spite of this limitation, it is clear that the introduction of graphical knowledge representation seemed to enhance epistemological access to science knowledge for high school learners; this is consistent with other studies (20, 21). A study with a larger sample size would, however, need to be carried out in order for benefits reported in this study to be confirmed, and also to shed more light on aspects which this study has not been able to address (see directions for future research).

Experiences during my research

While carrying out this study, there were a few challenges I encountered. At the beginning, I explored the option of identifying a conceptual graph tool that could be customized for my work. While there were several tools available, none of the ones I found (e.g. on the conceptual graph community website) was appropriate for my work. For this reason, I had to spend a lot of time reading and learning about different tools and how to apply them. I learned about visualization libraries, client libraries, graph databases, Jupyter notebook application. I also had to read widely to find out how to integrate the different tools together. Finding an open source customizable tool would have been very helpful. I may have been able to have addressed some further issues (e.g. the dynamic/interactive nature of the application – see future work in the next chapter) but for the time spent acquiring the knowledge I needed to develop the application.

During the experimental study, the principal and teachers of the high school involved were supportive though slow in responding. The learners responded very quickly and positively. I encountered some difficulties when carrying out the study as there was no computer laboratory available for use and internet access at the school. The alternative was to print out the graphs on paper and distribute to the learners. Although the essence of the application (graphical display of the knowledge) may have been captured by the hardcopy, it would still have been useful to have been able to observe how learners interacted with the application. This may have yielded interesting user experience data.

Chapter Five: Conclusion and Future work

This dissertation was motivated by a real world problem, that is, the observation that learners, especially those for whom English is not a home language, have difficulty understanding texts in their science textbooks there by reducing their ability for epistemological access to knowledge. We reviewed research pointing to a link between the performance in the matriculation examination results of learners in the Western Cape that have English as home language and those that do not (3). Such stratified performance clearly supports the link that has been argued to exist between proficiency in language and understanding of the curriculum as well as performance in assessment tasks (1).

In the specific context of scientific texts, we reviewed work on the peculiarities of scientific writing in English as a basis for:

(a) demonstrating the kinds of challenges or difficulties that learners would have, and

(b) understanding how knowledge representation might be used to intervene to attempt to address the challenges.

Halliday and Martin (4) identified six such peculiarities of scientific English. The first, **interlocking definitions**, has to do with definitions of terms in science textbooks sometimes containing so many other concepts that are used to define each other and that first have to be understood before the main concept being defined can be understood. We gave an example with the following definition of a circle in which a number of other concepts (radius, curve, distance, diameter) have to be understood:

> "A circle is a plane curve with the special property that every point on it is at the same distance from a particular point called the centre. This distance is called the radius of the circle. The diameter of a circle is twice the radius. The length of the circle is called its circumference" (4)

The second challenge, related to the first, has to do with **technical taxonomies**, that is, the fact that scientific concepts occur in clusters within which they are organized according to specific criteria of classification, e.g. relationship of superordination (e.g. x is a y) or relationship of composition (e.g. x is part of y). Regrettably the relationships underlying these clusters or taxonomies may not always be evident to the learner.

A third challenge is **lexical density,** that is, the number of words in a sentence that are technical terms or parts of terms. In scientific English, many of these words may occur within a noun group as in the following example:

> "**The conical space rendering of cosmic strips gravitational properties** applies only to straight strings" (4)

A fourth difficulty was described as **syntactic ambiguity** as in the following sentence:

> "lung cancer death rates are clearly associated with increased smoking" (4)

The verb group 'are clearly associated with' can be ambiguous. The expert may have enough background knowledge to determine that it means 'caused by'. A novice could process it as meaning 'causes' or 'leads to' (people/smokers upset by all the talk about lung cancer deciding to smoke/more). Equally ambiguous is the noun group that functions as the subject of the sentence. 'Lung cancer death rates' could mean:

- how many people die of lung cancer
- how quickly people die of lung cancer
- how quickly people's lungs die of lung cancer

The second noun group 'increased smoking' could mean:

- more people smoking
- people smoking more
- more people smoking more

Novices may not always have enough background knowledge initially to resolve these ambiguities.

A fifth challenge was described as **grammatical metaphor**, that is, the shift or change from one grammatical class (e.g. verb) to another (e.g. noun). In ordinary language use, reality tends to be expressed as Participants (animals, people, things, abstract objects), as Processes (verbs), and as Circumstances (how, adverbs). However, scientific English tends to convert many of these categories into nouns. Such conversion leads to loss of certain kinds of information and therefore to the kinds of ambiguity seen above. Instead of the underlying idea of ‘how quickly people die from lung cancer’, which comprises Circumstance (how quickly), Participants (people, lung cancer), Process (die), Preposition (from), academic English says ‘lung cancer death rate’. In this noun group, the original adverb (how quickly) has become the noun ‘rate’; the original verb ‘die’ has become the noun ‘death’; the original idea of ‘people’ has been lost and the preposition ‘from’ has also been lost. Such grammatical metaphor represents a view of the world in which virtually everything has turned into a noun (4). Thinking in this way may not always be easy for novices.

A sixth challenge, **semantic discontinuity**, was said to arise when writers of scientific texts “sometimes make semantic leaps, across which the reader is expected to follow them in order to reach a required conclusion” (4). Consider the following example:

> “Stem also transports water and mineral salts from the roots to the leaves and manufactured food from the leaves to all parts of the plants. It therefore has well developed xylem vessels, xylem trachieds and phloem” (39)

The information in the above example does not clearly state that xylem vessels, xylem trachieds and phloem are part of the stem. It also does not state explicitly what functions they perform.

An overview of computer science research on knowledge representation suggested there was a basis for applying knowledge representation formalism to life sciences curricular content in order to create a knowledge base system suitable for addressing challenges such as described above. Knowledge Representation (KR), also known as knowledge representation and reasoning (KRR), was seen to involve the analysis of knowledge (related to a real world issue) in order for this knowledge to be processed (i.e. reasoned with) by the computer or the human user. A number of principles underpinning KR were reviewed, including, notably, the principle that all KR is a “medium of expression and communication in which we tell the machine (and perhaps one another) about the world” (11). The implication of this principle was seen to be that KR must not only be adequate in expressing the representation, its expression had to also be understood by the interpreter (machine or human).

From the review, it became obvious that KR was indeed a suitable computer science framework for responding to the challenges learners have regarding understanding scientific texts. KR was seen as a means of "making things explicit" and "resolving ambiguities" (12). Even though processing by the machine is essential to the very conceptualization of KR, the above principle allowed for a view of KRs as language in which humans communicate with other humans, and whose effectiveness needs to be determined.

Guided in particular by the above principle, and also the human users (grade 10 learners) of the application that was to be developed, I chose the network representation scheme from among other schemes reported in the literature such as the logical, procedural and structured representation schemes (7).

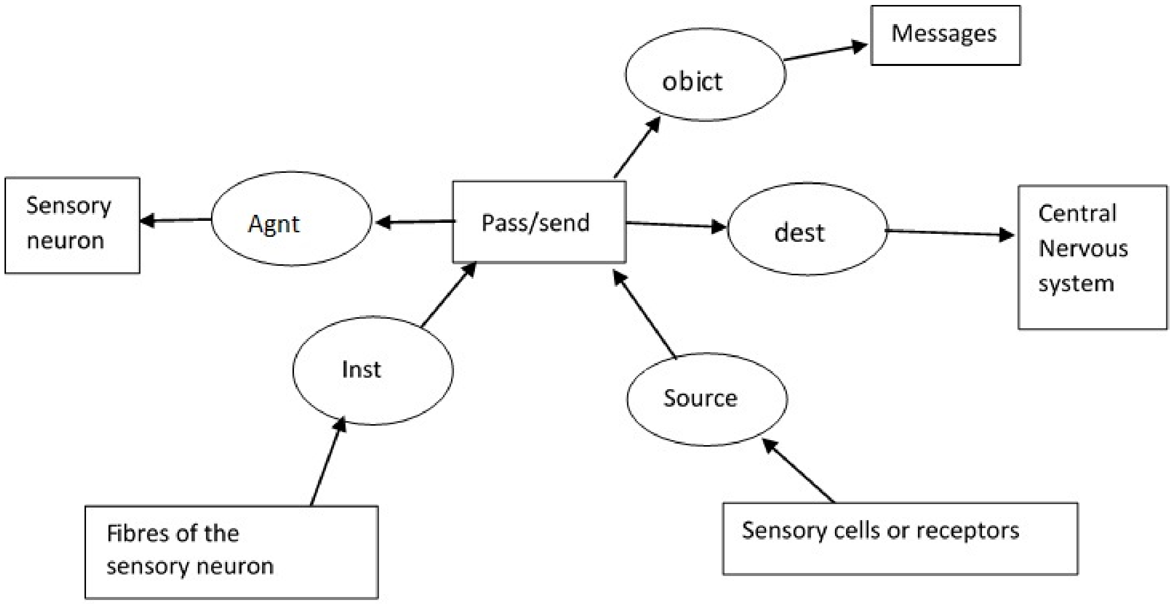
Within the network representation scheme which has such knowledge representation formalisms as semantic networks, Conceptual graphs, concept graphs (7), I chose Conceptual graphs as they are close to natural language expressions, allow for unambiguous expression, and make fewer processing demands on the human interpreter. It was seen that, in its display form, a simple canonical conceptual graph is bipartite, consisting of two concepts (represented by rectangles) and a relation in a circle/oval, as in the figure below:

The above graph reads as follows: "The relation of Concept_1 is Concept_2" or "The relation of Concept_2 is Concept_1".

In arguing for the relevance of this formalism to the real word problem underpinning the thesis, I pointed to how potential difficulties in a passage such as the one below could be addressed by the use of a conceptual graph representation.

> "Sensory neurons: along whose fibres messages are passed into the central nervous system from sensory cells or receptors"

It was suggested that learners could have difficulty answering a question like: Where are messages from sensory cells sent to? How are messages sent to the central nervous system? In contrast, the following conceptual graph in principle makes answering these questions easier.



The representation allows for understanding that:

- The agent of pass/send is sensory neurons
- The object of pass/send is messages
- The destination of (what is) passed/sent is central nervous system
- The source for (what is) passed/sent is sensory cells or receptors
- The instrument for passing/sending is fibres of the sensory neurons.

With this representation, answering the kinds of questions posed earlier (Where are messages from sensory cells sent to? How are messages sent to the central nervous system?) becomes easier.

Against this background, textbook passages on a topic in life sciences were identified, modeled using conceptual graphs and validated by a school teacher. To reflect the fact of learners typically drawing on English and their home languages in their attempt to understand curricular content, the relations were also translated into Kaaps, the variety of Afrikaans spoken in the local communities. The data was then used to populate an application with a graphical database which we had to develop. The application has a user interface with a visualization component integrated into it. For the graphical database, the Neo4j graphical database management system was used. The Neo4j database management system, like other relational databases, is manipulated by queries. It, however, uses the Cypher query language (CQL) instead of the SQL. The Jupyter notebook application was used in creating the user interface with Python. A connection was then created between Jupyter notebook and the Neo4j database through the use of "Py2neo client library", after which the entities (nodes) and relations were programmed using Python language in Jupyter. The entities and relations that were previously modeled were used to populate the graphical database. A graphical representation of a section of the text is obtained through the use of CQL queries which extracts the required set of nodes and relations and presents them as a sub-graph.

To test the impact of conceptual graph as a supplementary means of communicating curricular content to enhance understanding, an experiment was designed. Two groups of grade 10 learners were assigned to a control group (that would read text passages only, then answer seven questions) and an experimental group (exposed to both the text and the graphically modeled knowledge related to the text passages, and answering the same seven questions). Thereafter interviews were conducted to elicit the subjective experiences of the learners regarding the graphically modeled knowledge content. The hypothesis guiding the experiment was as follows: In the test to be conducted, there will be no difference in the performance scores of those learners who are exposed to the knowledge modeled in the application and scores of those learners exposed only to knowledge in text without the model.

On six of the seven questions, the experimental group performed better than the control group using descriptive statistics; this picture of performance was further confirmed using inferential statistics where we performed the F-test for variance and the T-test two-sample assuming equal variance on the control group and experiment group scores. The results were statistically significant and also showed that the experimental group performed better than the control group. Thus, the null hypothesis was rejected.

Data from the interview frequently referred to the graphically modeled knowledge making it easier to obtain information that was difficult to get from the text.

Although findings from larger studies (involving many more participants than the 41 learners used) are clearly required, the current study indicates that the implementation of graph-based knowledge systems is a promising means for intervening to enhance the understanding of English-language science textbooks by learners who may not be proficient in English.

Besides a larger pool of participants, future work should make it possible for the learners in the experimental group to access the graphical knowledge database on computers or on cell phones (both possible, but not done on account of limitations at this study's test site). Also, in future studies, more curricular topics should ideally be included rather than the one topic in the current study. Such studies may also consider the option of making the application dynamic and interactive. Integrating sound would also be worth considering.

References

1) Webb, V. (2004) African languages as media of instruction in South Africa: Stating the case. Language Problems & Language Planning, Vol. 28, Number 2, 2004, pp. 147-173.

2) Du Plooy, L and Zilindile, M (2014). Problematizing the concept epistemological access with regard to foundation phase education towards quality schooling. South African Journal of Childhood Education, Vol 4(1), pages 187-201

3) October, M. Medium of Instruction and its Effect on Matriculation Examination Results for 2000, in Western Cape Secondary Schools: A study of examination results in relation to home language and language medium. University of Cape Town .July 15, 2002.

4) Halliday, M. A. K and Martin, J.R. (1993) Writing Science: Literacy and Discursive Power. The Farm Press, London. ISBN: 0-75070-046-7 (chapter 4) pp. 69-86.

5) Chen, S.H and Jakeman, A.J. (2008) Norton, J.P Artificial Intelligence techniques: An introduction to their use for modeling environmental systems. Mathematics and Computers in Simulation Volume 78 pp. 379–400. Available: https://doi.org/10.1016/j.matcom.2008.01.028

6) Niwa, K; Sasaki, K and Ihara, H (1984). An Experimental Comparison of Knowledge Representation Schemes. Association for the Advancement of Artificial Intelligence (AAAI), Vol 5 Number 2. Available online: https://doi.org/10.1609/aimag.v5i2.435

7) Colton, S. (2005) Knowledge Representation. Available online:http://www.doc.ic.ac.uk/~sgc/teaching/pre2012/v231/leture4.html

8) Harmelen, FV; Lifschitz, V and Porter, B. (2008) Handbook of Knowledge Representation: Conceptual Graphs (chapter5), Elsevier, pp. 213-237. Available online: http://dai.fmph.uniba.sk/~sefranek/kri/handbook/handbook_of_kr.pdf

9) Sowa, J F. (1992) Conceptual Graphs as a Universal Knowledge Representation. Computers Math. Applic. Vol. 23, No. 2-5, pp. 75-93, Available: https://doi.org/10.1016/0898-1221(92)90137-7

10) Sowa, J F. (2000) Knowledge Representation: Logical, Philosophical and Computational Foundations. Brooks-Cole

11) Sowa, JF. (2000) Knowledge Representation: Logical, Philosophical and

Computational Foundations. Brooks-Cole. Reviewed by Stuart. C Shapiro. Available: https://aclweb.org/anthology/J/J01/J01-2006.pdf

12) Ahmad, K. (1995). Terminology in Advanced Microcomputer Applications (TAMA): Proceedings of the 3rd TermNet Symposium. International network for terminology Vienna, pp. 7-30

13) Robinson, W.R (1999). A view from the Science Education Research Literature: Concept Map assessment of classroom learning. Journal of Chemical Education. Today Volume 76, No.9. JChemEd.chem.wisc.edu

14) Novak, J. D. J. (1990) Res. Sci. Teach, Volume 27, pages 937–949

15) Jegede, O.J; Alaiyemola, F.F and Okebukola, P.A (1990). The effect of concept mapping on students' anxiety and achievement in biology. Journal of Research in Science Teaching (JRST), Volume 27, Issue 10**,** Pages 951-960 https://doi-org.ezproxy.uct.ac.za/10.1002/tea.3660271004

16) Nesbit, J.C and Adesope, O.O (2006) Learning with Concept and Knowledge Maps: A Meta-Analysis. Review of Educational Research (RER) Volume: 76 issue 3, page(s): 413-448. https://doi-org.ezproxy.uct.ac.za/10.3102/00346543076003413

17) Okebukola, P.A (1990). Attaining meaningful learning of concepts in genetics and ecology: An examination of the potency of the concept-mapping technique. Journal of Research in Science Teaching (JRST), Volume 27, Issue 5 Pages 493-504 https://doi-org.ezproxy.uct.ac.za/10.1002/tea.3660270508

18) Okebukola, P.A and Jegede, O.J (2006). Students' Anxiety towards and Perception of Difficulty of some Biological Concepts under the Concept-mapping Heuristic Research in Science & Technological Education, Volume 7, 1989 - Issue 1 Pages 85-92. https://doi-org.ezproxy.uct.ac.za/10.1080/0263514890070109

19) Chei-Chang Chiou (2008). The effect of concept mapping on students' learning achievements and interests. Innovations in Education and Teaching International Volume. 45, No. 4, pages 375–387, ISSN 1470-3297, ISSN 1470-3300 Doi: 10.1080/14703290802377240, http://www.informaworld.com

20) Sánchez, J; Cañas, A.J and Novak, J.D Eds (2010). Concept Map: A Strategy for enhancing reading comprehension in English as L2. Fourth Int. conference on Concept Mapping, Viña del Mar, Chile

21) Aleixandre, M.PJ; Brody', M.J; Tamir, P and Tomkiewicz, W (1996). The Nature of Knowledge in Biology and Its Implications for Teaching and Learning: In NATO ASI Series Advanced Science Institutes Series, 148.

22) Lucas, P J.F and van der Gaag, L C. (1991) Principles of Expert Systems

Centre for Mathematics and Computer Science, Amsterdam, Addison-Wesley.

23) Sowa, J.F. (1984) Conceptual Structures: Information Processing in Mind and Machine. London, UK, Addison-Wesley

24) Priss, U; Polovina, S and Hill, R (Eds.). (2007) An Introduction to Conceptual Graphs: ICCS 2007, LNAI 4604, pp. 1–14; Springer-Verlag Berlin Heidelberg

25) Conceptual Graphs: Online Course in Knowledge Representation using Conceptual Graphs. Available: http://cg.huminf.aau.dk/

26) Conceptual Graphs: Tools. Available: www.conceptualgraphs.org

27) Baget, JF and Mugnier, ML. (2002) Extensions of Simple Conceptual Graphs: the Complexity of Rules and Constraints. Journal of Artificial Intelligence Research. Vol. 16 pp. 425-465. Available: http://www.jair.org/media/918/live-918-2019-jair.pdf

28) Chu, S and Cesnik, B. (2008) Knowledge representation and retrieval using conceptual graphs and free text document self-organization techniques. International Journal of Medical Informatics. Volume 62, pp. 121–133. Available: https://doi.org/10.1016/S1386- 5056(01)00156-3

29) Bemauer, J. (1992) Conceptual Graphs as an Operational Model for Descriptive Findings. Institute for Medical Informatics, University of Hildesheim, D-3200 Hildesheim, Germany. Available; https://www-ncbi-nlm-nih-gov/pmc/articles/PMC2247526

30) Karalopoulos, A; Kokla, M and Kavouras**,** M. (2004) Geographic Knowledge Representation Using Conceptual Graphs: In: “7th AGILE Conference on Geographic Information Science” 29 April-1May 2004, Heraklion, Greece. Parallel Session 6.1- “Semantics II”

31) Aubert, JP; Baget, JF; Chein, M. (2006) Simple Conceptual Graphs and Simple Concept Graphs. Springer Berlin Heidelberg. ICCS: International Conference on Conceptual Structures, Aalborg, Denmark. Springer-Verlag, 14th International Conference on Conceptual Structures, LNCS (4068), pp.87-101

32) Prediger, S. Simple Concept Graphs: A Logic Approach
In: M-L. Mugnier, M. Chein(eds). (1998) Conceptual Structures: Theory, Tools and Application. Lecture Notes in Artificial Intelligence 1453, Springer Verlag, Berlin Heidelberg, 225-239

33) Sowa, J F. Conceptual Graphs. In: Handbook of Knowledge Representation, ed. by F. van Harmelen, V. Lifschitz, and B. Porter, Elsevier, 2008, pp. 213−237. Available online: http://www.jfsowa.com/cg/cg_hbook.pdf

34) Rajangam, E and Annamalai, C. (2016) Graph Models for Knowledge Representation and Reasoning for Contemporary and Emerging Needs – A Survey I.J. Information Technology and Computer Science, Vol.2, 14-22 Available: http://www.mecs-press.org/ijitcs/ijitcs-v8- n2/IJITCS-V8-N2-2.pdf; DOI/ 10.5815/ijitcs.2016.02.02

35) Agrawal,R; Gollapudi, S; Kannan, A and K. Kenthapadi. (2014) Similarity Search using Concept Graphs. Presented at the Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, New York, NY.

36) Sowa, J F. Semantic Networks. (1992) In: Encyclopedia of Artificial Intelligence, Wiley, 1987; second edition,. Available: http://www.jfsowa.com/pubs/semnet.pdf

37) Dieng, R. (1997) Comparison of Conceptual Graphs for Modeling Knowledge of Multiple Experts: Application to Traffic Accident Analysis. RR-3161, 1997. Available: https://hal.inria.fr/inria-00073528/document

38) Artificial Intelligence Articles: Semantic Nets.(2010) Available online: http://intelligence.worldofcomputing.net/knowledge-representation/semantic- nets.html#.WXfcNYGxXqB

39) Isaac, T; Chetty, S; Manganye, H T; Mpondwana, N L and White, L. (2012) Understanding Life Science (Grade 11). Third Edition, South Africa, Pulse Education Services. ISBN: 978-1-920192-26-6

40) Kress, G. Reading Images: Multimodality, Representation and New Media. Available: http://www.knowledgepresentation.org/BuildingTheFuture/Kress2/Kress2.html

41) Baker, C. (2011)Foundations of bilingual education and bilingualism (5th ed.). Clevedon, UK: Multilingual Matters. p. 288,.

42) Lewis, G; Jones, B and Baker, C. (2012) Translanguaging: developing its conceptualization and contextualization. Educational Research and Evaluation Vol. 18, No. 7, 655–670. Available: http://dx.doi/full/10.1080/13803611.2012.718490

43) Lewis, G; Jones, B and Baker, C. (2012) Translanguaging: origins and development from school to street and beyond, Educational Research and Evaluation: An International Journal on Theory and Practice, 18:7, 641-654. Available: http://dx.doi/full /10.1080/13803611.2012.718488

44) Jupyter Notebook. Available: www.jupyter.org

45) Cornell, W. E. (1991) Knowledge Representation and metaphor. Netherlands, Springer Science & Business Media.

46) Zaiontz, C. http://www.real-statistics.com/students-t-distribution/two-sample-t-test-equal-variances/

47) Peinl, R and Holzschuher, F. (2013) Performance of graph query languages: comparison of cypher, gremlin and native access in Neo4j. Published in: Proceeding EDBT '13 Proceedings of the Joint EDBT/ICDT 2013 Workshops Pages 195-204, Genoa, Italy ISBN: 978-1-4503-1599-9 Doi>10.1145/2457317.2457351

Appendixes

Appendix A: Ethical Clearance

UNIVERSITY OF CAPE TOWN

IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

University of Cape Town
RONDEBOSCH 7701 South Africa
E-mail: timm.hoffman@uct.ac.za
Telephone: + 27 21 650 5551

4 May 2017

Mary-Jane Antia

Department of Computer Science

"(Multilingual) Knowledge representation for enhanced epistemological access

Dear Mary-Jane Antia

I am pleased to inform you that the Faculty of Science Research Ethics Committee has approved the above- named application for research ethics clearance, subject to the conditions listed below.

- Implement the measures described in your application to ensure that the process of your research is ethically sound; and
- Uphold ethical principles throughout all stages of the research, responding appropriately to unanticipated issues: please contact me if you need advice on ethical issues that arise.

Your approval code is: **FSREC 24** – 2017

I wish you success in your research.

Yours sincerely

Signature Removed

Prof Timm Hoffman

Chair: Faculty of Science Research Ethics Committee

Cc: Prof Tommie Meyer (Supervisor)

Appendix B: Approval Letter from WCED

Western Cape Government
Education

Directorate: Research

Audrey.wyngaard@westerncape.gov.za
tel: +27 021 467 9272
Fax: 0865902282
Private Bag x9114, Cape Town, 8000
wced.wcape.gov.za

REFERENCE: 20170329 –9518
ENQUIRIES: Dr A T Wyngaard

Mrs Mary-Jane Antia
Department of Computer Science, UCT
Private Bag X3
Rondebosch
7701

Dear Mrs Mary-Jane Antia

RESEARCH PROPOSAL: (MULTILINGUAL) KNOWLEDGE REPRESENTATION FOR ENHANCED EPISTEMOLOGICAL ACCESS

Your application to conduct the above-mentioned research in schools in the Western Cape has been approved subject to the following conditions:

1. Principals, educators and learners are under no obligation to assist you in your investigation.
2. Principals, educators, learners and schools should not be identifiable in any way from the results of the investigation.
3. You make all the arrangements concerning your investigation.
4. Educators' programmes are not to be interrupted.
5. The Study is to be conducted from **19 April 2017 till 29 September 2017**
6. No research can be conducted during the fourth term as schools are preparing and finalizing syllabi for examinations (October to December).
7. Should you wish to extend the period of your survey, please contact Dr A.T Wyngaard at the contact numbers above quoting the reference number?
8. A photocopy of this letter is submitted to the principal where the intended research is to be conducted.
9. Your research will be limited to the list of schools as forwarded to the Western Cape Education Department.
10. A brief summary of the content, findings and recommendations is provided to the Director: Research Services.
11. The Department receives a copy of the completed report/dissertation/thesis addressed to:

    The Director: Research Services
    Western Cape Education Department
    Private Bag X9114
    CAPE TOWN
    8000

We wish you success in your research.

Kind regards.
Signed: Dr Audrey T Wyngaard
Directorate: Research
DATE: 13 July 2017

Lower Parliament Street, Cape Town, 8001
tel: +27 21 467 9272 fax: 0865902282
Safe Schools: 0800 45 46 47

Private Bag X9114, Cape Town, 8000
Employment and salary enquiries: 0861 92 33 22
www.westerncape.gov.za

Appendix C: Consent Form

DEPARTMENT OF COMPUTER SCIENCE

UNIVERSITY OF CAPE TOWN
PRIVATE BAG X3
RONDEBOSCH 7701
SOUTH AFRICA

RESEARCHER: Mary-Jane Antia
TELEPHONE: +27-78-2251464
E-MAIL: ANTMAR012@myuct.ac.za
URL: cs.uct.ac.za

Informed Voluntary Consent to Participate in Research Study

Project Title: (Multilingual) Knowledge representation for enhanced epistemological access

Dear Parents/Guardians and Learners: I write to request your kind consent for the study I am conducting, which is described below. Parents/guardians to please give their consent for their children/wards to participate.

Invitation to participate, and benefits: You are invited to participate in a research study involving high school learners. The study's aim is to develop a computer program which displays knowledge in high school textbooks in the form of graph, so as to minimize the negative effects which the use of academic English in textbooks sometimes has on learners who do not have English as home language. As learners, I believe that your experience would be a valuable source of information for my research, which will also be of benefit to you. At the end of the research, you and your teachers will have access to the knowledge graphs created in the program and you will also have learnt how to focus on some key elements of knowledge in your textbooks. I urge you to kindly give me your break time to carry out this study.

Procedures: In the first part of the research, you will be asked to do one of two things: either view a web page and study the graphical content or read a short text from your textbook; then you will be asked to answer some questions based on the content. I will mark the answers and your teacher will not have access to the results. In the second part of the research on the following day, you may be asked to participate in a 5 minute-interview.

Risks: There are no potentially harmful risks related to your participation in this study.

Disclaimer/Withdrawal: Your participation is completely voluntary; you may refuse to participate, and you may withdraw at any time without having to state a reason and without any prejudice or penalty against you. Should you choose to withdraw, the researcher commits not to use any of the information you have provided without your signed consent. Note that the researcher may also withdraw you from the study at any time.

Confidentiality: All information collected in this study will be kept private in that you will not be identified by name or by affiliation to an institution. Confidentiality and anonymity will be maintained as pseudonyms will be used.

What signing this form means: Parents/Guardians and Learners: By signing your section of this consent form, you agree to your child/ward participating in this study and the learner also agrees to participate. The aim, procedures to be used, as well as the potential risks and benefits of your participation have been explained verbally to you in detail, using this form. Refusal to participate in or withdrawal from this study at any time will have no effect on you in any way. You are free to contact me, to ask questions or request further information, at any time during this research.

Learners:

I agree to participate in this research (tick one box) ☐ Yes ☐ No _________(Initials)

Name of Participant_________________ (Signature) ___________________ Date_________

Parents/Guardians:

I give my consent (Name of Parent)_________________(Signature) ______________________

(Researcher) Mary-Jane Antia_ (Signature) _______________ Date_______

Appendix D-1: Texts and questions

ANGIOSPERM

TEXT 1

Angiosperm plants are the most varied and successful of all plants. Although most angiosperms are terrestrial, living in areas with adequate water supply (mesophytes), some angiosperms are hydrophytes, living in and around water, with still others being xerophytes, living in very dry areas.

QUESTIONS ON TEXT 1:

1) How many types of angiosperms are there in the text? ____

2) Mention the types of angiosperms found in the text. ____

   ____

3) Based on the text, explain to your friend how he or she can distinguish between the types of angiosperms that require water.

   ____
   ____
   ____

TEXT 2

Stem also transports water and mineral salts from the roots to the leaves and manufactured food from the leaves to all parts of the plants. It therefore has well developed xylem vessels, xylem trachieds and phloem.

QUESTIONS ON TEXT 2:

4) What makes it possible for the stem to transport minerals salts, water and manufactured food to different parts of the plant?

   ____
   ____

5) The xylem vessels, xylem trachieds and phloem can be found in the stem. True or False? (Write out the correct answer in the space below)

   ____

TEXT 3

The leaf's flattened form presents a large surface area for the absorption of light. It has mesophyll cells with chloroplasts containing chlorophyll to trap light energy. It has cuticle to reduce water loss by transpiration.

QUESTIONS ON TEXT 3:

6) Which of the following traps light energy? (circle the right answer)
   a. Mesophyll cells
   b. Chloroplast
   c. Chlorophyll

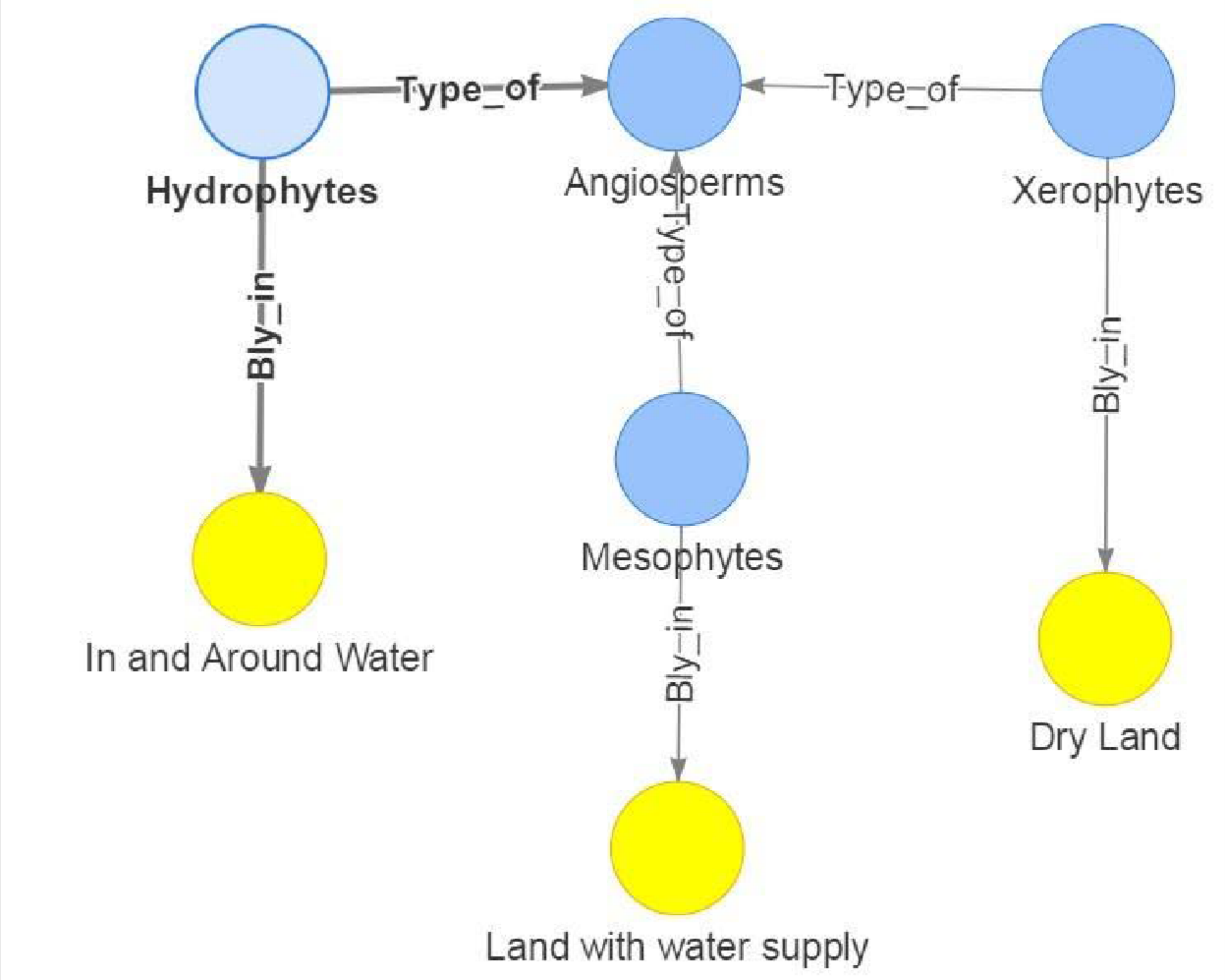
7) Mesophyll cells contain chloroplast. True or False? (Write out the correct answer in the space below)

   ____

Appendix D-2: Texts and questions

ANGIOSPERMS

TEXT 1 AND GRAPH 1



Angiosperm plants are the most varied and successful of all plants. Although most angiosperms are terrestrial, living in areas with adequate water supply (mesophytes), some angiosperms are hydrophytes, living in and around water, with still others being xerophytes, living in very dry areas.

QUESTIONS ON TEXT 1:

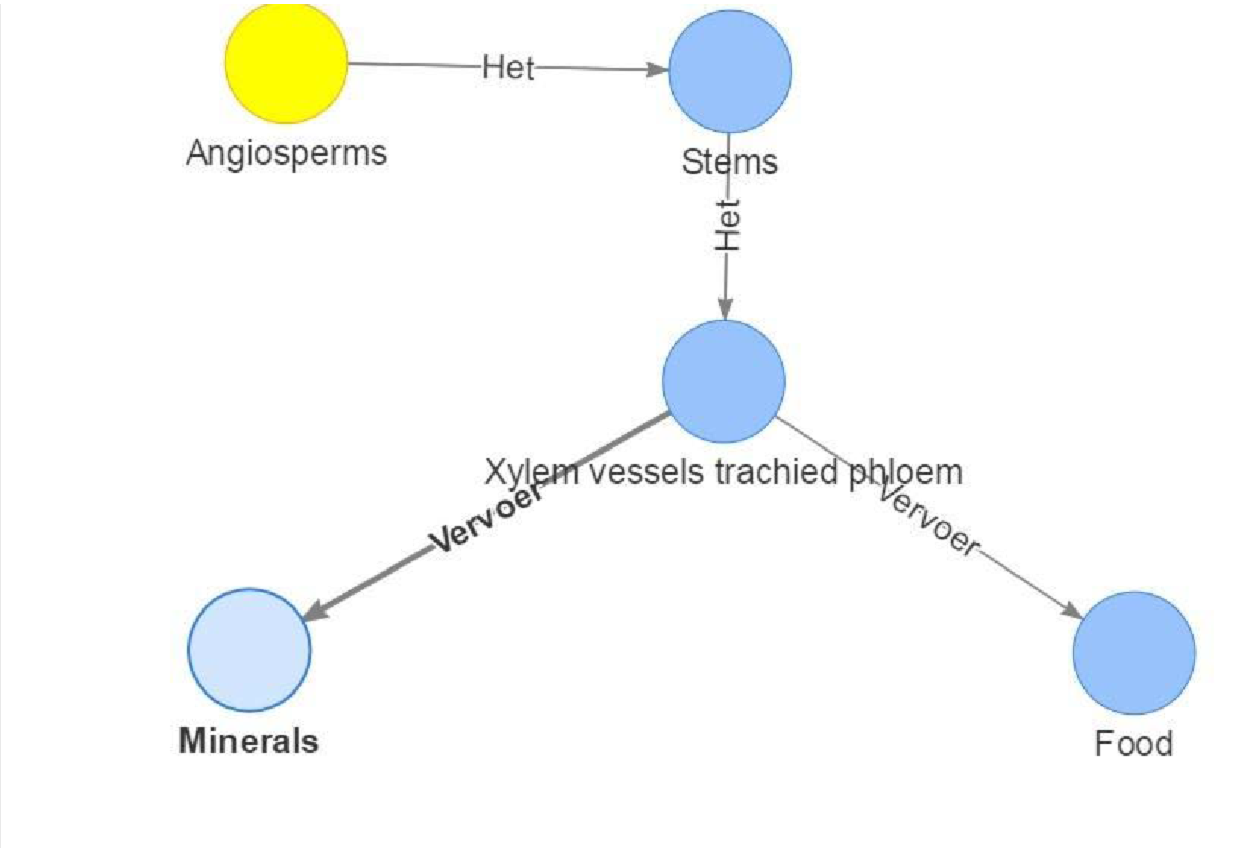
1) How many types of angiosperms are there in the text?
____

2) Mention the types of angiosperms found in the text.
____

3) Based on the text explain to your friend how he or she can distinguish between the types of angiosperms that require water.
____
____
____

Appendix D-3: Texts and questions (experimental group)

TEXT 2 AND GRAPH 2



Stem also transports water and mineral salts from the roots to the leaves and manufactured food from the leaves to all parts of the plants. It therefore has well developed xylem vessels, xylem trachieds and phloem.

QUESTIONS ON TEXT 2:

4) What makes it possible for the stem to transport minerals salts, water and manufactured food to different parts of the plant?

____

____

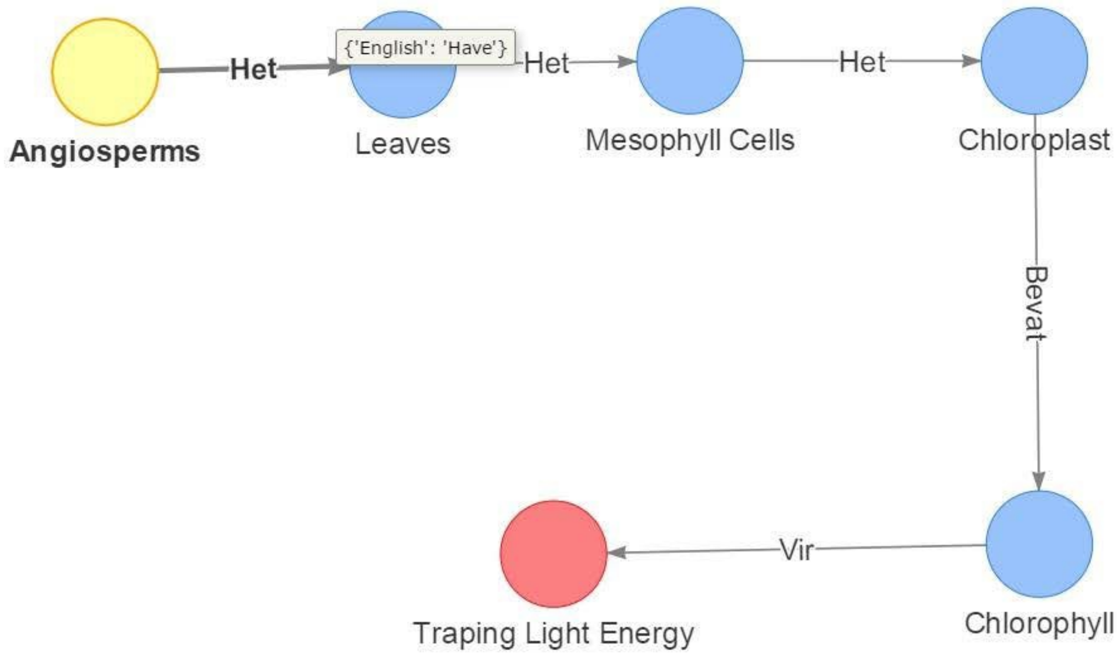
____

5) The xylem vessels, xylem trachieds and phloem can be found in the stem. True or False?

____

Appendix D-4: Texts and questions (experimental group)

TEXT 3 AND GRAPH 3



The leaf's flattened form presents a large surface area for the absorption of light. It has mesophyll cells with chloroplasts containing chlorophyll to trap light energy. It has cuticle to reduce water loss by transpiration.

QUESTIONS ON TEXT 3:

6) Which of the following traps light energy? (circle the right answer)
   a. Mesophyll cells
   b. Chloroplast
   c. Chlorophyll

7) Mesophyll cells contain chloroplast. True or false? (Write out the correct answer in the space below)

   ____