

**UNIVERSITY OF JYVÄSKYLÄ**

**USABILITY OF TERMINOLOGY MANAGEMENT  
PROGRAMS AND DATABASES - A SURVEY STUDY**

**A Pro Gradu Thesis**

**By**

**Mikko Lemmetti**

**Department of English 2001**

HUMANISTINEN TIEDEKUNTA  
ENGLANNIN LAITOS

Mikko Lemmetti

USABILITY OF TERMINOLOGY MANAGEMENT PROGRAMS AND  
DATABASES - A SURVEY STUDY

Pro Gradu -työ

Englantilainen filologia

Huhtikuu 2001

90 sivua

Tämän tutkielman tarkoituksena on tutustuttaa lukija sanasto- eli terminologiatyöhön, terminologian alan käsitteisiin ja teoriaan sekä kyselytutkimuksen avulla etsiä hyviä ja huonoja puolia terminhallintaohjelmistoista ja julkisista termipankeista Internetissä. Lisäksi pyritään tarjoamaan ehdotuksia terminhallintaohjelmistojen ominaisuuksien parantamiseksi ja julkisten termipankkien toimintojen selkeyttämiseksi. Tutkielman materiaali siis koostuu Internetissä olleen kyselylomakkeen antamista vastauksista ja niiden analyysistä. Koska vastauksia on vain 10 henkilöltä, tutkielman lähestymistapa on kvalitatiivinen.

Huomionarvoisia tekijöitä sanastotyössä ovat mm. termien muodostamistavat, terminologian johdonmukaisuus ja oikeellisuus sekä termien suhteet toisiin termeihin (esim. synonyymeihin). Kyselytutkimuksessa kysymykset kosketelevat juuri termipankkien sisältöä, terminologisten tietueiden muodostamista ja terminhallintaa yleisesti. Vastausten perusteella terminhallinta yrityksissä ja projektitasolla on hoidettu hyvin ja yksityishenkilöiden, esimerkiksi freelance-kääntäjien, kohdalla keskinkertaisesti. Yleensä laajemmat termipankit ja niiden muodostaminen asiakkaille oman käytön sijaan tarkoittavat, että termistöistä pidetään tarkempaa huolta.

Kyselytutkimuksen vastausten perusteella voidaan todeta, että kaupallisissa terminhallintasovelluksissa on parantamisen varaa: esimerkiksi indeksikenttien lukumäärää ei tulisi rajoittaa, helppokäyttöisyyttä (mm. nopea termien lisääminen) tulisi lisätä, ja ohjeistuksen tulisi olla riittävän laaja. Käyttöliittymien ulkoasu ja toimivuus osoittautuivat myös tärkeiksi, ja etenkin termitietueiden selkeä esitystapa helpottaisi termien sisäistämistä. Hakuominaisuudet olivat yhtä lailla olennaisia, ja termitietokantojen laajuus varsinkin Internetissä osoittautui varsin hyväksi.

Kyselylomakkeen tuottamia vastauksia tuli kvantitatiivista analyysia varten liian vähän, joten tuloksia ei voida myöskään yleistää. Lomakkeen rakenne olisi voinut olla hieman parempi, kysymykset laajempia ja vastausvaihtoehtoja enemmän. Siitä huolimatta kyselytutkimus poiki käyttökelpoisia vastauksia, joiden perusteella voitaneen antaa hyviä korjausehdotuksia tulevia termipankkeja ja terminhallintaohjelmistoja varten.

Asiasanat: terminology management, usability, qualitative analysis, improvements

## Contents

<b>1. AN INTRODUCTION TO THE THESIS.....</b>	<b>5</b>
1.1 THE OBJECTIVES AND RESEARCH PROBLEMS IN MORE DETAIL .....	6
1.2 PREVIOUS AND CURRENT RESEARCH ON TERMINOLOGY .....	6
<b>2. THE BASIC CONCEPTS OF TERMINOLOGY .....</b>	<b>8</b>
2.1 THEORY AND GENERAL CONCEPTS.....	8
2.1.1 <i>Basic Elements</i> .....	9
2.1.2 <i>Definitions, Concepts and Terms in Detail</i> .....	9
2.2 WHAT IS UNDERSTOOD WITH THE CONCEPT OF TERM .....	12
2.2.1 <i>Relationship Between Terms and Concepts</i> .....	12
2.2.2 <i>Relationship Between Terms and Contexts</i> .....	14
2.2.3 <i>What Constitutes a Good Term</i> .....	16
2.3 THE NEXT STEP IN TERMINOLOGY MANAGEMENT .....	17
<b>3. TERMINOLOGY MANAGEMENT .....</b>	<b>17</b>
3.1 TERM FORMATION.....	17
3.1.1 <i>When Terms Are Formed</i> .....	18
3.1.2 <i>How Terms Are Formed</i> .....	19
3.2 TERM BASE STRUCTURE .....	21
3.2.1 <i>Definitions As Part of Term Base Structure</i> .....	21
3.2.2 <i>Data Categories for Terminology Management</i> .....	23
3.2.3 <i>Data Category Specifications</i> .....	24
3.3 STANDARDS AND INTERCHANGE .....	26
3.3.1 <i>The ISO Standard</i> .....	27
3.3.2 <i>The SGML</i> .....	27
3.3.3 <i>The MARTIF Standard</i> .....	28
3.3.4 <i>The NTRF Standard</i> .....	30
<b>4. A QUESTIONNAIRE FOR GATHERING THE DATA, AND THE PREPARATION OF THE SURVEY.....</b>	<b>31</b>
4.1 QUESTIONNAIRE RELIABILITY, VALIDITY, AND REPRESENTATIVENESS.....	32
4.1.1 <i>External and Internal Reliability</i> .....	34
4.2 DESIGNING A QUESTIONNAIRE SURVEY .....	35
4.2.1 <i>Some Rules To Remember</i> .....	35
4.3 DEVELOPING AND TESTING THE QUESTIONNAIRE.....	36
4.2.1 <i>Questionnaire For the Web</i> .....	37
<b>5. THE SURVEY .....</b>	<b>38</b>

5.1 THE QUALITATIVE APPROACH .....	38
5.1.1 <i>The Number of Responses Received and Their Effect on the Analysis</i> .....	38
5.2 THE ACTUAL QUESTIONNAIRE .....	39
5.2.1 <i>The General Questions</i> .....	39
5.2.2 <i>The Specific Questions</i> .....	40
<b>6. THE RESPONSES UNDER SCRUTINY .....</b>	<b>48</b>
6.1 RESPONDENT NO. 1.....	48
6.2 RESPONDENT NO. 2.....	51
6.2.1 <i>A Closer Look at Euterpe</i> .....	51
6.2.2 <i>The Tapa Term Bank</i> .....	52
6.2.3 <i>Other Responses</i> .....	53
6.3 RESPONDENT NO. 3.....	56
6.4 RESPONDENT NO. 4.....	59
6.5 RESPONDENT NO. 5.....	62
6.6 A MID-POINT SUMMARY.....	66
6.7 RESPONDENT NO. 6.....	67
6.8 RESPONDENT NO. 7 - A COMPARISON WITH NO. 6.....	69
6.9 RESPONDENT NO. 8.....	72
6.9.1 <i>The TermStar</i> .....	73
6.9.2 <i>Other Responses</i> .....	74
6.10 RESPONDENT NO. 9.....	75
6.11 RESPONDENT NO. 10.....	77
<b>7. A SUMMARY OF THE DATA.....</b>	<b>79</b>
<b>8. A LOOK BACK .....</b>	<b>81</b>
<b>9. FUTURE SUGGESTIONS ON THE USABILITY OF TERM BANKS.....</b>	<b>84</b>
<b>10. CONCLUSION.....</b>	<b>86</b>



## 1. AN INTRODUCTION TO THE THESIS

The present study has two main goals. The first goal is to give a short introduction to the world of terminology management, its applications and relevant vocabulary. I will not assume that the reader already knows about terminology work and management; thus it is important that he/she would have at least some idea of what the present study will deal with. Moreover, to present results on certain aspects of terminology management without first acquainting the reader with the "science" itself would constitute a weak foundation for the study.

The introduction will be made by reviewing both English and Finnish literature on the subject, by clarifying examples, and by briefly reviewing one or two so-called *terminology management systems*, which are computer applications designed to collect, maintain and edit terminological information. The second goal is to present data about the actual use and management of different *term banks / terminological databases*. For this purpose, I launched a survey that utilized a questionnaire in order to gain information about the actual use and management of terminology databases. The questionnaire, whose various parts be dealt with in detail later, consisted of 25 various questions, and the responses from the questionnaire survey will be analyzed later. The questionnaire was sent out to terminologists, translators, database administrators, technical writers and to other people relevant to this field of study. The implementation of the questionnaire as well as a general view to the use of questionnaires will be elaborated in more detail later in the present study.

### **1.1 The Objectives and Research Problems in More Detail**

The primary intention in the present study is to find out what different users of terminology databases - or term banks - and terminology management systems think of their tools: what is good or what needs improvement, which kind of characteristics a term bank has and so on. I will also try to find out, for example, how well a terminology database serves the users' needs. With this I mean that not all term banks are well designed for the users; the banks can be too complex, too vague or otherwise fail to optimally give the users the information they need. Therefore, the primary concern will be the usability of terminology databases. Moreover, I will try to show whether there are any obvious trends toward building an ideal term bank. The concept of an ideal term bank is by no means clear, because what is ideal can be defined in many ways. However, there are recommendations and standards for terminology databases (n.b. terms *terminology databases* and *term banks* are used interchangeably in the present study), and some of these recommendations and standards will later be explained further.

### **1.2 Previous and Current Research on Terminology**

Lately, the demands for new terminology have been growing. This is due to ever-expanding commercialism, technology, and scientific knowledge. Hardly any person can have avoided stumbling constantly onto new terms and concepts. The field of terminology has been a subject of research all over the world for several decades. In Finland, the universities in Turku and Vaasa, for example, have been doing terminological research at least since the 1980s. Commercially, one of the oldest and most reliable institutions in terminology management, creation and standardization is *Finnish Centre of Technical Terminology*, which has been operating since 1974. The Centre, which serves as a source of technical and other terminology, has produced a comprehensive set of terminological literature, with one of its primary objectives being the establishment of terminological projects, the education in and promotion of the methods of terminology. (Nykänen 1999:7). In the United States, terminological research has longer traditions.

Although the actual theory of *terminology* has not been originally created in the United States, there are numerous academic institutions that work with terminological theories and applications all the time. One of those institutions is the Kent State University in Ohio, and one of the most experienced terminological

theorists and practitioners is Sue Ellen Wright, who has among other things compiled and edited the two volumes of *Handbook of Terminology Management* (later only HTM, HTM1, or HTM2). The first volume of the handbook deals with basic aspects of terminology management, and the second volume, which is at the moment of writing still in press, deals with more practical aspects of terminology management. Both volumes will be used as source material for reference in the present study, because they are quite comprehensive, and approach the field of terminology management from different angles.

The present study will not introduce specific philosophical theories of terminology. Neither will it intend to give the reader a comprehensive description of different schools of terminology because the present study is not a critical analysis on the theory or education of terminology itself, and because that would take the focus away from the actual subject of the paper. Rather, the study will introduce some aspects of terminology management presented in the different handbooks and in some Finnish publications. Among the topics discussed will be term formation, terminological definitions, terms and contexts, the standardization of terminology, and the guidelines of terminology management. These topics have been dealt with in the first volume of the *HTM*.

Another presentation here will be to the topics in the second volume of the Handbook, and those topics include: quality management for terminology, the criteria for evaluating terminology database management programs, data categories, and some practical considerations for a term bank, to name a few. As we move along, both Finnish and English approaches to terminology will be discussed, but most of the theoretical and pragmatic data will be based on Finnish conventions and theoretical approaches. The questionnaire will have only of Finnish respondents. It was necessary to limit the present study to one country, since I aim to illuminate the state of terminology management in Finland, and not expand the survey elsewhere. Moreover, for certain reasons, it is more relevant to gain data from sources that will probably and hopefully benefit the domestic domain.

The present study aims mainly to present the opinions and observations of different users of terminology and terminology tools, and hopefully give an overview of the current status of terminology management in Finland. For example, a manager in a

certain Finnish translation agency claimed that the overall quality of term banks in Finland is rather poor. This study will try to find out whether the claim is justified or not. The findings based on the questionnaire can hopefully be used to improve the quality of terminology databases. I will analyze the data qualitatively, because the number of received answers do not allow a quantitative approach and statistical analysis. The questionnaire itself is not necessarily based on any previous recommendation or a model questionnaire.

## **2. The Basic Concepts of Terminology**

In this chapter, I will be dealing with some theory and concepts of terminology: its “building blocks”, such as definitions, concepts and terms. All things in the universe have a meaning, and the human mind has a tendency to give meanings to things. An important job for a terminologist - a person who works with various issues of terminology - is to preserve and elucidate the meanings of different things, whether they are concrete or abstract, by using relevant terms, concepts and definitions. These are basically the area of *semantics*.

Semantics is a branch of linguistics that studies and aims to analyze and explain sometimes seemingly hidden, and different, meanings. The meanings can be related to anything from words to concrete objects to abstract notions, but if we concentrate on words that are first and foremost the means by which a linguistic meaning is communicated, we can see that choosing a right word for the proper referent is extremely vital in many ways. Semantics also studies at the communicative function of words.

### **2.1 Theory and General Concepts**

By *terminology*, I understand it to be a study of terms, as *biology* is a study of living organisms or *paleontology* is a study of fossils. Moreover, we also understand terminology as a special vocabulary, if you will, of a particular field of science, technology, and so on. One should not, therefore, confuse *terms* - the basic elements of terminology - with ordinary words. The distinction between terms and words - or *lexemes* - is necessary, because specific terms are not usually included in general lexicons, but rather in their own special ones. The semantic relations of terms do not differ significantly from the semantic relations of "general words". One vital aspect of terminology management is to understand these relations,

especially if term bank structuring involves using hierarchical presentation of terms, introducing synonyms for terms, and so on.

### **2.1.1 Basic Elements**

First, an introduction to the vocabulary of terminology, and to the terms of terminology will be in order. It is stated, for example, in the *Handbook of Terminology Work* (later *HTW*) (1988: 25—96) that there are at least three basic elements of terminology: *referents*<sup>1</sup>, *concepts*<sup>2</sup> and *terms*<sup>3</sup>. A fourth element will inevitably fall into the same group, namely *definition*<sup>4</sup>. To be exact, however, the field of terminology has adopted these elements from semantics. Concepts, terms, definitions, and referents have to be understood from a semantic point of view. Lyons (1977), for example, has conducted an extensive study on the interrelations and functions of these semantic elements. The next paragraph will deal with the terminology of those elements.

A *concept* can be seen as a unit of thought that combines various characteristics of an *object*, which can be either an abstract one or a concrete one. More specifically, if we think about an object such as a *vase*, its concept can tell that it is, for example: a vessel, inanimate, normally used for holding flowers, made of glass or ceramics, and so on. A definition is a statement that describes the concept, and a term is an expression that refers to the concept (Suonuuti 1997: 9). When we talk about terminology management and term bank structure, a *term* is the primary element on the basis of which all the database searches are basically made. However, a term is not itself sufficient to make a database entry. The term needs at least a good and accurate definition. The more comprehensive the definition is, the better it explains the term. However, the definition should be short and long enough to bear only the information that is needed and is relevant for the term at hand.

### **2.1.2 Definitions, Concepts and Terms in Detail**

Although the following discussion presents some elementary components of terminology as if they were only related to that particular field, the reader has to bear in mind that they have been present for a long time in the discipline of semantics. When we think of *definitions* in this context, they are a sort of attributes or characteristics that describe a certain individual entity. That entity might be, for

example, a bird, or an automobile. Any entity can be said to belong to a *class*, which may or may not contain other entities. In other words, a class presents a collection of individuals that by some reason belong to the same class, the reason being irrelevant. (Lyons 1977: 154).

Although Lyons presents classes and their members by means of a *set theory*<sup>5</sup> that is more familiar to mathematicians, further elaboration of the theory is not relevant for the present study. However, something has to be said about the set theory: the theory looks something like this: a class, let us say **T** (for Temperature scales), contains members *C*, *F*, and *K* (respectively Celsius, Fahrenheit, and Kelvin). The members would be represented in braces, as {C, F, K} and separated by commas. Therefore it can be said that "*C* is a member of the class **T**), and presented as follows: C · T.

We still have to have a way by which we define the actual membership of any class. One way, according to Lyons (1977: 158), is to list the members of a closed class, that is, a class whose members can be listed. On one hand, we cannot list all members of *numbers*, because there are infinite number of numbers (no pun intended). On the other hand, we can list the members of the class of *alphabets*.

Definitions can be subdivided into *extensional* and *intensional* ones. Lyons (1977: 158) states that a class can be defined on the basis of properties the members, or the terms, of the class share. That is called *extensional* definition. The *intensional* definition of a term would be that the term has some property or properties that qualify it as a member of a given class. In the field of terminology, the extensional definition of a concept is used as a delimiter that sets the boundaries for the referents of the concept (Suonuuti 1997: 16). According to Kalliokuusi (1999: 45), referents are “notions of a group of similar phenomena in the surrounding environment formed by people”. That is, we all have a general idea of different objects and abstractions in our environment and in our mind. We give them certain characteristics and values, and define familiar and new things partly through those characteristics and values. Or, the extensional definition of a concept can be defined as “the totality of the objects to which the concept refers” (Wright 1997: 339). Let us take an example: a **planet** is a concept, and its extensional definitions can be **Earth, Jupiter, Saturn** or in general, planets in our solar system or elsewhere (Bessé 1997: 63). That is, the concept of a planet can be extended to

cover all planets, regardless of their characteristics. In that light, Earth, Jupiter and Saturn could not be anything other than planets.

However, the extensional definition is not very descriptive. Therefore, an intensional definition is preferred in many instances. The intensional definition presents the general class of the defined concept, and intensional definitions separate themselves from the other concepts of the same class. In Bessé's example (1997: 69), a chemical compound known as **methionine** is defined as:

“a sulphur-containing amino-acid,  $\text{CH}_3\text{-S-CH}_2\text{-CH(NH}_2\text{)-COOH}$ , which must, for man as well as for other animals, be present in the food proteins. Apparently concerned, inter alia, in reactions involving methylation, such as synthesis of adrenaline and creatine (Chambers Technical Dictionary 1961).”

In the above situation, the definition explains both the function of that particular amino acid, that is, for what the amino acid is used, and its substance; what the amino acid is and what it is made of.

I consider there to be a rough division of concepts to *individual/specific concepts* and *generic concepts*, although in many instances, individual concepts mean the same as *particular concepts*, and *generic concepts* mean the same as universal concepts (Lyons 1977: 110). Whatever the word choice, these are considered general semantic concepts, and they have also been adopted by most of the terminologist. Lyons states that the entities that exist in the world are considered individuals, and the individuals have certain properties that then work as sort of a "concept-builders", i.e. they help to create and maintain the concept of an individual entity and its describing and classifying characteristics. Lyons goes on to describe what a concept is; he sees it as an idea, or a mental construct (a view shared also by many cognitive scientists) that the mind uses to acquire information about things, and to learn about them. Lyons' s description of a universal concept, then, is that the universal concept is associated with words like *man* (male) or *beautiful*, particularly when these kind of words are used to indicate that an entity can universally be classified as a man or considered to be beautiful.

In the world of terminology, then, and for example in the *Handbook of Terminology Work* (1988: 25—26), an individual concept is connected to only one referent or object, and in many cases, a name (or proper noun) expresses it. For instance, the individual concept **Gulf of Finland** refers only to the geographic extension of the Baltic Sea between Finland and Estonia. The generic concept is

formed by combining the common characteristics of individual referents, for example, all bodies of water that extend deep towards inland, and are of a considerable size, are called by the name **gulf**. Then, the name for the generic concept is a **term**. The next chapter is devoted to terms only; their different definitions and use.

## 2.2 What Is Understood With the Concept of Term

Suonuuti (1997: 25), for example, states that a term can consist either of one word or a compound word, or it can be a multi-word expression, for example **moisture**, **skyscraper**, and **fault line** respectively. Wright (1997: PPT1)<sup>1</sup> defines a term as “a word (or string) belonging to specialized areas of usage in one or more languages”, and “a prenegotiated word or term assigned to a specific concept in a specific subject field.” A term is not, in most cases, a word that can be found in a general dictionary. Moreover, a term can contain a group of letters not otherwise qualifying as words, and also of numbers or symbols. A significant number of commercial and technical terms are acronyms such as **NASDAQ**, **ASCII**, **CD-ROM** and so on. These acronyms stand for *National Association of Securities Dealers Automated Quotations*, *American Standard Code for Information Interchange*, and *Compact Disc Read Only Memory*.

### 2.2.1 Relationship Between Terms and Concepts

The relationships between terms and concepts are not easily described, and there are various philosophical views on how to call these relationships. However, there are acknowledged practices in semantics that help to understand how things relate to each other, and eventually, to their concepts. Everything can be said to have a meaning, or a *sense*, as Lyons (1977: 197) calls it. Moreover, he explains that there is a distinction between sense and *reference*. What I understand this to mean is that two separate expressions, for example, may refer to the same thing, but the expressions do not have the same sense. Having the same sense would therefore indicate *synonymy*.

One primary concern of structural semantics is the relations the sense has within lexemes (Lyons 1977: 270). These relations come in many forms, but not all of

---

<sup>1</sup> A PowerPoint Presentation n:o 1. No longer available on the internet. See Bibliography for details.



them are necessarily utilized in the field of terminology. However, it may sometimes be prudent to indicate what relationship a given term has with other terms. For instance, one could introduce a term **high beam**, meaning the headlights of a car that emit light a greater distance, as opposed to **low beam**. This opposition in semantics is called *antonymy*. In fact, giving an antonym for a term could prove useful in showing what the term *is not*, and possibly in helping to identify a previously unfamiliar term by showing an antonymic term that might be more familiar.

Terms, or lexemes, hardly occur in a vacuum. That is, there are numerous terms or lexemes that in some way are involved with others terms and lexemes. The involvement can be hierarchical, and this hierarchy is indicated by *subordination*<sup>6</sup>. This means that terms (I speak of terms from now on) have either a super- or subordinate term. For example, **hammer** has a superordinate term **tool**. Subsequently, a hammer can also have a subordinate term, for example, **carpenter's hammer**. A subordinate term is called a *hyponym*, and logically, the term higher in the hierarchy is a *hypernym*. The following paragraphs will deal with the subject at hand from the vantage point of terminology more than that of semantics.

In any language there are words that can have different meanings, even when their syntactic or phonological forms are the same. The same applies to terminology, where the terms can have different meanings depending on the subject field they belong to. For the purposes of terminology, the *HTW* (1988: 70—72) presents five different term - concept relations, but I will deal here with only four: *monosemy*, *polysemy*, *synonymy*, and *homonymy*. First, in monosemy, the term corresponds to only one concept and one concept corresponds to only one term. However, a complete monosemy in terminology is not very plausible, and therefore one term usually corresponds to at least two concepts. In that case, we talk about polysemic relations. In polysemy, the term corresponds to two or more concepts that are somehow related to each other. For example, a **foot** can mean a part of human body, a foot of a tree (as opposite to the top) or a foot of a bed (as opposite to the head).

When the terms correspond exactly to the same concept, they are synonyms. In terminology management, synonymy is considered a hindrance. Term users may

not necessarily know to which field a synonymic term belongs. For example, the term **condenser** belongs at least to the field of electronics (an electronic component), to the field of textile industry, and to the field of automotive engineering, to name a few. Therefore, if synonyms are not specifically linked to their respective fields in the terminology database, the database user will have a hard time deciding which term is the right one. Moreover, in many professions, experts will sometimes use different terms from the ones found in the term database. For instance, if a translator who uses a term bank deals with vocabulary found in rock drilling, they can find a source language term with a seemingly valid target language equivalence, but it may not be the one the professional drillers normally use. I will deal with the multilingual terminology management later on in the present study. Until now, we have only discussed monolingual terminology.

Finally, homonymy means that the terms can be the same in form, but differ in meaning. For example, a **hold** can be a place in a ship that is used to store cargo, or it can be a verb denoting to “keep something in one’s hand”. In fact, polysemy and homonymy are so close to each other that sometimes the distinction is impossible to make.

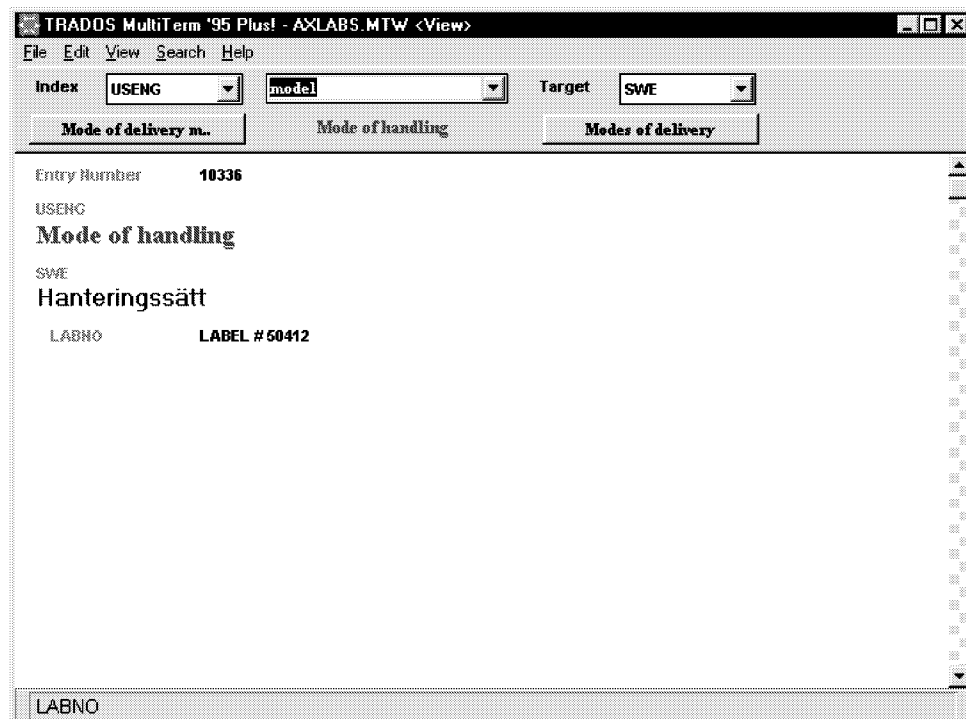
### **2.2.2 Relationship Between Terms and Contexts**

Any term is rather useless without a context to place it in. All terminological units, and in this case, terms, are connected first to a concept and then to a context. Terms usually go hand in hand with the concepts and contexts that are found in specific languages, for example in the languages of science, medicine, computer technology and so forth. When terms are formed - the actual process of term formation will be discussed later - one has to consider the context in which the term will appear. Usually, when terms are presented, they are presented in their actual contexts. In parallel to the use of a corpus database where certain words can be looked up with a *concordance*<sup>7</sup> search, the terminology databases should aim to give a real example of the use of the term, that is, to present the term in its actual context. In corpora and in terminology such examples help the user to visualize how, where and when a particular term (or a word in the corpus) is used. As Dubuc and Lauriston (1997: 81) point out, the context exemplifies a particular discourse carrying the term. Moreover, the context contains a sentence in which the term appears. The context can also be the length of several sentences or even of a

complete paragraph. However, the inclusion of an entire paragraph within a terminological entry is fairly rare, and it also would consume too much time and extensive amounts of storage space if all terms were to be presented in a whole-paragraph context.

Dubuc and Lauriston (1997: 82) claim that there are three types of contexts. First, there is an *associative context*, which is used only to show that the term is used in some specialized language. This context does not give any information about the concept that the term covers. Second, an *explicative context* will give a rough view of the concept that is covered by the term. This type of context is not entirely satisfying if a full explanation of the term's functions or characteristics is wanted. However, the explicative context is more informative than the associative context. Third, there is a *defining context*. The defining context contains descriptive information that is sufficient for the term both in quality and in quantity. This context also gives a good view of the concept covered by the term and therefore implies a true definition of the term.

In terminology management, it is important that when terminologists are building a new term bank, and adding terminological entries, they also spend enough time and effort to present terms in relevant contexts. Fairly often, however, many terms are



Picture 1: The user interface of MultiTerm (Microsoft)

not presented in any specific context; they are only given as they are, probably with just an equivalent target language term (see example below). This is the case in bi- or multilingual term banks. Therefore, one of the main criteria for a good term bank is the choice and inclusion of good contexts for the term. Of course, it depends largely on the manager and builder of a terminology database whether contexts are required or not. It also depends on the intended use of the database what is included in it.

### 2.2.3 What Constitutes a Good Term

When terms are chosen to be included in the database, one should make sure that they are linguistically correct and follow the norms of the given language (Suonuuti 1997: 25). This does not, however, make a term a "good" one in itself. The *HTW* (1988: 73—80) has some suggestions for a good term. A good term will give an illuminating image of a concept even for one who is not a subject-field expert. The *HTW* lists five most important features for a good term, which are elaborated below. I will explain the following definitions for a good term in my own words.

First, a good term should be *translucent*. This means that the most important features of the concept in question are seen in the term, i.e. the term describes the concept well. Second, the term should be *consistent*. A consistent term will form a clear system of terms with other terms, a system that describes the relations between concepts. Third, the term should be *appropriate*. In other words, this means basically that a term is not appropriate for a purpose when it creates false images. For example, in Finnish, one translation for a computer's storage device **hard disk** is **Winchester-levy**, 'Winchester disk'. There is a possibility that some people may get an illusion that the Finnish term has more to do with rifles than with computer accessories. Terminology experts do not encourage the use of negative-sounding terms or unnecessary loan translations. For example, in paper industry, the quality of the surface of any paper is indicated by its **smoothness**, not **roughness** (HTW 1988: 76).

Fourth, a good term should be *distinguishable* from other terms that are almost identical. For example, words like **wench** - **wrench** or other words that differ only by one letter may cause a mix-up especially in speech. Moreover, especially in Finnish language, compound words with a genitive-nominative combination may

be difficult to separate: compare for example **autohuolto** - **autonhuolto**, where the first term means a place where cars are repaired, and the second term means the actual process of repairing a car. Last, a good term should be *short*. When two terms are offered for use, usually the one that is the shortest is selected. If the term is too long, it tends to be shortened, and finally, a new unofficial term is created, and people start to use it. This is not the purpose of good terminology management.

### **2.3 The Next Step in Terminology Management**

In the next chapter I will discuss the actual processes and conventions of terminology management. The processes are viewed both from monolingual and multilingual perspectives. As the present study is concerned primarily with Finnish terminology work, the emphasis will turn more and more towards how terminology management is carried out in Finland. However, the upcoming discussion about terminology standards is based on international conventions.

The main areas in the next chapter will be term formation, definitions as the means of displaying important information, and the term equivalence. In Finland, and in other countries where a correspondence between source language - target language terms is vital, careful consideration is needed to find suitable *equivalents*<sup>8</sup> (or counterparts) for source language terms has to be given. This appears to be one of the main interests in Finnish terminology work.

## **3. Terminology Management**

In this chapter, I will introduce the laborious task of managing terminology, as well as various issues involved in the management process. The word *management* is used here to pertain to everything from the making of a term to the modifying of a terminology database. We talk about *terminology work*, which translates into Finnish as *sanastotyö*. However, terminology work is a larger field that contains terminology management. Therefore, these two are kept apart in the following text.

### **3.1 Term Formation**

A particular special language community, such as medical personnel, requires specific terms in order to communicate and develop communicative processes. Term formation is the process in naming the concepts that the special language community uses for those two purposes. Sager (1997: 25) claims that term

formation is different from general word formation, since term formation puts greater emphasis on pre-existing patterns and models to better facilitate communication and transmission of specific knowledge. Sager goes on to point out that term formation can be a process where entirely new combinations of phonemes are constructed. Making a new word or term that does not exist anywhere in the world is called *coinage*. In term formation, however, this method is somewhat rare. There are other, more suitable and preferable methods to form terms.

Normally, term formation uses existing lexical elements, basically words. There can be both simple and complex terms, the former consisting only of a single lexical element, and the latter consisting of compounds, affixed derivatives, phrasal units and so on. Some examples of more complex terms could be **fixed aerial direction finder** or **floor area of a dwelling**. It should be noted that simple and complex terms are always presented without an article, or if a term is a verb, without a preposition.

### 3.1.1 When Terms Are Formed

New information brings about new concepts. It also changes some old concepts. These new or modified concepts have to be expressed somehow, and term formation aims to do just that (HTW, 1988: 81). A concept can be expressed with different terms, but the ideal model of term formation is that one can come up with a term that stands out and is clearly separated from other terms that could arise. For example, in general language, the word **bucket** can have at least two different equivalents in the Finnish language, namely **sanko** and **ämpäri**, depending on the regional dialect. Where specific terminology is concerned, however, the variation should be minimal, and only one term for one specific concept should exist. One has to bear in mind, however, that different fields of study or different technological areas can indeed have some variation in the use of terms, that is, the field of medicine can use a different term for a concept than biology does, for instance. Therefore, different terms for one concept can co-exist, but preferably not within the same domain (field).

New terms are formed when there is no term for an existing and recognized concept in any language, and when large, consistent and specialized terminology is required. According to the *HTW* (1988: 81), new terms will not be formed, for instance, when there is a term for the concept in the target language, although in a

different form: **know-how** -> **osaaminen**, is preferred **taitotieto**; or **audio disc** -> **äänilevy**, not **audiolevy**. In spite of this rule, there are, in a sense, non-preferred terms around, and the reasons for using a contrived term for an existing one are numerous. For instance, a preference of one term over another might be due to the desire of a large company to develop their own terminology for their products to distinguish them from other companies' products. An example of this is a common word for an **audio disc**: **äänilevy**. An entertainment electronics corporation, however, could promote the use of **audiolevy** in their advertising, documentation and terminology.

### 3.1.2 How Terms Are Formed

The *HTW* (1988: 83) recognizes seven different ways to form terms. These methods are common in all European languages. However, some variation between American and European methods of term formation may exist. In this chapter I will deal with methods used, for example, in Finnish terminology work. The seven methods of term formation are *terminologization*<sup>9</sup>, *compounding*<sup>10</sup>, *derivation*<sup>11</sup>, *borrowing*<sup>12</sup>, *abbreviation*<sup>13</sup>, *conversion*<sup>14</sup> and *coinage*<sup>15</sup>. Terminologization means that there is a standard language word that is incorporated in some special language, or jargon. In jargon, the meaning is replaced by the definition of the concept. (HTW 1988: 84). For example, the word **depression** is usually connected to a physiological or mental state of a human being, as in “She is feeling depressed”, but in the terminology of trade depression means the period when economy is low and unemployment rate is high, and as a result, trade is not doing well, either. Moreover, the term for an everyday kitchen utensil, the **fork**, has been combined with words like **fork lift**, **fork joint**, **fork wrench** and so on to form terms that are used in industry, automotive technology, and mechanics for instance.

In compounding, two or more words are combined in order to form a new word that differs in meaning from the elements of the compound (Pyles 1993: 261). Therefore, for example, words or phrases like **sellout** (a noun) and **sell out** (a verb) do not mean the same: a **sellout** means a very successful sale of something, and to **sell out** means to play false or betray someone. In derivation, on the other hand, many derived words are formed by adding an affix (mostly a suffix) in a word. For example, words like **station** > **stationary**, **process** > a **processor** are the result of

derivation. Especially in Finnish, derivation is a fruitful way of making new words, and a significant number of everyday nouns, *derivatives*, have been derived from verbs. For instance, the noun **kampi** ('a **crank**') - has been (the exact etymological origins of these examples are not investigated) derived in Finnish from the verb **kammeta** ('to **crank**'), and the noun **poljin** ('a **pedal**') - has been derived from the verb **polkea** ('to **pedal**'), and so on.

Borrowing is a frequently used way to introduce new words into almost any language. In Finnish language, as well as in English, there are numerous borrowed words from various other languages. English has borrowed words from Latin, French, Greek, Spanish and so on, and Finnish from Swedish, Russian, and English, to name a few. The *HTW* (1988: 94) further defines three types of loans that are common in terminology work. These are *quotation loans*<sup>16</sup>, *special loans*<sup>17</sup>, and *general loans*<sup>18</sup>. Quotation loans are pronounced and written as they appear in the source language, for example: **management** (in trade and business), **layout** (in publishing and printing), and **mind mapping** (in business and education). Special loans are generally written and pronounced like any Finnish words, but they are clearly of foreign origin, that is, English: **reaktio** < **reaction**, **produktio** < **production**, **brunssi** < **brunch**, and so on.

General loanwords do not normally resemble their source language words at all, and these loanwords are fully adopted into Finnish language. Furthermore, there are so-called translation loans. The translation loans are translated into their target language. Translation aims to maintain the syntactic structure of the source language word in the target language word: **kantokyky** (Finnish) - **bärkraft** (Swedish) - **Tragvermögen** (German).

Acronymization is yet another way of creating new words and terms. The acronymic words or terms are a combination of the letters of several words in a compound word or in a phrase. Normally the letters are picked from the beginning of the words, but sometimes the new term requires the use of other letters as well. For example, the acronym **PDF** stands for **P**ortable **D**ocument **F**ormat, or **SIMM** as **S**ingle **I**nline **M**emory **M**odule. In Finnish, for instance, many of the acronymic terms come from the English language.

Conversion, on the other hand, is the process when the word class of a word is changed. This method is very common in English language, where one part of



speech changes to another, and in fact, this is known as the *functional shift* (Pyles 1993: 281). For example, when a word has been used as a verb, it is now used as a noun (or vice versa), or an adjective has come to be used as a verb. In the functional shift, the form - at least not in English - of the word does not change: to **trouble** (verb) -> **trouble** (noun), **dull** (adjective) -> to **dull** (verb).

### 3.2 Term Base Structure

As discussed briefly above in the present study, definitions are an important part of the structure of a terminological database. They can, in fact, define the overall quality of the database, and if the definitions are incomplete, incomprehensive or totally absent, the usability of the term bank decreases significantly. Therefore, depending on the effort put in at the creation phase of the term bank, the users can either benefit from the use of the term bank or find it insufficient for their purposes. That usually manifests as the need to use other means to get the required information by using other dictionaries, terminological databanks, or resorting to expert advice, which might add to the overall costs of a project.

There are many ways to display definitions, as there are definitions themselves. One way is presented by Strehlow (1997: 75), as he explains the usefulness of two-dimensional presentation of terms and their definitions. Strehlow claims in his presentation that the structural relations within any definition can be seen and understood promptly. Moreover, if terminological definitions are displayed consistently, analyzing them will be easier. The consistency can also improve the efficiency of standardizing the terminology as well as better the actual management of terminology. This can be considered as one of the cornerstones of a good term bank.

#### 3.2.1 Definitions As Part of Term Base Structure

There are many different ways to display definitions, and the examples given here follow those of Strehlow's (1997: 76—79). The format of definitions used in different occasions depends largely on the purpose of the terminological database. That purpose dictates which type of definitions are used and which are not. In the following picture the definition of the term **lion** can consist of several things which elaborate the substance and characteristics of the term. Normally, the information is presented in a *hierarchical* order where a larger class is shown first, and the

subclasses are shown consecutively, if they exist. Apparently this order is more suitable, since larger classes can only contain co-existing subordinate classes, but subordinate classes can hardly have more than one superordinate class. One reason for presenting information in hierarchical order is to help the users of term banks to understand how terms lie in larger contexts, and what properties the terms may have. That, consequently, can facilitate the understanding which terms may belong to the same classes, and so on. As the present study concentrates more on the aspects of technical terminology rather than natural sciences, this section will only elaborate the structural framework of definitions. The definitions and writing of technical terminology will be discussed later.

Strehlow uses five elements that constitute the definition of **lion** (see Fig. 3). They are *genus*<sup>19</sup>, *subgenus*<sup>20</sup>, *species*<sup>21</sup>, *differentia*<sup>22</sup>, *property*<sup>23</sup>, and *incident (or accident)*<sup>24</sup>.

78

**PRINCIPLES OF TERMINOLOGY MANAGEMENT**

<b>Term</b>	<b>:</b>	<b>lion</b>
<b>Genus</b>	<b>:</b>	<b>mammal</b>
<b>Sub-genus</b>	<b>:</b>	<b>cat family</b>
<b>(Species)</b>	<b>:</b>	<b><i>Panthera leo</i></b>
<b>Scientific name</b>	<b>:</b>	<b>large</b>
<b>Differentia</b>	<b>:</b>	<b>carnivorous</b>
		<b>tawny color</b>
		<b>mane</b>
<b>Incidental</b>	<b>:</b>	<b>mane frequently</b>
		<b>darker</b>

Figure 3: *Frame Representation of the Definition of "Lion"*

Picture 2: Handbook of Terminology Management, vol.1

These elements - or aspects - represent first the hierarchical information that puts the term in some sort of a category, and presents the possible subcategory/subcategories that belong to the main category. In this case, the main category is **genus**, and the subcategory is **sub-genus**. Where technical terminology is concerned, the main category could be **naval engineering**, and the subcategory **structural engineering**. The term in question could therefore be, for example,

**ballast port** (*examples mine*). Ballast is extra weight required in a ship to improve the ship's stability and steerability, and to make it float in a certain depth. In earlier days, for example, sand was used as ballast. The ballast port, then, is the port through which the ballast is loaded and unloaded.

The differentiating aspects, on the other hand, define the term in terms of its characteristics. Moreover, the incidental (or accidental) information is sometimes given to present the properties that do not necessarily apply to all members of the subcategory. The **differentia** of a lion shows what the animal is like compared to other animals in the vertebrate family: it is carnivorous - compared to some other mammals; it is large - compared to some other members of the cat family, and so on. The **incident/accident** variation is a property that has been observed in several cases, and therefore it can be included as valid information. The differentia of the **ballast port** could contain information that defines its look, size, material and actual placement in a ship. The incidental/accidental information could probably contain information on how the design of the ballast port is carried out in different engineering solutions.

In general, the model that Strehlow presents can be used as a template for almost any terminological definition. The next chapter will start to look at the more recent research on terminology. The theoretical background is based largely on the 2<sup>nd</sup> volume of the *Handbook of Terminology Management*. As regards the purpose of the present study, it is important that the emphasis moves toward the actual implementation of different systems of maintaining terminology.

### 3.2.2 Data Categories for Terminology Management

The terms, as mentioned earlier, are the building blocks of terminology databases. When a term is added in the database, it becomes a *terminological entry*. The entry, however, is much more than just the term. The whole entry, as will be seen later, is a combination of the term, its definition, perhaps the equivalent term in a target language, and so on. For example, one Microsoft's<sup>2</sup> software *glossary*<sup>25</sup> (a glossary is usually a collection of special vocabulary limited to one field in a list format) converted onto MultiTerm™ contains over 42 000 entries, and there are many other glossaries/term banks that can contain hundreds of thousands of entries.

Therefore, it is not a trivial thing to try to present the data in a readable and comprehensible format. The data has to be divided into categories, and each of those categories should have a logical place in the database, they should have a purpose, and the number and format of the categories should be carefully controlled in order to form a database with “no loose ends”. I intentionally use the word *should*, since many term banks do not meet these requirements. In the survey for the present study, questions will be asked whether the term banks that the respondents use really are logically constructed, whether the data categories serve the purpose they were designed to serve, and whether the number of entries is sufficient or not - or whether the terminology database is overloaded with non-essential information.

Wright (2000: 565) makes a distinction between *data fields* and *data categories*, although in colloquial use the word *fields* is used quite freely, and could thus be easily mixed with *data categories*. Moreover, Wright states that the individual units of data that are included in the database are called *data elements*. Data elements, in turn, are contained in data fields. The combined data fields form a data entry, or a terminological entry. The data categories used in a specific database depend on the use of the database, and their number or combinations are not constant. The *International Standardization Organization* (ISO) has formed several standards for the data categories used in terminological databases. In fact, the ISO standard 1087-2:1998 defines data category as “a class of terminological elements...viewed from a functional standpoint as instructions *for interpreting a given data field*.”

### 3.2.3 Data Category Specifications

One requirement of a data category is to protect the solidity of data in the databases, that the category has to be clearly specified, and that the people who add data into the databases follow the specifications given for the data category. Each data category has to have an individual name that should be used globally in the specified terminology management system. If the larger database is later updated or complemented with smaller terminology “blocks” one has to ensure that the additions do not use a different naming system for the same data category. If the

---

<sup>2</sup> Microsoft Corporation is one of the world's largest computer software manufacturers. Some of Microsoft's well-known computer programs is the operating system Microsoft Windows™.

components of the terminology database do not conform to the overall specifications, the integrity of the terminology database is compromised.

The definition of data categories is considered essential, since different terminology databases can handle the same data categories diversely. Therefore, the data entries should be consistent with the descriptions of data elements. Information in some databases may be divided into more subtle categories than in others, i.e. one database may work, for example, with only five categories whereas another database uses ten. Wright (2000: 569) states that this fluctuation in the division of information is called *granularity*<sup>26</sup>. This means that the smaller the “grains” of information, the greater the accuracy of data. In another words, if the data categories are presented too broadly, not all the vital information that should be contained in the entries is present when needed.

If there are only two data elements instead of, say, seven, in every whole entry, and those two elements must contain all the information that is needed to form a valid entry, the element size may grow too large. If one element consists of several lines of data, for example, a job applicant’s job resumé (a general example), it might be difficult to find a certain piece of information from among the whole lot of information. Moreover, if there is a database with several applicants for a job, and one wants to sort out only those applicants that have a college degree, the granularity is too large, and it prevents the precise manipulation of data. Where terminology databases are concerned, the difference in the grain size could be presented as follows (see the Tables):

Data element	Content
Term Definition	eyepiece a part of an optical instrument closest to the eye of a person, that he/she looks through, made of glass or plastic. Eyepieces are used for example in binoculars and microscopes.

Table 1: Granularity/Less granularity

In the Table 1 above (*example mine*), all the information on the term is put in a single large “container”, where a particular piece could be hard to find. However, when the content is split into smaller pieces, and indicated by additional data elements, the accuracy of database searches and of sorting the data, improves.

Data element	Content
Term	eyepiece
Purpose	a component in binoculars and microscopes
Material	glass or plastic
Physical location	closest to a person's eye through which he/she looks
Reference	Zeiss Z-ELM1000 Microscope User Manual, page 193. (an imaginary example)

Table 2: Granularity/More granularity

Wright also indicates that if there is too much repetition of information in the database, that is, if for example, the references to a source are included in every entry, it will consume much storage space and requires more from database management. Therefore, references like the reference "Zeiss" in Table 2, should appear only once in a bibliography, and even more so if the reference is large. There are also other ways to refer to the source of information, namely *links*. Links work as addresses that are accessed from the database or from outside the database, and they could be WWW site locations, links to other databases etc. The important thing about links is that instead of including all the source information in the terminology database with the entry, there is only a link to a *shared resource*. A shared resource could be, for instance, a large video archive or a separate database on a computer server.

### 3.3 Standards and Interchange

In this chapter, before I move on to discuss the present survey, I will briefly introduce some of the standards in terminology management and data interchange. Standards in management and interchange are vital, since without them there would be literally a jungle of different term bases, and these term bases would not be at all compatible. This may be the case even with standards, because only a fraction of all the term bases available today use the same standards. However, standards help in maintaining a logical structure of data, accommodate the transfer of information between different systems, form a level of unity between these systems and so on. I will give a brief description of some standards and, in a way, their sub-standards. The standards are presented in the following order: *ISO*, *SGML*, *MARTIF*, and *NTRF*.

### 3.3.1 The ISO Standard

One of the standards that have been formed to accommodate the needs for consistent terminology is ISO (*International Standardization Organization*) 12620. This standard does not, according to Wright (2000: 567), aim to standardize what category names should be used in different term banks, but rather to give guidelines about which category names should be used when the data in a certain terminology management system are transferred to another. The ISO 12620 requires that the data category names that do not comply with it should be converted to standardized names before the interchange, that is, the transfer from one system to another. The standard (ISO 12620:1998) is divided into three categories, which represent the most common types of information found in terminology databases. These categories include the term, the data related to the term, descriptive information related to the concepts from which the terms arise, and relevant data for term base administration (Wright 2000: 573).

The term-related information is further divided into smaller groups that help to identify the term. The identification could indicate, for example, whether the term is a phrase, a symbol, an equation, a phraseological unit (a noun, verb, adverb etc.), etc., that is, give grammatical information about the term, probable pronunciation guidelines, administrative information and so on. Where multilingual term bases are concerned, there are additional fields that present the source term in its target language form, its pronunciation, probable hierarchical relations, i.e. indicate whether the source/target term is a synonym, hypernym, or antonym etc.

The maintenance of multilingual term bases takes much more effort and requires more accuracy than the maintenance of monolingual term base. Therefore it is also vital that the target language representation of a term, its definitions, concepts and the like are at least as extensive as the source language representation. The ISO standard aims to ensure this as well to a certain degree.

### 3.3.2 The SGML

The letters SGML stand for *Standard Generalized Markup Language*, which is a standard that defines the structure of texts in a standardized presentation format. The SGML is based on the ISO 8879 definitions. This markup language uses special codes to carry information on the structure of the text, for example, its

headings, paragraphs, styles, formatting, and so on. (Jaakohuhta 1999: 773). The SGML format resembles the probably more familiar HTML (*HyperText Markup Language*), which is used in Internet documents. By utilizing so-called *tags*, the SGML can describe large structures in a text or even the smallest bits: the headings, subheadings, body texts, etc. or whether a word is a noun, an adjective, a verb, or whatever. The tags could look as follows: **<H1>**, **<H2>**, **<font>**, or **<Source>**, **<Target>** or **<termEntry>**. The purpose is that whenever a terminology database is constructed the standard tags are used. Then, when the information is transferred elsewhere, all terminology management systems that conform to SGML standards will have the data in exactly the same format, and there is no data loss due to incompatibility.

### **3.3.3 The MARTIF Standard**

The acronym MARTIF comes from the words *Machine Readable Terminology Interchange Format*. This standard is also an ISO standard (ISO 12200), although in many instances it is referred to by its acronym. The MARTIF standard was developed because the need to give terminology in electronic form has grown rapidly, and continues to do so. Moreover, this standard concentrates on presenting term data in a format that is readable by as many machines/computers as possible with as little incompatibility as possible. The MARTIF standard is also based on SGML, but it introduces a whole new set of features that were not originally in SGML. As MARTIF is purely a terminology interchange format, whereas SGML is used for other purposes as well, MARTIF uses indicators that are found only in MARTIF itself. However, these indicators, which are called *generic identifiers*, conform to SGML standards. For example, identifiers like **<term>**, **<termNote>**, and **<langSet>** are common to MARTIF.



The structure of a terminology entry, and the whole database is built of these small identifiers or elements. Larger elements can hold smaller elements, thus producing a hierarchy or a tree with several branches, that is, superordinate elements embed subordinate logical elements, and the result is a nested structure. The following example shows the actual structure of a MARTIF standard entry as it would appear if the source language were English, and the target language Finnish. The original version is from Melby (2000: 639).

```

<termEntry>
<adminGrp> <admin type='responsibility'> ISO TC 61, Plastics
</admin> </adminGrp>
<langSet lang=en>
<ntig>
<termGrp><term>thermal degradation</term>
<termNote type='partOfSpeech'>n</termNote></termGrp>
<descripGrp><descrip type='definition'>the entirety of all dele-terious
chemical modifications of plastic at elevated temperature
</descrip><note>It is essential to report the temperature and other
environmental conditions at which the phenomenon is studied.</note>
</descripGrp>
<ref type='relatedTerm' target='en47222'>aging</ref>
</ntig>

</langSet>
<langSet lang=fi>
<ntig>
<termGrp><term>l&auml;mp&ouml;hajoaminen</term>
<termNote type='partOfSpeech'>n</termNote>
</termGrp>
<descripGrp><descrip type='definition'>kaikki vahingolliset kemialliset muutokset muovissa
kohonneissa</descrip>
<note>On t&auml;rke&auml; ja&auml;, ett&auml; ilmoitetaan l&auml;mp&ouml;tila ja muut
ymp&auml;rist&ouml;olosuhteet, joissa ilmi&ouml;t&auml; tutkitaan.</note>
</descripGrp>
<ref type='relatedTerm' target='fr47222'>vanheneminen</ref>
</ntig></langSet>
</termEntry>

```

Original: Melby 2000, Adaptation mine.

I will explain only some of the tags found in the above table. The first tag *<termEntry>* is the superordinate element of the whole set, under which all the information relating to the entry is gathered. As we can see, the set ends to

`</termEntry>` tag, which is the ending tag. There must be both a starting and an ending tag. The `<adminGrp>` holds information of the field the entry is related to, which in this case is **plastics**. The `<langSet>` indicates the source and target languages of the term, and the `<ntig>` tag means *nested term information group*. Under the `<ntig>` indicator, all the information concerning the term (including the term) is placed. In the above example, the relevant information is the part of speech the term represents (*n* for a noun), and of course the term itself. This information is further nested inside the term information group by `<termGrp>`. The data category for a definition is presented by a description group `<descripGrp>`, under which the definitions in both the source and target languages are given. The `<langSet>` tag indicates the languages which are used in the entry.

As the SGML/MARTIF standard intends to be universal, it cannot hold special characters such as alphabets that, for example, are found in Finnish, Swedish, French or German. Therefore, these “special” letters have to be presented by a code that is understood by different systems regardless of their country of origin. The Finnish definition for “thermal degradation” as well as the target term looks strange with the letter codes (`&auml;` or `&ouml;`), but the final layout in the term base will look normal.

### 3.3.4 The NTRF Standard

The NTRF is a terminological standard especially intended for Nordic use. In fact, it means *Nordic Terminological Recording Format*. As the name implies, this standard has been developed in cooperation with terminology institutions of Finland, Sweden, and Norway. In spite of the fact that the NTRF is based on the SGML standard, it is rather different in its appearance. It is important to note that although the standard in question has been adopted in Finland for example with The *Finnish Centre of Technical Terminology* and as an official standard for terminology work, the NTRF is still apparently quite narrowly used. One terminology management system, namely Kielikone Corporation’s *TermIt*<sup>TM</sup> uses NTRF as its standard for term management and production.

The NTRF standard aims to make terminology management and term presentation *concept-oriented*. This means that the main emphasis is on concept systems, and one term corresponds to one concept, and consequently to one object. There is a difference between this concept-orientation and *text-oriented* (or term-oriented)

terminology work, in which the term is important from a translation point of view. The NTRF standard differs slightly from the actual format the TermIt™ terminology management system uses, namely in the subordination of different fields (data categories). According to TermIt's accompanying documentation (in a file), only certain fields in NTRF can be subordinated to each other, whereas in TermIt™ any field can exist inside another field, thus allowing more extensive nesting of information.

In my survey I will investigate which terminological standards each system has applied, what standards are followed, if any, and whether the users have found those standards inadequate or sufficient for their terminology work. In the above light, one could ask whether the relationships between different data categories allow efficient input and retrieval of information. The retrieval could be, for instance, sorting out segments which fulfill certain criteria (only synonyms for any term, terms that deal with economy, etc.), and perhaps forming a new database solely consisting of synonyms.

In the following chapter I will introduce some theoretical points of questionnaire research, questionnaire implementation and the methods used to gather the data for the present study. The subchapters will deal with questionnaire reliability, representativeness, and validity; the designing of a questionnaire in general, and especially for use in the Internet; some common rules of thumb on preparing a questionnaire, and so on. Almost any guideline on making a questionnaire can be taken into consideration even though the means by which a questionnaire survey can be launched might have changed over the decades. By this I mean that nowadays surveys are conducted increasingly in an electronic form, and not by sending out tons of letters to people of whom we are not even sure whether they receive or open the letters at all.

#### **4. A Questionnaire for Gathering the Data, and the Preparation of the Survey**

Questionnaires have been important tools in getting reliable information on various subjects, and they have been used for decades to collect people's opinions and preferences. I decided to use a questionnaire to get information on the usability of terminology management tools and specific terminology from different fields. Most of the questionnaires used in similar research have either been distributed by

mail or by a direct contact with the respondents. However, conducting a mail-based survey would be a both time- and money-consuming task. An electronic questionnaire that one can fill out any time and send back right after filling it is both a reasonably fast and a convenient way to get the information one wants. However, the data must be collected, sorted, analyzed and presented in spite of what method is used. Therefore I decided to construct an electronic questionnaire that can be filled in on the Internet. In that way the respondents can easily access the questionnaire form as well as fill it in and send it back with no extra cost and effort. In the following paragraphs I will describe the questionnaire itself, and also present some background information for this type of research.

#### **4.1 Questionnaire Reliability, Validity, and Representativeness**

One cornerstone of a functional questionnaire is the correct formation of the individual questions. The respondents should be able, as well as willing, to give answers to all of the questions in the questionnaire. The ability of the respondents to answer the questions stems from the assumption that the respondents have at least some kind of knowledge on the things that they are asked about in the questionnaire. Simply put, if a questionnaire regarding computer software is sent to milk farmers, for instance, it undoubtedly will produce very distorted data if any data at all. Therefore, it is important to devise an appropriate questionnaire for a proper purpose and audience. By appropriate I mean that if, for example, the Commissioners of the National Pensions Institute in Finland wanted to learn more about the quality of health care services, they do not send a questionnaire to someone who has not used health care services, and therefore is not able to answer a question "*How do you rate the proficiency of the nursing staff at your local health center?*" or some equivalent question. To argue this, the respondent might be able to answer that question of course, but his/her response is not based on facts or concrete evidence about the skills of the nursing staff. This means that the answer is not reliable.

Berdie (1986: 3) attests that in order to be *reliable*, a questionnaire has to be made of reliable items. According to him, a reliable item is one that every survey respondent will interpret in the same manner. If there is a possibility that any given question will not mean the same every time, or for every person, the question is not reliable - or at least the answers to that question. In other words, the same question

should be interpreted in the same way every time in spite of the person who reads the question.

The *representativeness* of a questionnaire survey, according to Berdie (1986: 4), means that the respondents who actually return the questionnaire and who have answered to all of the questions within, represent the general public. For example in Finland, a commercial television station may conduct a survey of 1,000 respondents. Afterwards, the station claims that the answers of those thousand people represent the whole nation. Basically, the purpose is to get a sample of people that could represent the population structure of the country. On the basis of the answers by those 1,000 respondents analysts deduce that whatever results arise, they are, with certain variation, the opinions of entire population. However, the representativeness of surveys also depends on the audience they are directed towards, as well as on what kind of data they are intended for.

The questionnaire survey for the present study is intended for a specific group of people, and it does not have to represent the opinions of a larger population. The purpose is to get data that represents, at least to some degree, the opinions of a selected group of professionals. These professionals consist of technical writers, translators, terminologists and so on.

The reliability of a survey means that a measurement will produce results that are not random (Mattila 1999: 60). That is, the results do not change if the same people are asked the same questions several times. The reliability of a questionnaire is the greater the less randomness its results have. By this I mean that if the questions are answered repeatedly in the same way, it is safe to conclude that the answers are reliable and not arbitrary. Nevertheless, there will be certain randomness in every survey. This randomness can stem from sporadic variations in the respondent's state of mind, health, and memory. These fluctuations do not distort the results of a survey, however, since different respondents will eventually even out each other's contribution to the error factor of the survey. Mattila (1999: 63) puts the validity of a survey as follows: "According to a general definition, the validity of a measurement describes how well the instrument used in a survey will measure precisely that concept that it is intended to measure." Sometimes the fact whether a measuring instrument (an individual question, for example) is functional or not can be discovered only after all the responses have been analyzed. If a question

produces unexpected answers, answers that cannot be comprehended, or no answers, the reason for that may lie in the fact that the instrument/question is not functioning the way it should. Let us say that the concept that is measured is customer satisfaction in a department store's sports department, and the information that is wanted deals with the salesclerks' willingness to serve. If, however, the questions are not specific enough, all types of answers might arise. One might ask: "What do you think about the sports section of our department store?" but fail to specify the question particularly to deal with the salesclerks. Thus, the measuring instrument will not fulfill its purpose.

#### **4.1.1 External and Internal Reliability**

Furthermore, the reliability of a survey can be divided into *external* and *internal* reliability. One of the reasons that affect the external reliability is whether the expected respondents respond at all. That may lead to a situation where not enough data could have been gathered for any tenable conclusions to be drawn from it. An ideal situation would be if the response rates exceed 90 per cent. However, in many cases, the response rates are much lower. If the same questionnaire is sent repeatedly, maybe enough answers will come back. On the other hand, if the questionnaire does not produce enough data on its first round, the blame may not be on the respondents but on the questionnaire itself.

Mattila (1999: 52) states that the topic of the survey also contributes to the low response rate. When the topic is interesting and important, the response rates tend to grow. The topic of my survey can be assumed to be both interesting and important, since there is a good chance that the data that arises from the survey will help in developing better terminology databases and terminology management. At worst, the results of this questionnaire survey will give a general view of the state of terminology management in Finland.

The internal reliability of the survey material depends on how accurate the answers from the individual members of the *sample*<sup>27</sup> are (Mattila 1999: 59). If the respondents give ambiguous answers, answer incorrectly, or answer deliberately differently than expected, then the material does not describe itself accurately. In statistics, however, a certain amount of acceptable errors is always calculated in. When the errors stay within set boundaries, the reliability of a survey can still be attained. By this I mean that there are always some errors in any data, but if the

errors are not significant, that is, if there are slight variations in answers here and there, they do not affect the final results of the survey.

## **4.2 Designing a Questionnaire Survey**

There are numerous points to consider when preparing a questionnaire survey. One of these points is to make clear who actually needs the information that the questionnaire survey produces: is it a company, an educational institution, the media or what? Any survey is inconsequential if the data have no meaning to anyone. The information to be gathered has to have a purpose, and the purpose of the present study is hopefully to help improve the standard of terminology management and interchange, produce better term banks, and help terminology users understand the importance of a uniform and consistent terminology management practices.

### **4.2.1 Some Rules To Remember**

Berdie (1986: 6—11) presents a five-point list whose different points are intended to help in designing a workable questionnaire. I will approach few of these points here. The first point in Berdie's list is *to know one's topic*. One has to be fully familiar with the topic of one's study underway before starting the actual construction point of the questionnaire. The second point is *to know one's people*. Berdie recommends that one should get input from potential respondents or other people that have knowledge about the subject of the survey. The third point in Berdie's list is apparently the most painstaking one, as it includes four things to be considered. The first thing is the consideration of money. The money issue becomes important when the preparation of a questionnaire survey includes much photocopying, mailing, printing, and so on. The costs of a mail-based questionnaire survey are, almost without exception, always higher than the costs of a web-based questionnaire. Nowadays the costs of almost any survey can be reduced to some degree by the help of electronic media, documentation, mail, and other forms of electronic communications. Setting up an internet page with a questionnaire, and informing relevant people about its presence, is much more inexpensive than sending out hundreds or thousands of paper questionnaires. Moreover, if one wants to maximize the probability of a high response rate, one may have to send a letter to the possible respondents prior to the actual questionnaire stating the purpose and function of the questionnaire and asking if the information gathered can be used,

and so on. Furthermore, a second questionnaire might be in order to ensure the fullest possible feedback. Of course, one need not rule out the use of preliminary announcements in the case of an electronic questionnaire, either. The simplest way would be to roughly determine which people should receive the actual questionnaire, and to send them either e-mail or notify them by some other means, for example by phone.

The third point, which is divided into four subpoints, deals with the issue of “*how much?*” The first subpoint is about money, the matter I already discussed above. However, in a general sense, financial issues can steer a research significantly. Therefore, a larger survey will almost inevitably involve larger sums of money. When university students conduct surveys, it is often the case that there is a very limited amount of money that can be used. Sometimes a company that could benefit from the results of the survey may offer to share the costs or take care of them completely.

The second subpoint deals with *time*. Berdie (1986: 8) has noticed that in many instances the whole study has not fitted inside the time frame that may have been set for it. He recommends that the best way to keep the time frame in realistic proportions and make the study fit in it is to work backwards from the expected completion date. That way one should be able to see if the separate stages of the study will actually be ready when they are intended to. Finally, one should decide how much work one should do to achieve a high response rate, or whether the amount of work is worthwhile. One should also be prepared to analyze the situation and act accordingly if the response rate is not what was expected.

### **4.3 Developing and Testing the Questionnaire**

Oppenheim (1966: 24—) also has some valid issues to consider when making questionnaires and conducting surveys. One of these issues is *pilot work*. He sees that most part of the pilot work falls to dealing with problems that the researcher thinks to be the most difficult to solve or settle from the beginning, as well as to things that are the most important for the survey as a whole. The pilot work in the present study is mainly a small-scale testing of the first draft of the questionnaire. The aim is to find out if there are any trouble areas in the questionnaire which may affect the outcome, or if there are some things to add or to leave out.



As the structure of the present questionnaire is concerned, it is, of course, made of a series of questions. Some of these questions are grouped so that they present a distinct section that can also be analyzed as a whole in addition to its individual questions. Oppenheim (1966: 37) suggests that a questionnaire should be both attractive and interesting to the person who is responding to it. He is right, since a questionnaire with a poor layout, arbitrary positioning of questions, insufficient filling instructions and so on, will most likely have a direct effect on the response rate. Pilot work is important in this case also, since the test subjects who see the first questionnaire may tell that the questionnaire is perhaps too short, long, vague, detailed etc.

#### **4.2.1 Questionnaire For the Web**

Designing a questionnaire for the web (or Internet) differs somewhat from the design of ordinary questionnaires. A printed questionnaire is a concrete form with one or more pages. When a person writes down answers on a paper questionnaire, the answers remain there unless someone wants to erase them or the paper gets lost, burns, or the proverbial dog eats it. The electronic questionnaire is more vulnerable to carelessness, unforeseen errors in soft- or hardware (computer programs, computers, network etc.). Most *forms* that can be viewed and accessed via Internet and a *browser*<sup>28</sup> work as follows: there are normally two buttons of which one sends the questionnaire as e-mail or by other means to the recipient who is in this case the person who collects the information from the questionnaires, and the other clears the form completely. If a respondent mistakenly sends out an incomplete questionnaire, the answers received from that questionnaire do not count as valid since all the required questions were not answered. Another fatal mistake would be that the respondent presses the button that resets the form, and all information will be lost. The only solution in both cases is to fill in the form from the beginning, which might be a frustrating task if the questionnaire is long, and the respondent has not sketched down his/her answers on a paper.

However, the electronic questionnaire is both relatively easy and a fast way to gather information from different sources. One does not have to send out letters, but rather to inform the respondents about the location of the questionnaire, and give instructions how to access it. It is relatively safe to assume that most people that work closely with terminology in some way will also have an access to

Internet. Moreover, the responses arrive by e-mail, and they are in a text format. The individual responses can be put in a table for comparison and analysis.

## **5. The Survey**

As mentioned earlier, it is practical to have any survey tested before it is really launched. Testing can be seen as the pilot for the survey, and during the testing phase the tools of the survey are fine-tuned to work optimally for a particular purpose. The overall analysis of the survey will be made using qualitative approach, although more solid conclusions could have been drawn if the responses had exceeded the meager 10.

### **5.1 The Qualitative Approach**

Qualitative data cannot be analyzed by statistical tools. Some of the questions in the present questionnaire will not produce numerical data, and therefore those questions have to be scrutinized in a different way. In fact, the number of responses, as mentioned earlier, was only 10. Therefore no statistical analysis is possible or purposeful. Showing percentages of the different variations to some of the preformed questions or any quantitative method would bear no significance. If the responses had arrived in tens or in hundreds, then the quantitative approach had been the only choice. Hopefully, the quality of the responses will, in the present study anyway, compensate for the quantity.

#### **5.1.1 The Number of Responses Received and Their Effect on the Analysis**

The respondents had two months to return the questionnaire. The date was set to 28<sup>th</sup> of February 2001. By that date, not more than 10 responses were received. That is not enough for statistical analysis, and it was disappointing to find out how little the survey really interested people. There could be many reasons for the lack of answers, and I will be considering that in more detail in **Chapter 7**. Therefore, I will go over every response individually, choosing the sections that have been answered, and which have relevant data. In the questionnaire surveys such as the present one, the statisticity of data is normally better, because the data lends itself to a variety of different analysis methods and ways to interpret it. I will, however, present the responses in some points in a list format, and then go over them more or less detailedly.

## 5.2 The Actual Questionnaire

The actual questionnaire consisted of 25 questions. Some of the questions, as mentioned, were open questions, so for those I cannot use statistical analysis. However, the answers to those questions will be dealt with in the present study to summarize similar or closely related answers. In the next subchapters I will introduce the questions, their answer possibilities, and their purpose in detail. During the testing phase I received some comments on how to improve and alter the questionnaire, and I also found some errors on the questionnaire that I had to correct before the survey was launched. The alterations were minor, and the errors were not significant, either. The most arduous task was to get some functions of the electronic form to work properly, and the feedback from the test questionnaire involved changing the wording of some questions.

### 5.2.1 The General Questions

The first five questions were general questions that defined the status of the respondent, his/her job description as well as some background information. **Question no. 1** asked whether the worked for a company or whether he/she was a private entrepreneur. If the response to **Q1** (questions and their respective numbers will later be referred to as Q1, Q14, etc. when appropriate) was *"I am working for a company"*, the responded had to answer to **Question no. 2**: *"If you work for a company, what is the line of business of your company?"* The response could be written as a free response, for example, *"network communications"*. The line of business largely defines what type of terminology the responded uses, so this question is relevant. Next, **Question no. 3** actually consisted of two questions: the first part asked, *"Which of the following best describes your own job, or occupational title?"* The preset answers were *a technical writer, a translator, a database administrator/administrator (in general), a terminology worker, a freelancer (any type), in a managerial position (for example, a professor), and a consultant*. In addition, one could also write another title in a given field, if none of the set choices were suitable.

The second part of **Q3** asked *"Do you construct terminology for your own purposes or for a client?"* Possible answers to this were obviously *for my/our own purpose* or *for a client*. **Question no. 4** inquired *"How often do you come into contact with terminology?"* For this question there were four possible answers to choose from:

*often, fairly often, seldom, and not ever.* The boundaries between *fairly often* and *seldom* are not very clear-cut, but one can hopefully deduce that *fairly often* is nevertheless more than *seldom*. In these types of questions it is difficult to use time as an indicator, for example, *once a week, every two months, once a year* etc. It is easier to just put it into words rather than count every instance one actually uses terminology. If the respondent had replied "*not ever*", that would not have prevented him/her from answering some of the questions later in the questionnaire, though. Finally, **Question no. 5** listed the possible languages that were used in the line of business of the company or in the work of a private entrepreneur: "*Which languages do you use/your company uses?*" Multiple languages were possible to choose from the given list: *Finnish, Swedish, English, German, French, Spanish, Russian, Chinese, Turkish, and Dutch.* It was possible to give another language as well.

### 5.2.2 The Specific Questions

Questions **6** to **24** were more specific, and the responses to those count as the core results of the survey. In **Question 6** the following was asked: "*What terminology management programs do you use/your company uses?*" Moreover, I wanted to know what publicly accessible terminology data banks the respondent/the company uses. The answers could be given in the form of a free-text list. Examples of the terminology management programs were TermIt, and MultiTerm, but other programs were possible as well. The public term bases could have been, for example *Euterpe* (the extensive term bank of the European Union), *Eurodicautom* (also a European data bank), and *Tepa* (a Finnish term bank). The answers could be given as a free-text list, and the number of individual programs was not limited.

Until now, the questions had not involved any subjectivity. **Question 7** was "*What pros and cons does your system (the term management system) have?*" The cons might have been, for example, too complex user interface (difficult to use) or insufficient instructions, whereas the pros could have been, for example, fluent term management at the program level (In this instance I mean that the user does not have to use any special tools to manage the terms, or to resort to script-

writing<sup>3</sup>). The respondent could again answer by just listing the good and not so good properties of the terminology management system.

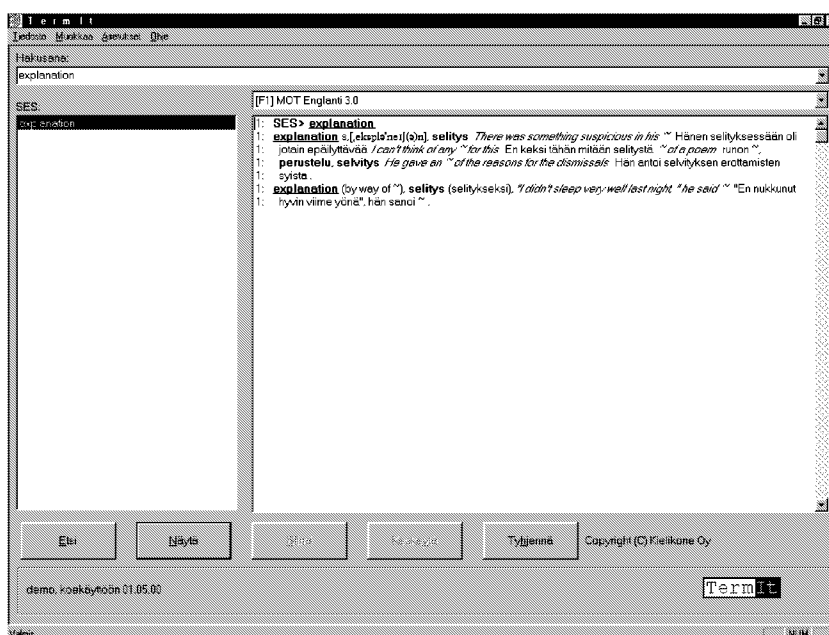
Although the free-text responses are harder to put into a statistical form, one can still hopefully find correlating answers, and thereby present them in numbers and show that, for instance, 23 of 76 respondents feel this or that.

**Question 8**, on the other hand, had only preset answer possibilities. In **Q8** I wanted to know about the overall usability of the programs. The respondent could list a total of seven different programs in their respective fields, if necessary. The usability could be graded as *very good*, *good*, *average*, and *poor*. As there are many aspects that contribute to the total usability of terminology management programs, **Q8** was not in itself sufficient to chart the properties of any program mentioned in the responses. Therefore more detailed questions were in order. However, the responses to this particular question depend mainly on how the individual user has experienced the programs he/she has been using or has used but does not do so any longer.

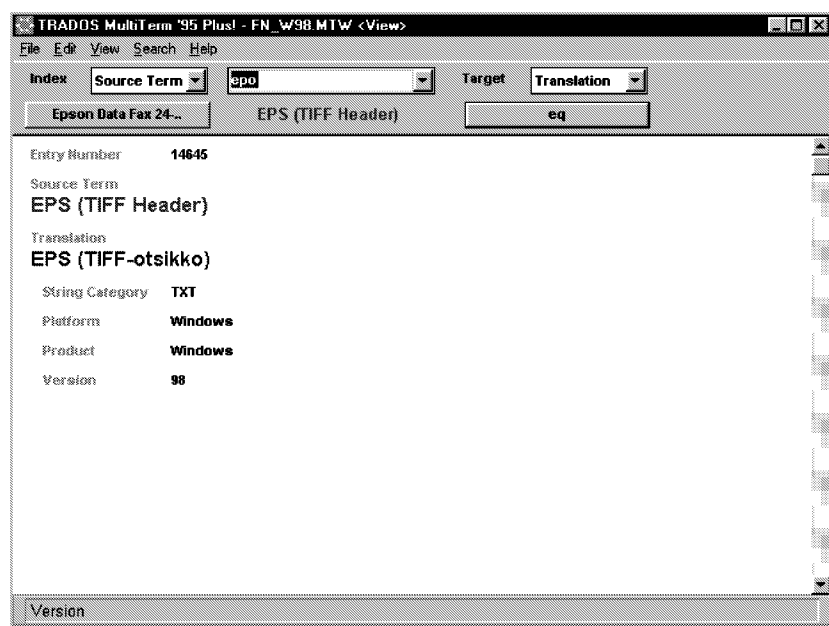
**Question 9** was as follows: *"Do you use your own models to make entries and categorize terms, headwords, definitions etc., or do you use the models that already exist in the terminology management program?"* This means that some, if not all, TM programs have preconstructed models at least in how the terminological entries are presented visually in the program. To help the respondent see what I meant I chose to include a picture of the interface of the TermIt™ program, and more specifically, the screen where the entries are given. In TermIt, the constituents (or data categories) of an entry are placed simply one upon the other in separate "boxes", whereas in MultiTerm, for instance, the whole entry shares the same screen. The data fields separate the source language term from its target language term, definition, synonyms, etc. in their own groups (see the pictures below).

---

<sup>3</sup> Scripting can mean here, for example, writing a set of instructions in the form of a small program that are then interpreted by the computer and by this being able to manage the terms in the database (add, remove, tag (=identify) etc.)



The picture on the left represents the structure of a term entry on Termit. There is a separate window for adding and modifying the entries, but this picture from the user interface. On the topmost field is a headword, and below that in a larger window are all the possible matches against the headword. On the right appear the target language translations for a term/word. All the data is represented top-down, with possible references etc.



This picture is from Trados MultiTerm, and we can see how different the user interface is. The small fields on the upper portion of the interface indicate the Index and Target fields that could include, for example, a source term, and a translation, as it is in this case. However, the Index and Target fields can contain other data categories as well. The buttons below the fields are for moving back and forth in the database, and the large window is reserved for the display of terms, their translations and other information. Contrary to Termit, different data categories can be indented to make the whole entry more readable.

Top picture (Picture 3): Courtesy of Kielikone/MOT Sanakirjasto

Bottom picture (Picture 4): Trados MultiTerm/MS Windows 98 glossary/Microsoft Corp.

From **Question 10** onwards I wanted to find out more about the actual structure of the respondents' term banks. As different purposes dictate the actual number of fields, or data categories, within an entry, and different companies may have their own needs for the number of separate fields, **Q10** asked how many fields there are in a normal entry. The fields could be, for example, *source language term*, *target language term*, *definition*, *context*, and *reference*. There may be many more fields; the fields can be named differently, and so on. One contributing factor to the compatibility of different term banks, or their contents, is the number of fields and their naming conventions. The respondents had to indicate the field count in one box, and give a list of other possible fields. In the questionnaire I talked about "compulsory" fields, which are the most necessary. Other fields can be added if the entry so requires.

One feature that also defines the usability of a term management program is how well the entries, or more specifically, the fields, can be added, removed and edited. In some cases, the term management program does not allow the actual managing of terms, or the fields, very fluently. The problems in managing the fields will be discussed in the results of the survey. The hands-on management was the main issue in **Question 11**: "*What is the manageability of the fields in terms of addition, deletion and revision?*" The possible answers were *good*, *average*, and *poor*. If the manageability was good, it means that there were no overall difficulties; if it was average, the user might have experienced some problems, and if the manageability was poor, the fields (or terms, or entries on a larger scale) were almost impossible to control.

Above, in **Q10**, the number of fields within an entry was inquired. **Question 12**, therefore, asked opinions on the number of fields: if there were *too much*, *enough*, or *not enough*. In **Q12** I have assumed that the number of fields is fixed. However, that is not always the case since there are of course TM's that can also facilitate the use of fields of almost any number of choice. For example, MultiTerm™ is such a program. Thus, if in **Q11**, the responses should concentrate on the *poor* end of the scale, one may be able to look for a correlation in **Q12** if the responses to **Q12** show that there are not enough fields.

As stated before, terms normally need a context. A sample context, for example, an extract from a literary source or equivalent is sometimes included with the term in

the database. The context the term appears in can also exist outside the database, and then it is appropriate to direct the user to that context. The reference to the source can be made by using *links* - more or less the same way as in the Internet (for example <http://links.to.net/source.html>), or the source can be referred to otherwise. **Question 13** delved into that: "*Does the record have a link or a reference to a place outside the term bank, where the term appears in context?*" The possible responses were *has a link always*, *has a link sometimes*, and *does not have a link*. Of course, the term bank record does not always require a reference to somewhere else, but it is helpful to know how the term is actually used. One has to bear in mind, however, that the source/context is only a single example of the usage of the term.

As discussed in Chapter 3.3, standardizing terminology is an important part of terminology management. Not all terminology workers adhere to standards, or they have adopted a so-called *ad hoc* standard, which basically means that an agreement on which terms to use or how to construct the database is made on case-by-case basis. **Question 14** was in fact a three-part question; first I wanted to know if there was a standard that the respondent's term bank used: "*Does the term bank adhere to a particular standard?*" The possible answers to this were simply *yes*, or *no*. The second part of **Q14** asked what the standard was, if any. If there was no particular standard, the response field could be left empty. The third part of the question was for possible reasons for the use of the given standard or the lack thereof.

**Questions 15** and **16** were related to the actual management of the term bank(s), more specifically to the number of people that control the term bank and have proper rights to do so. **Q15** thus asked, "*How many people are responsible for the management of the term bank?*" and **Q16** asked, "*Is everyone able to add, remove, and edit the records, or are the access rights restricted to particular individuals?*" The access to the terminology databases can be limited, so that unauthorized people cannot make changes to them. In translation and localization business, however, the whole translation personnel can have write permissions to the term banks, because more than often a term bank has to be started from scratch, and it is useful if as many translators as possible add new terms in the database. Of course, there has to be a mutual agreement among the translators which terms to add. The database administrator could and often will decide which terms to include in the final database. Therefore the administrator could be the only one who is able to



remove entries from the database, and the translators have only rights to add new suggestions (n.b. the added terms are always suggestions before they are approved.). The possible responses to **Q16** were: *everybody can make changes*, and *the access rights are limited*.

The usefulness of the term bank was under scrutiny in **Question 17**. How useful a term bank is, is often determined how many terms are found or if a specific term is found at all. The same is true with dictionaries; if a dictionary does not have a word that the user wants to find, the dictionary is not very useful. If one word out of 100 is not found from a dictionary, the dictionary is nevertheless 99 per cent useful. If, however, only 30 of 100 searched words are found, the situation is different. Therefore, the usefulness of a term bank is truly determined only after numerous searches and possible terms found. In **Q17** I did not ask how many of, say 20 searches, were matched, but rather how often a desired term was found: "*Can you find the term you want in the database?*" The available response choices were *always*, *nearly always*, *often*, *seldom*, and *never*. The reason why I did not require numeric information is that the term bank users do not necessarily count how many times they find a term that they are looking for. In the future, that could be another way to measure how useful a term bank is.

Defining or explaining the terms is a difficult task as well. A careful consideration has to be given to the information that is presented. It is not practical to give too detailed explanation for a term, but explaining the term vaguely or failing to explain it at all does not serve the user, either. As discussed in for example in **Chapter 2.1.2**, the definition for a term largely affects how the term is understood and whether it is suitable for the purpose at hand. **Question 18** asked, therefore, how the terms were usually explained. The responses were *too detailedly*, *accurately*, *well in general*, *vaguely*, and *not at all*. Once more, the issue in **Q18** may be difficult to resolve, since it is rather subjective whether the explanations are viewed as either too detailed or too vague.

The usefulness of the term is also measured in how well it describes the concept. For an expert in the field of the particular terminology, the term-concept relations are not, in my opinion, as crucial as for a non-expert. One criterion for a good term bank is that even an "ordinary person" is able to - at least to some degree - understand the term and its concept. **Question 19** was: "*Does the uninitiated*

*understand the term, i.e. how the term describes the concept?"* The respondents could choose from three alternatives: *the term describes the concept well*, *--describes the concept moderately*, or *--describes the concept poorly*.

The coherence of terms is also an important factor in the functionality of a term bank. If a concept has several different terms within the same term bank, the problem is which term to choose. There may be preferences for certain terms over others, and some concepts may have various terms according to the fields they are used in. With this I mean that all appearing terms may be valid, but some of them are used more often than others, and different companies and experts might hold on to specific terms. For example, in earth drilling industry, there may be terms that almost look like slang to a layperson when compared to the standard language in general. A better example could be chemistry where the chemicals used commercially have a so-called *trade name* in addition to their chemical compound name. Even though a chemical could be referred to by its chemical name, in many cases the trade name is used. In cases like this, it is fully acceptable to have more than one term for a concept. **Question 20**, then, was as follows: *"Are the terms formed consistently: are there several terms for a concept that are used inconsistently, or do you use consistent terms for the same concept?"* The respondents could choose from the following responses: *same concepts have different terms that are used inconsistently*, and *the terms are consistent*.

**Question 21** concerned the interchangeability of the term banks between the company (if any) and its clients (if any). Problems in terminology interchange can be caused by, for example, incompatible standards or by the pure fact that two terminology management programs cannot read each other's data format. The responses to the question *"Are your term banks compatible with your clients' term banks?"* could be as follows: *usually compatible*, *seldom compatible*, and *not compatible*. In addition, there was a text field where the respondents could elaborate the possible problems that might have been evident with the interchangeability or compatibility of different terminology databases. 100 per cent compatibility would be, of course, optimal, but considering the number of different terminology management programs in the market, it is likely that some incompatibility always occurs.

As mentioned earlier in this chapter, a term bank can be a constantly changing entity; translators and terminologists add, remove, and edit the terms in the database. There are, however, term banks that are introduced to a user as a complete package in which all the terms that appear in the database have been approved. This can be once more compared with a general dictionary; all the user can do is to look up a word/term, and use what is found. Therefore there are both ready term banks, and what I would call *open term banks* in which one can introduce new terms. **Question 22** asked about this: "*Do you form your own terms, or do you use prepared terms that have already been validated?*" In retrospect, it should be noted that the wording of **Q22** was somewhat awkward, when the possible answers are given. There were, in fact, two questions in one, which really cannot be answered together with *regularly*, *seldom*, or *never*. The feedback to **Q22**, as well as a retrospective analysis on other questions will be seen later on.

**Question 23**, "*Are the right forms, spelling and grammaticality of terms being taken care of?*" dealt with the quality management of terminology. It is not enough that the terms are properly explained, but it is also important that the spelling and grammaticality of every record is checked and corrected, if necessary. A designated proofreader or another person who is competent should make sure that proper linguistic forms are used. Four pre-formed possible answers were: *being taken care of well*, *taken care of fairly well*, *taken care of poorly*, *not taken care of at all*.

That was all for the preset questions. The last two questions, **Question 24** and **Question 25** were reserved for additional comments. **Q24** was as follows: "*What other comments do you have regarding the functionality, structure and maintenance of the term base/term bases?*" The field for the response could also be left empty if nothing came to mind. I included this free-form question in the questionnaire to supplement the preceding data with possible responses that could not be given in the other sections.

As will be seen in some of the responses I will be dealing in the next chapter, the questionnaire had some parts that required fine-tuning. **Chapter 4** discussed the design of a questionnaire, and some of the problems that might lie in it. Feedback and suggestions for improving the questionnaire was asked in the final question, **Q25**.

## 6. The Responses Under Scrutiny

In this chapter I will be reviewing the received responses. As will become evident, the respondents represent various professions from academic to administrative. The total number of responses was only 10, although I anticipated more. I informed the possible respondents by e-mail, and believed that whenever a company or the like was concerned, one contact would produce several answers. Among the e-mail recipients were large domestic companies, and institutions that have a large personnel. In addition to contacting the companies directly, I also posted the notice about the questionnaire survey on two discussion groups, one of which dealt with translation, and the other dealt with terminology. I will review every question that was answered to, first by indicating the number of the question, then presenting the answer, and finally discuss the answer more or less in detail.

### 6.1 Respondent no. 1

The first respondent works for a company, as a terminology worker in civil service. She prepares terminology for clients, and the primary languages used are *Finnish, Swedish, English, German, and French*. Q6 asked which terminology management programs or other terminology databases are used. The responses were

- *MultiTerm, Eurodicautom, and TIS.*

MultiTerm was introduced earlier in the present study, but Eurodicautom and TIS were not. Eurodicautom is an Internet-based multilingual terminology database that is upheld by the *European Commission's Translation Service* (the links to Eurodicautom's Web site appear after the bibliography). Presently Eurodicautom contains over 5.5 million terms. TIS (Terminological Information System) is a "simplified version of the terminological database used by terminologists and translators working in *The General Secretariat of the Council of the European Union*"<sup>29</sup> (TIS).

**Question 7:** *What pros and cons does your system (the term management system) have?*

**Response:** MultiTerm is a good tool for terminology management when operating in a network environment. It does not lend itself well to terminology projects, however, because adding partial record sets during the project is difficult. Moreover, trying to print the data out is also difficult, and one has to convert the data into the RTF format before anything can be done.

When the respondent talks about the network environment, it means that there is a possibility in MultiTerm to use it over the network. That way the participants in a

project, for example translators, can share a common database that exists on a network server. This obviously facilitates the use of a large database, and guarantees that the database is up-to-date for all the users that share it. Adding record sets once the database has been established may indeed be difficult. One *can* add individual entries, but those entries have to follow the configurations that have been made during the definition stage of the database. In the definition stage the *index fields*, that is the fields that include, for example, the source term, the target term, the definition etc. by which the searches are conducted, and the data is sorted. To **Question 8** about the overall usability of the TM program, the respondent answered *average*.

**Question 9:** *Do you use your own models to make entries and categorize terms, headwords, definitions etc., or do you use the models that already exist in the terminology management program?*

**Response:** Yes.

**Question 10** did not produce any answers regarding the number of fields making up an entry.

**Question 11:** *What is the manageability of the fields in terms of addition, deletion and revision?*

**Response:** Average.

**Question 12:** *The number of fields (that is, if there are enough fields).*

**Response:** Not enough.

**Question 13:** *Does the record have a link or a reference to a place outside the term bank, where the term appears in context?*

**Response:** Has a link always.

**Question 14** asked if the term bank adhered to a particular standard. The response was *Yes*, but somehow the respondent did not indicate which standard the term bank really used. This information would have been relevant.

**Question 15:** *How many people are responsible for the management of the term bank?*

**Response:** 1 to 3.

**Question 16:** *Is everyone able to add, remove, and edit the records, or are the access rights restricted to particular individuals?*

**Response:** Access rights are limited.

The response to **Question 17** about how often a term is found in the database was *often*. This can mean that the database is well constructed, and that it is quite extensive. Combined with the response to **Question 18** that dealt with the explanation of terms, the response being that the terms were explained *well in general*, at this point the used TM program would receive an average grade.

However, there are still many points to consider before any conclusions can be drawn. Then, we move on to **Question 19**: *Does the uninitiated understand the term, i.e. how the term describes the concept?* It appears that the term describes the concept *moderately well*. There may thus be some variation to the descriptions; it is partly a matter of decision whether to maintain the same accuracy and comprehensiveness in all entries, but in my mind it would be appropriate that every entry had approximately the same amount of information. That contributes both to the overall appearance of the database but also lessens the need to further elaborate the concept by some external reference, i.e. the information is already there.

As for the consistency of terms in **Question 20**, the response was that the terms were used consistently, which of course is a good thing. In the case of large databases, and with many people being able to alter the data, the consistency of terminology has to be kept in a tight leash. Quality management and assurance is vital in consistency issues as well, and not only in language-wise. The terms may be grammatically and syntactically correct, but if the database contains much inconsistency, it will undermine the usability and reliability of any terminology.

**Question 21:** *Are your term banks compatible with your clients' term banks?*

**Response:** Usually compatible.

**Question 22:** *Do you form your own terms, or do you use prepared terms that have already been validated?*

**Response:** Seldom.

Since **Q22** was not very well put, I have to make an educated guess to which part of the question the respondent answered. Therefore I take that the terminology workers in the respondent's company do not form their own terms very often. In terminology projects there is often a large workgroup that goes over new suggestions, possible terms that have already been introduced, decides on word forms, etc. Term formation was discussed in **Chapter 3.1**.

As mentioned, quality management of terminology is important. A database that is both consistent and correct linguistically is easy and trusted source of information, as a large and well-known general dictionary.

**Question 23:** *Are the right forms, spelling and grammaticality of terms being taken care of?*

**Response:** They are being taken care of well.

Finally, **Question 24** afforded an unrestricted review of properties regarding terminology management in general. The response explained well why, for instance, not all term records in the system the respondent uses are not equally comprehensive:

"The projects we are involved in include the forming of concept analyses and normative recommendations for the use of the terminology. Moreover, we are engaging in **descriptive surveys**, which causes that not every record is necessarily as complete as every other record."

Descriptive surveys will apparently function as guidelines for the possible record construction. Because there were no detailed explanation of those surveys I have to deduce that the terminology workers are trying to find out which various concepts are used and by whom, and then forming term-concept-definition relations that are most suitable. Again, this is but a speculation.

## 6.2 Respondent no. 2

The second respondent (at this point I have to emphasize that the order of the responses is arbitrary) also works for a company in state administration. He is a database administrator/manager, and the terminology is for his personal use, and it is used often. The primary languages used are *Finnish, Swedish, English, German, French, Spanish, and Russian*. In addition to these languages the respondent indicates that "all" languages are involved in his line of work.

**Question 6:** *What terminology management programs do you use/your company uses?*

**Response:** MultiTerm 95, Euterpe, Eurodicautom, Tapa, TIS.

### 6.2.1 A Closer Look at Euterpe

Some of the terminology databases have already been introduced briefly in Chapter 5.2.2, but the Euterpe database and Tapa may need a further look. Euterpe is a multilingual terminology database of the *European Parliament*. It is prepared in cooperation with the *Trados Corporation*, and in fact, Euterpe is based on MultiTerm as well. The main difference is that the database is accessible through the Web. Euterpe is a complete collection of terminology of the EP, containing over 195,000 entries in 11 languages. The Euterpe search page (links after Bibliography) allows the selection of source and target languages, the layout format (bilingual, full, languages only, source-target), and the search results can be seen in the result window. As extensive the database is, it is not, however near

comprehensive at least if Finnish language is concerned. Apparently the emphasis on translated terms lie somewhere else, but for example a few random searches produced no results:

Source language: English. Term: *Friction drag*. Subject: Transport

Target language: Finnish. Term: Not found.

Source language: English. Term: *Frise aileron*. Subject: Safety, transport

Target language: Finnish. Term: Not found.

It is, of course, premature to say whether the database would produce more matches with an extensive search, but this at least shows that there is much to improve before the Euterpe database is complete. Nevertheless, new terms are constantly added, and hopefully the existing terms get their target language translations and definitions as soon as possible.

### 6.2.2 The Tepas Term Bank

The Tepas term bank is maintained by the *Finnish Centre of Technical Terminology* in cooperation with the *Library of Helsinki University of Technology*. Tepas is an online terminology database, and its multilingual terminology entries amount to about 100,000. The search form is available on the internet (see the URL after Bibliography). Searching terms is quite simple: there are 1 to 2 possible headwords that can be searched with Boolean algorithms AND and OR. The search can be performed as a general lookup, or the results can be browsed through. The results can also be sorted alphabetically by five different languages or by the source of reference. Also a free-text search can be made, and the searched text can be cut, followed by an asterisk (\*). Then all terms starting with a given string will be returned.

When the same search criteria was given to Tepas, the following results were given:

Source language: English. Term: *Friction drag*.

Target language: Finnish. Term: Kitkavastus. Subject: Water conservation.

Source language: English. Term: *Frise aileron*. Subject: Safety, transport

Target language: Finnish. Term: Not found.



As we can see, the search with Tera produced 50 per cent better results, but the subject was different from what Euterpe indicated. A *Frise aileron* is an aeronautical term referring to a wing mechanism that is used to control the movements of the airplane.

### 6.2.3 Other Responses

The second respondent indicated that also other terminology databases are in use, but the ones mentioned above are the most common.

**Question 7:** *What pros and cons does your system (the term management system) have?*

**Response:** MultiTerm is a clear and simple tool for searching terminology, and adding material into the database is relatively simple. One problem, though, is that the language fields are not sufficient, since MultiTerm will allow only 20 index fields.

There will be a new version of MultiTerm, but the present version is indeed limited in the number of index fields. Moreover, MT is a so-called *16-bit application* that does not allow long filenames or long index field strings. That is a major hindrance. The respondent uses MT as a project tool, but he states that it is not suitable for project work, because it "requires a lot of orientation". Overall, though, the respondent grades MultiTerm usability as *Good*.

**Question 9:** *Do you use your own models to make entries and categorize terms, headwords, definitions etc., or do you use the models that already exist in the terminology management program?*

**Response:** Yes.

**Question 10:** *The number of fields.*

**Response:** 15 and up.

**Question 10b:** *Which fields do you use?*

**Response:** Source, spelling, definition, explanation (apparently different from definition), additional information, context, example, comment, country, project, person, phenomenon, organization, field, branch of administration, language, grammar, style, reliability, relation, registry date, registrar.

As we can see, the database in this case uses much more fields than the previous one. Moreover, as indicated at the end of the response sheet I received, the database that the respondent's company uses may have fields ranging from 15 to 500, depending on the record. Probably not all of the fields are used in every entry, but the amount of information an entry can show, is quite extensive. Some fields, as the *project*, *person*, *reliability*, *registry date*, and *registrar* apparently help to determine why the entry was made, who made it, and whether the information

entered can be trusted. For example, the project that produced the entry can be traced and used for reference.

**Question 11:** *What is the manageability of the fields in terms of addition, deletion and revision?*

**Response:** Good.

**Question 12:** *The number of fields (that is, if there are enough fields).*

**Response:** Not enough.

**Question 13:** *Does the record have a link or a reference to a place outside the term bank, where the term appears in context?*

**Response:** Has a link always.

The responses **12** and **13** show that: first, there is the same problem with the database that the number of fields that can be given, is limited. There is, therefore, clearly a need for improvement. Second, the links and references with the entries help to find the source of information, additional information, and possibly other relevant data. This would be considered as a plus.

**Question 14:** *Does the term bank adhere to a particular standard?*

**Response:** No.

**Question 14c:** *Please give reasons why or why not.*

**Response:** We aim to follow terminological standards whenever possible, but we have to make adaptations according to our current needs.

**Question 15:** *How many people are responsible for the management of the term bank?*

**Response:** 1.

**Question 16:** *Is everyone able to add, remove, and edit the records, or are the access rights restricted to particular individuals?*

**Response:** Access rights are limited.

**Question 17:** *Can you find the term you want in the database?*

**Response:** Nearly always.

**Question 18:** *How are the terms explained?*

**Response:** Well in general.

**Question 19:** *Does the uninitiated understand the term, i.e. how the term describes the concept?*

**Response:** Describes the concept moderately well.

**Question 20:** *Are the terms formed consistently: are there several terms for a concept that are used inconsistently, or do you use consistent terms for the same concept?*

**Response:** The terms are consistent.

As we can see in the response to **Q14**, the respondent mentions no particular standard. Many companies apparently make their own adaptations to better suit their own purposes. One of these companies is the Kielikone Ltd. Mainly, Kielikone makes use of the NTRF standard, but has altered some properties of it. The terminology database in this response appears to be quite extensive, but the

actual rate of acceptable matches remains unclear. It is, therefore, the result of the effort people put in to add the terms in the database. When the purpose and field terminology of *any* database is determined, a better view for what to incorporate in the database can be formed. In other words, adding terms that are frequently used, clearly facilitates the future searches and yields better results. Moreover, when the terms are explained well generally (some exceptions notwithstanding), and the terms describe the associated concepts *moderately* well, it can be said that at least the database in this second response is reasonably well formed. Good databases and good terminology management programs do not mean the same thing, however. The consistency of a database may stay in control if the number of people who can change the contents of a database is limited. On the other hand, if only one or two people maintain the database, the content may go unchecked or without critical analysis.

**Question 21:** *Are your term banks compatible with your clients' term banks?*

**Response:** Usually compatible.

**Question 22:** *Do you form your own terms, or do you use prepared terms that have already been validated?*

**Response:** Regularly.

**Question 23:** *Are the right forms, spelling and grammaticality of terms being taken care of?*

**Response:** Taken care of fairly well.

The last pre-formed answer set produced the results above. Compatibility seems to be reasonably good, but since I have not asked who the clients usually are, I cannot determine how widely the databases in this case are circulated, and how frequent the interchange of the databases actually is. The response to **Q22** is not, once again, clear: does it mean that the company forms their own terms or not. The questionnaire design has not been very successful in this particular question. Finally, the quality management of the databases in question works, according to the respondent, reasonably well. It may be that the person who responded is also responsible for the quality of the databases that are used. Again, it is a good thing if a person who maintains the database can also check the grammar, syntax and other linguistic aspects, but if that person is not very competent in linguistic matters, the overall quality of the database may degrade.

### 6.3 Respondent no. 3

We move on to the third response set. This response set came from a certain ministry of Finnish government. **Question 1**, "*Do you work for a company, or are you a private entrepreneur?*" allowed only two possible answers, *namely for a company*, or *an entrepreneur*. Therefore government bureaus and the like fall into the company category, although the first question could have included more choices. As in the previous response set, this respondent is in an administrative position, a terminology worker, often involved in terminology matters. The terms are prepared for clients. In addition to that, some of the databases the ministry produces are published.

The primary languages this respondent has to use are *French* and *Swedish*. It is rather surprising that only two languages are mentioned, especially when English is not. However, this may be due to the fact that in the European Union, a large number of texts - and therefore, terminology - are in French. Moreover, Swedish being the second official language of Finland, sometimes the languages a terminology worker uses involve fewer languages, and other languages, than what might be expected. To avoid too much repetition of questions, I will from now on only indicate the number of the question I am presenting.

**Question 6/Response:** MultiTerm, Eurodicautom, Tapa

**Question 7/Response:** Maintaining synonyms is difficult, and one cannot print out synonyms in a desired order when using the RTF format. The instructions in MultiTerm are not comprehensive enough.

**Question 8/Response:** Program: MultiTerm, Usability: Good.

The MultiTerm appears to be a very common terminology management program. One reason may be that the MultiTerm is inexpensive, quite compact, and it is designed to work mainly with a translation software, although it is possible to use it without other translation products. The Eurodicautom and Tapa term banks were already mentioned, but they also seem to be used by many companies and bureaus. The Eurodicautom database is most likely used often because of the European Union. The council of the EU has declared, for example, that all documents in the Union be translated into the target language of each member country. Therefore the logical solution to find accurate information regarding the terminology used in the Union is to use, for instance, Eurodicautom.

The MultiTerm terminology management program receives one more critical blow from this respondent: the management of synonyms is poor. I would deduce, however, that not only synonyms are difficult to print out, but also other individual fields may be hard to filter and print out. In addition to this problem, the respondent tells us that the instructions of the program are insufficient. MultiTerm has an *online help* (a help file on the user's hard drive that can be opened in the program), which is a very basic introduction to the functions and properties of the program. There is also a *tutorial demo* that can be downloaded from the Trados Web site, and a downloadable user guide (if one does not have the programs on a CD-ROM). The user guide has 335 pages, and it is hard to imagine why the instructions would be unsatisfactory. The tutorial demo, on the other hand, seems quite simple: for example, the MT user interface is introduced after a "Getting Started" section, then there is an introduction how to search the database in different ways, introductions to filtering and adding entries, and finally a demonstration in creating a new database. The demo may be adequate for an overview, but more detailed instructions cannot be found in it.

- Question 9/Response:** Yes.
- Question 10/Response:** Approximately 16.
- Question 10b/Response:** Classification number of the record, country, field, branch of administration, additional information, source of definition/explanation, "see also" (additional reference), recommendability of term, unchecked/checked/result of terminology project/standard (status).
- Question 11/Response:** Average.
- Question 12/Response:** Enough.
- Question 13/Response:** Has a link always.

From the responses from the second and the third respondent we see that the number of fields is roughly the same in the respondents' databases (the first respondent did not give the field count). The naming of fields is not very similar, although there are one or two instances that may count as identical fields. The sources of definitions and explanations - apparently they are separate fields - are most likely used as a verification of the data: the source can be checked for validity as well. Moreover, indicating the sources instead of just giving a definition may help to better see the relationship between terms and their contexts. However, if there are too many references, that is, if the user is constantly directed away from the database, the database may become a reference book with no actual data in

itself. The user should be able to find most of the desired information within the database at hand.

The *status* field is commonly used at least in localization business; when the client is releasing a new or an updated product, it needs a translation for the documentation, and often for the software as well. There are normally previous translations with existing terminology, which is usually approved. The terms that have been approved can have the status approved, or checked, as in the database of our third respondent. In my opinion, it would be better to have only one field, namely the status field, and to use the *unchecked/checked/result of terminology project/standard* or equivalent as attributes. That would reduce considerably the number of index fields, especially when MultiTerm does not allow more than 20 index fields.

Responses to questions **11** to **13** show that the database is reasonably well constructed and functional, but a further analysis of the functionality of the database is, according to the respondent, impossible, since the terminology they are creating for the database is their first. This means most likely that the MultiTerm database is their first; other databases have been created and used before.

**Question 14/Response:** No (does not follow any standard).

**Question 14c/Response:** The structure of the database has originally been designed for language counseling, which requires that information can be searched quickly using many different criteria.

There appears to be a slight discrepancy between the fact that the terminology is prepared for clients, and that the database does not follow any standard. I would gather that standardizing the format of the data would serve the clients better than optimizing the speed of the searches. However, the saying "time is money" could be behind the response to **Q14c**: the clients pay only for the information, not the time it takes them to receive it. This is even more true when one considers service numbers; many service numbers are no longer toll-free - the service is charged by the time used, and therefore less time means lower costs.

**Question 15/Response:** 1.

**Question 16/Response:** Access rights are limited.

**Question 17/Response:** Often.

**Question 18/Response:** Accurately.

**Question 19/Response:** The term describes the concept well.

**Question 20/Response:** The terms are consistent.

On the basis of the three responses reviewed so far it seems that the term bank management is very limited. This may be the result of reorganization of tasks and concentration of responsibilities. The terms appear to be found often, and they are explained accurately. Furthermore, they describe the associated concepts well. Therefore it can be said that the term bank in this instance is also fairly well organized and maintained. As discussed earlier in the present chapter, even the term bank in the third response set is not without flaws. To the question of compatibility, **Question 21**, the response was *usually compatible*. **Question 22**, *Do you form your own terms, or do you use prepared terms that have already been validated?* was left unanswered. That is because the wording of the question was ambiguous, and the respondent saw that she could not answer it. Finally, to **Question 23** on quality management, *Are the right forms, spelling and grammaticality of terms being taken care of?* the response was that they are taken care of well.

As a summary at this point, there have been no significant flaws in the terminology databases or TM programs I have discussed. There have not been any outstanding positive features, either. It is regrettable, though, that the respondents have evaluated only the features of a program, and not the web databases. Answers regarding the usability of the Euterpe, Eurodicautom, and other server-side term banks would have been welcome already.

#### **6.4 Respondent no. 4**

Unlike the previous respondents, the fourth respondent was a private entrepreneur. Moreover, not a terminology worker, but a translator. In spite of, or because of that, he was also often involved with terminology. The languages used were Finnish, Swedish, and English, and the terminology management programs/terminology databases were once again MultiTerm, Eurodicautom, and Tapa. The Eurodicautom database was graded *good*, and MultiTerm was graded *average*: there was "nothing to complain" about Eurodicautom, but MT according to the respondent, learning to use MultiTerm takes much time. Apparently the latter is not an easy program to use. However, if we look back to the response to **Question 7** in **Chapter 6.3**, it was mentioned that the instructions in the program were not adequate. This is perhaps one reason why MultiTerm requires much

learning. The basic functions of the program can be learned, in my opinion, even without instructions, but when one should use various filtering methods, input models, and such. I will not explain these in the present study, because this paper is not about MultiTerm, although some illumination on the functions of the program has been necessary. **Question 9** about producing and using own models as opposed to prepared models produced no answer, so we will not know whether the models in the MultiTerm are those of the program or the user. In Eurodicautom and Tapa, there is obviously no possibility for an outside user to create their own models or fields. Next, I will present some responses that were given, in a list as before:

- Question 10/Response:** 3.  
**Question 11/Response:** Average.  
**Question 12/Response:** Enough.  
**Question 13/Response:** Has a link always.  
**Question 14/Response:** No.

As we can see, there are 3 fields in the database (MultiTerm), which is much less than in the databases in the previous answers. The respondent did not indicate, though, what those fields were. The manageability of fields is, however, only average. This is surprising, since one could assume that having fewer fields would make the management of the database easier. There must be some other factors present that prevent the database from being well managed, but they do not become evident. It looks also that there are enough fields, although being so few. Perhaps there is no need for more fields, especially when the respondent is working in his private business and does not have to share the database with others or to produce terminology services for clients. There are also links to external sources, which in this case probably work as a personal reference. Because the database is used privately, the need for a standardized format may not be so great.

- Question 15/Response:** 1.  
**Question 16/Response:** Access rights are limited.  
**Question 17/Response:** Seldom.  
**Question 18/Response:** Well in general.  
**Question 19/Response:** [The term] describes the concept moderately well.  
**Question 20/Response:** The terms are consistent.

Now, if we look the answers above, some of them explain themselves: since there is only one person in charge of the terminology management and the database, the



access rights are also limited, although there would be no need to speculate on that, and the limitations are therefore nominal. The response to **Q17** was interesting, though. To the speculation of why a person who maintains a database by himself would fail to find the terms he needs I would say that probably not much effort has been put to adding terms in the database, or the database has been constructed so that the searches do not produce the results that are desired. This is one disadvantage of not having a larger team to work with terminology, and it may be difficult for only one person to decide what to have in a database, let alone control the overall quality and quantity of terms. The responses to **Questions 18** and **19** do not differ from the previous respondents' answers, and therefore cannot confirm or disprove the claim that more people in charge of terminology databases were better. The response to **Question 20** is again fairly obvious, since when one person constructs a database for his/her own use, there should not, at least, be much or any variation in terminology as regards their consistency.

**Question 21/Response:** *[The term banks are] usually compatible [with the clients' banks].*

**Question 22/Response:** *[I use own terms] seldom.*

**Question 23/Response:** *[The right forms, spelling and grammaticality of terms are being] taken care of fairly well.*

I have used square brackets in this chapter to refresh the memory of the reader in case they have the idea of what the answers regarded. The response to **Q21** did not correlate with **Question 3b**, because the respondent did not indicate if the terminology is only for his own use, or if it was also prepared for clients. However, the term banks, according to the respondent, are usually compatible with clients' term banks. Furthermore, the respondent attested that he seldom uses his own terms. Finally, the response to **Q23** shows that the right forms, grammaticality, and spelling of terms in the database are checked and corrected, if necessary, reasonably well.

To sum up, the fourth respondent seems to have a database of average quality: the state of actual management of the user's term bank falls roughly in the same category as in the previous response sets, there are considerably fewer fields, and the terms are not found nearly as often as in the data banks used by the first three respondents. When we look at the significance of the present response set for the possible improvements of different term banks, this set did not bring about any new observations or ideas. Therefore, I will move on to the 5<sup>th</sup> response set in the next chapter.

## 6.5 Respondent no. 5

If the previous responses were not very fruitful, the same cannot be said about the following ones. The fifth respondent works for a company in an executive position, and his main interests lie in *technical network communications*. He is in contact with terminology fairly often – less than a terminologist or a translator would be. The main languages the respondent is dealing with are Finnish, Swedish, English, German, French, Spanish, and Russian. The primary terminology management programs used are MultiTerm, and *MOT*. In addition to that, the respondent uses the Eurodicautom web database.

The *MOT* program is another language product in the Kielikone Company's product line, and it is much like TermIt, but in my opinion, more of a dictionary than a terminology management program. The dictionary-like features of *MOT* will be seen in the following responses. First, however, I will deal with the responses concerning MultiTerm. The pros in MultiTerm were as follows:

- Graphic interface.
- The exported data can be formatted. The conversion can be made to HTML format, for example.
- Links from one term to another.
- The built-in input models speed up working with the database.

The graphic interface is nowadays almost a given, but there have been programs quite recently that were text-based when, for example, MS Windows or Macintosh operating systems were not as common as the MS-DOS system. The interface in MultiTerm could, in my mind, be more modern, but probably the emphasis has been paid more to the functionality of the program as to the appearance. The second point in the above list speaks about formatting the data. When a database structure and entries are based on a recognized standard, such as the SGML, the conversion to another standardized format, as the HTML, works generally without any problems. These two standards are very closely related, and in fact, the *HyperText Markup Language* has been developed from the *Standard Generalized Markup Language*.

The actual exporting takes place after the export definitions have been made. In the definition phase, the user indicates, for example, which fields to export and by which criteria. The exporting, and the preparations for it, could be simpler to the

user, although the present respondent did not mention that. If one wants to export data from MultiTerm, it will not be possible off the cuff.

Linking the terms to one another takes place with the help of *cross-references*. The cross-references can be created when creating a new entry, or by adding the cross-reference to an existing entry. When a referenced term is clicked, the program searches and displays a match, if a match exists. The *input models*, on the other hand, are basically like document templates in a word processor; the models facilitate the adding of new terms, especially when importing large records from an external file.

There was much good in MultiTerm, but there were also some cons as well:

- The manageability of synonyms especially when working with multilingual terminologies.
- When there are two or more homonyms, for example 'shaft' (as an elevator shaft) and 'shaft' (an axle), how to mark the entries and their target language equivalents.
- Changing the search direction from the target language to the source language and vice versa, when the other language has more than one equivalent, will become problematic. As a default, there is only one equivalent for each term.

The problem with synonyms is evident here also, and homonymy is even more problematic. One possible solution is to mark each term with a single consecutive number, or with a "main entry" number followed by a "subentry" number divided by a period. Below is an example from the *Collins Cobuild English Language Dictionary* (1987: 896):

**3 Matter** is **3.1** written material, especially books and newspapers

**3.2** part of a magazine, newspaper, book, film, etc.

**3.3** the physical part of the universe consisting of solids, liquids and gases.

This is a reasonably clear way to indicate the different manifestations of the term, but there are many issues to resolve when making entries such as the one above. For example, the terminologist should decide which definition is the most important, and how many different definitions to a homonymic term should there be. When the term base is designed for some specific field, for example, automotive engineering, the solution could be that the primary definition is for that particular field, and possible other definitions came next. Moreover, by numbering

each definition, the user could search only the "second definitions" if desired, and so on.

The respondent indicated that the project group had to resort to scripting in order to solve the problem that arose when the search direction was changed in MultiTerm. Scripts are small programs that instruct the computer or the application to perform certain tasks. Different programs understand different scripting languages, and for instance, WWW browsers could use the *JavaScript* or plain *Java*<sup>TM</sup> languages for different functions. Scripting can be very helpful, and often very powerful, way to automatize an execution of some tasks, and although it takes time to write a working script, the time is saved afterwards when the users do not have to manually perform operations such as searches and sorting. As regards the usability of a terminology database, it is not very productive having to resort to a sort of a detour to accomplish something that should be possible already in the program.

The opinions on the MOT dictionary were as follows:

- The Pros:** Search capabilities, the searched word can be cut (for example, lich\*, and so on), the search is continuous - that is - the program searches the words while writing.
- The Cons:** The program is very "non-visual".  
The technical glossaries of Kielikone are in the form of 'Source term' - 'Target term', without explanations.

Indeed, the search capabilities in MOT are quite good. If the user wants to find all words/terms starting with a specific string, the program will return all words that begin with that string. However, if the user wants to search words that end with a specific string (for example, \*ion), MOT cannot perform the search. Compared with the other Kielikone product, TermIt, the latter *can* search with a wildcard<sup>30</sup> character, and return all matching words. This function would be very useful in MOT as well. The visual appearance of the MOT program, on the other hand, is very simplified. The search window is divided into four sections: a field for the headword/search word, a drop-down list for the installed dictionaries, on the left hand side there is a list for the found words, and on the right is the list of translations. Because MOT is a dictionary more than a term base, there are no definitions for the words. Printed dictionaries usually give examples of usage, but MOT does not do even that. Therefore, MOT cannot be considered as an ideal term bank, although the actual databases could be very comprehensive.

The Eurodicautom database received the following evaluation:

<b>The Pros:</b>	Multilingual, support for technical terminology, the terminology comes from authorized terminology producers, well defined terms, indicated sources.
<b>The Cons:</b>	Long search times in the Web.

It is quite clear that when a terminology database is created for a vast international user group and is made under the control of the European Union, the properties of the database have to be thought through extremely well. The Eurodicautom system contains currently 36 different term bases from agriculture to transport, so it is easy to understand that constructing a terminology collection is an enormous task. I tried to look up few terms earlier in the present study from another terminology database, and decided to try to look up the same term in Eurodicautom as well:

Source language: English. Term: *Frise aileron*. Subject: aviation.

Target language: Finnish. Term: *Frise-tyyppiset siivekkeet*.

The search returned a translation where the proper noun 'Frise' was attached to a descriptive particle, and 'aileron' received the target language equivalent. In itself the target language term would have been insufficient for others than those who have knowledge in aviation. However, the source term included also an alternative: 'anti-yaw aileron'. The yaw axis is the vertical axis of the airplane, but aviation technology need not be explored further here.

As indicated, the terms in Eurodicautom come from validated sources, and they are checked thoroughly before adding them in the system. Furthermore, the Eurodicautom term bank has a feedback form in which users can comment on any terms that they encounter, give suggestions for corrections and additions. In this sense, Eurodicautom is also a project tool with all the users as potential project contributors. As responses to **Question 8**, the three term banks mentioned received the following grades: *MultiTerm - Very Good*, *MOT - Average*, *Eurodicautom - Good*. The following list shows the responses to **Questions 9—23**. I have included a short verbal reference to the respective questions in parentheses.

<b>Question 9/Response:</b>	Yes. (Own models)
<b>Question 10/Response:</b>	17+5. (Field count normally + additional fields)
<b>Question 11/Response:</b>	Average. (Manageability of fields)
<b>Question 12/Response:</b>	Enough. (Number of fields)
<b>Question 13/Response:</b>	Always. (How often terms are found)

<b>Question 14/Response:</b>	Yes. (Is there a standard)
<b>Question 14b/Response:</b>	NTRF etc. (Which standard)
<b>Question 15/Response:</b>	1 - 2 (No. of people responsible for the term bank)
<b>Question 16/Response:</b>	Access rights are limited.
<b>Question 17/Response:</b>	Often. (Found terms)
<b>Question 18/Response:</b>	Accurately. (How terms are described)
<b>Question 19/Response:</b>	The term describes the concept well.
<b>Question 20/Response:</b>	The terms are consistent.
<b>Question 21/Response:</b>	Usually compatible. (Term bank compatibility)
<b>Question 22/Response:</b>	Never. (Use of own terms)
<b>Question 23/Response:</b>	Taken care of fairly well. (Grammaticality etc. of terms)

The fifth respondent has indicated that his company uses their own term bank models, which is not at all surprising, since the company's clientele is varied, and the same models would not necessarily be suitable for everyone. The term banks serve users better if they allow multiple field structures and addition of relevant data categories depending on different needs. For example, more extensive presentation of synonyms could be desired in some instances, in other instances one could need alternative definitions if an entry consisted of more than one target term, etc.

The respondent's databases are apparently comprehensive enough, and they have enough fields, although their manageability is only average. The NTRF standard, modified to a certain degree, though, ensures that the data is usually compatible with the databases of the clients of the company. If the exported data from MultiTerm is not directly compatible with other data formats, it is modified so that it can be further read with the most common text processors and spreadsheet programs such as MS Excel, and MS Word. In general, the term banks mentioned by the fifth respondent are well organized, comprehensive, usually compatible, and fairly well quality-controlled. The possible improvements concerned the same TM programs as in the four preceding response sets, and especially the MultiTerm program.

## 6.6 A Mid-point Summary

To save space and the readers' time, I will, from this response set forward, show the responses in a similar list format as before, but with the exception that now I only comment on some of the answers. This is because most of the remaining

responses are identical or almost identical with those already reviewed, and I see no reason to repeat them over and over again. What I will do is to summarize the findings after the last respondent, and, when a response to a particular question differs from the mainstream, I will comment on that. At this point of reviewing the responses it is perhaps appropriate to say that the terminology databases and TM programs have been - and will be - roughly the same, few exceptions notwithstanding, as well as their evaluations. It is obvious that different companies and different private entrepreneurs maintain their terminology databases in different ways, and it is difficult to give a prescription on how to maintain them even better. Therefore, it is probably best to settle for presenting observations and opinions. However, when the possible improvements concern a particular program rather than a practice, the responses speak for themselves.

### **6.7 Respondent no. 6**

The sixth respondent works as a terminology worker for a company, which is actually an association involved in terminology. The terminology is prepared for customers, and the primary languages are Finnish, Swedish, English, German, and French. In addition to the usual TermIt, TEPA, and Eurodicautom term banks, the respondent uses the *Termium*® data bank. Termium is a language product of *The Translation Bureau of Canada*, a terminology database which contains 3,000,000 bilingual terms in English and French on a CD-ROM that can be purchased. Therefore, Termium is not a public term bank. Unfortunately, the sixth respondent did not share any views on this terminology tool.

The respondent's company has their internal term base which runs with TermIt. According to the respondent, the database is easy to use, and it is working very well. The downside is that TermIt cannot produce glossaries that can be printed in a printing press. The program does not allow printing on a printer, either. Another minus with the TermIt program is that it does not allow addition of pictures. MultiTerm, on the other hand, allows. Having images with a definition is sometimes very useful. For example, the printed version of the *Moderni Lääketieteen Sanasto (A Modern Terminology of Medicine, 1988)*, has many entries with illustrations. The old saying that one picture tells more than a thousand words is true in terminology as well; if a verbal definition is not considered sufficient, it should be elaborated with a picture. Moreover, the respondent

indicated that she could also use so-called *internal links* which are links from the terms inside the definitions to the term's term record. This was not very clearly put, but I would see it like this: when a definition of one term includes other terms that also exist in the database, there should be a link from those other terms to their own entries. Obviously, a cross-reference is in question here. Again, MultiTerm has the cross-reference possibility. The cross-references, however, cannot appear within a definition, but rather as individual fields. Next, I will give a list of the responses to **Questions 8—23:**

<b>Question 8/Response:</b>	Program - TermIt, Usability - Very Good. Program - Eurodicautom, Usability - Good.
<b>Question 9/Response:</b>	Yes.
<b>Question 10b/Response:</b>	The number of fields varies database by database. Normally there are following fields: Terms (languages FI, SW, EN, DE, FR), synonyms, headword, definitions, annotation, contexts, source, creation date, creator, change date, changer, comments. Also the subfields: genus, domain, source of field.
<b>Question 11/Response:</b>	Good.
<b>Question 12/Response:</b>	Enough.
<b>Question 13/Response:</b>	Always.
<b>Question 14, 14b/Responses:</b>	Yes, NTRF.
<b>Question 15/Response:</b>	8
<b>Question 16/Response:</b>	Everyone is able to add, change, and edit the data in the database.
<b>Question 17/Response:</b>	Often.
<b>Question 18/Response:</b>	Well in general.
<b>Question 19/Response:</b>	The term describes the concept moderately.
<b>Question 20/Response:</b>	The terms are consistent.
<b>Question 21/Response:</b>	Usually compatible.
<b>Question 22/Response:</b>	Regularly. (Own terms)
<b>Question 23/Response:</b>	[The grammaticality, etc. of terms is] being taken care of well.

The usability of TermIt has been graded to be better than in the previous instances. The Eurodicautom appears to be a good terminology database, but the main tool in the respondent's company is TermIt. In addition to the programs mentioned, the company uses also the MOT dictionary among other electronic dictionaries, but those programs were not graded. As we can see, the used fields cover a variety of different information, and the terminology bases are standardized by the NTRF standard. Because the respondent's company is heavily involved in collecting and making terminology, the terminology management tools and their functionality are extremely important. The amount of data that the company has to handle requires



that there are several terminology workers who have access to the databases the company produces.

The respondent emphasized the importance of being able to create terminology that can be printed in one form or another. As I mentioned, the TermIt program cannot print the data, and in MultiTerm, first one has to export the desired data, then format it, and only after that print it. Therefore one vital improvement in a terminology management program should be the ability to print or to export to a format that is printable. One possible format could be the PDF format, since it is both very portable, and the text and image quality can be optimized for printing in a printing press as well. The export could be done as a regular print job with the required data selected. If we take an ordinary word processing program, there is a possibility to print the documents to a PostScript or a printer file. The PostScript format and the printer file format (.ps and .prn) can be converted to the PDF format.

Printing was not the only suggested improvement. The respondent also expressed the need for incorporating a drawing program with the TM tool. That way the terminologist, translator, etc. could easily draw concept charts within the TM tool, and include the charts with the term record. Having a drawing tool in a TM program could be a welcome addition, but I am concerned what would happen if too many features that exist in word processors, graphic design tools, desktop publishing software, and so on are added to the terminology management programs. The obvious trend with the graphic design tools has been that many new features have been added from one version to another because the users have requested them. That way the programs have become enormous in terms of used hard disk space, computer memory and resources, and bloated beyond a reasonable size. Hopefully the voice of the user is heard with caution when it comes to terminology management software.

### **6.8 Respondent no. 7 - A Comparison with no. 6**

The seventh respondent also works for the same company as the previous respondent. The term bases that the present one is involved with, use Finnish, Swedish, English, German, French, and Russian. The TM programs and term banks that the respondent uses are TermIt, Eurodicautom, Tapa, and other term banks available on the internet. She graded the TermIt program as *good*, the

Eurodicautom database also *good*, and the Tapa term bank received only an *average* evaluation. There were thus no significant differences to the previous respondents' evaluations on these programs and databases. The TermIt, however, seemed to lack the same features as in the sixth response set. The user expressed a desire to have links between different term records (cf. MultiTerm), and also to have an ability to add pictures and other helpful material. This helpful material could be, for example, *concept charts* (graphic illustrations of relations between terms, concepts, definitions, other terms, etc.).

It is interesting to compare the answers between the respondents 6 and 7. I decided to use a table to show the two responses. I apologize for the repetition of information.

Question	Respondent no. 6	Respondent no. 7
9. <i>Do you use your own models to make entries and categorize terms, headwords, definitions etc., or do you use the models that already exist in the terminology management program?</i>	Yes.	No.
10. <i>How many fields there are in a normal entry?</i>	Varies database by database.	4 - 6.
11. <i>What is the manageability of the fields in terms of addition, deletion and revision?</i>	Good.	Average.
12. <i>How would you describe the sufficiency of fields?</i>	Enough.	Enough.
13. <i>Does the record have a link or a reference to a place outside the term bank, where the term appears in context?</i>	Has a link always.	Has a link always.
14. <i>Does the term bank adhere to a particular standard? If yes, which standard.</i>	Yes. NTRF.	Yes. I do not know.
15. <i>How many people are responsible for the management of the term bank?</i>	8.	1 - 2.
16. <i>Is everyone able to add, remove, and edit the records, or are the access rights restricted to particular individuals?</i>	Everyone is able to add, change, and edit the data in the database.	Access rights are limited.
17. <i>Can you find the term you want in the database?</i>	Often.	Nearly always.
18. <i>How are the terms usually explained?</i>	Well in general.	Well in general.
19. <i>Does the uninitiated understand the term, i.e. how the term describes the concept?</i>	The term describes the concept moderately.	The term describes the concept well.
20. <i>Are the terms formed consistently: are there several terms for a concept that are used inconsistently, or do you use consistent terms for the same concept?</i>	The terms are consistent.	The terms are consistent.

21. Are your term banks compatible with your clients' term banks?	Usually compatible.	Seldom compatible.
22. Do you form your own terms, or do you use prepared terms that have already been validated?	Regularly.	Regularly.
23. Are the right forms, spelling and grammaticality of terms being taken care of?	Being taken care of well.	Being taken care of well.

Table 3: A comparison of Response sets 6 and 7.

The first difference is evident in **Q9**: One respondent indicated that own models are used, and the other that they are not used. A possible explanation could be, for instance, that either one of the respondents saw model creation as an individual task rather than something that concerned the whole company. Another explanation could be that the 6<sup>th</sup> respondent answered *yes* to *do you use your own models*, and the 7<sup>th</sup> respondent answered *no* to *do you use the models that already exist in the terminology management program*, which, in fact, means that the latter respondent also answered affirmatively. My apologies for the poor phrasing of the question that caused this ambiguity. In **Q10**, however, the 6<sup>th</sup> respondent said that the number of fields varies, whereas the 7<sup>th</sup> respondent gave a clear figure: 4 - 6. The reason for this may simply be that the terminology workers of the company work with many different databases, and in some databases the field count could indeed vary so that no fixed number cannot be given, whereas other databases have a more regular count.

Another difference, now perhaps in subjective opinion, is in **Q11**: the 6<sup>th</sup> respondent graded term manageability as *good*, but the 7<sup>th</sup> respondent graded it as only *average*. The discrepancy in this case may again have been caused by the use of different TM tools. One respondent may use more, for example, MultiTerm, and the other, say, TermIt. I would have thought, though, that the responses had been vice versa, because if the number of fields were more fixed, it would facilitate the management of those fields. The next partial difference was found in **Q14**. Both respondents indicated that there is a standard that the term banks use, but the 7<sup>th</sup> respondent did not know what the standard was. Probably the latter does not have to be aware which standard the database uses, but this seems somewhat far-fetched, since compared with the responses to **Q21**, the term banks that the 7<sup>th</sup> respondent has used were indicated as being *seldom compatible* with the clients term banks, whereas the 6<sup>th</sup> respondent said that the term banks have been *usually compatible*. Maybe the variation between different terminology projects and tools is significant,

after all, for example, the Finnish Centre of Technical Terminology is involved in several terminology projects, and has to deal with terminology from a multitude of different fields that there is bound to be certain fluctuation both in the term bases as well as the TM tools.

The differences between the two respondents continue in **Q15-Q17**: first, the 6<sup>th</sup> respondent indicated that there were eight people responsible of the term base management, and the 7<sup>th</sup> respondent had only one or two. Moreover, according to the former respondent, all of those eight have change privileges to the database, whereas with the latter the access rights are limited. Second, the question of how often the terms are found produced two slightly different answers: *often*, and *nearly always*. These responses could be different only because the boundary between the concepts *often* and *nearly always* is not very clear, and in fact, they could mean the same. The choice of words depends on the individual.

Finally, the overall usability of the term banks in both cases discussed above seems to be good. However, one difference remains: the 6<sup>th</sup> respondent indicated that the descriptiveness of concepts by terms was only moderate (**Q19**), whereas the 7<sup>th</sup> respondent claimed that the terms described the concepts well. This, again, could be the result of differences between term base sizes and constructions, and who have been responsible of the writing of entries. One has to bear in mind, though, that what one person thinks to be good is not necessarily as good for the other.

### **6.9 Respondent no. 8**

The eight respondent is a private entrepreneur, a translator who often needs terminology for her own use. She is specialized in Finnish-German-Finnish translations, and the TM tools/databases she uses are mainly MultiTerm, Eurodicautom, Tapa, and a newcomer, *Transit*. Transit is in fact a translation memory tool whose sister product, *TermStar* is more of a terminology management program. The respondent mentioned, however, only the Transit translation memory program, but I would like to concentrate briefly on the TermStar program in the next subchapter.

### 6.9.1 The TermStar

The TermStar and Transit are translation and terminology management products of a multinational company, the STAR. The company is involved, for example, in translation and localization industry, and terminology management. TermStar is a MS Windows application, and among its many features are the following (citations slightly altered from the web site <http://www.star-ag.ch/products/termstar/termfea.htm>):

- A full set of data fields is predefined according to the latest findings in terminology research. User-definable fields are also available. Use all of the available fields or any convenient combination of fields. Any pre-defined field can be re-named.
- Simultaneous search in multiple physical databases. The terms from each database are automatically merged and sorted on the display, while still physically residing in their original separate database files.
- Concept-oriented: TermStar opens a new data record for every concept, not for every term. A concept can contain multiple terms and language versions for one definable object, characteristic or action.
- Easy creation of extracts and/or subsets of the terminology database (e.g. for the creation of a list of abbreviations, specialized dictionaries on subject areas, project dictionaries).

As we can see from the above list, the TermStar appears to be a very sophisticated terminology management program. According to a translator in a Finnish translation agency in Turku, TermStar outweighs MultiTerm in several places: for example, TermStar has a more illustrative user interface, browsing through different terminologies is easier (cf. the second list item above), and different fields with different terminologies can be brought together to form a project-like combination. Moreover, TermStar has understood the need for synonym management, illustrations, and printing properties which MultiTerm has yet to demonstrate. Unfortunately, TermStar is also a heavier product than MultiTerm, as well as more expensive. Because at the time of writing I have no personal experience in the STAR products, I cannot give more detailed review of the TermStar program. Next, however, back to the responses from the respondent number 8.

### 6.9.2 Other Responses

The MultiTerm program, then, received a grade *good* in usability, although the respondent stated that the recording of terms takes a lot of work, and term recognition is not working if the terms are not in their right form. Most likely the term recognition feature is connected with the translation memory program than can search terms from the terminology database while translating. If a term appears in the translated text in a different form (for example, *mikrophone* -> *microphone*) from the term in the database, the recognition will fail. MultiTerm is capable of so-called *fuzzy search* where terms with some similarity are looked up. However, the fuzzy search is a feature which cannot be activated from the translation memory program, but has to be used from MultiTerm itself. The Transit program, on the other hand, received the grade *good* as well, but the user interface was said to be somewhat inconvenient. Other term bases or terminology management programs were not mentioned. Next, a list of the rest of the responses, and a review after the list.

<b>Question 9/Response:</b>	Yes.
<b>Question 10/Response:</b>	10.
<b>Question 10b/Response:</b>	Source of term, annotations, client, field, context.
<b>Question 11/Response:</b>	Average.
<b>Question 12/Response:</b>	Enough.
<b>Question 13/Response:</b>	Has a link always.
<b>Question 14/Response:</b>	No.
<b>Question 15/Response:</b>	1.
<b>Question 17/Response:</b>	Often.
<b>Question 18/Response:</b>	Well in general.
<b>Question 19/Response:</b>	Describes the concept moderately.
<b>Question 20/Response:</b>	The terms are consistent.
<b>Question 21/Response:</b>	Seldom compatible.
<b>Question 22/Response:</b>	Seldom.
<b>Question 23/Response:</b>	Taken care of fairly well.

Since **Question 9** was not worded accurately enough, I had to conclude that the respondent meant that she uses her own models to make the entries. This assumption is backed up by the fact that the fields she gave in the response to **Q10b** are not directly found in, at least, MultiTerm. Compared to TermIt, which has over 10 pre-named fields to choose from, MultiTerm has only few. Both MultiTerm and TermIt allow the changing and addition of fields in the existing

field list. In general, the manageability of the term bases in the response set 8 was deemed average, and the generally consistent terms described their concepts moderately. Since there is only one person responsible for maintaining the terminology, no need for access restrictions are necessary. As far as the usability of term bases was concerned, improvements to the recording of terms and user interface were hoped.

### 6.10 Respondent no. 9

As a translator in a private enterprise, the 9<sup>th</sup> respondent uses terminology for his personal purposes in Finnish, Swedish, English, and, exceptionally, *Danish* and *Norwegian*. The programs and term bases that the respondent uses are MultiTerm, Euterpe, and Eurodicautom. The only TM program that received a grade was MultiTerm, and the respondent thought it to be only average. He attested that the program is hopelessly archaic, and database management and updating requires a considerable amount of work. Moreover, the product development lags behind, which is true because the current Edition 3 of the MultiTerm program does not differ significantly from the Edition 2 in the most important aspects that are immediately apparent to a user. Most of the shortcomings of MultiTerm have already been discussed.

The respondent had also some positive points to give: MultiTerm operates swiftly (the search times are short, for example), and the import function helps reasonably easy database importing from other formats. The other formats, though, are currently only an ANSI<sup>31</sup> text format (.txt), and an ASCII text format (.tx8). However, one can create an importable database, for example, in MS Word with the required header information, convert the document to the ASCII format, and then import the data into MultiTerm. Most commercial applications support conversions to a plain text format. The rest of the responses are shown below:

<b>Question 9/Response:</b>	Yes.
<b>Question 10/Response:</b>	4.
<b>Question 10b/Response:</b>	Target term, source term, co-text, additional information.
<b>Question 11/Response:</b>	Poor.
<b>Question 12/Response:</b>	Enough.
<b>Question 13/Response:</b>	Has a link always.

The field count is not very high in this database, but it is probably sufficient for the needs of the respondent. An interesting data category, namely *co-text*, is found among the categories. In most of the cases I have seen, the field is *context*, but both expressions are valid, usually referring to the same thing: the surroundings of the term. Co-text, according to Crystal (1997: 96), is a term used by some British linguists in order to make a difference between a linguistic context (= co-text), and the situational environment (= context).

MultiTerm received its most critical review by the 9<sup>th</sup> respondent. For the first time, the management of terms was graded *poor*. Some of the users are evidently more critical than others, but the respondent did not give any concrete suggestions for improvements. His only wish was that adding the terms should be possible directly, and not through several stages, but that wish captured the essence of an easy term management. In my opinion, the trend should be towards more visual planning and organizing of terminology records, for instance, in a way of desktop publishing where the page elements can be dragged and dropped anywhere on the page. This free-form management where the results are seen immediately is possible also with word processors, and it is called the WYSIWYG feature. The recently coined acronymic term monstrosity is short for *What You See Is What You Get*, and it means simply that. In fact, MultiTerm is closer to the WYSIWYG concept than TermIt, for example, but still far away from ideal.

<b>Question 14/Response:</b>	No.
<b>Question 15/Response:</b>	1.
<b>Question 16/Response:</b>	Everyone is able to add, change, and edit the data in the database.
<b>Question 17/Response:</b>	Often.
<b>Question 18/Response:</b>	Well in general.
<b>Question 19/Response:</b>	The term describes the concept moderately.
<b>Question 20/Response:</b>	The terms are consistent.
<b>Question 21/Response:</b>	Usually compatible.
<b>Question 22/Response:</b>	Seldom.
<b>Question 23/Response:</b>	Being taken care of well.

There is nothing new in the data of the responses to **Questions 15—23**, and the only noteworthy thing is the response to **Q21** when compared with **Q14**: although the respondent has not been following any standard in creating the databases, the clients' databases were indicated to be usually compatible with the respondent's



database. To clarify this, I would first like to straighten the possible misunderstanding that might have been caused earlier: translators do not only create term bases for clients (creating terminology is not the sole domain of terminologists) - the clients have often their own databases which they send to the translators to aid in the translation process, and to keep the terminology consistent and following the clients' conventions.

If the compatibility seems strange in the light of the "missing standard", it is not probably so strange any longer when we talk only about the terminology management software, and not the term base itself. Therefore, there is always a compatibility between the TM programs if the client's and the translator/terminologist's software are the same. The difference then lies in the structure of the databases, and in the variation in field names, field count, etc. Next, I will move on to the last respondent, after which I will draw a summary on the observations from the 10 respondents.

### **6.11 Respondent no. 10**

The last respondent is a professor in a Finnish university. In addition to his academic career he is a technical writer, and the latter profession involves dealing with terminology with clients. The languages that the respondent uses are Finnish, Swedish English, German, French, and Russian, and the terminology management programs and databases are TermIt, MultiTerm, Eurodicautom, Tepas, and *Walter3*. The last term base, more specifically *Valter3*, is available in the internet with a slightly different name (see the link after bibliography), and it is a project of the *Department of Translation Studies* of the University of Helsinki. The *Walter3* is a relation database that, according to the respondent, does not have limitations to the number of data categories that can be used in the database (response to **Q10**).

In *Walter3*, the user can make searches from the available domains (for example, *medicine, law, government*) on a term, text, or on a concept basis. A part of the term, concept, or a text can be also searched. The basic features of the database are available to every internet user, but the advanced features require a username and a password. Although the respondent graded the *Walter3* system as *very good*, the first impressions could be quite different to other users not familiar with it. Since I have no fuller experience on the *Walter3* term base, it is better left for future to

analyze the features of the system. The other term base management program was MultiTerm, but it received only an average grade.

- Question 9/Response:** Yes. (Use of own models)  
**Question 11/Response:** Good. (Term manageability)  
**Question 12/Response:** Enough. (Data categories)  
**Question 13/Response:** Has a link always. (Reference to the context)  
**Question 14/Response:** No. (No standard)  
**Question 14c/Response:** There is a possibility to import and export terms that are represented in different standardized formats.

I have to assume that the respondent is referring to the Walter3 system, first because of the response to **Q10**, second because Walter3 was mentioned as the primary term base in **Q8** (see the question phrasing on page 41), and third because of the response to **Q14c**. The data categories are not limited, which allows the user to add any field that is necessary for the current purposes. The links are rather only references to the sources of context or the source of the definition. The import/export functions, however, are not available to "non-members", that is, for people that do not have proper access rights.

- Question 15/Response:** 1. (People responsible for the use of the database)  
**Question 16/Response:** Access rights are limited.  
**Question 17/Response:** Often. (Found terms)  
**Question 18/Response:** Well in general. (Term explanation)  
**Question 19/Response:** Describes the concept moderately. (Concept descriptiveness)  
**Question 20/Response:** The terms are consistent.  
**Question 21/Response:** Usually compatible. (Data compatibility)  
**Question 22/Response:** Regularly. (Term formation)  
**Question 23/Response:** Taken care of fairly well. (Quality control)

The remaining responses (15 - 23) show that the term base in question is of an average quality, at least for the time being. The terms are found often, but not always; they are explained generally well, and describe their concepts moderately well. The Walter3 term base appears to be concentrated more on the field of language and translation, since random searches with words and terms related to these fields, as well as to the field of terminology, produced more matches than searches with words/terms related to, for example, technology. I have to assume in **Question 22** on term formation that the respondent regularly uses pre-existing terminology. The references to the sources of terms and their definitions seem to

confirm the assumption. In general, the Walter3 term base offered no exceptional new insights to the field of terminology management, but it has a potential to become a very usable term base.

This concludes the analysis of the responses to the questionnaire. In the next chapter, I will briefly summarize the observations and reviews by the respondent. The **Chapter 8** will have a backward glance on the present study, as well as discussions on the questionnaire on the basis of the feedback from the respondents.

## **7. A Summary of the Data**

In this chapter, I will summarize the observations from the 10 different respondents. Although the amount of data did not lend itself to a statistical analysis, the terminology management programs and databases that the respondent evaluated have shown that there is indeed something to improve as regards the usability of term banks.

As a total, there were 10 different term banks (I count an internet database including several individual term banks as one) that the respondent mainly used. Of those 10, the most common ones were MultiTerm, Eurodicautom, Euterpe, Tepas, and TermIt. By far, *MultiTerm* was the most widely used TM program (9 of 10 respondents), and the *Eurodicautom* database was used by all (10 of 10) respondents. The Finnish *Tepas* was also very popular: 7 of 10 respondents have used it.

Because MultiTerm was the most popular TM program, it received most of the criticism. First, the data (terms) could not be properly exported or printed in a desired format. The ability to print and export was a feature that has yet to be incorporated in MultiTerm. Another TM program, namely the Transit/TermStar, had sophisticated export and printing features, although the demo version I received from the STAR Company did not allow the exploration of these features. Second, the management of synonyms was also inadequate in MultiTerm. Third, the actual entry of terms into the database was considered too complex and slow. Compared, for example, to TermStar, where the *Rapid Entry* mode makes on-the-fly term entries possible, MultiTerm lags behind enormously. Moreover, the restrictions to the number of possible fields in the databases were seen as a thing which should be corrected.

The most commonly used internet-based term bank was Eurodicautom. Because the amount of data in the entire database, the terms that I searched in random were usually found in Eurodicautom. However, one has to bear in mind that also TermIt and MOT are quite good terminology management programs, because both programs can have many separate terminology bases. These bases include, for example, terminologies of technology and trade in various languages, such as French, Swedish, English, Russian, etc. The MOT program, on the other hand, has a number of optional dictionaries as well, for example, financing and business, electricity, steel industry, and so on. The special dictionaries have to be ordered separately, though, so they are not available immediately to everyone. Another obvious disadvantage with the non-internet terminology management systems was that the possibility of using graphics was almost non-existent. TermIt and MOT did not have the feature at all; MultiTerm allowed having some graphics with the entries, but the graphics files have to exist in a directory on the computer hard drive or in the network, and they cannot be embedded.

There were also positive points in the TM programs that have to be mentioned. For example, MultiTerm was considered a good terminology management tool when multiple people operated it over the network; one respondent thought of it as a simple and clear tool for retrieving terms, and another respondent stated that the lookup functions in the MOT program were flexible (word-fragment search). In the public term bases, the Eurodicautom was good because it was multilingual and very extensive. As stated earlier in the present study, one main criteria of a good terminology database is how comprehensive it is, and how often the required terms are found. The Euterpe database is fairly comprehensive as well, and what makes it even more usable is that the users can download the current database in their computer, and use it with a special lookup module (a small program) that can be downloaded as well.

As mentioned earlier, standardizing terminology makes better interchange of data possible. In spite of that, not all respondents indicated that they have any kind of standard in their database. That is not to say that the TM program itself would not follow any standard, but the question is more about whether the structure of the database is such that it can be easily imported in another TM program. The terminology databases on the Internet are usually based on some standards, either SGML, MARTIF, NTRF, or the like, and the trend should be that more and more

databases that companies or individuals create, followed a recognized standard. A so-called *blind interchange* is possible, but it is not certain whether all data is compatible.

Different respondents had different opinions on the functionality of the terminology management programs and term bases, and some of the responses were quite subjective. Moreover, there were not enough responses to expand the observations to present a wider scope of users. Some conclusions can be drawn from the responses, however. First, most of the currently available TM programs need to be worked on and improved, although there are programs that have clearly succeeded in product development. Second, the term bases on the internet are usually graded on how much data they contain, how well they are organized, and how professionally the terminology is maintained.

By and large, the present survey has, in my mind, fulfilled the expectations that were imposed on it - at least to a degree. The purpose was to find what pros and cons there were in terminology management programs and terminology databases, and that is what was rather successfully accomplished. One primary criteria of a meaningful scientific study is that its results can be generalized. This, however, was not possible in the present study. The responses that were dealt with in **Chapter 6** will hopefully give at least some indication towards the focal points in future development of terminology management programs. Unfortunately, the public term bases received too little criticism in order to be properly evaluated. We have to test for ourselves the quality and usability of different internet-based term banks. A good comparison can be made between, for example, the international Eurodicautom database and a domestic Walter3 or Tepas term banks.

In the next chapter, I will take a look back, and evaluate the present study briefly. I will also introduce some feedback from the respondents. **Chapter 9** is reserved for some suggestions on the usability of term banks, after which the present study will be completed in **Chapter 10**.

## **8. A Look Back**

In this chapter I will introduce the feedback from the respondents about the questionnaire, and discuss the probable reasons why the survey did not produce as

many responses as was expected. First, however, the feedback on the questionnaire.

In the questionnaire I assumed that the term bases consisted of records of fixed size, and one respondent also noticed that. He indicated that the record sizes could vary considerably depending on the information that is to be put in. Also, the level of information varies, because the term bases are built in different ways, and it is not always possible to have every record as all-inclusive as every other record. The respondent also considered the questionnaire too limited, and thought that there could have been more possible responses than there was. I admit that the questionnaire was somewhat limited in almost every sense, but for the most part there were reasons for this. One reason was that if I had included more pre-formed answers and selections, the questionnaire would probably have become so large that even the 10 respondents would have had difficulties filling out the form because they would probably have become exhausted or frustrated. In addition, the questions were formed so that the respondents might have had the idea that record sizes in the databases were fixed, although that was not my intention.

Another respondent said that she could not see the text she typed in the text field as a whole. That is clearly a technical problem, since the so-called *word wrapping* setting did not move the text to a new row when the right edge of the text box was reached. That problem will be corrected in future questionnaires, and it was not very significant problem after all. What was significant was the formation of certain questions. In **Q9**, "*Do you use your own models to make entries and categorize terms, headwords, definitions etc., or do you use the models that already exist in the terminology management program?*" there were actually two questions in one. This type of question is extremely difficult to answer since the respondent cannot necessarily decide to which part I wanted him/her to answer. Answering *yes* would mean that the respondent uses own models, and not pre-existing models, or, that he/she does not use own models but pre-existing models. The *no*-response would also be difficult to decode. In **Q22**, "*Do you form your own terms, or do you use prepared terms that have already been validated?*" there was the same problem. I could have either made two separate questions instead of one, formed the questions differently, or formed the response choices so that the respondent could have chosen which part of the question to answer. For example: *yes, we form our own terms*, or *no, we do not form our own terms*.

It is clear in retrospect that the questionnaire could have been better. However, it worked. I cannot say whether there would have been more responses, had the questionnaire been sent by ordinary mail. An internet-based questionnaire will continue to defend its position, although the same response rate is probably not possible as with more ordinary questionnaires. It would be a fruitful research topic to study which type of questionnaire elicits the best results, but the present study was not for that purpose. The purpose was to get information on the state of terminology management and the usability of terminology management systems, and that was, at least in my opinion, accomplished. Whether the results will produce any improvements to future term bases and TM software will be seen. When more and more studies like the present one are conducted, the manufacturers and terminologists will gain a better understanding on what is going on in the world of terminology, what users really need, and what they do not need in a TM system.

The reasons why the questionnaire did not produce as many responses as I should have hoped or as should have been necessary to make the present study statistically significant can be various. One reason could be that I limited the survey within the boundaries of Finland as well as the fact that I did not contact hundreds of potential respondents. What I assumed when I sent an e-mail to a large company or institution was that the recipient of the e-mail message would have spreaded the information to other workers of the company, and one contact would have produced responses from several people. Furthermore, the discussion groups I sent the message about the survey were available to many people. However, the interest in the questionnaire seemed to be exceptionally low. Apparently there are too many questionnaires going around so that people get frustrated to answer most of them. Moreover, many terminologists and translators have their hands full of work already even without having to answer to questionnaires. I did not start to call people, either, to make them answer. One reason was that I wanted to save expenses. Moreover, it was not certain whether the respondents were reachable by phone even during the office hours. I assumed that most of the potential respondents read their e-mail regularly or followed the newsgroups; that way I could cover more ground than by making "random" phone calls.

**Chapter 9** will give suggestions for improving the terminology management systems and databases. Some of the suggestions are based directly on the responses

and some are my personal opinions or conclusions. That is, they do not present the opinions of anyone else.

## **9. Future Suggestions on the Usability of Term Banks**

This chapter will tie all the strings together, and suggest how the term banks and the TM programs could/should be improved in the future. First of all, the developers should be aware that the needs for good terminology and terminology management software increase every day. Good and consistent terminology is vital since people have to deal with a variety of information from a variety of fields. The educators need to stay up-to-date with the information they share with their students, technical and special terminology have to serve the needs of manufacturers, customers, terminologists, localizers, and so on. The European Union will produce tons of new information that include terminology of legislation, agriculture, aviation, various scientific disciplines, to name a few. The terminologists of the Union will have to handle new and unfamiliar terms as well as old ones now more than ever, and the translators are facing an enormous task in producing translations for the languages of the member countries.

The usability of term banks is a combination of many factors. It is not enough that the term bank is extensive, although that is one of the major goals of a truly useful one. The terms have to be properly collected, and presented accurately and consistently. Furthermore, a usable term bank should contain enough information on the contexts in which the terms are used, their definitions should be sufficient and precise, and the semantic relationships with other terms (by synonymy, etc.) should be indicated when necessary. Overdefinitions (defining too comprehensively) and irrelevant information should be avoided to maintain a clear and compact construction of the entries. The semantic relations might sometimes need the help of an illustrative picture, and therefore a terminology management program should have a feature that allows the addition of graphics. The need for graphics was also indicated by the respondents, as well as a possibility to make printed dictionaries, or print out selected portions of the term base for instance for proofreading.

The number of data categories should not be limited either. The user should have a freedom to use as many categories as required, especially when there are different terminologies and situations that call for different size records. A single term does



not always need a complex construction of data categories: a sufficient entry might consist of just the source and target terms in desired languages, a definition, and a reference to the context the term is found. When, for example, one needs to know who has made the entry, what is the subject of the term, whether the term is validated etc., it should be possible offer that information.

In the internet, the terminology databases serve a wider scope of users than do the individual programs. Therefore it is essential that the databases are clear to use and have sufficient instructions. In many cases, though, the visuality and fluency of operation is far from ideal, albeit that the data in the database were extensive. As in the electronic questionnaires, the user of an internet terminology database types the terms in the fields. For instance, all the fields in a term lookup screen should have proper *legends*<sup>32</sup> that show what the fields are for, and the possible selections should be clearly indicated so that the user does not have to guess how the lookup works. In a sense, the interface of the database should be basically foolproof.

Both types of terminological databases - private and public - need quality control. Private databases here mean those that are made for customers. It is, however, more difficult to maintain quality if there is only one person responsible for the management of terminology. Therefore, at least one people in addition to the database administrator should check the consistency and accuracy of the terminology, although in one-person private enterprises that is not always possible. A freelance translator, for example, should have more than one source of reference for the terms he/she uses in translations in order to make sure that the terms are correct. The safest way is to use validated term services and databases such as the Eurodicautom or Tepas.

One final suggestion for improvement deals with the terminology standards. Good standards, such as the MARTIF or NTRF are already available, but not used nearly enough. Terminology conventions, for instance, help to increase the awareness of standardized terminology and data formats. Terminology projects should concentrate on developing term records in a standardized format to facilitate the interchange of data, and ease the processing of data by making it compatible with the most common database management systems. Basically, all that is required is little programming.

These are the crucial points for improving the quality of TM systems and databases. However, there are many more aspects to the usability of terminology management, but only a fraction of them can be approached in the present study. Hopefully at least some of the main issues were dealt with here, and hopefully the results benefit at least some people involved in terminology management. More studies will be conducted, and they could produce more generalizable results than the present one.

## 10. Conclusion

I have summarized most issues concerning the present study in the two previous chapters. I have not, though, reviewed the research process or my own success/lack of success in the research. In general, the completion of the thesis has taken a little over a year, which is probably the average time these types of treatises take to write. I did not experience any significant problems during the research process, but in the beginning there were some difficulties in finding the right approach. I launched the survey with a anticipation to find enough statistical data. This was not, however, possible since the questionnaire did not produce enough responses.

One of the most interesting things in the whole study was the construction and implementation of the web-based questionnaire, firstly because I have not done that before, secondly because I decided to build the questionnaire from scratch and even make a script that processed the internet form. I looked for some examples in earlier questionnaires, of course, but they were not directly applicable to my purposes. I received important feedback from the respondents, although there could have been more. In general, the respondents seemed to be satisfied with the questionnaire. The small-scale pilot produced necessary suggestions for the final version of the electronic form.

There is plenty of material on terminology and terminology management, but unfortunately it was not readily available. For instance, the *Handbook of Terminology Management*, vol. 2 has appeared in print only recently, and the first volume had to be borrowed as an interlibrary loan. Although the University of Jyväskylä is not involved in terminology studies, the two volumes would be a valuable addition to the books of the library. I have a strong wish that more literature dealing with terminology and translation should be purchased by the libraries of the university.

A professor at the State University of Kent, Ohio, Ms. Sue Ellen Wright has been an indispensable source of information and support. Without her, vital points of the present study would not have been written. As the field of terminology management was rather unknown to me before the writing of the thesis, I had to acquaint myself with it. Hopefully after the completion of the present study I will be at least a little more knowledgeable on the issues of terminology than before. Furthermore, I have had the chance to learn more about terminology management programs and databases previously unfamiliar to me. That will surely help me in my work as a freelance translator.

## Bibliography

- Berdie, Douglas R. 1986: *Questionnaires - Design and Use*. New Jersey: The Scarecrow Press.
- Bessé, Brunno de 1997. *Terminological Definitions*. In *Handbook of Terminology Management*, vol. 1. Philadelphia: John Benjamins B. V.
- Collins Cobuild English Language Dictionary* 1987. London&Glasgow: William Collins&Sons Ltd.
- Crystal, David 1997. *A Dictionary of Linguistics and Phonetics*. Cornwall: Blackwell Publishers.
- Dubuc, Robert & Andy Lauriston 1997. *Terms and Contexts*. In *Handbook of Terminology Management*, vol. 1. Philadelphia: John Benjamins B. V.
- Heikkilä, Tarja 1998: *Tilastollinen tutkimus*. Helsinki: Edita.
- Hirsjärvi, Sirkka & Pirkko Remes, Paula Sajavaara 1997: *Tutki ja kirjoita*. Tampere: Kirjayhtymä.
- Jaakohuhta, Hannu 1999: *Suuri tietotekniikan tietosanakirja*. Jyväskylä: Gummerus.
- Kalliokuusi, Virpi 1999: *Määrittelyn monet kasvot*. In *Toimikunnista termitalkoisiin*. Helsinki: Yliopistopaino.
- Kuhmonen, Kaisa (ed.) 1999: *Toimikunnista termitalkoisiin: 25 vuotta sanastotyön asiantuntemusta*. Helsinki: Yliopistopaino.
- Lyons, John 1977: *Semantics I*. Cambridge: Cambridge University Press.
- Mattila, Hannu 1999: *Lomakekysely asiakaspalautteen hankkimismenetelmänä*. Esimerkkinä Kelan asiakaspalvelututkimus. KELA: Sosiaali- ja terveysturvan tutkimuksia 48. Helsinki: Kansaneläkelaitos, tutkimus- ja kehitysyksikkö.
- Melby, Alan & Klaus-Dirk Schmitz and Sue Ellen Wright 2000: *Terminology Interchange*. In *Handbook of Terminology Management*, vol. 2. Philadelphia: John Benjamins B. V.
- Nykänen, Olli 1999. *TSK - 25 vuotta sanastotyön asiantuntemusta*. In *Toimikunnista termitalkoisiin*. Helsinki: Yliopistopaino.
- Oppenheim, A. N. 1966: *Questionnaire Design and Attitude Measurement*. New York: Basic Books, Inc.
- Pyles, Thomas & John Algeo 1993: *The Origins and Development of English Language*. USA: Harcourt Brace College Publishers.
- Sager, Juan C. 1997. *Term Formation*. In *Handbook of Terminology Management*, vol. 1. Philadelphia: John Benjamins B. V.
- Strehlow, Richard A. 1997: *Frames and the Display of Definitions*. In *Handbook of Terminology Management*, vol. 1. Philadelphia: John Benjamins B. V.
- Suonuuti, Heidi 1997. *Guide to Terminology*. Nordterm 8. Helsinki: Invapaino.
- The Handbook of Terminology Work - Principles and Methods of Applied Terminology Science 1988. Tekniikan sanastokeskus/SFS Handbook n:o 50. Jyväskylä: Gummerus
- Vala, Ursula & Heli Hartikainen, Matti Klockars 1988. *Moderni Lääketieteen Sanasto*. Hanko: Recallmed.

Wright, Sue Ellen & Gerhard Budin 1997. *Handbook of Terminology Management, vol. 1. The Basic Aspects of Terminology Management*. Philadelphia: John Benjamins B. V.

Wright, Sue Ellen & Gerhard Budin 2000. *Handbook of Terminology Management, vol. 2. Applications-oriented Terminology Management*. Philadelphia: John Benjamins B. V.

Wright, Sue Ellen 1997. *Basic Principles of Terminology Management*. A PowerPoint Presentation. Kent State University, Ohio USA: Wright and Associates.

Internet links mentioned in the text

**Eurodicautom/Info:** <http://eurodic.ip.lu/cgi-bin/edicbin/EuroDicWWW.pl?page=Support/presentationfr>

**TIS (Terminological Information System):**  
<http://tis.consilium.eu.int/utfwebtis/frames/introfsEN.htm>

**Euterpe:** <http://muwi.trados.com/>

**TEPA:** <http://www.tsk.fi/>

**TERMIUM:**  
[http://www.translationbureau.gc.ca/pwgsc\\_internet/english/03\\_tools/03\\_termium.htm](http://www.translationbureau.gc.ca/pwgsc_internet/english/03_tools/03_termium.htm)

**TermStar/WebStar:** <http://www.star-ag.ch/webTerm/webterm.htm>

**Walter3:** [http://www.rosetta.helsinki.fi/~kkl\\_test/](http://www.rosetta.helsinki.fi/~kkl_test/)

---

<sup>1</sup> Referent - tarkoite

<sup>2</sup> concept - käsite

<sup>3</sup> term - termi

<sup>4</sup> definition - määritelmä

<sup>5</sup> set theory - joukko-oppi

<sup>6</sup> subordination - alisteisuus

<sup>7</sup> concordance - yhtäpitävyys, vastaavuus, konkordanssi

<sup>8</sup> equivalent - vastine

<sup>9</sup> terminologization - termittäminen

<sup>10</sup> compounding - yhdistäminen

<sup>11</sup> derivation - johtaminen

<sup>12</sup> borrowing - lainaaminen

<sup>13</sup> abbreviation - lyhentäminen

<sup>14</sup> conversion - konversio

<sup>15</sup> coinage - tekosanat (here)

<sup>16</sup> quotation loans - sitaattilainat (here)

<sup>17</sup> special loans - erikoislainat (here)

<sup>18</sup> general loans - yleislainat

<sup>19</sup> genus - suku

<sup>20</sup> subgenus - alasuku

<sup>21</sup> species - laji

<sup>22</sup> differentia - tunnusmerkki

<sup>23</sup> property - ominaisuus

<sup>24</sup> accident - satunnainen ominaisuus (here)

<sup>25</sup> glossary - erikoissanasto

<sup>26</sup> granularity - rakeisuus (here)

<sup>27</sup> sample - otos

<sup>28</sup> browser - selain (e.g. Internet Explorer, Netscape)

<sup>29</sup> the General Secretariat of the Council of the European Union - Euroopan unionin neuvoston pääsihteeristö

<sup>30</sup> wildcard - jokerimerkki (for example, \*, ?, #)

<sup>31</sup> ANSI - American National Standards Institute

<sup>32</sup> legend - seloste