

Contrastive Analysis of English and Polish Surveying Terminology

Contrastive Analysis of English
and Polish Surveying Terminology

By

Ewelina Kwiatek

**CAMBRIDGE
SCHOLARS**

P U B L I S H I N G

Contrastive Analysis of English and Polish Surveying Terminology,
by Ewelina Kwiatek

This book first published 2013

Cambridge Scholars Publishing

12 Back Chapman Street, Newcastle upon Tyne, NE6 2XX, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Copyright © 2013 by Ewelina Kwiatek

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-4438-4410-1, ISBN (13): 978-1-4438-4410-9

TABLE OF CONTENTS

List of Illustrations	viii
List of Tables	x
Preface	xii
Introduction	xiii
Chapter One.....	1
Creating a Termbase for Surveying Terminology	
1.1 Theoretical backgrounds	
1.1.1 The mainstream approach	
1.1.2 Sociocognitive approach	
1.1.3 FrameNet	
1.1.4 Conclusions	
1.2 Surveying	
1.2.1 Differences in the naming of the field	
1.2.2 Classification systems	
1.2.3 Development of the classification system for surveying	
1.3 Selection of data categories	
1.3.1 Overview of data categories	
1.3.2 Issues in selecting and organising data categories	
1.3.3 Description of data categories	
1.4 Software aspects of terminology management	
1.4.1 Criteria for evaluating terminology management systems	
1.4.2 Terminology management solutions	
1.4.3 Design of the surveying termbases in MS Access	
Chapter Two	68
Methodology of Term Collection	
2.1 Approaches to term collection	
2.2 Creation of the corpus	
2.2.1 Design of the corpus	
2.2.2 Collecting texts for the corpus	
2.3 Term extraction tools	
2.4 Concordances	

Chapter Three	94
Analysis of Terms	
3.1 Linguistic processes	
3.1.1 Remarks on Polish phonology and morphology	
3.1.2 Word formation processes	
3.1.3 Multi-word units	
3.2 Terminological processes	
3.3 Summary	
Chapter Four	154
Analysis of Concepts	
4.1 Theories of meaning	
4.1.1 Referential theory	
4.1.2 Communicative theory	
4.1.3 Mentalist theory	
4.1.4 Semantic relations	
4.1.5 Componential analysis	
4.2 Ontologies	
4.2.1 Software for creating ontologies	
4.2.2 Ontologies as knowledge organisation systems	
4.3 Conceptual mismatches	
4.3.1 Case study 1: theodolite vs transit	
4.3.2 Case study 2: level vs spirit level	
4.3.3 Case study 3: surveying vs geodesy	
4.3.4 Case study 4: surveying assistant vs chainman	
4.3.5 Case study 5: aspect of projection, tangency and case	
4.3.6 Case study 6: mapping methods	
4.3.7 Case study 7: system of route classification in the UK and Poland	
4.3.8 Case study 8: land registration vs cadastre	
4.4 Translational strategies for dealing with conceptual mismatches	
4.5 Classification and solution of translation problems resulting from conceptual mismatches	
4.5.1 Case study 1: transit vs teodolit reiterationjny	
4.5.2 Case study 2: levels	
4.5.3 Case study 3: surveying vs geodesy	
4.5.4 Case study 4: surveying assistant vs chainman	
4.5.5 Case study 5: aspect of projection and tangency	
4.5.6 Case study 6: mapping methods	
4.5.7 Case study 7: public rights of way	
4.5.8 Case study 8: land registration vs cadastre	
4.5.9 General discussion	
4.6 Summary	

Chapter Five	291
Conclusion	
Abbreviations	293
Glossary	296
Bibliography	300
Index	320

LIST OF ILLUSTRATIONS

- Figure 1-1 Termbase record from the MS Access surveying termbase in the form view
- Figure 2-1 A schematic presentation of the four logical possibilities for hits and misses in term extraction
- Figure 2-2 Default settings in MultiTerm Extract
- Figure 3-1 Structures of the neoclassical compound *trigonometry*
- Figure 3-2 Typology of multi-word expressions.
- Figure 4-1 Ogden's and Richards' (1949) semiotic triangle
- Figure 4-2 Saussure's model of the sign
- Figure 4-3 Semiotic triangle in the mentalistic theory
- Figure 4-4 Meronymy system
- Figure 4-5 The relation of hyponymy
- Figure 4-6 Draft concept system in CAOS.
- Figure 4-7 Representation of the noun *table* in WordNet
- Figure 4-8 Representation of the verb *walk* in WordNet
- Figure 4-9 Representation of the adjective *intelligent* in WordNet
- Figure 4-10 Representation of the noun *krzesło* in Polish WordNet
- Figure 4-11 The entry for interjection in GOLD (2010)
- Figure 4-12 Transit
- Figure 4-13 Theodolite
- Figure 4-14 Aspects for cylindrical and planar projections
- Figure 4-15 Tangency
- Figure 4-16 Choropleth map of the population of Albania
- Figure 4-17 Cartogram of the total population of countries in the world
- Figure 4-18 Cartogram of the Gross Domestic Product of countries in the world
- Figure 4-19 Signposting of public rights of way
- Figure 4-20 Definitive Map of the Parish of Eythorne, Kent
- Figure 4-21 Waymarking of foot trails in Poland
- Figure 4-22 Signposting of car trails in Poland
- Figure 4-23 The tourist trails in the Świętniki Górne region
- Figure 4-24 Specimen from the property register
- Figure 4-25 Specimen from proprietorship
- Figure 4-26 Specimen from charges register
- Figure 4-27 Title plan
- Figure 4-28 Extracts from a land and building register and from a cadastral map
- Figure 4-29 Parallel tree diagram
- Figure 4-30 The project-down approach
- Figure 4-31 Hyperonymic approach
- Figure 4-32 No Interlingua

Figure 4-33 EuroWordNet

Figure 4-34 No Interlingua

Figure 4-35 The concept lattice for horses

Figure 4-36 Scalar diagram for colours in English and Welsh

LIST OF TABLES

- Table 1-1 Extension of the classification system
- Table 1-2 Juxtaposition of surveying fields in the UDC, Britannica and in Bannister et al. (1998)
- Table 1-3 Missing surveying fields and codes in the *EB*
- Table 1-4 Sample monolingual terminological record according to Cabré's guidelines
- Table 1-5 Sample monolingual terminological record according to Sager's guidelines
- Table 1-6 Monolingual terminological record created on the basis of Wright's theory
- Table 1-7 The term record created specifically for the requirements of surveying terminology on the basis of existing term records
- Table 1-8 Term record for nüvi® 1200 pocket street GPS with minimum granularity
- Table 1-9 Term record for nüvi® 1200 pocket street GPS with increased granularity
- Table 1-10 Settings of data categories in MS Access
- Table 2-1 A quantitative analysis of methods of text acquisition used to compile surveying corpora in English and Polish
- Table 2-2 Number of term candidates and number of actual terms in English corpora
- Table 2-3 Number of term candidates and number of actual terms in Polish corpora
- Table 3-1 Classification of prefixes
- Table 3-2 Classification of subordinate endocentric verb-containing compounds
- Table 4-1 Componential analysis of English furniture items
- Table 4-2 Transit vs theodolite
- Table 4-3 Theodolite in Polish
- Table 4-4 Levelling instruments and devices
- Table 4-5 Parts of levelling instruments
- Table 4-6 Levelling instruments in Polish
- Table 4-7 Levelling instruments' parts in Polish
- Table 4-8 Names of the field in English
- Table 4-9 Concepts behind the field names in Polish
- Table 4-10 Surveying assistant vs chainman
- Table 4-11 Aspect of projection vs tangency
- Table 4-12 Types of thematic maps
- Table 4-13 Maps and mapping methods in English
- Table 4-14 Mapping methods in Polish
- Table 4-15 Kartogram in the Polish concept system
- Table 4-16 Concept system of the public rights of way in the UK
- Table 4-17 Land registration system in the UK
- Table 4-18 System of registering rights to land in Poland
- Table 4-19 Cadastral division of Poland
- Table 4-20 Definition of words for horses
- Table 4-21 Analysis of definitions of horses

Table 4-22 Equivalents for *transit* in Polish

Table 4-23 English terms based on *level* referring to levelling instruments and their parts along with their Polish equivalents

Table 4-24 Polish equivalents for *dumpy*, *tilting* and *wye level*

Table 4-25 Matching English and Polish surveying terms

Table 4-26 Matching English and Polish terms referring to mapping methods

Table 4-27 Translation of English national institutional terms into Polish

PREFACE

This book is based upon my PhD thesis at Swansea University, submitted in February 2012. Initially, I would not have thought of undertaking my PhD had my supervisor, Pius ten Hacken, not encouraged me to do so. When I came to Swansea in 2006, I only planned to stay for a year to do an MA in Translation with Language Technology. I already had a background in surveying (MSc in Geodesy and Cartography from Wroclaw University of Environmental and Life Sciences) and a passion for technical translation. While in the final stage of working on my MA dissertation in Swansea, Pius suggested I could combine the two in the research on surveying terminology. I am extremely grateful for his guidance and providing me with constructive criticism through the entire PhD project. I would also like to thank my second supervisor, Alison Williams, for painstakingly going through a very rough version of each chapter of my thesis and the entire manuscript, for correcting my clumsy formulations and indicating points that are ambiguous or not explicit enough to the reader.

I am also grateful to the Department of Languages, Translation and Media at Swansea University as I was able to benefit from their infrastructure and to Andrew Rothwell, the Head of the Department, for his comments and suggestions regarding the software aspects of terminology management. Thanks are also due to the Department of Computational Linguistics at Copenhagen Business School, and in particular to Hanne Erdman Thomsen, for giving me access to CAOS. In addition, I would like to thank my colleagues from the University with whom I had stimulating discussions over the years.

Finally, this research could not have been completed without the support of my family – my husband Karol, who helped me with many technical aspects of this research, our daughter Sara, who turned out to be a very cooperative baby, and my parents and parents-in-law.

INTRODUCTION

This book provides a comparison and analysis of surveying terminology in English and Polish. The purpose of the book is three-fold: firstly, to investigate how surveying terms are created and how they are named in English and Polish, secondly, to analyse concept systems of the two languages with respect to surveying terminology, and finally, to indicate the areas of surveying in which terminological and conceptual differences occur, the factors that trigger them and translation strategies which are used to solve them.

I have chosen surveying terminology for my research field as I have been educated and have professional experience in this field, which allowed me to discover that there are many terms which do not have equivalents in the target language, as well as concepts which occur only in one language. The name of the field itself is both intriguing and ambiguous. The field is commonly referred to as *surveying*, or most recently *geomatics*, in the Anglo-Saxon countries, while in the continental European tradition, it is called *geodesy* (*geodezja* in Polish). Arguably, the name *surveying* is too general because it indicates that anything can be surveyed, e.g. *literature survey*. I would personally opt for the name *land surveying*, which was used in the past. However, the paradigm changed in the Anglo-Saxon tradition and the current name of the field is *surveying*. Thus, to be consistent with the English conventions, I will refer to the field as *surveying* and will consider it as equivalent to *geodesy*.

Surveying terminology is an under-researched area and publications referring to surveying terminology in English and Polish are scarce. To my knowledge, there are four bilingual surveying dictionaries for English and Polish. Although they are called dictionaries, they are rather glossaries as they typically provide terms and their equivalents. The oldest one, the English-German-Polish dictionary by Tatańczyk (1991) includes 5,310 entries. A modified and expanded version of this dictionary was published on CD-ROM in 2005 (Tatańczyk, 2005). It includes 8,500 entries and covers various subfields of surveying and related disciplines, e.g. astronomy, civil engineering, physics, photography, photointerpretation, GPS, geology, computer science, mathematics, mining, remote sensing and optics. Apart from giving equivalent terms in the target language, the

dictionary also provides gender and part of speech specification for German terms.

A second surveying dictionary is the Polish-English, English-Polish dictionary of terms from the fields of surveying, cartography and real estate by Downarowicz and Leśniok (2006). It includes 30,000 entries arranged in alphabetical order. It includes only terms and their equivalents in the target language.

The Internet dictionary of geomatics by Gaździcki (2005) is much more specialised than the other two dictionaries. However, at the same time it is quite limited as it covers only the part of the surveying domain that deals with state-of-the-art measurement technologies and computations. The dictionary is based on the printed version of *Leksykon geomatyczny - Lexicon of Geomatics* (Gaździcki, 2002). The lexicon was made available on-line in 2004 and updated in 2005. It includes both English and Polish entries. However, the Polish part of the dictionary contains more information, as, apart from equivalents, the dictionary also gives definitions for terms which are not available for English entries. Terms in the two languages are not the same; the Polish part includes more entries than the English part.

The English-Polish, Polish-English dictionary by Hycner and Szortyka (2005) is organised differently as it is divided into three parts. Part 1 is devoted to technical subfields of surveying such as geodetic surveying, cartography, GPS, photogrammetry and remote sensing, while part 2 is more legal-based, covering terms referring to cadastre. Part 3 includes names of the subfields of surveying and related disciplines. Each part is arranged in alphabetic order both in English and in Polish.

The above-mentioned dictionaries, although quite useful to surveyors, do not provide sufficient information for technical translators as they neglect problematic terms. Entries where conceptual mismatches occur are either simplified by providing a direct translation or omitted. For this reason, these dictionaries are of little help to those who are not experts in the field.

These dictionaries are the only publications on surveying terminology in English and Polish. There are no publications on terminological research in the surveying field. This type of research is required to enrich the content of surveying dictionaries and make them more useful for translators and technical writers, and also for specialists in the surveying field. Surveying terminology needs to be approached more systematically, preferably using a corpus-based study that provides evidence for documenting terms and concepts in addition to term-related and concept-related information.

My aim is to shed light on surveying terminology by combining terminological knowledge with expertise in the field in order to create two monolingual concept-oriented termbases with explicit information on concepts. These termbases will be used to identify conceptual mismatches and offer solutions for dealing with them. Data collected in these termbases will be analysed in the book to identify the differences between the terminological and conceptual systems in English and Polish.

The termbases include concepts mainly from three subfields of surveying: geodetic surveying, cartography and GPS, which have been selected out of ten subdomains of surveying with a reason. The first two fields are quite traditional and developed independently in different countries, whereas GPS is a relatively new field and has many recently-added terms. The methodology I developed for dealing with conceptual mismatches is extendible because I cover all types of problems that are expected in the field of surveying. It can be used to expand the research to the remaining seven subfields of surveying.

The book is structured into two parts. Part one is concerned with the compilation of termbases which document surveying terms and concepts (chapters one and two), and part two focuses on the analysis of terms and concepts in these termbases (chapters three, four and five).

After this introduction, chapter one describes how monolingual surveying termbases with translation equivalents were created in English and Polish. It discusses different approaches to terminology and terminological databases and their usefulness for creating surveying termbases (1.1). After that, it moves to a discussion of the problems with the name of the field and elaborates a classification of the field into a number of subfields (1.2). Next, selection of data categories for the surveying termbases is described (1.3), followed by a discussion of the termbase design and selection of the software (1.4).

Chapter two describes how surveying terms were collected. First, it discusses what approaches to terminology collection are available and selects those which are useful in this project (2.1). Then, it presents how the surveying corpora were designed and compiled (2.2) and how terms were extracted from them (2.3). It ends with a discussion of concordances (2.4).

Chapter three analyses how terms were created and how they were named. Starting from surveying termbases described in the course of previous chapters section (3.1) elaborates on linguistic processes which include word formation, and section (3.2) discusses term naming.

Chapter four concentrates on the meaning of terms. First, it examines various theories of meaning and their features (4.1). Then, it presents

ontologies and their role in the representation of concept systems and discusses semantic relations between concepts (4.2). After that, conceptual mismatches in surveying termbases in English and Polish are described (4.3), followed by the presentation of translation strategies for dealing with them (4.4). Then, a classification of translation problems caused by conceptual mismatches is provided and solutions to selected conceptual mismatches are offered (4.5). The last section (4.6) concludes the whole chapter. Chapter five provides a summary of the research.

CHAPTER ONE

CREATING A TERMBASE FOR SURVEYING TERMINOLOGY

The aim of this chapter is to describe the working methods used in the creation of a termbase of surveying concepts. First, different approaches to terminology and terminological databases are discussed and their potential application for generating a surveying termbase is evaluated (1.1). Then, problems with the name of the field are discussed and a classification of the field into a number of subfields is developed (1.2). Next, an overview of data categories, their features and attributes is provided and individual decisions are made about which data categories to include in the surveying termbase. Choices for writing up terminological records are made at this stage (1.3). Finally, different software packages for terminology management are compared in order to select the one to be used in this project. These considerations are followed by a description of the design of the termbase in the selected software package (1.4).

1.1 Theoretical backgrounds

Terminology as a scientific discipline started to develop in the 1930s, when scholars from Russia, Austria and Czechoslovakia became aware of the proliferation of terms and the diversity of forms as well as relationships between terms and concepts (Cabr , 1999, p. 7). Modern terminological research was commenced by the engineer Eugen W ster (1898-1977). His doctoral thesis (1931) was the first work on terminology. It outlined a new approach to terminology as the author did not focus on compilation of the specialised vocabulary or standardisation of existing terminology, but was interested in establishing principles for the creation of new terms (Pearson, 1998, p. 9).

His interest in the theory of terminology, known as the traditional approach to terminology, appeared more than thirty years later with the publication of *Die vier Dimensionen der Terminologiearbeit* ‘The four dimensions of terminology work’ (1969), in which he presented for the first

time four aspects of terminology work: the special subject field, the languages, the purpose and the degree of abstraction. His overall approach to the theory of terminology, *Einführung in die Allgemeine Terminologielehre und terminologische Lexikographie* 'Introduction to the General Theory of Terminology and to terminological lexicography', was published posthumously in 1979 (Cabr , 1999, p. 225).

The traditional approach represented by W ster and adopted by many later prominent terminologists was challenged by the corpus-based approach which evolved in the 1980s (Sager, 1990, p. 56). The traditional and corpus-based approaches constitute the mainstream approach to terminology, which is described in section 1.1.1. An alternative approach was developed by Rita Temmerman and is known as the sociocognitive theory of terminology. This approach is characterised in section 1.1.2. Apart from the traditional approach and the sociocognitive approach, there is a third approach to terminology, based on the FrameNet formalism. This approach evolved in parallel with the formation of the traditional approach and is presented in section 1.1.3.

1.1.1 The mainstream approach

The traditional approach to terminology is often referred to as the *general theory of terminology* (Pearson, 1998, p. 10). It is based on the theory developed by an Austrian engineer, Eugen W ster. In his theory of terminology, W ster (1979, p. 2) argued that the proper description of terms differs from the proper description of general language words. He suggested that work on terms differs from work on general language words in three respects. First, terminology work starts from the concept. Concepts exist independently of terms and any expression used to designate them should be considered in isolation from their labels or terms. Concepts are mental abstracts to which labels are assigned. The second distinction which W ster makes refers to vocabulary. W ster believes that terminologists are interested only in vocabulary and are not concerned with the theory of morphology or syntax. Traditional terminologists were not interested in examining terms in use as they only wanted to establish what they represented. The third distinction W ster makes is about standardisation. Terminologists are concerned with imposing norms for the use of language. Their objective is to fix and standardise meaning in order to avoid confusion. This is achieved by creating a standardised collection of terms (W ster, 1979, p. 2).

W ster's approach to terminology is concept-oriented or onomasiological. It has been applied by many later terminologists. Wright (2001b, p. 579)

claims that concept orientation is a recognised international standard for terminology databases. Warburton (2001, p. 687) adds that effective management of synonyms and equivalents requires them to be linked through the concept. Concept-oriented or terminographic collections consist of multiple records linked by a concept and are usually stored in a multidimensional structure, such as a database, with multiple access points to each record.

Wüster's approach was not ideal, as in many cases terminologists start their work not from concepts but from terms, which they find in the text or corpus of texts they examine. They first come across word forms and then they try to establish their meanings. The approach they take is thus corpus-based or semasiological.

The distinction between onomasiology and semasiology is a traditional one in continental structural semantics and in the Eastern European tradition of lexicographic research (Geeraerts, 2006, p. 37). It is quite hard to establish which of these two perspectives came first.

The onomasiological tradition seems to be older as it was used between the post-classical European written culture (c.800) and 1700 in various genres of text including non-alphabetical glosses, glossaries and dictionaries (Hüllen, 1999, p. 406). However, the term *onomasiology* was coined only in 1902 by the German linguist Adolf Zauner in his dissertation on body-part terminology in Romance languages (Grzega, 2002, p. 1021).

On the contrary, the term *semasiology* was created between 1822 and 1824 by a German scholar, Christian Karl Reising (1792-1829), who introduced an architecture of grammar that comprises an explicit semasiological component alongside traditional elements such as etymology and syntax (Schmitter, 2008, p. 575). Thus, in Reising's view, semasiology indicated the meaning of the word form.

Reising's ideas were popularised by his followers, Friedrich Haase (1808-1867) and Ferdinand Heerdegen (1845-1930). The term *semasiology* was also applied in England and the USA as well as in German-speaking countries. Its dominance lasted until it was replaced by *semantics*. The use of the term *semantics* by Charles R. Lanman (1850-1941), a scholar of Sanskrit at Harvard, in a lecture in the USA in 1894 marks the beginning of the gradual replacement of the term *semasiology*. In France, by contrast, the term *sémasiologie* was never in contention. The term which was established there was *sémantique*, introduced by Michael Bréal (1832-1915) in 1897 (Schmitter, 2008, p. 584). The term *sémantique* was adopted all over the world, while the popularity of the term *semasiology* decreased and it remained in use only in German-speaking countries. The

term semasiology ceased to denote all aspects relating to the theory of meaning and was instead used with reference to a specific semantic perspective (Schmitter, 2008, p. 585), which starts from forms and looks for their meaning.

Two terms which received recognition in different parts of the world are *semiotics* and *semiology*. Both terms have the Greek word *semeion* ‘sign’ as their etymological source and refer to the study of signs. They have different histories, however. The term *semiotics* has an American origin as Charles Sanders Peirce (1839-1914), an American philosopher and logician, used several variants of the Greek word in his works: ‘semeiotic’, ‘semeiotic’ and ‘semiotic’. The diffusion of ‘semiotics’ as the currently accepted form began in the mid-1960s (Nuessel, 2006, p. 193). The term *semiology* has a French origin as Ferdinand de Saussure, a Swiss linguist, used the French expression *sémiologie* to name the study that deals with social production of meaning from sign systems (Saussure, 1916/1969, p. 68). The term *semiology* travelled to the United States under the influence of Saussure’s linguistic theories, while *semiotics* travelled to Europe and became the preferred designator of the field today (Nuessel, 2006, p. 193).

Referring back to *semasiology*, Karpova (2006, p. 709) claims that *semasiology* in its current sense was developed in the 1980s and covers various aspects of a word’s semantic structure; from the simple correlation of a word and a concept, to the theory of reference and a description of a hierarchically organised structure of lexical entries. The development of computer technologies in the 1980s which facilitate the storage and management of large corpora in electronic form revolutionised the semasiological approach which became corpus-based. The semasiological perspective overshadowed the onomasiological one as most dictionaries use a universal alphabetical order for arranging entries (Hartmann, 2006, p. 669). Entries are derived from the corpus which provides other information on entries such as examples, grammatical data and semantic relations. The semasiological approach is applied by a number of terminologists including Jennifer Pearson, Juan C. Sager and Maria Teresa Cabré.

The corpus-based approach must not be confused with the corpus-driven approach. The fundamental distinction between these two methods was introduced by Tognini-Bonelli (1996). It was made on the basis of the purpose for which the corpus is being used. In the corpus-based approach, corpora are used mainly to “expound, test or exemplify theories and descriptions” (Tognini-Bonelli, 1996, p. 1). The corpus works as a repository. It is used, for example, to validate existing categories or different

applications, to test a tagger or a parser (Tognini-Bonelli, 1996, p. 1). On the other hand, the corpus in the corpus-driven approach is more than a repository of examples to back pre-existing theories. It is also used to discover new facts in order to refine the hypothesis (Pearson, 1998, p. 49).

Initially, concept-oriented and corpus-based approaches were in conflict but in time they were integrated and nowadays they are often used together and complement each other. Van der Vliet (2006, p. 62) claims that a system of concepts, which is an intermediate structure between terms and domain knowledge, should be built by combining the top-down approach using domain knowledge and the bottom-up approach using a corpus. Terminologists using the top-down approach should be familiar with the subject field before they start structuring the field as it helps to build a concept system and establish relations between concepts. The top-down approach is concept-oriented (or onomasiological) as it starts from concepts and looks for their names. The candidate terms and collocations that can be linked to various concepts are extracted from a corpus of texts relevant for a particular domain by using the bottom-up approach.

The bottom-up approach starts from words and looks for their meanings. It relies on the corpus, which is examined to find the terms and to observe the way terms combine with other terms in compounds, collocations and sentences. Therefore, this approach is referred to as corpus-based or semasiological. In the corpus-based approach, the terms and their combinatoric properties are a basis for collecting and structuring domain knowledge (van der Vliet, 2006, p. 62).

Apart from strictly corpus-based or concept-oriented approaches, some terminologists, for example Wright (2001b, p. 552), suggest a third solution, which is based on the combination of both approaches. A corpus is the source of the information and a starting point for the research. A terminologist who applies this solution gets a list of terms as a result of the extraction process. These terms need definitions. While defining these terms and providing examples that document the use of the terms, the terminologist encounters new terms in definitions and examples and becomes aware of the concepts they represent. Then, he starts working in the opposite direction, from concepts to their meanings, looking at the concept structure of a particular domain and relations between concepts such as synonymy, hyperonymy, holonymy. The list of terms obtained in the extraction process is only a reference list. Terms from this list are usually entries in the termbase, but the number of terms in the termbase will be much higher than the number of terms on this list, as the terminologist aims to create a complete network of concepts and has to define all new concepts that occur in definitions and examples.

1.1.2 Sociocognitive approach

The sociocognitive approach to terminology was developed by Temmerman (2000). This approach relies on *socioterminology*, which is a relatively new trend in terminology that tries to get the study of terminology back to the study of real language usage (Boulanger, 1995, p. 197). Socioterminologists state that language is not suited for standardisation as it changes all the time. They also believe that social aspects, alongside cognitive ones, should be considered in terminological theory and practice. Temmerman reacts against the traditional onomasiological approach to terminology, against traditional definitions of concepts (consisting of an intensional or extensional definition) that reflect the position of the concept in a concept system, and against the univocity principle which states that there should be a one-to-one correspondence between a concept and a term. Temmerman strongly supports a corpus-based approach which is the starting point for her analyses. She uses a corpus of texts on the life sciences to carry out an empirical study of categorisation and lexicalisation processes. On the basis of her findings, she questions the validity of traditional approaches to terminology and suggests an alternative, which is inspired by the cognitive sciences. Temmerman (2000) bases her research and theory on such paradigms as hermeneutics (which is the main source of post-modernism represented by Derrida) and the cognitive approach in semantics, which both react against structuralism.

Temmerman (2000, p. 1-9) examines the term *concept* and possibilities for describing its meaning. She suggests that terminologists should start from *units of understanding* instead of *concepts*. She claims that *concepts* along with *categories* are two kinds of *units of understanding*. Only a few units of understanding do not have prototype structures and could therefore be named concepts (e.g. intron). The ones which have a prototype structure are categories (Temmerman, 2000, p. 43).

Temmerman (2000, p. 95) uses Idealised Cognitive Models (ICMs) to study units of understanding in the field of the life sciences. ICMs were discovered and described by Lakoff (1987) who, inspired by Fillmore's frame semantics, wrote a book entitled *Women, Fire and Dangerous Things. What Categories Reveal about the Mind*. The main message delivered by this book is that people organise their knowledge by means of structures called *idealised cognitive models* (ICMs) and category structures and prototypes are by-products of that organisation. Temmerman (2000) builds on Lakoff's theory and claims that ICMs may be perceived as conceptual prerequisites for understanding the meaning. ICMs are complex structures consisting of units of understanding. The intracategorical or internal structure (aspects, facets) of the units of understanding, as well

as its intercategory structure (the relationship of the given unit of understanding with other units of understanding with the same frame of mind), depend on the ICM within which a unit of understanding has been identified (Temmerman, 2000, p. 96).

Temmerman also believes that terminologists should replace traditional definitions by templates of meaning description. In traditional terminology, concepts are defined on the basis of necessary and sufficient characteristics (Temmerman, 2000, p. 226). In sociocognitive terminology meaning description can be presented via a template. There will be different types of templates and different modules of information within the template depending on whether the unit of understanding is an umbrella category, an entity or an activity. Different information modules can vary in information importance on a scale from 0 to 2 (from 0=irrelevant to 2=prominent). Temmerman also believes that synonymy and polysemy are functional in the process of understanding and that there does not need to be a one-to-one correspondence between the concept and the term. Polysemisation can occur due to a change in the world as new technology evolves or may be caused by a change in the understanding of the category.

Temmerman explores the link between the structure of understanding a category and the process of lexicalisation. She argues that two opposing forces are at work when categorisation takes place within a language community (Temmerman, 2000, p. 133). One of these forces is the urge for univocity and the other one is the urge for diversification (due to the evolution of categories and units of understanding over time, which results in polysemy and synonymy). She also analyses the mechanisms behind the urge for a new and better understanding of terms. Temmerman's hypothesis states that these mechanisms are related to, and inspired by, metaphorical reasoning (Temmerman, 2000, p. 158). Metaphor is a multidimensional phenomenon. Until the late 20th century it was used only to describe a literary figure of speech, but recently its use has widened to philosophy, psychology, linguistics and other cognitive sciences.

Lakoff and Johnson (1980) develop a new approach to the study of metaphor. They state that metaphor is present in everyday life, not just in language but also in thought and action. They also argue that human conceptual system is fundamentally metaphorical in nature. The central aspect of their argument is that metaphor is a kind of thinking or conceptualisation not limited to language. Language, however, helps to observe how metaphor works. They use a conceptual metaphor (or metaphorical conceptualisation), ARGUMENT IS WAR, to demonstrate

how a concept can be metaphorical and can structure an everyday activity. The conceptual metaphor ARGUMENT IS WAR may be encountered in (1):

- (1) Your claims are *indefensible*.
 He *attacked every weak point* in my argument.
 I *demolished* his argument.

The words in italics refer to war. Lakoff and Johnson (1980, p. 4) suggest that the examples given above provide evidence that our culture conceptualises arguments through the war metaphor and therefore the way in which people conduct arguments is conditioned by the way in which they conduct wars.

Lakoff and Johnson's view on metaphor, its role in everyday life and relation to language, was recognised and elaborated by many other linguists. For example, Knowles and Moon (2006, p. 4) claim that metaphor is a basic process in the formation of words and meanings. Concepts are lexicalised through metaphor. Many senses of polysemous words are metaphors of different types, e.g. *jewel* is a metaphor for something valuable, *fox* is a metaphor for a cunning, wily person.

Temmerman carried out a study of metaphorical lexicalisations or part of metaphorical ICMs (m-ICMs) in the language of the life sciences. The aim of her study is to prove the non-arbitrariness of the sign in the life sciences, in the sense that metaphoric models result in lexicalisations. She wants to prove that terms are motivated. It is a very interesting approach as it opposes the Saussurian principle of the arbitrary character of linguistic signs, which says there are no motivated links between the signifier and the signified (Saussure, 1916/1969, p. 66). The main outcome of Temmerman's study is a description of metaphorical models that explain the phenomenon of metaphor at the lexeme, category and domain levels.

Having analysed features of traditional principles and new propositions, Temmerman (2000, p. 223) develops the following set of alternative principles to traditional terminology:

1. Sociocognitive terminology starts from units of understanding - most of which have a prototype structure.
2. As understanding is a structured event, a unit of understanding has an intercategory and intracategory structure and it functions in cognitive models.
3. Depending on the type of unit of understanding and on the level and type of specialisation of sender and receiver in communication, what is more essential or less essential information will vary.

4. Synonymy and polysemy are functional in the progress of understanding and therefore need to be described.
5. Units of understanding are constantly evolving; cognitive models (e.g. metaphorical ICMs) play a role in the development of new ideas which implies that terms are motivated.
6. Sociocognitive terminology supports a combined semasiological and onomasiological approach to terminography. It takes into consideration the role of metaphorical idealised cognitive models and synonymy and polysemy in the process of understanding. It replaces traditional definitions with templates for meaning description. The distinction between concepts and categories is central within sociocognitive terminology. Temmerman's approach questions the validity of traditional terminology as, in practice, terminologists start their work from concepts rather than from a list of terms.

Temmerman's approach, in a slightly modified version called termontography, is currently used at the Erasmushogeschool in Brussels, where she is a lecturer (Thelen & Steurs, 2010, p. 260). Termontography is a multidisciplinary approach in which theories and methods for multilingual terminological analysis of sociocognitive theory are combined with methods and guidelines for ontology engineering (Centrum voor Vaktaal en Communicatie, 2009). In this approach, a clear distinction is made between conceptual modelling at a language-independent level and language-specific analysis of units of understanding. A key view in termontography is that knowledge analysis should precede the methodological processes which are generally conceived as the starting-points in terminography, i.e. the compilation of a domain-specific corpus of texts and the understanding and analysis of the categories that occur in a certain domain. This view is supported by the fact that the aim of terminological databases is to represent in natural language those items of knowledge or *units of understanding* which are regarded as relevant to specific purposes, applications or groups of users. In termontography, the units of understanding and their intercategorical relations are structured in a common knowledge base or categorisation framework. On the one hand, this framework supports the information-gathering phase during which a corpus is developed. On the other hand, it allows terminographers to establish specific extraction criteria to define what should be considered a 'term'. Moreover, pre-defined knowledge also influences the terminographer's working method and the software tools that will be employed to support that working method.

1.1.3 FrameNet

FrameNet is a computational lexicographic project whose purpose is to represent information about the semantic and syntactic properties of English words and encode this information in a database (FrameNet, 2008). This information is extracted from large electronic corpora using manual and electronic techniques. Although FrameNet was not developed for terms, the database and its construction technique may be very useful in describing them (van der Vliet, 2006). FrameNet is based on the theory of Frame Semantics, whose central idea is that word meanings must be described in relation to semantic frames.

The linguistic basis of FrameNet is Fillmore's theory of Case Grammar (Fillmore, 1968). FrameNet itself was developed in two stages: FrameNet I and FrameNet II (FrameNet, 2008). The FrameNet database consists of a lexical database and an annotation database. The lexical database contains entities of the following types:

- frames: named data structures which are used to represent a concept in a domain, e.g. *survey*;
- frame elements (FEs): kinds of entities that can participate in the frame, e.g. FEs of the frame *survey* are: *contractor*, *object*, *purpose*;
- lexemes: sets of forms taken by a single word (e.g. the English verb lexeme *measure* has four word forms: *measure*, *measures*, *measured*, *measuring*);
- lemmas: particular forms of the lexeme that are chosen by convention, e.g. *measure* is the lemma of the verb lexeme *measure*. The lemma is the canonical form of a lexeme and in lexicography it is usually also the citation form or headword by which it is indexed;
- lexical units (LUs): associations between lemmas (units of form) and frames (units of meaning). LUs correspond to dictionary senses.

The following types of relations between frames and frame elements exist in FrameNet:

1. *Frame Inheritance*. Frame Inheritance is a relation between frames which depends on the fact that a more specific frame inherits from a more general frame, e.g. the *Surveying_tools* frame evoked by the verbs *measure*, *set out* and *level* inherits from the more general

Surveying frame that has the same FEs, e.g. *method*, *participants*, *accuracy* and *purpose*.

2. *The Subframe Relation.* Subframes describe subevents which are part of complex events described in terms of frames, e.g. the *cadastral* frame is a complex event and it consists of such subframes as: *ownership*, *location*, etc.
3. *The 'Uses' Relation.* This relation is similar to inheritance but less strictly defined. In this relation, frame *A* uses frame *B* and frame *A* need not have a FE corresponding to each FE of frame *B*, e.g. the *Plan* frame uses the *Map* frame, but whereas the *Map* frame has a FE *projection*, the *Plan* frame does not.
4. *The 'See also' relation.* The 'see also' relation is a pointer from one or more frames to another frame whose definition contains a discussion of differences among the frames in the group. For example, the two uses of words like *load* in *She loaded the wagon with hay* and *She loaded the hay onto the wagon* are treated in the two frames *Filling* and *Placing* respectively (Baker, Fillmore & Cronin, 2003, p. 287). The detailed discussion of differences between these two frames is located in the definition of *Filling* and there is therefore a reference, 'see also' relation, from *Filling* to *Placing*. There is, however, no systematic solution to the problem of relating these two words. It is done on an ad hoc basis.

The annotation database contains sentences and their annotation. For each lemma, there is a set of annotation layers for frame elements, phrase types, grammatical functions, etc. The lemma, along with its layers, is represented by an entry in the Annotation Set table, which links a sentence, a subcorpus and a LU.

The annotation process depends on labelling sentences using labels that indicate various semantic and syntactic properties. It is an interactive and semi-automatic process. The Semantic Annotator program presents the sentences one at a time and provides menus that identify the available syntactic phrase types, the grammatical function of the constituents and the frame element names for the target frame. The operator's role is to identify the sentence constituents corresponding to frame elements, and to tag them according to their grammatical function (subject, object, oblique), their phrase type (NP, PP, VP, etc.) and their frame element (Lowe, Baker & Fillmore, 1997).

The lexical database may include semantic types, which were mentioned above and which are used to recognise some aspects of the meaning, e.g. positive and negative connotations. It can also contain notes

which are used to record questions and problems that arise when the lexical or annotation database are created.

However, certain types of lexical information are not covered in the FrameNet database compared to dictionaries. The FrameNet database does not include phonological, morphological or etymological information about the words in the database. It does not provide information about lexical relations, e.g. synonymy, antonymy or hyponymy, either. Finally, it does not offer any statistical data about the frequency of occurrence of syntactic patterns or about LUs.

A FrameNet database is implemented as a MySQL database. Three types of interface have been developed for this database. One is for browsing, the second is for searching and the third is for updating the database. The FrameNet database is much more informative and systematic than a paper dictionary or even an average electronic dictionary. There is no fixed limit to the number of examples. It also has room for information on combinatorics, e.g. which preposition follows a certain verb and which verb to use in combination with a noun. The use of frames facilitates presenting this information as systematically as possible. A FrameNet database may be used for the production of dictionaries and in various projects in the domain of language technology.

Examples where FrameNet formalism is used include the development of multi-lingual lexica, word sense disambiguation and machine translation. The formalism is also used in Natural Language Processing (NLP) systems that perform question answering, information retrieval and automatic semantic parsing (Baker & Sato, 2003). FrameNets have been built for other languages such as German, Spanish, Japanese or even Polish. While the FrameNets in the first three languages are very well developed and have their entries linked with English-based FrameNet lexical entries to arrive at a contrastive FrameNet lexicon, the Polish FrameNet is only a small project whose aim is to describe a subset of about 200 Polish verbs using Frame Semantics. The Polish project is described by Zawisławska (2010).

One of the most practical applications of FrameNet is the *Kicktionary* created in 2006 which is a domain-specific trilingual (English, German, and French) lexical resource of the language of Soccer, created by the FrameNet visitor, Thomas Schmidt from Germany. Kicktionary relies on Frame Semantics and uses semantic relations in WordNet style as an additional layer of structure (FrameNet, 2008). The lexicon currently contains around 1,900 lexical units which are organised in 104 frames and 16 scenarios. Each LU is illustrated by a number of examples from a

multilingual corpus of football match reports from the UEFA website (Schmidt, n.d).

The idea of FrameNet databases is also applied in the *Referentiebestand Nederlands (RBN)* which is a multi-purpose lexical database of Dutch (van der Vliet, 2007). The case for using FrameNet in a terminological context was made by van der Vliet (2006, p. 57) who suggests that the FrameNet approach provides a way of describing the semantics of terms in a frame-based way by linking them to concepts. The use of frames has many advantages. It leads to the establishment of an explicit link between terms and the domain knowledge, increases consistency by the systematic description of domain knowledge and provides the possibility of basing a definition on a selection of the represented knowledge, which leads to more flexible definitions (van der Vliet, 2006, p. 61). A database with frame-based description has better retrieval possibilities. It can be entered through terms but it additionally benefits from the slots and fillers which are used to search for individual terms or sets of terms.

1.1.4 Conclusions

When the mainstream approach, the sociocognitive approach and FrameNet are compared, it may be noticed that they differ quite significantly. The mainstream approach is a combination of the onomasiological and corpus-based approaches. It originated as a traditional onomasiological approach but after the expansion of the use of computers in the 1980s it was followed by the corpus-based approach with which it was later integrated (ten Hacken, 2010b).

The insights from the sociocognitive approach are more radical. Temmerman (2000) uses the corpus-based approach as a tool to obtain textual evidence confirming the limitations of the traditional approach. She reacts critically against the onomasiological perspective, traditional definitions of terms and univocity. She suggests that the semasiological approach and templates for units of meanings would be much more efficient. She also recognises polysemy and synonymy as useful in terminology.

FrameNet is a contribution to terminology of a different type from the traditional corpus-based or sociocognitive approaches. It includes guidelines on how to build a concept system based on frames and presents the advantages which a database with a frame-based description of domain knowledge offers. The FrameNet approach is lexicographic but it can also be applied in terminology.

1.2 Surveying

The domain of surveying is very challenging for terminologists because of discrepancies in the naming of the field and its complex structure. The name of the field has evolved over time from *land surveying* to *surveying* and *geomatics*. A lack of international agreement as to how the field should be named and different naming conventions, both in continental Europe and the Anglo-Saxon countries, has increased the number of different labels used for the field. Furthermore, the field of surveying can be divided into a number of subdomains, e.g. cadastral surveying, plane surveying. Different classification schemes provide an overview of the field, but they often differ significantly in the subfields they specify.

In this section I will look at differences in the naming of the field and analyse a range of labels under which the field is known (1.2.1). I also compare and contrast classification schemes in surveying (1.2.2) and I attempt to develop a uniform classification system for this field (1.2.3) based on the analysis in the first two subsections.

1.2.1 Differences in the naming of the field

The term *surveying* is quite ambiguous as the field is named, perceived and understood differently in different countries, with notable differences existing between the continental European tradition and the Anglo-Saxon tradition. The field is named *surveying* (formerly *land surveying*) in the Anglo-Saxon tradition (Ghilani & Wolf, 2008, p. 1) and *geodesy* in the European tradition (Hycner & Dobrowolska-Wesołowska, 2008, p. 16). I will try to provide definitions and examples of usage of these terms in order to find differences in the meaning between them.

I will now look at how these terms are described in a number of textbooks on surveying, encyclopaedias and standards. In the case of *surveying*, I have an article by Lyman and Wright (2009) in the *Encyclopaedia Britannica* and a surveying textbook by Bannister et al. (1998). According to Lyman and Wright (2009), *surveying* is a method of making relatively large-scale, accurate measurements of the Earth's surface. Its principal modern uses are in the fields of transportation, building, land use, and communications. Surveying is divided into plane surveying, which deals with mapping small areas, such as a building site or a parcel, and geodetic surveying, which focuses on mapping large areas of the globe, e.g. the territory of a country. The definition of surveying by Bannister et al. (1998) and the division of surveying into geodetic and