CHAPTER THREE

RESEARCH METHODOLOGY AND FRAMEWORK OF THE ANALYSIS

3.0 Introduction

Chapter three gives an outline of the corpus and the methodology of the research which uses a Hallidayan systemic-semantic model of language (described in the preceding chapter) to look at the distribution of lexical items in the context of two fundamental topics in chemistry. The lexical items were analysed for their cohesive force in the text. The Hallidayan systemic-semantic model was described in chapter two of this dissertation. This chapter also outlines how Martin's (1981,1989 & 1992) theoretical framework which is complementary to Hallidayan linguistics was used to label the semantic relationships between lexical items in the text. Martin's theoretical framework will be given a more detailed treatment in the following pages. This chapter also demonstrates how a text may be explicated for its cohesive relations and how

contrastive features may be captured in system diagrams to conceptualise some of the observations and theories in chemistry.

3.1 Selection of data

I was particularly interested to do a linguistic analysis in chemistry texts as this would give me an opportunity to apply linguistic principles to explicate the language of chemistry. I had the confidence and also the knowledge to carry out an analysis of this nature as I had been initially trained in the field of chemistry at the bachelor's degree level and as such have been sufficiently exposed to the literature in this field during my undergraduate years. Towards this direction I used chemistry textbooks that represent the genre of introductory and general textbooks typical of the kind first and second year students pursuing a course in chemistry would use as a source of reference. I also chose to work with these books as they had the qualities of being general or introductory.

3.2 Collection of data

After consultation with some academicians in the field of chemistry, I looked through all the holdings on general or introductory chemistry textbooks stored in the computer in the Universiti of Malaya library. I noted down about 40 call numbers and proceeded to locate

books on the shelves. Altogether I came across eleven texts which were revised editions published between 1988 and 1993 and noted that most of the writers had affiliations with the United States. The books had a number of things in common; they covered more or less the same number of topics, the topics were arranged in more or less the same order and the length of the chapters on the topics were more or less the same. For instance, the chapter on "gases" covered the Gas Laws, The Ideal Gas Equation, The Kinetic Molecular Theory, Dalton's Law of Partial Pressures, Graham's Laws of Diffusion and Effusion and Deviations from Ideal Behaviour. Similarly, the chapter on "Atoms and The Atomic Theory" covered the Atomic Theory, The Structure of the Atom, X-Rays and Radioactivity, Mass Relationships of Atoms, Molecules and Chemical Formulas, Ions and Ionic compounds, Percent Composition by Mass of Compound and Laws of Chemical Combination.

Since the eventual aim was to try and conceptualise some observations and theories on a scale of delicacy (a scale to make finer the options available), only two chapters but not more were chosen for the analysis. The focus of my research was to describe the language of chemistry in terms of the selection of vocabulary. It was a question of detailed indepth content analysis of chemistry texts. The two chosen chapters had far more similarities in coverage than

differences and whatever differences centred around additional information surrounding the important chemistry concepts. Situational constraints too did not permit the handling of more chapters as all analyses were done manually.

3.3 The Corpus

The two chapters studied were Chapter Two of the topic "Atoms and The Atomic Theory" from page 32 to 56 from the book General Chemistry written by Petrucci and Harwood and Chapter Ten of the topic "Gases" from page 301 to 329 from the book Chemistry: The Central Science written by Brown and Le May. Illustrations and examples which are supplementary information, historical explanations, review questions, problems, equations, key word explanations, summaries and exercises are excluded from the analysis. The classification of the topic "Atoms and The Atomic Theory" into general subtopics and specific subtopics is outlined in Table 3.3.1 and the classsification of the topic "Gases" into general subtopics and specific subtopics is outlined in Table 3.3.2 The length of text examined at any one point in time is the subtopic.

Table 3.3.1

Atoms and The Atomic Theory

2-1 Early Chemical Discoveries and The Atomic theory
Law of Conservation of Mass
Law of Constant Composition
Dalton's Atomic theory

2-2 Electrons and Other Discoveries in Atomic Physics
The discovery of Electrons
X-rays and radioactivity

2-3 The Nuclear Atom
Protons and Neutrons

2-4 Chemical Elements
Isotopes
Isotopic Masses

2-5 Atomic Masses

2-6 The Avogadro Constant and the Concept of The Mole

2-7 Using the Mole Concept in Calculations

Table 3.3.2

Gases

10-1 Characteristics of Gases

10-2 Pressure

10-3 The Gas Laws

10-4 The Ideal-Gas Equation

10-5 Dalton's Law of Partial Pressures

10-6 Molecular Weights and Gas Densities

10-7 Quantities of Gases involved in Chemical Reactions

10-8 Kinetic-Molecular Theory
The Ideal-Gas Equation

10-9 Molecular Effusion and Diffusion; Graham's Law
Graham's Law of Effusion
Mean Free Path and Thermal Conductivity

10-10 Non-Ideal Gases: Departures From the Ideal-Gas Equation
The van der Waals Equation

3.4 Operational definitions of key concepts used in the research

The following operational definitions of key terms are consistent with a Hallidayan view of language. Lexis is a term used to refer to the content words of a language. These content words belong to the group of open class items.

In text analysis, units of vocabulary displaying a recurrent pattern are called lexical items. This recurrent pattern is the semantic relation between one lexical item and another. For example, the lexical items "compound", "substance", "oxygen" and "hydrogen" display some kind of a recurrent pattern in the text. Therefore, a lexical item is the smallest unit of meaning which enters into a cohesive relation with another lexical item in the text. The lexical items of a text are distributed along the syntagmatic and paradigmatic dimensions. Since the lexical items explicated from the text are a manifestation of semantic units, the explication is not restricted to only single lexical items but is extended to include groups and phrases as well. For example, "substance", "ions", and "highly" are identified as lexical items which are single morphemes whereas "fundamental particles", "minute indestructible particles", "fixed numerical ratios" and "law of constant composition" are identified as lexical items which are groups or phrases.

A semantic relation is the kind of sense relation between two lexical items that exert a cohesive force with each other. Semantic relations may be realised along the syntagmatic and paradigmatic contexts. Taxonomic labels are used to identify the type of semantic relation formed between one lexical item and another. The relations of synonymy and antonymy are two examples of taxonomic relations.

A lexical string or a lexical set is constructed of a group of lexical items; each lexical item displays a cohesive relation with a succeeding and preceding item. Each lexical item is actually an instance of a substitute for the preceding item. We may also view a lexical string as made up of a distribution of alternatives. The research methodology explains how lexical items constructing a string are coded. The lexical items forming the strings may be related to each other through the taxonomic relations of hyponymy (general/specific), meronymy (part/whole), repetition, synonymy and antonymy. The terms lexical string and lexical set are used interchangeably in this thesis.

A semantic field is a group of lexical items defining a concept or idea. For example, the lexical items electron, proton and neutron form a group of items which define the concept fundamental particles of atoms which is termed as one of the semantic fields within the confines of the text analysed. Another example would be the semantic

field states of matter whose members are the lexical items temperature, pressure and volume. The concept of semantic field is further elaborated in chapter four.

Another important concept in Hallidayan linguistics is the concept of delicacy which is used to denote further possible groupings of items. According to Halliday (Halliday & Martin 1981 : 27), "the linguist analysing a text can select a point beyond which he takes account of no further distinctions and can specify the type of relation between different systems in which he is interested."

To elaborate, consider the following system network shown in Figure 3.4.1 which describes the behaviour of fundamental particles of matter in an electric field.

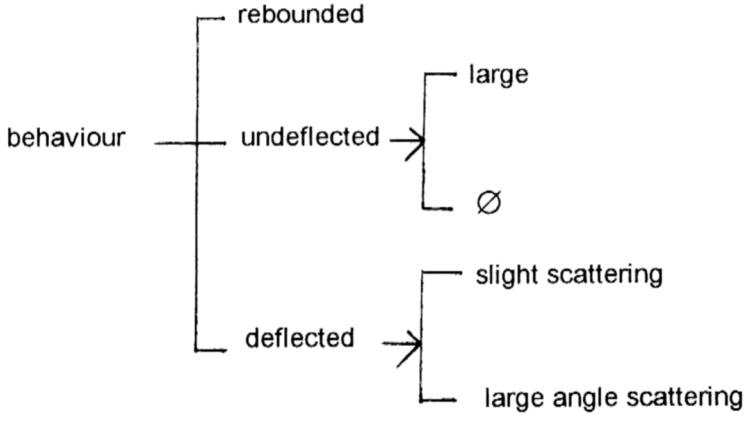

Fig. 3.4.1

It can be noted that the system has more delicate realisations towards the extreme right of the system compared to the extreme left of the system. The entry condition behaviour motivates the features

rebounded, undeflected and deflected which are all gradable lexical items in opposition with each other. The features undeflected and deflected further realise more delicate options on a cline. This carries the meaning that items may be further differentiated or intensified. The feature undeflected forecloses the options large and ∅ and the feature deflected forecloses the options slight scattering and large-angle scattering. These options are the most delicate items of the network. On the other hand, the feature rebounded does not abstract anymore delicate options and thus the system may not be extended any further here.

3.5 Methodology

The analyses were carried out in three stages to interpret the lexical organisation of the text. I labelled the first stage **the scanning stage**. The text was examined in order to have a general overview of the vocabulary items from the open class set of items. This overview enabled me to circle in pencil lexical items that appear to be semantically related to other lexical items in the text. The text was also marked for the boundaries of the stretches to be examined. All sentences were numbered 1 to n.

The second stage was called **the analytical stage**. Appropriate semantic relationships were established for the lexical items that are cohesively related to each other and these were coded onto paper. The coding of the lexical items always began with the first sentence. The first sentence was examined for lexical items that collocate or are in a taxonomic relationship with each other. These were written down on paper with some space between them from left to right along a horizontal line. The second sentence was examined to see which meanings of lexical items cannot be recovered on their own but only through the interpretation of an item from sentence one. In other words, the lexical items of sentence two refer retrospectively to the lexical items of sentence one. The lexical items of sentence two were coded to occupy the space below the lexical items of sentence one.

Following this, sentence three was looked at to see if it constitutes any lexical items that are related to lexical items from sentence two. The same process was repeated until all the sentences in the text had been accounted for. Each time a sentence was examined the immediately preceding sentence was looked for a semantic relationship. Each lexical item is tied to the preceding item by a tie and several ties form a lexical string. The lexical items were then grouped into lexical sets labelled L1 to Ln. These reflect various

semantic fields within the stretch of text under observation. All the lexical strings branch downwards and display lexical items along the syntagmatic context (when seen along the horizontal axis) and paradigmatic context (when seen along the vertical axis). The process by which lexical sets were arrived at appeals to domain knowledge, inferences, intuitions about the meanings of words and their verification with a chemical dictionary or encyclopedia. The analyses of all the stretches of texts are presented in chapter four and the theoretical construct underlying the analysis will be discussed in the next section.

The following figure, Figure 3.5.1 shows how the lexical items were coded onto paper in **the analytical stage** of the analysis.

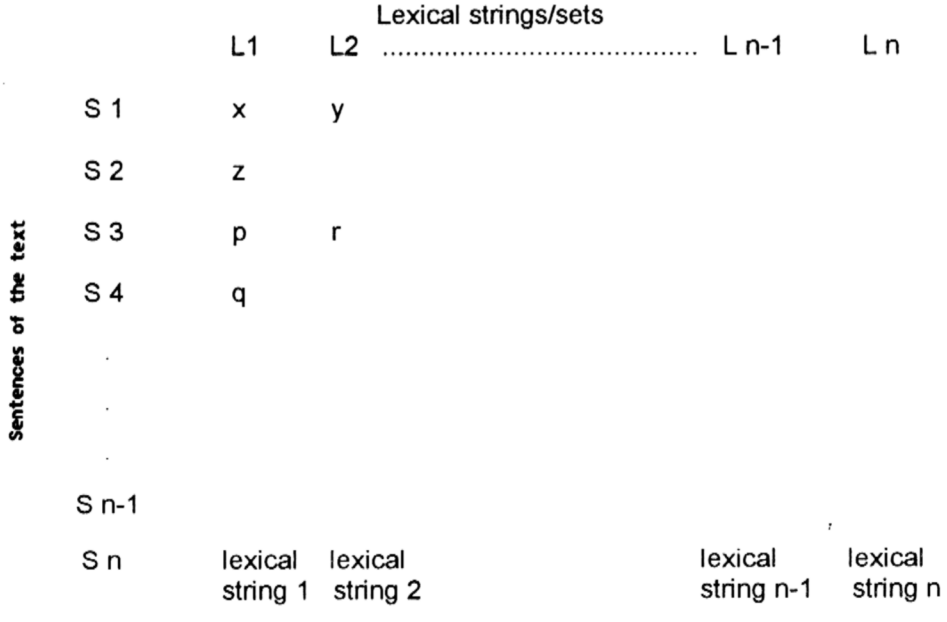

Fig. 3.5.1

Cohesively related lexical items are labelled for the semantic relationships they espouse using Martin's (1981, 1989 & 1992) categories. Figure 3.5.2 shows Martin's analytical framework for categorising collocation and taxonomic relations. The kinds of taxonomic relations are hyponymy, cohyponymy, repetition, synonymy, antonymy, meronymy and comeronymy.

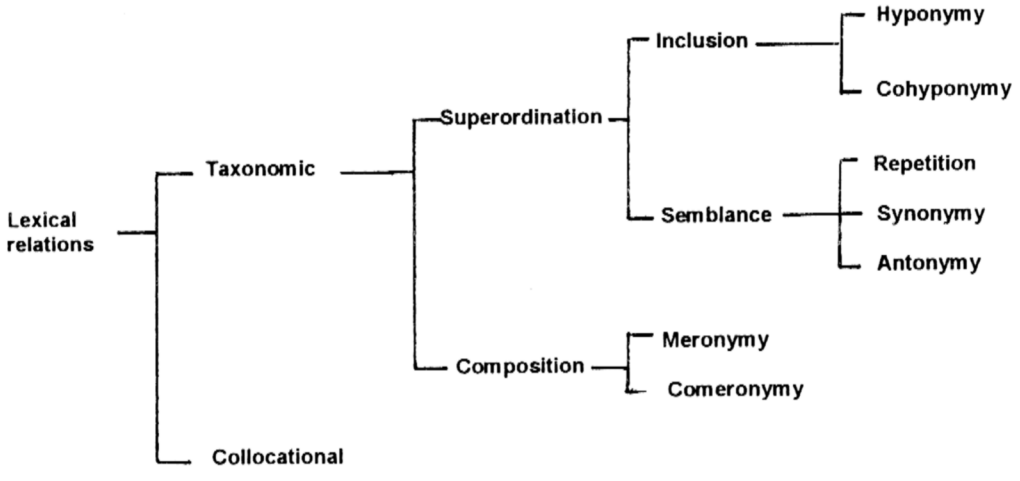

Fig. 3.5.2

The categories of cohesive relations are used in the following senses.

A relationship of hyponymy is established between two lexical items when one is a subclass of the other. For example, alpha rays is

a hyponymy of electromagnetic radiation, atomic mass of mass and carbon-12 of isotopes. The subclass which is the subordinate term or specific term is defined with reference to the class which is the superordinate or general term.

A finer distinction between hyponymy and hyperonymy can actually be made that is a hyponymy relation is formed when the class precedes the subclass in the string and a hyperonymy relation is formed when the subclass precedes the class in the string. However, the relation of hyponymy is established for both class/subclass or subclass/class relations, in order not to complicate the lexical string.

Lexical items are related through the relationship of co-hyponymy when they are subclasses of the same class. For example, alpha rays, beta rays and gamma rays are hyponyms of the superordinate electromagnetic radiation and are therefore cohyponyms of each other. Similarly, carbon and oxygen are hyponyms of element and are also related to each other through a relationship of cohyponymy.

Lexical items having similarity in meaning or complementary meaning are semantically related through a relationship of synonymy with each other. For example, in chemistry, mass, atomic weight and atomic mass unit are different ways of describing the weight of a

substance and are therefore linked to each other through a relationship of synonymy.

Martin (1992 : 301) observes that "the only true synonymy" is the relationship of repetition. A lexical item may be repeated in the same morphological form or may be repeated in a variant form. Inflexional variants are lexical items of the same grammatical class. For instance, mass and masses and nucleus and nuclei are different forms of the same words. Derivational variants are lexical items which become a new word altogether and change the grammatical class. For example, "compressibility" is derived from "compress" and "heating" is derived from "heat." However, since the difference between a derivational and inflexional variant is not always easy to identify, identical repetitions, derivational variants and inflexional variants are all subsumed under the broad term repetition.

Another interesting feature worth noting is that in chemistry symbols are used to represent chemistry phenomena. For example, α particle represents alpha particle and **C-12** represents carbon atom. These symbols may also be termed as repetitions of the lexical items they represent although in different forms.

When two lexical items have opposite meanings the relationship of antonymy is formed. Antonyms such as slight scattering and large-angle scattering, or increasing and decreasing or

expands and contracts are gradable in the sense that there are degrees of differences in terms of intensity. Non-gradable oppositions such as measured quantity and calculated quantity or real behaviour and ideal behaviour do not display differences in terms of intensity. Nevertheless, both the gradable and the non-gradable items are lexical items which may be contrasted with each other.

Both meronymy and comeronymy relationships are of the compositional kind. A relationship of meronymy is established between two lexical items when one is composed of or made up of the other. For example, neutron is a meronymy of nucleus. When both lexical items are parts of the same whole a relation of comeronymy is established. Since both protons and neutrons make up the nucleus, protons and neutrons are comeronyms of each other. Another example is that the ideal-gas equation is composed of both Boyle's Law and Charles' Law. Boyle's Law and Charles' Law are meronyms of the ideal-gas equation and are comeronyms of each other.

The category of collocation depends on the association between lexical items which are likely to co-occur in a similar environment or which mutually predict each other's company. For example, carbon-12 atoms predict carbon-13 atoms, combustion predicts air and atomic theory predicts fundamental laws. In other words, carbon-12 and carbon-13 are likely to occur in the same

contextual environments and are termed collocates of each other. The same may be said of the other two examples.

A lexical string need not be restricted to just one kind of relationship. This means that it may be constructed of both superordination and composition kinds of taxonomic relations. For example, a lexical string may be constructed of both hyponymy and meronymy relations or repetition and hyponymy relations.

The following extract from subtopic 10.9 of the chemistry chapter analysed which is chapter ten on "Gases" illustrates how lexical relations are established between lexcial items. It shows how the meaning of a lexical item may be recovered by reference to a preceding item.

Extract from subtopic 10.9 of the chapter on "Gases." (Brown & Le May : 322)

(S1) We have already made reference to the fact that the molecules of a *gas* do not all move at the same **speed**. (S2) Instead, the molecules are distributed over a range of **speeds**, as shown for *nitrogen* at two different **temperatures** in Figure 10.12. (S3) The distribution of **molecular speeds** depends on the mass of the gas molecules and on **temperature.**

"Molecular speeds" (S3) is in a hyponymic relationship with "speeds"(S2) and "nitrogen"(S2) is also in a hyponymic relationship with "gas"(S1). "Molecules"(S2) is in a semblance relationship denoting repetition with "molecules"(S1) and "temperature" (S3) is

related to "temperatures" (S2) in a similar way. Such taxonomically related lexical items create cohesion in the text and are organised along the paradigmatic dimension of the text. The concept of collocation may also be explored in this stretch of text. In S1 molecules, gas and speed are collocates, in S2 molecules, speeds, nitrogen and temperature are collocates and in S3 molecular speeds, mass , gas molecules and temperature are collocates. The collocates predict items likely to occur in the same environment along the syntagmatic dimension of the text.

The third step of the analysis which I shall term, **the conceptualising stage** is to represent the lexical items in a system network in order to show the gradable or comparable lexical items as a set of options. For example, the system effusion realises the features rapid or slow. This may be presented in a simple system network such as that shown in Figure 3.5.3. A gas may effuse rapidly or slowly under certain conditions.

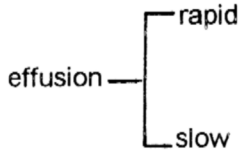

Fig. 3.5.3

3.6. Framework of the analysis

The approach I have used to explore the lexical organisation of the text is largely based on the work by Firth (1957), Halliday (1961, 1966,1978 & 1985), Halliday & Hasan (1976), Hasan (1980, 1984 & 1987), Martin (1981,1987, 1989 & 1992), Halliday & Martin (1981), Halliday & Fawcett (1987) and Fawcett (1988).

The view of language underlying the methodology used is that an explication of lexis at the syntagmatic dimension alone will not suffice to show the text as a cohesive piece of discourse. The organisation of lexical items at both the syntagmatic and paradigmatic dimensions must be taken into account to study the text as a semantic unit. In this approach, a holistic representation seen as multi-linear is favoured instead of uni-linear.

Collocation and taxonomy are two broad categories used to label lexical relationships in the text. Cohesion is achieved in the text when succeeding parts of the text depend on preceding ones to recover their meaning. A lexical item which is in a cohesive relationship with another is related through semantic means and not through structural means when the text is looked at for its cohesive properties across sentence boundaries. Items along the syntagmatic axis may collocate or may be taxonomically related to each other and these items constitute the structure of the sentence. Items along the

paradigmatic axis may be taxonomically related to each other and these constitute the system of the text.

As the text unfolds before the reader, lexical items in succeeding sentences are related to preceding ones and these appear as alternatives the writer uses to convey meaning. These alternatives are always distributed along the paradigmatic dimension of the text and are therefore taxonomically related and not collocationally. The concept of collocation which is about the association of lexical items in similar contextual environments does not involve the paradigmatic dimension of the text but only the syntagmatic dimension. Therefore the taxonomic labels which are used to establish the semantic relationship of the lexical items are given more importance in this research.

The semantic stratum of the text gives us information about the organisation of the lexis and how cohesion is achieved. The lexicogrammatical stratum gives us information about the collocational possibility of lexical items and also enables the application of systemic theory to display lexical items which can be intensified or compared.

3.7 Conclusion

This chapter has described the research methodology and outlined the theoretical framework used in the analysis of lexis in the study. The following chapter presents the findings of the study using the methodology and theoretical framework just illustrated. The findings concern taxonomic analyses and system networks.