Masaryk University
Faculty of Arts

Department of English and American Studies

English Language and Literature

Kamila Králíková

Basic Aspects of Terminology Management
Bachelor's Diploma Thesis

Supervisor: PhDr. Jarmila Fictumová

2010
I declare that I have worked on this thesis independently, using only the primary and secondary sources listed in the bibliography.
Acknowledgements

I owe a debt of gratitude to my supervisor, PhDr. Jarmila Fictumová, for her valuable help and kind guidance, support and patience. I would also like to thank PhDr. Kateřina Gašová and Ing. Petr Jandek from Moravia Worldwide and to Mr. Tom Smith from SDL International for providing me with helpful materials and advice.
Table of Contents

Table of Contents ........................................................................................................... 1
List of Abbreviations ........................................................................................................ 2
Glossary of Terms ............................................................................................................ 3
1. Introduction ................................................................................................................ 5
   1.1 Current situation on the translation and localization market ............................. 8
2. Terminology .................................................................................................................. 8
   2.1 Where does terminology come from? ............................................................... 9
   2.2 What is terminology? .......................................................................................... 9
   2.3 The fundamental pillars of terminology work .................................................. 11
      2.3.1 Objects ....................................................................................................... 11
      2.3.2 Concepts and characteristics ..................................................................... 12
      2.3.3 Terms, definitions and subject fields ......................................................... 14
   2.4 Terminology management .................................................................................... 15
3. Why to use terminology? ............................................................................................ 16
   3.1 Need for consistency ........................................................................................... 16
   3.2 Perils of ignoring terminology management .................................................... 18
4. How to create terminology? ........................................................................................ 19
   4.1 Terminology extraction ....................................................................................... 19
      4.1.1 Manual extraction ....................................................................................... 22
      4.1.2 Automatic extraction ................................................................................. 23
      4.1.3 Statistical vs. linguistic method of terminology extraction ....................... 25
      4.1.4 Demonstration of the statistical and linguistic term extraction using SDL
         MultiTerm Extract 2009 and SDL PhraseFinder 2005 ...................................... 26
         4.1.4.1. Extraction using SDL MultiTerm Extract 2009 ................................. 27
         4.1.4.2 Extraction using SDL PhraseFinder 2005 ........................................... 32
      4.2 The role of a terminologist ............................................................................... 36
   4.3 Getting the right terms to the right people .......................................................... 37
      4.3.1 Offline sharing .............................................................................................. 37
      4.3.2 Online sharing ............................................................................................. 38
5. Globalization, translation, localization and related issues ........................................ 39
   5.1 Globalization ......................................................................................................... 40
   5.2 Internationalization ............................................................................................. 40
   5.3 Localization ........................................................................................................... 41
   5.4 Challenges of localization .................................................................................... 42
      5.4.1 Cultural and colloquial references ............................................................... 43
      5.4.2 Geopolitical issues ....................................................................................... 44
      5.4.3 Colours ........................................................................................................ 45
      5.4.4 Graphics, icons, symbols ............................................................................ 45
      5.4.5 Currency and number format ..................................................................... 46
      5.4.6 Date and time format .................................................................................. 46
      5.4.7 Name and address format ........................................................................... 47
      5.4.8 Abbreviations and acronyms ..................................................................... 47
      5.4.9 System of measures ..................................................................................... 48
6. Standards and standardization organizations ............................................................. 48
7. Conclusion ..................................................................................................................... 51
Abstract .......................................................................................................................... 53
Anotace ............................................................................................................................. 54
Primary sources ............................................................................................................... 55
Bibliography ..................................................................................................................... 55
Appendix .......................................................................................................................... 62
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>COPRAS</td>
<td>The Cooperation Platform for Research and Standards</td>
</tr>
<tr>
<td>I18N</td>
<td>Internationalization</td>
</tr>
<tr>
<td>IAMLADP</td>
<td>The International Annual Meeting on Language Arrangements, Documentation and Publications</td>
</tr>
<tr>
<td>ICTSB</td>
<td>Information and Communication Technology Standards Board</td>
</tr>
<tr>
<td>ISO</td>
<td>International Organization for Standardization</td>
</tr>
<tr>
<td>L10N</td>
<td>Localization</td>
</tr>
<tr>
<td>LGP</td>
<td>Language for general purposes</td>
</tr>
<tr>
<td>LISA</td>
<td>The Localization Industry Standards Associations</td>
</tr>
<tr>
<td>LSP</td>
<td>Language for special purposes</td>
</tr>
<tr>
<td>NASA</td>
<td>National Aeronautics and Space Administration</td>
</tr>
<tr>
<td>OPE</td>
<td>One-person enterprise</td>
</tr>
<tr>
<td>S.I.</td>
<td>The International System of Units</td>
</tr>
<tr>
<td>SL</td>
<td>Source language</td>
</tr>
<tr>
<td>TCP/IP</td>
<td>Transmission Control Protocol / Internet Protocol</td>
</tr>
<tr>
<td>TKB</td>
<td>Terminology knowledge base</td>
</tr>
<tr>
<td>TL</td>
<td>Target language</td>
</tr>
<tr>
<td>TSP</td>
<td>Translation Service Provider</td>
</tr>
</tbody>
</table>
Glossary of Terms

**Globalization (g11n).** The process of dealing with business issues associated with launching a product or service globally.

**Internationalization (i18n).** The process of generalizing a product or service in such a way that it will not need to be re-designed for a particular locale.

**Language for general purposes (LGP).** Language used in everyday life for common communication and discussion on ordinary topics in various common situations.

**Language for special purposes (LSP).** A language that serves experts in a particular subject field for communication and discussion on issues concerning their specialised branch of knowledge.

**Lingua franca.** A language that is adopted as a common language between persons speaking different languages.

**Locale.** A set of characteristic features defining a particular language or region. It covers character encoding, writing system, keyboard layout, fonts, date, time and monetary formats and others.

**Localization (l10n).** A process of adjustment of a software product or service to a specific market or culture so it seems to be naturally produced in that region. It considers language, culture, customs and the characteristics of the target locale.
**Standardization.** The process of creating, establishing and adopting norms to achieve a consensus on terminology used in various areas of public concern.

**Term extraction.** The process of distillation of specialised vocabulary from a given corpus, either manually or automatically.

**Termbase.** A database or structured collection of terms, usually multilingual. It contains information about the terms, e.g. parts of speech, subjects, concepts, definitions, context sentences, acronyms and details.

**Terminology.** Terminological data or a specialised vocabulary used in a particular subject field.

**Terminology management.** A group of specific and specialised tasks primarily concerning handling, recording and presentation of terminological data for special purposes. It involves maintaining databases, creating and updating glossaries, creating new terms and finding equivalents of terms in a target language.

**Translation.** The process of converting texts from a source language (SL) into a target language (TL) while keeping the meaning of the original in order to convey the same message in the target language.
1. Introduction

The aim of this diploma thesis is to provide beginning translators and localizers with a basic manual how to address the issues of terminology and its management. It pursues the field of software localization and translation of technical texts with respect to specialised terminology. It provides readers with a basic overview of what terminology and its management deal with and suggests steps that should be taken into consideration before starting terminology work. The thesis should help beginning translators avoid some pitfalls of localization with respect to terminology. It also points out the importance and significance of terminology and proper management and suggests procedures that can be employed in order to streamline translation and localization processes.

Chapter 1 provides readers with a brief overview of impacts that terminology and communication have on society. It also describes the current situation on the translation and localization market with respect to terminology and its management.

Chapter 2 deals with theory and definitions. It explains terminology from the theoretical point of view and important factors that have to be addressed.

Chapter 3 discusses the issue of terminology and terminology management from the point of view of consistency. It explains why consistent terminology is needed and what consequences the inconsistent use of terminology could have.

Chapter 4 describes methods of terminology extraction and demonstrates the process using specialised terminology extraction tools. It also discusses the role
of a terminologist and summarizes advantages and disadvantages of two ways of terminology sharing.

Chapter 5 explains widely debated expressions, describes differences between translation and localization and points out the issue of internationalization. The chapter also advises beginning translators and localizers on problematic areas in localization that require close attention.

Chapter 6 briefly introduces the standards relating to translation, localization and terminology and organisations which develop such standards or participate in the effort.

Our developed civilisation relies on information technologies. Chemistry, mechanical engineering, aeronautics, transport, medical research – these are only a few examples of fields in which terminology plays a primary role. Even manned space flights are in large part driven by software. A great deal of human work is driven by computers, machines or devices and thus terminology is inevitably becoming an integral part of our lives. Some fifty years ago, we could only have dreamt about space flights. What was a desire half a century ago becomes reality today. We even consider tourists to be sent to the space for pleasure, for example. With the technological development and progress and the spread of the Internet, the human race is not comprised of separate communities any more. For all that, hundreds of nations and various languages exist and every single language has its own specifics concerning syntactical or morphological aspects and the use of specialised expressions according to their needs and geography. It is needless to say that proper communication is
essential for understanding between nations, particularly at the level of international cooperation.

Translators, localizers and proofreaders very often meet projects where both the source and target texts are inconsistent with respect to terminology. Such texts make translation and subsequent review rather challenging because translators or proofreaders are forced to look up suitable equivalents of terms in dictionaries, on the web, or – in the worst-case scenario – to take a guess. They therefore spend a lot of time doing other activities than translating or reviewing. This decreases the productivity of translators, localizers or proofreaders and increases the risk of poor results and dissatisfaction of the customer. And of course, these problems also carry higher financial costs for all parties involved. In some cases, translation or localization of poor quality, full of inaccurately selected terms, can lead to misunderstanding, bringing discredit to customers and may result in sanctions and financial recourse. Nowadays, more and more companies – and language service providers as well – realize the role that terminology management plays in everyday business, but unfortunately there are still companies that do not care or that are not aware of potential consequences. Since the world is driven by computers and information technologies, using consistent and standardized terminology becomes a necessity, not a luxury. For example, civil aviation has adopted English as a lingua franca and it is not difficult to imagine what could happen if air-traffic controllers and pilots of planes did not use standardized terminology when communicating. Thousands of human lives could be lost.
1.1 Current situation on the translation and localization market

In the beginning of this year, SDL International conducted Terminology Survey 2010, setting a target to obtain business professionals’ and translators’ opinions on the use of terminology and its management from within the business and translation processes. SDL International surveyed both customers and translators. In total, there were 1,707 responses. The survey revealed that even though 93% of business respondents considered brand consistency important for their business and were aware of the impact terminology has on their brand, only less than a half of respondents had a process in place to manage terminology. Similar results are shown in the field of translation. 92% of the translators surveyed considered terminology to be a very important part of the translation process and their everyday work. They also admitted they spent a lot of time and effort on managing terminology correctly and that the activities relating to research of terminology had negative impact on their productivity.

2. Terminology

The object of this chapter is to introduce the issues of terminology in general; the chapter provides readers with theoretical background and elements that closely relate to terminology and its management in practice. This chapter familiarises readers with the most important facts that have to be kept in mind before starting any terminological work.
2.1 Where does terminology come from?
Although it may sound strange, terminology has accompanied humankind from the very beginning. When a prehistoric man first made a tool to kill animals for food and gave it a name, he unwittingly created the first specific-purpose object, bound it to a particular concept, created a specialised term for it and thus in fact, with a slight overstatement, laid the foundations of a modern field of study. The first trail-blazers in the field of modern terminology were, among others, Andreas Vesalius\(^1\) (1514-1564), who is often referred to as the founder of modern human anatomy, or Carl von Linne\(^2\) (1707-1778), the father of botanical and zoological systematic nomenclature. Even though their terminology is more than five and three hundred years old respectively, it is still valid and used today.

2.2 What is terminology?
Terminology management primarily deals with terminological data, or terminology. When searching for a definition of terminology, a deadlock can be easily reached. There are many definitions and explanations of what terminology is. They vary, depending on the source, author, field of study in which it is used, or even on an individual point of view. It could be “a system of words used to name things in a particular discipline”\(^3\), “the study of concepts

---

\(^1\) [http://www.britannica.com/EBchecked/topic/626818/Andreas-Vesalius](http://www.britannica.com/EBchecked/topic/626818/Andreas-Vesalius)


\(^3\) [http://wordnetweb.princeton.edu/perl/webwn?s=terminology](http://wordnetweb.princeton.edu/perl/webwn?s=terminology)
and their representations in specialised language\(^4\), or “a set of designations belonging to one special language” (ISO 1087-1:2000). We can also find that terminology is “a specialised vocabulary of a field, the nomenclature”\(^5\). So, what definition of terminology is correct? In fact, any of the definitions mentioned above cannot be considered wrong. Broadly speaking, it does not matter how terminology as such is defined – it is more important what lies behind. The basic and principal function of terminology is to transfer knowledge, in the strict sense of the word to transfer specialised knowledge. The primary purpose of terminology is to enable professional communication. Were it not for professional communication, no knowledge transfer would be possible. And without knowledge transfer humankind would not be able to develop, neither intellectually nor materially. No education would exist; no professional research would be possible to conduct. Terminology ensures that we communicate in the right way and that we can understand each other. It represents a building block of an LSP and forms an integral part of it. Herbert Picht defines language for special purposes\(^6\) as “totality of all linguistic means used within a limited professional area of communication in order to ensure communication between people working in this area”. (24 May 2010: 4) In other words, an LSP is a special language which serves experts in a particular field for communication on issues relating to their specific subject field. Although an LSP is usually and broadly understood as technical and specialists are considered to be professionals, for some cases this necessarily does not have to apply. If a


\(^5\) http://www.termbases.eu/

\(^6\) based on Lothar Hoffmann’s model
woman is interested, for example, in beading and beadwork, she uses an LSP with specialised vocabulary relating to her hobby, such as *lobster claw, earnut, slide lock, dichroic glass pendant, headpin, bead tip* or *jumpring*. It can be said that “any language that is used to discuss a specialised subject can be considered an LSP” (Bowker and Pearson, 27). This has a logical implication that an infinitive number of LSPs can exist and each of them can have its own specialised terminology. Generally known as *terms*, the specialised vocabulary undeniably plays an important role in an LSP and represents a significant feature of it.

### 2.3 The fundamental pillars of terminology work

Objects, concepts, specific characteristics, subject fields, definitions and terms represent the fundamental pillars on which any terminological work stands. The more solid the pillars are, the higher quality and better results are achieved. The mentioned features closely relate to each other and are interconnected. Without them there would be no proper terminology work or management possible and thus no usable glossaries or term banks can be created. In this subchapter these features are discussed in more detail.

#### 2.3.1 Objects

For terminology work, defined in ISO 1087-1:2000 as “work concerned with the systematic collection, description, processing and presentation of concepts and their designation”, a point of departure is an object. No work with terminology
is possible without objects because they, whether real or existing in our mind, form reality. “They are fixed in time and space; they cannot change without losing their original character. A change will convert them into other objects”, writes Picht (25 May 2010: 3) Objects can be broadly subdivided into three general groups: material objects (e.g. a computer, a gold coin, a tree), immaterial objects (e.g. a currency rate, a trip plan) or imaginary objects (e.g. a naiad, an elf). Material objects exist independently of human beings, even though they could be produced or created intentionally or unintentionally by a human. Immaterial objects are described as “object[s] without physical form whose existence and relation to space and time is given through the imagining subject” (Picht, 25 May 2010: 6) and imaginary, or maybe rather imagined, objects exist just in our mind; they can represent our ideas about perfect holidays, for example.

2.3.2 Concepts and characteristics

Concepts directly relate to objects and their designations, and they are inseparable of them. As Pearson mentions, “concepts [...] are abstract entities isolated from text” (Pearson, 2). Existing only in our mind, concepts represent an idea of a particular object, but this idea depends on every single person since each individual can imagine a different object under a specific concept. Let us explain this on the example of the object car. Of course, it is generally known that a car is “a road vehicle for one driver and a few passengers” (MacMillan Dictionary). However, every single person’s idea about a particular car may be (and usually is) different; for instance, for a young single city
woman a personification of a car could be a slightly bigger, bright-coloured shopping trolley in contrast to a village farmer whose projection of a car will be probably a powerful, all-terrain vehicle. The concept is generally considered to be independent on a particular language. Pearson argues that “concepts exist [...] independently of any language” (Pearson, 10). This point of view is true only to some extent because concept systems of a SL and a TL do not always overlap and they can be (and actually are) influenced by the social or cultural background. Let us take the word *rice*. In English or Czech expressions for different types of rice can be found – some of them are white rice, long-grain rice, parboiled rice or husked rice, for example – but Asians, for whom rice forms a fundamental constituent part of their diet and is vitally important, know dozens of various types of rice. And each type of rice is bound to a certain concept, i.e. to a purpose for which the specific type of rice is used. Without determining an accurate concept which a particular term is bound to, translators and localizers can easily get to a crossroads and choose to continue in a wrong way in their work.

To describe concepts, determinative characteristics are used. They represent building blocks of knowledge. In ISO 1087-1:2000, a characteristic is defined as “abstraction of a property of an object or a set of objects”. To be able to understand a concept and to discern the specific concept from concepts associated with it, the use of characteristic is essential. The characteristic *long-grain* may differentiate concepts of *rice*. 
2.3.3 Terms, definitions and subject fields

As already stated above, terms or terminological units represent specialised vocabulary used in a particular subject field. ISO 1087-1 defines a term as “verbal designation of a general concept in a specific subject field”. In fact, virtually anything could become a term – starting from product and company names (Windows, Google), abbreviations (NY for New York or DC for direct current), through acronyms (GPS for Global Positioning System, AD for Alzheimer disease) up to marketing taglines or slogans (Just do it\(^7\); Melts in your mouth, not in your hands\(^8\)). To avoid ambiguousness and vagueness, each term should be clearly defined. As ISO 1087-1:2000 states, a definition is “representation of a concept by a descriptive statement which serves to differentiate it from related concepts”. As such, a definition describes the meaning of a term in relation to a specific subject field and must not contain synonyms of the term. Let us take a term *mouse*, for instance. MacMillan English Dictionary gives two definitions, each relating to a different subject field:

1. a small furry animal with a long tail
2. a small object that you move in order to do things on a computer screen

In the first example the subject field concerns the animal kingdom, while the second definition relates to a technical sphere - to be more precise, to information technology. The subject field has to be specified because failing that the term could be easily misapplied in translation or localization. As the example of a definition shows, the English term remains the same but the

\(^7\) Nike (1988)  
\(^8\) M&M Candies (1954)
concept it represents is completely different, although the striking visual resemblance between the mouse-animal and the mouse-device can be implied. Hand in hand with globalization and development of innovative and modern technologies and processes, there goes the need for creation of new concepts and terms.

2.4 Terminology management

Terminology work that is “concerned with the recording and presentation of terminological data” (Schmitz, 3) is called terminology management. It creates the foundations for all technical writing. Terminology management involves a range of demanding and time-consuming tasks, such as extracting terms, maintaining terminology databases, creating new glossaries, updating and maintaining existing glossaries, finding right equivalent terms in a TL, creating new terms and many other procedures. “Terminology management starts in the source language (SL), and then moves to the target languages (TL) and in the end all the work and data gets consolidates into one termbase” (Warburton, 8). This in essence follows that both SL and TL terminologists are involved and they have to cooperate on a single database which contains the SL terms and as many TL terms as needed. The SL terminologists deal with the source terminology, i.e. they identify terms, provide concepts and write definitions in cooperation with experts in the specific subject field; the TL terminologists build on the output of the SL terminologists’ work, i.e. they work with a list of terms set as key terms by the SL terminologists. In fact, the TL terminologists “just” find and add suitable equivalents in the TL to the same database or glossary.
3. Why to use terminology?
Having already set basic building blocks connected with terminology from the theoretical point of view and having been provided with a brief summary of the issue, the importance of terminology and terminology management is now addressed. This chapter explains why terminology should be created, maintained and managed properly; it focuses on consistency of both SL and TL content and discusses correlation between them.

3.1 Need for consistency
Consistency is a magic formula and at the same time it represents a nightmare for any person involved in technical writing. Where there is no consistency, translation of low quality (speaking about quality higher than low is completely inappropriate in such cases) can be expected. The use of the right terminology is crucial for the correct understanding and communication and hence translation quality. However, the quality of the translation goes hand in hand with a consistent and high-quality source text. In his article Dunne argues that “defective terminology represents a potentially significant source of risk in communication” and an incorrect, inconsistent or ambiguous source text and thus translation as well can lead to misunderstandings or incomprehension in communication. “Terminological inconsistency also causes confusion and undermines usability” (Dunne, 33). For instance, an ambiguous translation of an instruction manual for a medical device can result not only in a damage of the device, difficulties in operating or impossibility to use the device properly but even in serious or fatal injuries of people. “When terminology is inconsistent
across and within products, these products can become more difficult to use,” says Dunne (33). And Kara Warburton from IBM explains: “Because IBM products are developed in different laboratories, similar or identical concepts of objects have come to be associated with different terms. For instance, the terms shadow cursor, scale line cursor, grid cursor and scale cursor all refer to the same concept, but in different products” (Dunne, 33). The SDL Terminology Survey 2010 shows that only 12% of the business respondents acknowledge their terminology to be consistent throughout their business. And on the other hand, a total of 96% of the translators surveyed states that they meet with inconsistencies in the source content9. Keeping terminology clearly defined and consistent is not as easy as it could seem. In fact, this is a demanding, time-consuming and challenging task but sparing no effort on such work is worthwhile in the end. Maintaining consistency within terms in a TL is as important as in a SL. Let us suppose a company needs to localize specialised software and to translate thousands of pages of accompanying documentation and marketing materials in a very short time. The volume of work does not enable to hire one translator, localizer and proofreader rolled into one; there will be necessary to use services of several translators, localizers and proofreaders. Although being on the same level of proficiency in translation and localization, the translators, localizers and proofreaders may (and usually do) have different opinions on some target terms. The only possible and viable solution of keeping consistency within the texts being translated and software being localized represents the right and clear terminology, managed properly

9 40% frequently and 56% occasionally
from the very beginning – in optimal case, key terms are resolved before launching the translation and localization phases of the project.  

3.2 Perils of ignoring terminology management  
The ground rule for successful terminology work says that solely one term exists for one concept and solely one concept denotes one term. A situation when different terms for the same concept are used can lead to confusion and misunderstanding. The same applies vice versa – using one term for denoting multiple concepts also entails ambiguity and confusion. As has been said earlier, a concept system of a SL language does not always have to match a concept system of a SL. In many cases, a single term in a SL may correspond to multiple terms (but within different concepts or subject fields) in a TL or vice versa. One of the most infamous examples of such situation happened to a leading software company a few years ago. The company contracted out a localization of its product relating to customer relationships management from English into Czech. As the purpose of the use of the application indicates, the subject field was human resources management. English uses the same terms male and female both in human resources and in electrical engineering; however the concepts are different in these two subject fields. Unfortunately, this does not apply to Czech where the different terms for the different concepts are used. A confusion of these two concepts in Czech led to serious

10 Unfortunately, the theory is usually very different from the practice where the ideal case is when terminology is managed at least somehow.
11 The service pack could be found at the address http://support.microsoft.comkb/952149; a newspaper article concerning this issue can be found here http://digiweb.ihned.cz/c1-26537220-microsoft-se-omluvil-za-preklad-muz-neni-kolik-a-zena-neni-zdirka
mistranslation of the product – the terms *male* and *female* should have been translated as *muž* and *žena* instead of using *kolík* and *zdířka* which relates to electrical engineering. The company had to issue a special service pack to correct this mistranslation which, in the last instance, cost a lot of money and time, not speaking about sullying the company’s good name. “Without controls, inconsistent and inappropriate terms infiltrate product user interfaces, documentation, packaging, marketing materials, and support Web sites”. (IBM)

4. **How to create terminology?**

This chapter deals with the ways in which terminology is created and shared. It describes some techniques and procedures used nowadays to achieve a final list of key terms that are eventually included in a glossary or a termbase. The chapter explains advantages and disadvantages of manual and automatic extraction of terms candidates and discusses the role of a terminologist in creation a list of term candidates and validation of them in order to get a useful and applicable list of specific terms.

4.1 **Terminology extraction**

Terminology extraction refers to a process of selection words and phrases with exact and/or special meanings concerning to a specific subject field in a particular corpus or corpora. For purposes of this thesis, let us define corpora to be a collection of technical texts, i.e. software strings, documentation, user manuals, marketing materials, web pages etc. First of all, it is necessary to
gather as much information as possible, in ideal case all the materials that are intended to be translated. Then, on the basis of the volume of collected materials, TL terminologists have to decide on the method that will be employed for term extraction – be it manual or automatic using a specialised software tool. Using a manual way of extraction is possible, provided that the project to be managed is not very large and a relatively small number of short texts exist. This is a typical scenario for an OPE who need not deal with large-scale projects and cannot afford investing into specialised software. To get a picture of how the process of extraction works, regardless of being manual or automatic, let us suppose this specialised text relating to camera operation exist:

**Digital Tele-Converter**

*The focal length of the lens can be increased by an equivalent of 1.7x or 2.1x. This enables a faster shutter speed and less chance of camera shake than the zoom (including digital zoom) used by itself at the same zoom factor. The images may appear coarse, however, depending on the combination of recording pixel (p. 74) and digital tele-converter settings used.*

The example below demonstrates what happens if all specialised terms from the text are removed:

The focal length of the lens can be increased by an equivalent of 1.7x or 2.1x. This enables a faster shutter speed and less chance of camera shake than the zoom (including digital zoom) used by itself at the same zoom factor. The images may appear coarse, however, depending on the combination of recording pixel (p. 74) and digital tele-converter settings used.

---

12 The whole user manual for the Canon camera can be downloaded from http://www.retrevo.com/support/Canon-SD1400-IS-Digital-Cameras-manual/id/23635cd577/t/2/
The text becomes completely incomprehensible not only to ordinary readers but even to experts proficient in the subject field. However, when elements of a common language, or of a LGP, are removed, the specialised content of the text is partly revealed and the text immediately becomes clear and to some extent comprehensible:

**Digital Tele-Converter**

- focal length
- lens
- increased
- equivalent
- 1.7x
- 2.1x
- faster shutter speed
- camera
- zoom
- (digital zoom)
- zoom factor.
- images
- coarse,
- recording pixel
- digital tele-converter settings

The words and phrases remaining after the removal of common language elements are called *term candidates*. At this point the mentioned words and phrases have not represented the actual terms yet. The list of term candidates has to be later evaluated by a terminologist who decides on what words and compounds will become the key terms that should be included in a glossary and strictly followed during translation or localization. The TL terminologist is also responsible for finding equivalents of the source terms in a TL. Using the verb *find* instead of *translate* is appropriate here since terms, by definition, are language elements that cannot be translated in the real sense of the word. The TL terminologists do not translate or localize the terms; they find suitable
equivalents for them in the TL. The role and work of the TL terminologists is essential because obscure, ambiguous or unclear technical terms can affect lots of spheres of business, for instance classifications of products or successful search for products.

4.1.1 Manual extraction

Until the mass arrival of personal computers in the 1980s and later on with the global spread of the Internet, all terminology work and terminology management had to be done manually. Translators kept their terminology written by hand or using a typewriter on paper index cards and stored in large-sized file cabinets. They had to sit and study in a library for long hours and find information or definitions in specialised / technical literature, journals, magazines, dictionaries etc. At that time, the prerequisite for successful terminology management and thus quick and fruitful search of a term was the meticulous work of a translator. It is not necessary to stress that such work must have been really demanding and time-consuming. Nowadays, when terminologists choose a manual approach to terminology extraction, they apply almost the same steps. The disadvantage of this method consists in the fact that it is time-consuming and requires a human who performs all this work. The terminologists have to read through the materials that are to be translated and based on their own consideration they manually select and highlight term candidates (as in the example mentioned above). But at the same time, the terminologists have a direct and relatively high control over what is going to be
the term candidates. Thus almost all trash\textsuperscript{13} could be eliminated at this initial stage. The manual extraction also enables the terminologists to identify morphological (declined nouns or conjugated verbs, for example) and syntactical variations\textsuperscript{14}. The output of their work is usually a simple word list in Word or Notepad or an Excel spreadsheet full of columns and rows, which is often confusing, especially when many items exist. According to Massion, “after a while, the manual terminology extraction (single/multiple languages) or the simple administration of Word or Excel lists is no longer sufficient for a professional terminological work” (49). Obviously, the manual extraction scenario can work for OPEs; however for TSPs, “who regularly need to extract several tens of thousands of technical terms” (Massion, 49), that kind of solution becomes unfeasible.

4.1.2 Automatic extraction

The new and constantly evolving information technologies allow terminologists to perform at least some of the mentioned tasks concerning terminology extraction automatically, using computers and specialised tools but they are still far from perfect. Despite new trends and improvements in this field, terminology management still remains a challenge for all the parties involved, especially for the terminologists, because there is still a lot of work that has to be done manually by humans. A specialised software tool helps the terminologists extract term candidates from the corpora thus – unlike the

\textsuperscript{13} i.e. expressions occurring frequently, such as „Please try to...”

\textsuperscript{14} Example: To enable both parallel and serial connection... – noun phrases parallel connection and serial connection are both term candidates
manual extraction – the only task the terminologists are supposed to perform is the evaluation of the resulting list and the determination of the key terms. Compared to the manual extraction, the obvious advantage of the automatic extraction at the early stage is time. The prerequisite for automatic extraction is the existence of information in an electronic form and a suitable format. The electronic materials are then fed into a chosen software tool and by a few mouse clicks the term candidates are pre-selected automatically without the need of human intervention. The electronic documents can be entered into a software tool by anybody; no terminologist is needed. In large organisations or LSPs that employ localization teams consisting of several people, this is usually done by project technicians. The terminologist does not have any control over what is chosen. Criteria for the term candidates’ selection are given by the software setup. If the technicians are not familiar enough with using such software, there is the danger that the programme is set inappropriately and some important term candidates may be omitted because the pre-selection of the term candidates is done automatically by a software programme itself. The disadvantage of automatic extraction is that the resulting list of the term candidates will in high probability contain a lot of trash which has to be later eliminated by a human. Anne Seerup speaks from her personal experience as a freelance translator and localizer in the field of medical and IT-related materials: “The process is so tedious because you have to go through this LONG list of terms and verify and delete. This takes hours to do.”15 The software programme also does not fully recognize morphological or syntactical

15 http://www.proz.com/forum/software_applications/96347-terminology_extraction_software.html
variations. However, these specialised term extraction tools have been developed based on industrial functional standards and the output can be easily converted into a format suitable for spreadsheets or databases. And last but not least, the OPEs usually find the price of such software very high in comparison with what they can benefit from.

4.1.3 Statistical vs. linguistic method of terminology extraction

As far as terminology software is concerned, there are many software development companies and software vendors. No matter who the vendor or developer is, only two types of software programmes for this specialised task exist, based on whether a statistical or linguistic method is employed for the extraction. The tools with the statistical approach focus on the frequency of expressions in the corpora being analyzed (i.e. how often the specific expression occurs in a corpus). The more frequent the expression, the higher the probability that the expression is a term. This inevitably brings a lot of trash in the resulting list of term candidates, despite the possibility of defining a stopwords list which is a list of words that will be ignored by the programme during the extraction (i.e. sentence connectors, adverbs, prepositions etc.). The statistical tools work for all the languages. On the other hand, the tools using the linguistic method are more efficient but they are limited only to certain languages and can be used only for languages they were created for. Through an intelligent algorithm, language rules and lexical databases, the linguistic tools analyze the content of materials to be translated and determine the words
and phrases that could be term candidates. The statistical method is ideal for large and small document processing, while the linguistic approach works best for small-scale document processing because it usually consumes more memory and computing time than the tools with the statistical approach.

4.1.4 Demonstration of the statistical and linguistic term extraction using SDL MultiTerm Extract 2009 and SDL PhraseFinder 2005

This subchapter briefly demonstrates the usage of two well-known terminology extraction tools, SDL MultiTerm Extract 2009 and SDL PhraseFinder 2009. SDL MultiTerm Extract 2009 is based on the statistical approach to extraction and SDL PhraseFinder 2005 employs the linguistic engine behind. The demonstration is performed on the text called Using SPSS for Windows which is covered in a publicly available book (see Primary Sources section). SPSS is the computer software used for analysis of quantitative data for social scientists.

Since the text consists of approximately 6,000 words and thus is relatively short, terminologists can first perform an optional step. They would read through the text and highlight potential term candidates manually. This procedure can make it easier for the terminologist to decide on the length of the term candidates (i.e. how many words should be included as a minimum and maximum number of words per term during the software setup). It also helps to verify that the settings of the programme are correct. (For the purposes of this demonstration, the manual highlighting has been done only on a short excerption of the text; the whole text is attached to the thesis as Appendix).
Beginning SPSS
To start SPSS, double click on the spsswin icon on your computer screen. If there is no icon, → the Start button in the bottom left-hand corner of your screen. From the menu of programs, → SPSS for Windows. A follow-on menu will appear, from which you should select SPSS 11.0 for Windows. When SPSS loads, you may be faced with an opening dialog box with the title 'What do you want to do?' and a list of options. Many users prefer to disable this opening box. It is not important in relation to the following exposition, so → Cancel. You will then be faced with the SPSS Data Editor. This is made up of two components: Data View and Variable View. In the following discussion, these two screens are referred to as the Data Viewer and the Variable Viewer. You move between these two viewers by selecting the appropriate tab at the bottom of the screen. The Data Viewer is in the form of a spreadsheet grid into which you enter your data. The columns represent variables—in other words, information about characteristics of each person in the gym study sample. Until data are entered, each column simply has var as its heading. The rows represent cases, which can be people (as in the example you will be working through) or any unit of analysis. Each block in the grid is referred to as a 'cell'. Note also that when the data are in the SPSS spreadsheet, they will look different, for example, 1 will be 1.00. [...]

4.1.4.1. Extraction using SDL MultiTerm Extract 2009
1. Launch the SDL MultiTerm Extract 2009 and choose to create a new project.

Then select a type of the project. In this case, Monolingual Term Extraction Project is used because the term candidates are extracted from the English file. Specify the project name and location.
2. Since the project is new and no termbase exists, click the No Termbase radio button.
3. As the next step, browse for the file from which a list of term candidates should be extracted.

![Image of the New Project Wizard dialog box]

4. The following dialog box of the wizard demands close attention. The options Minimum and Maximum term length refer to the number of words which will be extracted for each term candidate. Usually, the option Silence/noise ratio should be left as is because it is balanced by default. After clicking on the Stopword Lists button, the terminologists can select their own stopword lists. If a specific file is not selected, the default stopword list file is used. This file includes words (function words, for example) which will be ignored during the extraction and will not be included in the list of term candidates.
5. Clicking on the *Finish* button displays the list of the extracted term candidates. In comparison with the text excerpt above, it is obvious that the software has been set correctly as the list contains expressions chosen manually.
The bottom left *Term Properties* pane allows the terminologist to validate all term candidates as terms, to add synonyms (if any exist) or acronyms, or delete them if they are not applicable. The terminologist can enter definitions of terms here as well. This pane also makes it possible to create context sentences that can be later included in the glossary. The bottom right *Concordance* pane shows the expression, selected in the *Term* pane, in context. These context sentences are obtained from the source file.

6. The list of term candidates then can be exported to a tab-delimited text file (TXT) which looks as follows:
This file can be later easily converted into an Excel spreadsheet, for example, and completed with the TL equivalents, notes, definitions etc. based on the individual business requirements and needs.

4.1.4.2 Extraction using SDL PhraseFinder 2005

1. Launch SDL PhraseFinder 2005 and select the type of extraction. In this case, the monolingual extraction is selected for the same reasons as above.
2. Click on the *Add File(s)* button and browse for the file from which a list of phrases should be extracted.

3. In the next dialog box, click on the first radio button – *Analyze each file completely*.
4. Click on the *Maximum Number of Words per Term* checkbox and select the number. In this case, the same number of words (i.e. 4) is selected as in the example above. Then check the box *Function words* in the group *Ignore* and in the group box *Ignore Capitalization* check the boxes *First Uppercase (except start of sentence)* and *ALL UPPERCASE*. These settings ensure more accurate results.

5. After clicking on *OK* the following dialog box displays:
Since the extracted phrases contain the expressions chosen manually, the setup of the software is correct. The main pane makes it possible to view the candidate terms, edit them and identify the ones that are valid terms. Context sentences with the extracted phrases are displayed in the bottom right pane. If a particular phrase is not applicable, right-click on the item and select *Remove*. The item is removed from the list. To validate an item as a term, right-click on it and select *Validated*. The item changes its colour to yellow. If an item is edited, it changes its colour to light blue.

6. After the validation, the resulting list of phrases can be exported through the command *Export* in the menu *File*. Several formats can be selected, for example TXT which can be opened and further edited in Excel or any word-processing editor.
4.2 The role of a terminologist

Hand in hand with realizing the full potential of terminology management, companies also start to appreciate the value a good terminologist represents for them; “the task of the terminologist is now recognised as a bona fide job in many companies“, says Gouadec (113). The job of a terminologist is tough, demanding and unfortunately in many cases still underrated. R. Woyde mentions that “most translators are generalists and NOT subject-matter experts”. Terminologists usually work as translators and localizers at the same time and only “the few practising ‘full-time’ terminologists are to be found in international organisations or in major translation companies” (Gouadec, 113).

In order to streamline and speed up the process of translation or localization while keeping consistency of key terms, translators should be provided with a glossary containing exact and validated terminology. And this is the time when a terminologist whose “main role [...] is to ensure that translators have access to reliable, complete and coherent terminological resources” (IAMLDAP)\(^\text{16}\) comes into play. Lynne Bowker mentions: “As a precursor to performing terminological research, a terminologist must compile a corpus of texts that will then be mined for terminological data.” (3) With a slight overstatement, translators translate and terminologists – apart from translating – do all the remaining work connected with terminology. After the pre-selection of the term candidates, the terminologists have to read through the whole list of term candidates, eliminate the unnecessary expressions and decide on what term

\[\text{http://www.iamladp.org/PDFs/joint_training_ventures/job_profiles/EU_profiles/GenericSkillsDraftTerminologist.pdf}\]

36
candidates finally become terms. Having selected the terms, terminologists are then responsible for the identification of corresponding concepts and for writing definitions of the key terms chosen. Then terminologists find equivalents of SL terms in a TL text. Since terminologists, just as translators, are usually generalists rather than subject-matter experts, they often cooperate with skilled experts proficient in the specific subject field in terms of writing the definitions or the determination of the concepts. Terminologists’ work also involves building terminology knowledge base (TKB) with proper data structures and creating termbases or glossaries for the translators. They deal with language queries posted by translators and “advise the translators on the best terminology to use, particularly when standard terms, local usage or neology are involved” (Gouadec, 133).

### 4.3 Getting the right terms to the right people

This chapter deals with aspects of terminology sharing. It describes advantages and disadvantages of online and offline terminology sharing.

#### 4.3.1 Offline sharing

It is obvious that if a company wants to keep consistency within a translation project, it has to distribute the right terms to all translators, localizers, proofreaders and reviewers involved in a translation or localization chain. It is not always easy as it could seem at first sight, because terminology evolves continuously (and sometimes very quickly) and new terms and concepts are
coined. This is caused by the common practice to start translation or localization during the development of a product and not when the product is completely ready to be introduced to the market. Beyond doubt this is not ideal but the pressure of keen competition is so high that such procedure is essential. Translators and localizers need easy and quick access to the terms they are searching for and companies are in need of employing a way that enables the dynamic change of terms and disseminating them to all members of the translation team.

Let us suppose a company has created a glossary in a form of a termbase and now it wants to share it with all people involved in the translation or localization project. One of the methods the company can use is offline sharing which basically means that the glossary is sent by e-mail or uploaded to a server from where the translators can download it. When any term is changed, an update glossary has to be issued and the process of uploading the glossary, notifying users of the new glossary by e-mail and downloading it has to be repeated. In order to ensure the consistency and the updated glossary, this process has to be repeated again every time when even a minor change occurs. The users usually end up with many glossaries in their computers, they are confused and the risk of using a wrong or even deprecated term is magnified.

4.3.2 Online sharing

The second way of sharing a termbase is an online method. A company dedicates a secured server on which the termbase is placed and provides
access to it in real time to corresponding users. The terminologists work directly with the server termbase and implement changes and new terms as they emerge. When any term changes or a new one is coined, it is immediately reflected in the termbase on the server and the users are constantly provided with the most updated glossary. Since many users access the server simultaneously, this method is usually feasible for LSPs or companies which can afford to invest in proper hardware and ensure secure access without compromising the company data. The prerequisite for using such solutions for translators is a fast and reliable connection to the Internet. Although a fast connection to the Internet is a must today and is almost a standard everywhere, there may be some translators “of the old school” who prefer to work without meeting this requirement.

5. Globalization, translation, localization and related issues

This chapter addresses the phenomenon of globalization from the point of view of localization. It explains the differences between translation and localization and briefly touches on the issue of internationalization. It overviews the most important areas of localization that have to be taken into consideration and underlines why the usage of consistent and standardized terminology during localization plays a vital role in conveying the message. Drawing on interesting examples from real life, the chapter points out the most problematic areas that must be taken into account and dealt with carefully during localization and even before the process of development of a product or service starts.
5.1 Globalization

Globalization is the word that affects the whole world nowadays. Focusing on a small local market is not simply enough for a company which wants to be successful and profitable. Giant multinational companies with offices and branches all over the world and employing people of utterly different cultures come into existence. But globalization inevitably entails national and cultural awareness as well. When a company decides to go global and starts to sell its products or services all over the world, it has to consider a huge number of cultural, geopolitical or social aspects and differences. With respect to financial, time and competition aspects, it is hardly imaginable that such a company would develop its products or services for every single market separately. The products and services have to be developed and produced with internationalization in mind.

5.2 Internationalization

Internationalization is often abbreviated as I18N, where 18 denotes the eighteen letters between i and n. The Localization Industry Standards Association (LISA) defines internationalization as follows:

*Internationalization is the process of enabling a product at a technical level for localization. In other words, an internationalized product does not require remedial engineering or redesign at the time of localization. Instead, it has been designed and built to be easily adapted for a specific market after the engineering phase.*
I18N is a procedure that needs to be done just once, while the localization step has to be repeated, based on the number of languages or locales in which a product or service is intended to be produced. Schmitz points out that “the more stuff you push into I18N out of L10N, the less complicated and expensive the process becomes” (6). It basically means that a product or service has to be developed in such a way that could be described as one size fits (almost) all, so that a product or service can be easily tailored to specific markets or languages. This approach does NOT always apply. For instance, the Ethiopian calendar contains thirteen months while the calendar widely used in Europe only has twelve. But this has to be taken into account when preparing a product or service for localization through internationalization.

5.3 Localization

The Localization Industry Standards Association (LISA) explains the issue of localization on its website as follows:

Localization is often treated as nothing more than ‘high-tech translation’, but this view does not capture its importance, its complexity, or what actually takes place during localization. It also hides the fact that localization must be integrated with other business processes if it is to be effective. Localization is an integral part of globalization, and without it, other globalization efforts are ineffective.17

Broadly speaking, translation is generally referred to as “the activity of changing spoken or written words into a different language” (MacMillan Dictionary). In

17 http://www.lisa.org/Localization.61.0.html
other words, when translating, the SL texts are transferred into the TL with keeping the meaning and purpose of the original text. While the expression *translation* is widely understood, the expression *localization* (abbreviated as L10N, where *L* stands for the number of letters between *l* and *n*) is relatively new and localizers very often meet with the question asking for explanation. According to Schmitz, localization refers to “adapting a product or service to a local or regional market” (Schmitz, 3) with the aim to keep appropriate linguistic, cultural and technical requirements. LISA, as a leading authority in the field of standardization, defines localization as “the process of modifying products or services to account for differences in distinct markets”. Esselink (1) further explains: “The term *localization* is derived from the word *locale*, which traditionally means a small area or vicinity. Today, *locale* is mostly used in a technical context, where it represents a specific combination of language, region and character encoding.” For instance, the Spanish speakers in Mexico use a different *locale* than the Spanish speakers in Spain. Localization of a software product involves translation not only of the software itself, but also of paper and online documentation, help, tutorials, websites etc. The aim of localization is to adapt the software to a target market in the way that a purchaser in France, for example, would perceive the product as having been developed and produced in France.

### 5.4 Challenges of localization

During localization, cultural, geopolitical, social and language aspects have to be taken into consideration based on the specific country or locale into which
the particular product or service is localized. In this subchapter a closer look is taken at the most important and complex issues.

5.4.1 Cultural and colloquial references

This is a delicate area that has to be dealt with very carefully. What is suitable for one market could be inappropriate or even offensive for another.

Tom Edwards in his article mentions the cultural faux pas made by Ben & Jerry’s company in Ireland in spring 2006. “Over in Ireland, an uproar was caused in spring 2006 by the Ben & Jerry’s ice cream brand when it released a flavour called “Black and Tan”. In the United States, this name refers to a drink that is prepared with Guinness and ale; but in Ireland the name refers to an infamous auxiliary unit of the British army, soldiers who were known for particularly undisciplined and violent behaviour in dealing with the local people during the Anglo-Irish war of the early twentieth century – quite the negative historical connection.” (Edwards, 22)

Other problems may arise from colloquial differences. Edna Ditaranto mentions in her article that “extreme awareness of and attention to the difference between cultures, even when translating documents into a language that is spoken in multiple countries, such as Spain (spoken in 21 countries) or Portuguese (11 countries)” (45). She gives an example of using such an innocent word as *papaya* which “in most places refers to a fruit but in Puerto Rico it refers to female genitalia.” (46)
In August 2008, the famous Swedish furniture company amused its potential Czech customers. IKEA is known for naming its products after towns and villages to make them more attractive for the customers. A name of the Danish town Hoven was chosen for the name of a rug. Unfortunately in Czech, this word exactly corresponds to the accusative plural of a vulgar designation for human or animal excrements. Since a process of automatic concatenation of words was applied, IKEA Czech organized a clearance sale of excrements (“Doprodej Hoven” in Czech). The management of the Czech branch realized that, apart from a few jokers, nobody was interested in buying this rug and they had to ask IKEA headquarters in Sweden to rename the product. Naturally, the renaming for the Czech language was approved and the sale of the rug increased.

5.4.2 Geopolitical issues

Needless to say that this is the area to which meticulous attention should be paid, particularly during internationalization. How to deal with such hot issues like illustrations of Tibet? Should it be illustrated and referred to as a part of China, which lays claim to this territory? For obvious reasons, experts on geography and international relations have to be consulted.

5.4.3 Colours
It is common knowledge that different cultures interpret the colour symbolism differently\textsuperscript{19}. In Western cultures the black colour symbolizes death and mourning, while in Spain it is the colour of ceremonial significance. The white colour in Europe is connected with purity, innocence and virginity, but in the Orient it is the colour of mourning. The European medieval world considered yellow to be the colour of dishonour: prostitutes, beggars, unmarried mothers and lepers had to wear a yellow mark. This symbolism of the yellow colour was also misused by the Nazis when they enjoined the Jews to wear the yellow Star of David. In Western countries a yellow flag marked the places of disease and quarantine. On the other hand, in China this colour is the national colour dedicated to the emperor.

5.4.4 Graphics, icons, symbols
Websites use national flags very often to identify language mutations. It could be noticed that this practice has become a standard one and only a few website creators are aware of the issues arising from it. For instance, the national flag of Spain is used to indicate a Spanish mutation of the content. Spanish, however, is not spoken only in Spain in Europe but also in more than 20 other countries, such as Belize, Mexico, Peru, Venezuela, Argentina etc. The Spanish-native speakers in these countries could feel offended. Gestures used as icons pose another problematic area. Graphical illustrations showing body parts, such

\textsuperscript{19} http://kreativity.blog.cz/0910/symbolika-barev; http://www.onlio.com/clanky/psychologie-barev-3.html#axzz0rh7xsUO3
as hands or feet, may be found offensive in some cultures or countries. Identically, using icons representing homophones, such as an image of paws to designate the verb pause, can work in some languages but not in the others.

5.4.5 Currency and number format

As opposed to the US or the United Kingdom, in continental Europe the currency symbol comes after the numbers. For instance, twenty five pounds (£25) is written 25 £. Different decimal separators are used in different countries. While a dot is used as a decimal separator (ten thousand is written 10,000.00) in the US, the Czech typographical rules require a comma to be used (10 000,00).

5.4.6 Date and time format

In Czech, the correct format of dates is day, month, year (dd/mm/yyyy). In the US, example among others, the month precedes the day (mm/dd/yyyy). In English-speaking countries, the twelve-hour format of time is used (the period from midnight to midday is referred to as A.M., the period from midday to midnight as P.M.). The Czech language uses the twenty four-hour format of time (i.e. eight o’clock in the morning is written 8,00; eight o’clock in the evening is written 20,00).
5.4.7 Name and address format

Usually, the first name is placed before the surname, as in Kamila Kralikova. The Hungarians, however, expect their first names to come after their surnames (i.e. Kovacs Laszlo). Problems may arise from the differences in address formats. In the UK, for example, it is common to use letters in ZIP codes unlike the Czech Republic where only numbers are used.

5.4.8 Abbreviations and acronyms

When localizing abbreviations and acronyms, localizers have to be aware of the fact that acronyms and abbreviated phrases may be considered offensive in some countries. A good example, on which the issue with acronyms and abbreviations can be illustrated, despite the fact that it is not from real life, comes from Red Dwarf\(^{20}\), a famous British science fiction sitcom:

> Erm, I think we're all beginning to lose sight of the real issue here, which is: what are we going to call ourselves? Erm, and I think it comes down to a choice between "The League Against Salivating Monsters" or, my own personal preference, which is "The Committee for the Liberation and Integration of Terrifying Organisms and their Rehabilitation Into Society." Erm, one drawback with that -- the abbreviation is CLITORIS.

Apart from having possible offensive or pejorative meaning, the acronyms can also represent words which might be reserved for and denote a different product or service in the TL. For example, looking up the acronym CAT in the Acronym Finder\(^{21}\) on the web returned 197 meanings.

---
\(^{20}\) [http://www.reddwarf.co.uk/guide/index.cfm](http://www.reddwarf.co.uk/guide/index.cfm)
\(^{21}\) [http://www.acronymfinder.com/CAT.html](http://www.acronymfinder.com/CAT.html)
For this reason, the acronyms – except for the ones that are universally recognized and well-known (such as NATO or UNESCO) – should be explained or broken down in parentheses.

### 5.4.9 System of measures

Although most countries of the world have already adopted the international metric system of units of measurements (S.I.), in the Anglo-Saxon countries (the US or the UK, among others) the imperial measures are still officially used. The imperial units of measures are also normally used in aviation. The inconsistency in the use of systems of units of measurements may cause a lot of problems. In 1999, the Mars Climate Orbiter\(^{22}\) crashed and, according to NASA, this was caused by “the failed translation of English units into metric units in a segment of ground-based, navigation-related mission software”.

### 6. Standards and standardization organizations

This chapter briefly describes the effort to establish standards for translation, localization and terminology. It also mentions the organisations and associations that participate in or even directly develop the standards.

To achieve a consensus on terminology used in various areas of public concern, standards have to be created and established. There are several types of standards, for example product standards, functional standards or terminology standards. The object of standards is to ensure the quality within

\(^{22}\) [http://sunnyday.mit.edu/accidents/mco991110.html](http://sunnyday.mit.edu/accidents/mco991110.html)
products and services, to secure their efficiency, safety, reliability and interoperability, to reach the exchangeability and reusability (for instance, pipe fittings or data) and through functional standards (e.g. TCP/IP protocols) to enable the use of technologies. Since corporations and companies all over the world nowadays realize the significance of information and account it to be their highly-valued assets, the natural necessity of standardization of language and information has emerged.

Standardization refers to an activity leading to the establishment and adoption of standards, or norms. It is based on open and voluntary involvement of organisations. Independent organisations and associations cooperate on the development of standards for all areas of public concern. The Cooperation Platform for Research and Standards (COPRAS), a support initiative introduced by European standards organisations, defines standardization and its aim as follows:

*The primary aim of standardization in the current social and economic context is to help encouraging the free movement of goods. Standardization will help to remove technical barriers, open up new markets, and enable new economic models. It helps to create economies of scale while at the same time increasing opportunities for product differentiation and competition and services. Consequently, standardization may help establish compatibility and interoperability, it may enable market self-regulation, and guard the safety and health of citizens.*

This platform was initiated in cooperation with the ICT Standards Board (ICTSB), whose main role is to consider and later create standards and
specifications based on the requirements of concrete markets and competent sources.

Founded in 1947 with the aim of enabling problem-free exchange of goods, the International Organisation for Standardization, better known as ISO, is the non-governmental and world’s largest developer and publisher of international standards with the headquarters in Geneva, Switzerland. It unites national standards institutions from more than 160 countries and serves as a bridge between the public and private sectors. It has developed and maintained the ICS (International Classification for Standards). This classification system is used for technical standards and it functions as a structure for catalogues and databases of technical standards and normative documents. It has 99 top-level divisions but only 40 of them are currently in use. The rest is reserved for subjects that could come into existence in future. As far as terminology and language issues are concerned, the ISO TC/37 group of standards applies. The standards cover principles and methods, terminographical and lexicographical working methods, systems to manage terminology, knowledge and content, language codes and others. The work on the development and harmonization of standards is continuing all the time. All the standards are available for purchase through the ISO’s websites.

The website of the Localization Industry Standards Association (LISA) (which is a member of the ISO) is a useful source of information relating to globalization, localization and industry standards. The LISA joins globalization professionals and it offers helpful services and recommendations to its
members. LISA membership is a paid service and thus suitable for corporations or middle and large businesses, not for OPEs.

7. Conclusion

The thesis first presents the influence of terminology and its management on communication and comprehension. It points out the importance of consistency within source and target texts and shows how inconsistency can affect the quality of translation and the productivity of translators. It describes the situation that currently prevails on the translation and localization market and it also provides the theoretical background.

The next part of the thesis focuses on more practical aspects. It points out the issue of consistency of terms with respect to usability of the localized products or services. It presents arguments for the creation and use of terminology and also mentions consequences which ignoring terminology can cause. It provides the readers with two ways of terminology extraction – statistical and linguistic methods and describes their advantages and disadvantages. Using well-known software, a demonstration of both methods is presented. In connection with term extraction, the role of a terminologist is addressed as well and the job of a terminologist and the possibilities of terminology sharing are described in detail.

The final part of the thesis is divided into two parts. The first part explains the concepts of globalization, internationalization and localization. The second part is devoted to the challenges that localizers have to face in their everyday
work. It summarizes the most interesting issues of localization work and illustrates them on examples from practice.

The last chapter of the thesis presents the term *standardization*. It explains why the use of standards is recommended and provides the readers with background information about the leading organisation in this branch of knowledge.

The main objective set for the thesis was to provide readers – and beginning translators or localizers above all – with a basic handbook of issues relating to terminology and its management. Since standardized and consistent terminology plays a key role in today’s society and communication, it has to be dealt with carefully and in appropriate ways. This thesis familiarizes readers with background information concerning terminological work and, drawing on the examples from practice, it explains the most problematic areas of localization. The purpose of this thesis is not to dispirit the readers but to encourage them to embark on proper terminology management. Through consistent and suitable terminology management, they will be able to streamline their work and produce better quality during the process translation and localization. Adding the value to their work, they will also increase customers’ satisfaction and thus help them promote their brands.
Abstract

The aim set for this diploma thesis called Basic Aspects of Terminology Management has been to make beginning translators and localizers familiar with the issue of terminology and its management. The purpose of this thesis is to provide basic instructions how to proceed with the creation and management of terminology.

The thesis itself first presents the readers with the importance of terminology and also of the situation which predominates on the current translation and localization markets. The readers are provided with basic knowledge which should be acquired before they start to work with terminology. The thesis stresses the necessity of keeping both source and target texts consistent in terms of terminology.

In the following part of the thesis, the readers are made familiar with two ways of terminology extraction – manual and automatic. The process of the automatic extraction is demonstrated on two commonly used specialised software tools, each of them employing a different approach to the extraction. Further the important role of a terminologist in the creation and management of terminology is mentioned and advantages and disadvantages of both online and offline terminology sharing are described.

In the final part of the thesis terms, such as globalization, internationalization or localization are explained, and their interconnectedness is illustrated. This part of the work is particularly aimed at practical aspects of terminology from the point of view of software localization. It points out challenging areas that have to be dealt with very carefully during translation or localization. This part of the thesis also contains brief and useful information about standards relating to terminology and its management and about organizations involved in standardization and in creating, establishing and adopting the standards.
Anotace

Cílem předkládané diplomové práce s názvem *Basic Aspects of Terminology Management* je seznámit začínající překladatele s problematikou terminologie a její správy. Účelem této práce je poskytnout čtenářům základní návod, jak postupovat při vytváření a správě terminologie.

Práce nejprve seznámuje čtenáře s důležitostí a významem terminologie a také se situací, která v současné době panuje na trhu překladu a lokalizace. Uvádí základní teoretické vědomosti, které by si měli překladatelé osvojit dříve, než začnou pracovat s terminologií. Práce zdůrazňuje nutnost zachování konzistence jak ve zdrojovém, tak i v cílovém textu.

V další části se čtenáři seznámi se dvěma způsoby extrakce terminologie – ruční a automatickou. Proces automatické extrakce terminologie je demonstrován na dvou nejrozšířenějších specializovaných softwarových nástrojích, které používají rozdílné přístupy k extrakci termínů. Práce zmíňuje důležitou roli terminologa při tvorbě a správě terminologie a popisuje výhody a nevýhody sdílení terminologie v režimu online a offline.

Závěrečná část práce vysvětluje pojmy, jako je globalizace, internacionalizace a lokalizace, a jejich vzájemné propojení. Tato část je zaměřena především na praktické aspekty terminologie z pohledu softwarové lokalizace a seznámuje čtenáře s oblastmi, u nichž je třeba při lokalizaci dbát zvýšené opatrnosti. V této části čtenáři také nalezou základní informace o standardech souvisejících s terminologií a její správou a o organizacích, které se zabývají vytvářením standardů a standardizací.
Primary sources


Bibliography


ENCYCLOPÆDIA BRITANNICA. 21 June 2010.

http://www.britannica.com/EBchecked/topic/626818/Andreas-Vesalius

ENCYCLOPÆDIA BRITANNICA. 21 June 2010.


creation-and-use-of-terminology-in-localization/


Healey, Jim. 2007 “What does it take to be a good translator?“. *Multilingual* July/August 2007: 29-36


http://www-01.ibm.com/software/globalization/topics/terminology/terms.jsp


Edwards, Tom. 2007. “Cultural impacts from headlines”.
*Multilingual* July/August 2007: 21-23

*Multilingual* April/May 2007: 39-42


“Ikea nabízela „Doprodej Hoven“, kontroverzní koberce pak přejmenovala”.


http://www.iso.org/iso/catalogue_detail.htm?csnumber=20057

Jones, Andrew. 2007. “Scheduling successful localization projects”.
Multilingual July/August 2007: 49-51

http://www.lisa.org/Homepage.8.0.html


Massion, François. 2007. “Terminology management: a luxury or a necessity?”.
Multilingual April/May 2007: 47-50


“Microsoft se omluvil za překlad. Muže není kolík a žena není zdířka”.

Monahan, Sonia. 2009. “Measuring QA to improve translation cost and speed”.
Multilingual July/August 2009: 48-50

http://www.onlio.com/clanky/psychologie-barev-3.html#axzz0rh7xsUO3

Pantel, Patrik and Dekang Lin. 2001. "A Statistical Corpus-Based Term Extractor" In E. Stroulia and S. Matvin (eds.) Advances in Artificial


*Tecom, German professional association for technical communication and Information-Development*. 2009. 2 January 2010. http://www.tekom.de/index_neu.jsp?


Wright, Sue Ellen, and Bernard Budin [Comp.]. 1997. *Handbook of Terminology*
2001chicago/presentations.html
Appendix

Using SPSS for Windows

CHAPTER GUIDE 243

Introduction 244

Getting started in SPSS 244
Beginning SPSS 244
Entering data in the Data Viewer 244
Defining variables: variable names, missing values, variable labels, and value labels 247
Recoding variables 249
Computing a new variable 250

Data analysis with SPSS 251
Generating a frequency table 251
Generating a bar chart 251
Generating a pie chart 253
Generating a histogram 253
Generating the arithmetic mean, median, standard deviation, and range 254
Generating a contingency table, chi-square, and Cramer's V 254
Generating Pearson's and Spearman's rho 257
Generating scatter diagrams 257
Comparing means and eta 257
Generating a contingency table with three variables 257

Further operations in SPSS 261
Saving your data 261
Retrieving your data 261
Printing output 261

OVERVIEW 262
QUESTIONS FOR REVIEW 262

CHAPTER GUIDE

In order to implement the techniques that you learned in Chapter 11, you would need to do either of two things: learn the underlying formula for each technique and apply your data to it, or use computer software to analyse your data. The latter is the approach chosen in this book for two main reasons.

• It is closer to the way in which quantitative data analysis is carried out in real research nowadays.
• It helps to equip you with a useful transferable skill.
You will be learning SPSS for Windows, which is the most widely used package of computer software for doing this kind of analysis. It is relatively straightforward to use. We will be continuing to refer to the techniques introduced in Chapter 11 and will continue to use the gym survey as an example. This chapter largely operates in parallel to Chapter 11, so that you can see the links between the techniques learned there and the use of SPSS to implement them.

Introduction

This chapter aims to provide a familiarity with some basic aspects of SPSS for Windows, which is possibly the most widely used computer software for the analysis of quantitative data for social scientists. SPSS, which originally was short for Statistical Package for the Social Sciences, has been in existence since the mid-1960s and over the years has undergone many revisions, particularly since the arrival of personal computers. The version that was used in preparing this section was Release 11. At the time of writing, this was the current version of SPSS, though Release 12 was about to be released. The book's website will include any significant changes to instructions to using SPSS that result from this and subsequent releases. In particular, Release 12 contains significant changes to the operation of Chart Editor which is discussed later in this chapter. From this point on, when referring to SPSS for Windows in the text, it will be called simply SPSS. The gym survey used in Chapter 11 will be employed to illustrate SPSS operations and methods of analysis. The aim here is to introduce ways of using SPSS to implement the methods of analysis discussed in Chapter 11.

SPSS operations will be presented in bold, for example, Variable Name: and Analyze. Names given to variables in the course of using SPSS will also be presented in bold italics, e.g. gender and reasons. Labels given to values or to variables are also in bold but in a different font, e.g. reasons for visiting and male. Box 12.1 presents a list of basic operations in SPSS. One further element in the presentation is that a right-pointing arrow—→—will be used to denote 'click once with the left-hand button of your mouse'. This action is employed to make selections and similar activities.

Getting started in SPSS

Beginning SPSS
To start SPSS, double click on the spsswin icon on your computer screen. If there is no icon, → the Start button in the bottom left-hand corner of your screen. From the menu of programs, → SPSS for Windows. A follow-on menu will appear, from which you should select SPSS 11.0 for Windows. When SPSS loads, you may be faced with an opening dialog box with the title 'What do you want to do?' and a list of options. Many users prefer to disable this opening box. It is not important in relation to the following exposition, so → Cancel. You will then be faced with the SPSS Data Editor. This is made up of two components: Data View and Variable View. In the following discussion, these
two screens are referred to as the Data Viewer and the Variable Viewer. You move between these two viewers by selecting the appropriate tab at the bottom of the screen. The Data Viewer is in the form of a spreadsheet grid into which you enter your data. The columns represent variables—in other words, information about characteristics of each person in the gym study sample. Until data are entered, each column simply has var as its heading. The rows represent cases, which can be people (as in the example you will be working through) or any unit of analysis. Each block in the grid is referred to as a 'cell'. Note also that when the data are in the SPSS spreadsheet, they will look different, for example, 1 will be 1.00.

Entering data in the Data Viewer
To input the data into the Data Viewer, make sure that the top left-hand cell in the grid is highlighted (Plate 12.1). If it is not highlighted, simply click once in that cell. Then, type the appropriate figure for that cell—that is, 1. This number goes directly into that cell and into the box beneath the toolbar. As an alternative to using the mouse, many people find it easier to use the arrow keys on their keyboard to move from cell to cell. If you make a mistake at any point, simply click once in the cell in question, type in the correct value, and click once more in that cell. When you have finished, you should end up in the bottom right-hand cell of what will be a perfect rectangle of data. Plate 12.2 shows the Data Viewer with the data from the gym survey entered (though only part of the set of data is visible, in that only the first twenty-two respondents and ten of the twelve variables are visible). The first row of data contains the coded answers from the completed questionnaire in Box 11.2. In order to proceed further, you will find that SPSS works in the following typical sequence for defining variables and analysing your data.

Box 12.1 Basic operations in SPSS for Windows

The SPSS Data Editor. This is the sphere of SPSS into which data are entered and subsequently edited and defined. It is made up of two screens: the Data Viewer and the Variable Viewer. You move between these two viewers by selecting the appropriate tab at the bottom of the screen. The Data Viewer. This is the spreadsheet into which your data are entered. When you start up SPSS, the Data Viewer will be facing you. The Variable Viewer. This is another spreadsheet, but this one displays information about each of the variables and allows you to change that information. It is the platform from which you provide for each variable such information as: the variable name; a variable label; and value labels (see below).

The Output. When you perform an analysis or produce a diagram (called a 'chart' in SPSS), your output will be deposited here. The Output Viewer superimposes itself over the Data Editor after an analysis has been performed or a chart generated. A Variable Name. This is the name that you give to a variable, e.g. gender. The name must be no more than eight characters. Until you give a variable a name, it will be referred to as varOOOOO1, etc. When the variable has been given a
name, it will appear in the column for that variable in the Data View window. It is generated from the Variable Viewer.

A Variable Label. This is a label that you can give to a variable but which is not restricted to eight characters. Spaces can be used, e.g. reasons for visiting. The Label will appear in any output you generate. It is generated from the Variable Viewer.

A Value Label. This is a label that you can attach to a code that has been used when entering data for all types of variables other than interval/ratio variables. Thus, for varOOOOO!, we would attach the label male to 1 and female to 2. When you generate output, such as a frequency table or chart, the labels for each value will be presented. This makes the interpretation of output easier. It is generated from the Variable Viewer.

• Missing Values When you do not have data for a particular variable when entering data for a case, you must specify how you are denoting missing values for that variable. Missing values are generated from the Variable Viewer.

• Recode. A procedure that allows codes or numbers to be changed. It is especially helpful when you need to combine groups of people—for example, when producing age bands.

• Compute.... A procedure that allows you to combine two or more variables to form a new variable.

• Analyze. This is the point on the menu bar above the Data Editor from which you choose (via a drop-down menu) which method of analysis you want to select. Note that, whenever an item on a menu appears with a right-pointing arrowhead after it, this means that, if you select that option, a further menu will follow on.

• Graphs. This is the point on the menu bar above the Data Editor from which you choose (via a drop-down menu) which chart you want to select.

• Chart Editor. When you produce a graph, you can edit it with the Chart Editor. To activate this editor, double-click anywhere in the graph. A small chart editor window will appear and your main graph will appear opaque until you exit the Editor. From the Editor, you can make various changes and enhancements to your graph.

1. You make a selection from the menu bar at the top of the screen, e.g. -> Analyze.
2. From the menu that will appear, make a selection, e.g. -> Descriptive Statistics.
3. This will bring up a dialog box into which you will usually inform SPSS of what you are trying to do—e.g. which variables are to be analyzed.
4. Very often, you then need to convey further information and to do this you have to —-> a button that will bring up what is called, following Bryman and Cramer (2001), a sub-dialog box.
5. You then provide the information in the sub-dialog box and then go back to the dialog box. Sometimes, you will need to bring up a further sub-dialog box and then go back to the dialog box.

When you have finished going through the entire procedure, —» OK. The toolbar beneath the menu bar allows shortcut access to certain SPSS operations.
Defining variables: variable names, missing values, variable labels, and value labels

Once you have finished entering your data, you need to define your variables. The following steps will allow you to do this:

1. Go to the Variable View tab at the bottom of the Data Viewer (opens the Variable Viewer shown in Plate 12.3).

2. To provide a variable name, click on the current variable name (e.g. var00003) and type the name of the name you want to give it (e.g. reasons). Remember that this name must be no more than eight characters and you cannot use spaces.

3. You can then give your variable a more detailed name, known in SPSS as a variable label. To do this, go to cell in the Label column relating to the variable for which you want to supply a variable label. Then, simply type in the variable label (i.e. reasons for visiting).

Then you will need to provide 'value labels' for variables that have been given codes. The procedure generally applies to variables that are not interval/ratio variables. The latter, which are numeric variables, do not need to be coded (unless you are grouping them in some way). To assign value labels, go to the Values column relating to the variable you are working on. A small button with three dots on it will appear. Click the button. The Value Labels dialog box will appear (Plate 12.4). Select the box to the right of Value and begin to define the value labels. To do this, enter the value (e.g. 1) in the area to the right of Value and then the value label (e.g. relaxation) in the area to the right of Value Label. Then click Add. Do this for each value. When you have finished, click OK.

5. You will then need to inform SPSS of the value that you have nominated for each variable to indicate a missing value. In the case of reasons, the value is 0 (zero). To assign the missing value, go to the cell for this variable in the Missing column. Again, click the button that will appear with three dots on it. This will generate the Missing Values dialog box (Plate 12.5). In the Missing Values dialog box, enter the missing value (0) below Discrete missing value: and then click OK.

In order to simplify the following presentation, reasons will be the only variable for which a variable label will be defined.

Recoding variables

Sometimes you need to recode variables—for example, when you want to group people. You would need to do this in order to produce a table like Table 11.3 for an interval/ratio variable like var00002, which we will give the variable name age. SPSS offers two choices: you can recode age so that it will be changed in the Data Viewer, or you can keep age as it is and create a new variable. This latter option is desirable whenever you want to preserve the variable in question as well as create a new one. Since we may want to carry out analyses involving age as an interval/ratio variable, we will recode it so that a new variable, which we will call agegp, for age groups, will be created. The
The aim of the following operations is to create a new variable—agegp—which will comprise five age bands, as in Table 11.3.

1. Transform -> Recode -> Into Different Variables [opens Recode into Different Variables dialog box shown in Plate 12.6]
2. age ~Y • button [puts age in Numeric Variable->Output Variable: box] -> box beneath Output Variable Name: and type agegp -›• Change [puts agegp in the Numeric Variable->Output Variable: box] -› Old and New Values... [opens Recode into Different Variables: Old and New Values sub-dialog box shown in Plate 12.7]
3. the circle by System- or user-missing and by System- missing under New Value, if you have missing values for a variable, which is the case for this variable
4. circle by Range: Lowest Through and type 20 in the box —> box by Value in New Value and type 1 —> Add [the new value will appear in the Old- >New: box]
5. first box by Range: and type 21 and in box after through type 30 —> box by Value in New Value and type 2 —> Add
6. first box by Range: and type 31 and in box after through type 40 -› box by Value in New Value and type 3 -› Add
7. first box by Range: and type 41 and in box after through type 50 -› box by Value in New Value and type 4 -› Add
8. circle by Range: through highest and type 51 in the box -› box by Value in New Value and type 5 -› Add —> Continue [closes the Recode into Different Variables: Old and New Values sub-dialog box shown in Plate 12.7 and returns you to the Recode into Different Variables shown in Plate 12.6]
9. -›OK

The new variable agegp will be created and will appear in the Data Viewer. You would then need to generate value labels for the five age bands and possibly a variable label using the approach described above.

Computing a new variable
A person's total amount of time spent in the gym is made up of three variables: cardmins, weimins, and othmins. If we add these up, we should arrive at the total number of minutes spent on activities in the gym. In so doing, we will create a new variable totalmin. To do this, this procedure should be followed:
1. Transform -> Compute... [opens the Compute Variable dialog box shown in Plate 12.8]
2. under Target Variable: type totalmin
3. select suM[numexpr, numexpr,...] from the list underneath Functions: and click on the button with an upward-pointing arrowhead to send it into the box underneath Numeric Expression:  
4. from the list of variables at the left, —> cardmins -› • button [puts cardmins in box after SUM]; -› weimins -› • button [puts weimins in box after ,cardmins]; -› othmins -› • button [puts othmins in box after,weimins]
5. -›OK
The new variable totalmin will be created and will appear in the Data Editor.
Now at last, we can begin to analyse the data!

Data analysis with SPSS

Generating a frequency table
To produce a frequency table like the one in Table 11.2:
1. -> Analyze -> Descriptive Statistics -> Frequencies... [opens the Frequencies dialog box shown in Plate 12.9]
2. -> reasons for visiting -> • button [puts reasons for visiting in Variable[s]: box]
3. ->0K
The table will appear in the Output Viewer (see Plate 12.10)

Note that in the Frequencies dialog box, variables that have been assigned labels will appear in terms of their variable labels, but those that have not been assigned labels will appear in terms of their variable names. This is a feature of all dialog boxes produced via Analyze and Graphs (see below).

Generating a bar chart
To produce a bar chart like the one in Figure 11.2:
1. -> Graphs -> Bar .. . [opens Bar Charts dialog box]
2. -> Simple ->Summaries for groups of cases -> Define [opens Define Simple Bar: Summaries for Groups of Cases sub-dialog box shown in Plate 12.11]
3. reasons for visiting -> • button by Category Axis -> [reasons for visiting will appear in the box] -> N of cases beneath Bars Represent [if this has not already been selected, otherwise continue without doing this]
4. ->0K

Generating a pie chart
To produce a pie chart like the one in Figure 11.3:
1. -> Graphs -> Pie .. . [opens the Pie Charts dialog box] -> Summaries for groups of cases -> Define [opens the Define Pie: Summaries for Groups of Cases sub-dialog box]
2. ->reasons for visiting -> • button by Define slices by [reasons for visiting will appear in the box] -> N of cases beneath Slices Represent: [if this has not already been selected, otherwise continue without doing this]
3. ->0K

In order to include percentages, as in Figure 11.3, double-click anywhere in the chart in order to bring up the Chart Editor. The chart will appear in the Chart Editor and the main figure will become opaque. Then -> Chart and then -> Options... and then place a tick by Percents [there should also be a tick by Text].
Your chart will be in colour, but, if you only have access to a monochrome printer, you can change your pie chart into patterns, which allows the slices to be clearer. At the end of the next section, there is a description of how to do this.

Generating a histogram
Producing a histogram like the one in Figure 11.4 is somewhat more complicated. The standard procedure for generating a histogram is —> Graphs —> Histogram and then selecting the relevant variable. This procedure will generate a very good histogram but SPSS will define the age bands. If you want to define the bars yourself, you should follow the steps for the one produced for Figure 11.4, which involved following the steps entailed in generating a bar chart:
1. -» Graphs -» Bar .. . [opens Bar Charts dialog box]
2. —» Simple —> Summaries for groups of cases —> Define [opens Define Simple Bar: Summaries for Groups of Cases sub-dialog box shown in Plate 12.11]
3. -Y agegp —» • button by Category Axis [agegp will appear in the box] -» N of cases beneath Bars Represent [if this has not already been selected, otherwise continue without doing this] -» OK
4. after the bar chart appears in the SPSS for Windows viewer, double-click anywhere in the body of the figure; this will bring up the SPSS for Windows Chart Editor shown in Plate 12.12 -» Chart -» Bar Spacing... [opens the Bar Spacing sub-dialog box]
5. in the small box to the right of Inter-Bar Spacing replace the figure in the box with 0 -» OK [closes the Bar Spacing sub-dialog box and returns you to the SPSS for Windows Chart Editor shown in Plate 12.12]
6. -» File -» Close

A new bar chart will appear in the SPSS for Windows Chart Editor, which will be the same as the one produced in Figure 11.4. This procedure essentially entails producing a bar chart but with the spaces between bars removed so that a histogram is generated. This procedure allows you to define your own bars.

When the SPSS for Windows Chart Editor is open, all figures can be edited, so that, for example, colours can be changed or patterns inserted. This can be very useful if you do not have access to a colour printer.
Figures 11.2, 11.3, and 11.4 were produced by changing the bars or slices to white and then introducing patterns. When in the SPSS for Windows Chart Editor, you can experiment by -» Format and then selecting from the choices there. This procedure applies to all charts.

Generating the arithmetic mean, median, standard deviation, and range
To produce the mean, median, standard deviation, and the range for an interval/ratio variable like age, the following steps should be followed:
1. -» Analyze -» Descriptive Statistics —> Explore... [opens the Explore dialog box]
2. ➔ age ➔ • button to the left of Dependent List: [puts age in the Dependent List: box] ➔ Statistics under Display ➔ OK

The output will also include the 95 per cent confidence interval for the mean, which is based on the standard error of the mean. The output can be found in Table 12.1.

Generating a contingency table, chi-square, and Cramer's V

In order to generate a contingency table, like that in Table 11.4, along with a chi-square test and Cramer's V, the following procedure should be followed:

Table 12.1 Explore output for age (SPSS output)

<table>
<thead>
<tr>
<th>Explore</th>
<th>Case Processing Summary</th>
<th>Explore</th>
<th>Case Processing Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>Valid Missing Total</td>
<td>N</td>
<td>Percent N Percent N Percent</td>
</tr>
<tr>
<td>ACE</td>
<td>89 98.9% 1 1.1% 90 100.0%</td>
<td>Descriptives</td>
<td>Statistic Std. Error</td>
</tr>
<tr>
<td>AGE</td>
<td>Mean 33.5955 .9420</td>
<td>95% Confidence Lower bound 31.7235</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Interval for mean Upper bound 35.4675</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5% Trimmed mean 33.3159</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Median 31.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Variance 78.971</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Std. Deviation 8.8866</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Minimum 18.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maximum 57.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Range 39.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Interquartile Range 14.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Skewness .446 .255</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kurtosis -.645 .506</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. ➔ Analyze ➔ Descriptive Statistics ➔ Crosstabs [opens the Crosstabs dialog box shown in Plate 12.13]
2. ➔ reasons for visiting ➔ • button by Row[s] [reasons for visiting will appear in the Row[s]: box] ➔ gender ➔ • button by Column [s]: [gender will appear in the Column [s]: box] ➔ Cells [opens Crosstabs: Cell Display sub-dialog box shown in Plate 12.14]
3. Make sure Observed in the Counts box has been selected. Make sure Column under Percentages has been selected. If either of these has not been selected, simply click at the relevant point. ➔ Continue [closes Crosstabs: Cell Display sub-dialog box and returns you to the Crosstabs dialog box shown in Plate 12.13]
4. ➔ Statistics... [opens the Crosstabs: Statistics sub-dialog box shown in Plate 12.15]
5. → Chi-square → Phi and Cramer's V → Continue [closes Crosstabs: Statistics sub-dialog box and returns you to the Crosstabs dialog box shown in Plate 12.13]

6. → OK

The resulting output can be found in Table 12.2.

If you have a table with two dichotomous variables, you would use the same sequence of steps to produce phi.

Generating Pearson's r and Spearman's rho
To produce Pearson's r in order to find the correlations between age, cardmins, and weimins, follow these steps:
1. → Analyze → Correlate → Bivariate... [opens Bivariate Correlations dialog box shown in Plate 12.16]
2. → Y age → • button → cardmins → • button → weimins → • button [age, cardmins, and weimins should now be in the Variables: box] → Pearson [if not already selected] → OK

The resulting output is in Table 12.3.

To produce correlations with Spearman's rho, follow the same procedure but instead of selecting Pearson, you should → Spearman instead.

Generating scatter diagrams
Scatter diagrams, known as scatterplots in SPSS, are produced in the following way. Let us say that we want to plot the relationship between age and cardmins. There is a convention that, if one variable can be identified as likely to be the independent variable, it should be placed on the x axis, that is, the horizontal axis. Since age is bound to be the independent variable, we would follow these steps:
1. → Graphs → Scatter... [opens the Scatter Plot dialog box]
2. → Simple [usually this has been automatically selected] → Define [opens the Simple Scatterplot sub-dialog box shown in Plate 12.17]
3. → cardmins → • button by Y Axis: → age → • button by X Axis: → OK

The scatter diagram can then be edited by bringing up the SPSS for Windows Chart Editor. For example, the type and size of the markers can be changed by clicking anywhere in the chart in the Chart Editor and then → Format and then → Marker...

Comparing means and eta
To produce a table like Table 11.5, these steps should be followed:
1. → Analyze → Compare Means → Means... [opens the Means dialog box shown in Plate 12.18]
2. → cardmins → • button to the left of Dependent List: → reasons for visiting → • button to the left of Independent List: → Options [opens the Means: Options sub-dialog box]
3. -> A nova table and eta underneath Statistics for First Layer —» Continue [closes the Means: Options sub-dialog box and returns you to the Means dialog box shown in Plate 12.18] —» OK

Generating a contingency table with three variables
To create a table like that in Table 11.7, you would need to follow these steps:
1. -> Analyze —» Descriptive Statistics —» Crosstabs ... [opens the Crosstabs dialog box shown in Plate 12.13]
2. -> othsourc —» • button by Row[s] [othsourc will appear in the Row[s]: box]
3. -> age3 [this is the name given to the newly created variable with age recoded into three categories] —» • button by Column [s]: [age3 will appear in the Column [s]: box] —» gender —» • button beneath Previous [gender will appear in the box underneath Layer 1 of 1] —» Cells [opens Crosstabs: Cell Display sub-dialog box shown in Plate 12.14]
4. Make sure Observed in the Counts box has been selected. Make sure Column under Percentages has been selected. If either of these has not been selected, simply click at the relevant point. —» Continue [closes Crosstabs: Cell Display sub-dialog box and returns you to the Crosstabs dialog box shown in Plate 12.13]
5. —» OK
The resulting table will look somewhat different from Table 11.7 in that gender will appear as a row rather than as a column variable.

Further operations in SPSS

Saving your data
You will need to save your data for future use. To do this, make sure that the Data Editor is the active window. Then, —» File —» Save As... .
The Save Data As dialog box will then appear. You will need to provide a name for your data, which will be placed after File name:. I called the file 'gym study'. You also need to decide where you are going to save the data—for example, onto a floppy disk. To select the destination drive, —» the downward pointing arrow to the right of the box by Save in. Then choose the drive to which you want to save your data. Then —» Save.

Remember that this procedure saves your data and any other work you have done on your data—for example, value labels and recoded variables. If you subsequently use the data again and do more work on your data, such as creating a new variable, you will need to save the data again or the new work will be lost. SPSS will give you a choice of renaming your data, in which case you will have two files of data (one with the original data and one with any changes), or keeping the same name, in which case the file will be changed and the existing name retained.

Retrieving your data
When you want to retrieve the data file you have created, —» File —» Open The Open File dialog box will appear. You then need to go to the location in which you have deposited your data to retrieve the file containing your data and then
Open. A shortcut alternative to this procedure is to -> the first button on the toolbar (it looks like an open file), which brings up the Open File dialog box.

Printing output
To print all the output in the SPSS for Windows Output Viewer, make sure that the Viewer is the active window and then -> File -> Print The Print dialog box will appear and then --> OK. To print just some of your output, hold down the Ctrl button on your keyboard and click once on the parts you want to print. The easiest way to do this is to select all the elements you want in the output summary in the left-hand segment of the Output Viewer shown in Plate 12.10. Then bring up the Print dialog box. When the Print dialog box appears, make sure Selection under Print Range has been selected. The third button on the toolbar (which appears as a printer) provides a shortcut to the Print dialog box.

OVERVIEW
The aim of this chapter has been to introduce the ways in which SPSS can be employed to implement the techniques learned in Chapter 11. Learning new software requires some perseverance and at times the work put in does not seem to be worth the learning process. But it is worth it. It would take you far longer to perform calculations on a sample of around 100 than to learn the software. If you find yourself moving into much more advanced techniques, the time saved is even more substantial, particularly with large samples. One final point is to remind you of the desirability of becoming familiar with SPSS before you begin designing your research instruments, because it is advisable to be aware of difficulties you might have in presenting your data in SPSS at an early stage.

QUESTIONS FOR REVIEW

Getting started in SPSS
• Outline the differences between: variable names, variable labels, and value labels.
• In what circumstances might you want to recode a variable?
• In what circumstances might you want to create a new variable?

Data analysis with SPSS
Using the gym survey data, create:
• a frequency table for exercise,
• a bar chart and pie chart for exercise and compare their usefulness;
• a histogram for cardmins;
• measures of central tendency and dispersion for cardmins;
• a contingency table and chi-square test for exercise and gender,
• Pearson’s r for age and cardmins;
• Spearman’s rho for carduse and weiuse;
• a scatter diagram for age and cardmins;
• a comparing means analysis for totalmin and reasons for visiting.