TERMINOLOGY AND LEXICOGRAPHY
RESEARCH AND PRACTICE

Terminology and Lexicography Research and Practice aims to provide in-depth studies and background information pertaining to Lexicography and Terminology. General works will include philosophical, historical, theoretical, computational and cognitive approaches. Other works will focus on structures for purpose- and domain-specific compilation (LSP), dictionary design, and training. The series will include monographs, state-of-the-art volumes and course books in the English language.

Series Editors
Marie-Claude L'Homme
Ulrich Heid

Consulting Editor
Juan C. Sager

Volume 5

Kyo Kageura

The Dynamics of Terminology.
A descriptive theory of term formation and terminological growth.

TERMINOLOGY AND LEXICOGRAPHY
RESEARCH AND PRACTICE

Terminology and Lexicography Research and Practice aims to provide in-depth studies and background information pertaining to Lexicography and Terminology. General works will include philosophical, historical, theoretical, computational and cognitive approaches. Other works will focus on structures for purpose- and domain-specific compilation (LSP), dictionary design, and training. The series will include monographs, state-of-the-art volumes and course books in the English language.

Series Editors
Marie-Claude L'Homme
Ulrich Heid

Consulting Editor
Juan C. Sager

Volume 5

Kyo Kageura

The Dynamics of Terminology.
A descriptive theory of term formation and terminological growth.

THE DYNAMICS OF TERMINOLOGY

A DESCRIPTIVE THEORY OF TERM FORMATION AND TERMINOLOGICAL GROWTH

KYO KAGEURA
National Institute of Informatics, Tokyo

Acknowledgements

I started the research described in this book around 1990, when I began working on conceptual descriptions of term formation patterns with Professor J. C. Sager at the Centre for Computational Linguistics, UMIST. Although the direction of my research has changed gradually and gained in more mathematical technicality since then, all the theoretical aspects of this book have their roots, either explicitly or implicitly, in my intensive and intriguing discussions with Professor Sager. Just how indebted I am to him for his constant encouragement and understanding, I find it difficult to express.

From around 1995, as my theoretical standpoint on the study of terminology took shape, I felt the necessity to augment my conceptual approach with a quantitative approach on the descriptive front. I had the luxury of thinking intensively about this in 1996, when I stayed with the Natural Language Processing group at the Department of Computer Science, the University of Sheffield. I would like to thank Professor Yorick Wilks and all the staff there.

One of the most important triggers for the development of an interpretative framework for the quantitative part of my work was the International Quantitative Linguistics Conference held in Helsinki in 1997, where I had the opportunity to learn LNRE models from Dr. R. Harald Baayen of the Max Planck Institute of Psycholinguistics, Nijmegen, the Netherlands. This helped me to finalise the work described in Part III of this book. For that, I owe him a great deal.

I knew the work of the "French connection" of computational terminology well from their publications, but after meeting them face-to-face in 1998 benefited greatly from the stimulation their intellectual input gave to my work. Among them, I especially wish to thank Professor Christian Jacquemin of CNRS-LIMSI and Professor Béatrice Daille of IRIN, the

University of Nantes. My research cooperation and many discussions with them clarified what was lacking in my research agenda, and thus supported this work indirectly.

The concrete descriptions of the formation and growth of the terminology of Japanese documentation in this book were made possible by Professor Yasuhiro Ohta of Bunkyo University, Mr. Hajime Sasaki and Professor Shunsaku Tamura of Keio University, who have made pioneering efforts in listing the Japanese terminology of documentation. I greatly appreciate their work.

Many people generously gave of their time to listen to my ideas and discuss some of the topics contained in this book. Among them, I would particularly like to express my gratitude to my colleagues Dr. Atsuhiro Takasu, Dr. Akiko N. Aizawa, Professor Teruo Koyama, Dr. Masaharu Yoshioka and Dr. Koichi Takeuchi at the National Institute of Informatics, Japan, where I currently work, my former students Keita Tsuji and Fuyuki Yoshikane, Dr. Satoshi Sekine of New York University, Dr. Sandra-Yamilet Santana, Professor Hiroshi Nakagawa of the University of Tokyo, and Dr. Toru Hisamitsu of Hitachi. My thanks also go to Professor Masao Nagasawa of Aichi Syukutoku University and Professor Eisuke Naito of Toyo University, who introduced me to the interesting research area of information studies, which led me to terminology.

In the process of writing this book, Professor Sager, Dr. Marie-Claude L'Homme of the University of Montreal, and Professor Béatrice Daille read the manuscript in detail and offered many relevant comments and criticisms, not all of which I could accommodate in the present work. Ms. Irma Nunez and Ms. Stephanie Coop read and proofed the draft carefully.

Finally, I would like to thank the many other people too numerous to mention who have also played a role in the research reported in this book.

31 May 2002
Tokyo, Japan

Contents

Introduction

In recent times there has been a growing interest in the study of technical terms (henceforth simply "terms" for succinctness), as can be witnessed by the publication of textbooks (Cabré 1993; Felber 1984; Picht & Draskau 1985; Sager 1990), of collections of papers (Rey 1995; Sonneveld & Loening 1993) and of a journal (*Terminology*), as well as regular conferences on terminology such as *Terminology and Knowledge Engineering* and *Computerm*. Despite this fact, the study of terminology, i.e. the theoretical and applied study of terms as coherent systems of lexical items endowed with a singular creative dynamism, is as yet neither clearly defined nor is there general agreement about its scope.

A related problem is the fact that, while work concerning what is traditionally known as the "theory" or "principles" of terminology is pursued simultaneously with, but independently of, terminology-related NLP applications, little effort is being devoted to theories underlying the descriptive analysis of terms. Besides, most of what currently passes for a theoretical foundation of terminology amounts to little more than a simplified, a priori theory of conceptual structures supported by largely prescriptive principles of what "should be" rather than what is the actual usage of terms.

This situation seems to reflect basic characteristics of terms, i.e. terms manifest themselves as concrete linguistic objects within a specialised discourse and their number is constantly growing. The fact that terms are first and foremost concrete linguistic objects makes it difficult to define the theory of terms at a proper level of abstraction. Many so-called theories about terms are really only theories of something — for instance, of concepts — that can be used to describe terms. In addition, many studies treat only a very limited number of terms, mostly for exemplification. The fact that terminology (and the number of terms) is constantly growing, on the other

hand, fosters application-oriented studies of the computational treatment of terms, but without satisfactory theoretical and/or descriptive foundations.

In short, theories of terms — as opposed to theories of something for describing terms — are missing in the academic study of terminology. On the side of applications or descriptions, there is a lack of solid descriptive studies, as opposed to the application-motivated processing of terms. Given this situation, research in the field of terminology needs to be broadened to include concrete descriptive analyses of terminology based on an explicitly stated theoretical position; only through the accumulation of this type of study does it become possible to consolidate the theory of terminology. This book is an attempt in this direction, focusing on the phenomenon of term formation and terminological growth.

The organisation of the book

The book is divided into four main parts. Part I (Chapters 1 and 2) is devoted to clarifying the author's view of the object of the study as well as defining his theoretical standpoint. On the basis of close observation of terms and examinations of the existing theoretical studies of terms, it will be argued that a theory of terms or terminology should deal with the terminology of a domain in its totality, because it is only with respect to individual domains that the very concept of "term" is consolidated. It will also be argued that a theoretical study of a terminology should be accompanied by the descriptive study of a terminology, for proper descriptive studies *are* theories of terms. The concept of the dynamics of terminological growth is then introduced, and the overall framework is illustrated by means of the description of conceptual patterns of term formation, complemented by the analysis of quantitative regularities of terminological growth.

Parts II and III are devoted to the detailed development of the theoretico-descriptive framework for the study of the dynamics of terminology and the required concrete description of the actual manifestation of this dynamics. Throughout these two parts, the Japanese terminology of documentation (introduced below) is used in the analyses.

As a first step towards the characterisation of the dynamics of terminology, Part II (Chapters 3–6) is devoted to the description of the conceptual patterns which determine the formation of terms within the chosen domain. The basic aspects to be observed at the conceptual level are the relationships

between terms and their constituent elements, the relationships among the constituent elements, as well as the type of conceptual combinations used in the construction of the terminology. In Chapter 3, the basic descriptive framework, as well as the elements necessary for the description of the conceptual patterns underlying term formation, are discussed. Chapter 4 is devoted to a presentation of the conceptual categories discovered in the analyses of terms. Chapter 5, then, examines the conceptual relationships between constituent elements of terms. Chapter 6 describes the characteristics of term formation patterns in the field of documentation, which are based on the concrete descriptive devices introduced in Chapters 4 and 5.

The description of conceptual patterns of term formation, to be examined in more detail, has a logical limitation. If, by describing conceptual patterns, we try to give necessary and sufficient conditions for the formation of terminology, we end up listing all the combinations of linguistic items in the terminological data used in the study. This, however, obscures the observation of the dynamics and would reduce the study to a natural history of existing terms. The description of conceptual patterns, therefore, must necessarily remain somewhat general, at the level where the broad regularities of term formation patterns in a domain can most properly be described. This immediately leads to the loss of the fine granularity of the description.

Part III (Chapters 7–9) explores the quantitative analysis of the patterns of terminological growth, which compensates for the limitation of the description of conceptual patterns of term formation and thus completes the description of the dynamics of terminology. There is a statistical method to observe the growth patterns of lexical items in their entirety within a certain category or set of lexical items. Applying this method, it becomes possible to give a detailed description of the dynamics of the potential growth patterns of the terminology of a domain. In Chapter 7, the statistical method is presented together with factors that should be taken into account in the application of the method. Chapter 8 is, then, devoted to describing the growth patterns of constituent elements within the terminology. Chapter 9 details the growth patterns of terms for each subset of terms, the formation patterns of which were observed in Chapter 6.

Part IV (Chapter 10) concludes the study. By examining the theoretical standpoint introduced in Part I with respect to the findings of the concrete description carried out in Parts II and III, it evaluates what was achieved and what should be explored further.

The data

The present study takes the position that a theoretical work on terminology should, as a logical requirement, be accompanied by a concrete description of the terminology of a domain. As such, Parts II and III, together with theoretical and methodological discussions, present the results of analyses of terminology. For the concrete analyses, Japanese terminological data in the field of documentation are used, a field of which the author has an in-depth knowledge. The data are taken from Wersig & Neveling (1984), the Japanese version of Wersig & Neveling (1976), a small but representative terminological glossary in the field of documentation[1]. This is a multilingual glossary with the indication of related terms.

A few normalisations were applied to the Japanese entries of Wersig & Neveling (1984), such as the normalisation of orthographic and minor notational variants. Also, a single non-noun simple entry, 自動 (automatic), was omitted. As a result, 1,228 terms were obtained.

There are no mechanically applicable criteria for delimiting constituent elements or morphemes in Japanese, as Japanese lacks boundaries between linguistic units such as spaces or hyphens. So the morphemes of terms were identified manually, based on the criterion originally introduced by Nomura & Ishii (1988), which reflects an average Japanese speaker's intuition about morphemes and was successfully applied to large-scale analysis of Japanese terms. The method is briefly described as follows:

1. A minimal element is defined as the minimal linguistic element which bears a meaning in current Japanese.
2. According to the origin of linguistic elements, i.e. *wago* (original

---

[1] The idea of collecting terms from articles and texts was examined but discarded because the available texts depended too much on the circumstantial tendencies and the data was sparse from the point of view of collecting a representative set (sample) of terms. Among reference sources such as glossaries, Wersig & Neveling (1984) was a semi-optimal choice from the point of view of representativeness and balance of terms, partly because no better alternatives existed. There were three other glossaries when the study was started: Monbusyou (1958) was too old and too prescriptive; Young (1988), being a translation of Young (1983), was too biased to the U.S. library services and some Japanese terms were artificially coined; and JIS Series (1989) was incomplete and, again, too prescriptive. A new glossary, JSLIS (1997), which is better in its coverage and timeliness, has appeared since, but by the time of its publication, the author's analysis based on Wersig & Neveling (1984) had already been completed.

Table 1. *Number of terms by number of morphemes.*

Length of terms	Number of terms in the data	(%)
1 morpheme	246	(20.0%)
2 morphemes	621	(50.6%)
3 morphemes	283	(23.0%)
4 morphemes	61	(5.0%)
5 morphemes	15	(1.2%)
6 morphemes	2	(0.2%)

Table 2. *Quantities of morphemes.*

Running morphemes	Different morphemes	Average use
2668	830	3.21

Japanese), *kango* (originated in Chinese) and *gairaigo* (originated in [mainly] Western languages), the morphemes are defined as follows:

a. *Wago/gairaigo*: A minimal element constitutes a morpheme. But if a minimal *wago* element is combined with a *wago* affix, this unit is treated as a morpheme. A small number of Roman alphabet acronyms are treated as *gairaigo*. The number of *wago* morphemes is 38 by type and 83 by token. An example of *wago* morphemes is 手 (hand). The number of *gairaigo* morphemes is 199 by type and 602 by token. An example is コンピュータ (computer).

b. *Kango*: A first-order combination of two minimal elements in the terms constitutes a morpheme. However, a minimal *kango* element attached to a morpheme (in an affix-like manner) is also treated as an independent morpheme. Arabic numbers are treated as *kango*, as their readings are mostly the same as *kango* numbers. The number of *kango* morphemes in the data is 589 by type and 1979 by token. An example is 図書 (book).

c. *Kango* and *wago*: A first-order combination of a minimal *kango* and a minimal *wago* Chinese character element is treated as a single morpheme. This is a modification from the original rule by Nomura & Ishii (1988). The number of mixed morphemes is 4 both by type and by token. An example is 係員 (staff).

Japanese "morphemes" as defined above are closer in status to words that make up English compounds in that they mostly have independent mean-

ings (Kageura 1994). Table 1 shows a breakdown of the terms by number
of morphemes. Table 2 shows basic quantitative information about mor-
phemes in the data. The full lists of terms and morphemes are given in
Appendices C and D, respectively.

Part I

Theoretical Background

Chapter 1

Terminology: Basic Observations

In this chapter, the basic status and nature of terminology within language is clarified. Then, the traditional approach to the study of terminology is summarised and critically examined, and some recent developments in the study of terms are briefly introduced. This chapter is intended to give readers the basic background against which the theoretical position of the present study is outlined. This will be elaborated in Chapter 2.

1.1   Basic observations

1.1.1   *Terms and related notions*

Any discussion about the basic status, nature and function of terms within language must start with a provisional definition of "term" and immediately related concepts. According to Bessé, Nkwenti-Azeh & Sager (1997) — a compact and convenient glossary of the expressions of this field — "term" and "terminology" are defined as follows:

term : A lexical unit consisting of one or more than one word which represents a concept inside a domain.
terminology : The vocabulary of a subject field.

Two sets of expressions are important in these definitions, i.e. "lexical unit" and "vocabulary" on the one hand and "concept inside a domain" on the other. "Lexical unit" and "vocabulary" are conventional linguistic notions defined adequately in many dictionaries of linguistics. They need not concern us further at the moment.