



National Library
of Canada

Bibliothèque nationale
du Canada

Acquisitions and
Bibliographic Services Branch

Direction des acquisitions et
des services bibliographiques

395 Wellington Street
Ottawa, Ontario
K1A 0N4

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file / Votre référence

Our file / Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

**Towards Knowledge-Base Systems
for Translators**

by

David R. Miller

**School of Translation and Interpretation
University of Ottawa**

Under the supervision of

Douglas Skuce

Department of Computing Science

and

Brian Harris

School of Translation and Interpretation

Thesis submitted to the

School of Graduate Studies and Research

University of Ottawa

**in partial fulfilment of the requirements for the degree of
Masters of Arts (Translation)**



National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file / Votre référence

Our file / Notre référence

THE AUTHOR HAS GRANTED AN IRREVOCABLE NON-EXCLUSIVE LICENCE ALLOWING THE NATIONAL LIBRARY OF CANADA TO REPRODUCE, LOAN, DISTRIBUTE OR SELL COPIES OF HIS/HER THESIS BY ANY MEANS AND IN ANY FORM OR FORMAT, MAKING THIS THESIS AVAILABLE TO INTERESTED PERSONS.

L'AUTEUR A ACCORDE UNE LICENCE IRREVOCABLE ET NON EXCLUSIVE PERMETTANT A LA BIBLIOTHEQUE NATIONALE DU CANADA DE REPRODUIRE, PRETER, DISTRIBUER OU VENDRE DES COPIES DE SA THESE DE QUELQUE MANIERE ET SOUS QUELQUE FORME QUE CE SOIT POUR METTRE DES EXEMPLAIRES DE CETTE THESE A LA DISPOSITION DES PERSONNE INTERESSEES.

THE AUTHOR RETAINS OWNERSHIP OF THE COPYRIGHT IN HIS/HER THESIS. NEITHER THE THESIS NOR SUBSTANTIAL EXTRACTS FROM IT MAY BE PRINTED OR OTHERWISE REPRODUCED WITHOUT HIS/HER PERMISSION.

L'AUTEUR CONSERVE LA PROPRIETE DU DROIT D'AUTEUR QUI PROTEGE SA THESE. NI LA THESE NI DES EXTRAITS SUBSTANTIELS DE CELLE-CI NE DOIVENT ETRE IMPRIMES OU AUTREMENT REPRODUITS SANS SON AUTORISATION.

ISBN 0-612-00487-2

Canada



UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

Acknowledgments

Writing this thesis has been a long and difficult, but ultimately worthwhile, experience. The contributions of many people deserve acknowledgment. First, of course, I wish to thank my two thesis directors, Douglas Skuce of the Computer Science Department, who is the Director of the Artificial Intelligence Laboratory, and Brian Harris of the School of Translation and Interpretation (STI). Both have been invaluable sources of critical advice and inspiration, and both have been very patient with me. Another person who, in many respects, was an unofficial third thesis director was Ingrid Meyer, who divides her time between the STI and the AI Lab. She acted as an independent advisor who could always be counted on for helpful suggestions and encouragement.

Several people generously gave of their time to help with the experimental part of the thesis. The testing of my knowledge base would not have been possible without the cooperation of Lynne Bowker, Karen Eck and Heather Moores, all graduates of the STI, who acted as my test subjects. Helen Meubus, the English Editor at the Bank of Canada, graciously agreed to provide an independent evaluation of the results.

Thanks are also due to Claude Henin, the Assistant Dean of the Undergraduate Office of the Faculty of Administration, who took time out of his busy schedule to look over my knowledge base. I have promised him a bilingual glossary of the terminology stored in it.

During the many hours I spent at the AI Lab, I benefited from the company of as pleasant and helpful a group of colleagues as exists anywhere. In particular, Timothy Lethbridge deserves recognition as someone who was always ready to put aside his work to help with a problem. Karen Eck helped prepare the hard-copy graph of my knowledge base.

Last but not least, I would like to thank my wife Violet. This thesis simply would never have been completed had it not been for her support, encouragement and willingness to look after the kids.

Abstract

The aim of this thesis is twofold. First, on a theoretical level, it aims to examine knowledge from two standpoints: translation theory and artificial intelligence (AI). Second, on a more practical level, it aims to explore the applicability of research in the branch of AI known as knowledge engineering to the eventual development of knowledge-base systems for translators (KBSTs).

Understanding is a critical stage of the translation process. Understanding takes place through the interaction of the meaning of the text with the reader's prior knowledge. The "opaqueness" of a text, or its resistance to understanding, often reflects a difference between the translator's knowledge profile and the knowledge profile that the author of the original text assumed for his reader. Consequently, knowledge maximization can be considered a viable translation strategy for countering opaqueness. When translators work without sufficient knowledge, they are forced to fall back on a "transcoding" or word-bound approach, which, although it sometimes produces acceptable results, is likely to produce an unidiomatic text and is much more prone to serious translation errors. The goal of a KBST, therefore, must be to provide enough knowledge to allow the translator to engage his "interpretative" or meaning-based mode of translating.

Knowledge and knowledge representation are central issues in Artificial Intelligence (AI). Knowledge engineers and developers of knowledge-based term banks both rely on conceptual analysis in building their systems. A knowledge-management system developed at the University of Ottawa called CODE (*Conceptually Oriented Description Environment*) offers a knowledge acquisition and retrieval environment that can be adapted to the needs of translators. A knowledge base on stock-market options, called optionCODE, was developed by the present author using this system to explore the principles and problems of designing and using KBSTs. An informal experiment demonstrated that translators using this knowledge base as their sole source of knowledge to translate a text concerning options performed as well as, if not better than, a translator using traditional sources.

KBSTs have the potential for incorporation into other systems, most notably translator's workstations, but also machine translation systems, document-production systems and distributed-knowledge networks. While there are a number of theoretical and methodological issues connected with the design and implementation of KBSTs that require further investigation, KBSTs will eventually become an important knowledge acquisition tool for translators.

Résumé

Le présent mémoire de maîtrise comporte deux volets, l'un théorique, l'autre, pratique. Le premier examine la notion de connaissance sous deux angles, soit celui de la traductologie et celui de l'intelligence artificielle. Le second explore la pertinence d'appliquer les résultats des recherches effectuées dans le domaine de l'intelligence artificielle connu sous le nom de génie cognitif à l'élaboration de systèmes à base de connaissances à l'intention des traducteurs (SBCT).

La compréhension du message est le pivot de l'acte traductionnel. Elle résulte de l'interaction qui s'opère entre la signification du texte et le bagage de connaissances du lecteur. L'«obscurité» d'un texte - les entraves à sa compréhension - tient souvent à l'écart qui existe entre le profil de connaissances du traducteur et celui que l'auteur prête à ses lecteurs. La maximisation des connaissances apparaît donc comme une solution envisageable pour pallier ce problème d'obscurité des documents. Lorsque les connaissances qu'il a d'un domaine sont limitées, le traducteur n'a d'autre choix que de se tourner vers le «transcodage», c'est-à-dire la transposition fondée sur les formes linguistiques. Bien que cette approche donne parfois des résultats acceptables, les traductions ainsi produites sont rarement idiomatiques et risquent fort de comporter des erreurs graves. Par conséquent, un SBCT devra être conçu de façon à fournir à ces derniers les connaissances qui leur permettront de recourir à une approche «interprétative», c'est-à-dire fondée sur le sens.

Les connaissances et la représentation des connaissances sont au coeur de la recherche sur l'intelligence artificielle. Les spécialistes du génie cognitif et les concepteurs de banques de terminologie à base de connaissances appuient leurs travaux sur l'analyse notionnelle. Un système de gestion des connaissances mis au point à l'Université d'Ottawa et appelé CODE (*Conceptually Oriented Description Environment*) représente un outil d'acquisition et d'extraction des connaissances pouvant être adapté aux besoins des traducteurs. L'auteur du présent mémoire s'en est servi pour créer une base de connaissances portant sur le marché boursier des options, appelée optionCODE, ce qui lui a permis d'explorer les postulats et les problèmes que posent la conception et l'utilisation des SBCT. Une expérience informelle a permis de démontrer que des traducteurs n'utilisant que cette base de connaissances comme source d'informations pour traduire un texte sur les options s'en sont aussi bien sortis, sinon mieux, qu'un traducteur recourant aux sources traditionnelles.

L'un des avantages des SBCT est qu'ils peuvent être intégrés à d'autres systèmes, en particulier le poste de travail du traducteur, mais aussi à des systèmes de traduction automatique, de production de documents ainsi qu'à des réseaux répartis de base de connaissances. Bien qu'il reste à élucider un certain nombre de problèmes de conception et de mise en oeuvre, les SBCT ont un avenir prometteur comme outil majeur d'acquisition des connaissances dans le domaine de la traduction.

Throughout this document, the masculine pronouns "he," "him" and "his" have been used in a generic sense to avoid awkward "he/she" constructions. They are to be interpreted as applying equally to men and women. Translation is one field where the sexes are clearly on an equal footing.

Towards Knowledge-Base Systems for Translators
Table of Contents

Acknowledgments	i
Abstract	ii
Résumé	iii
1 Introduction	1
1 Genesis and Aim of Thesis	1
2 What is a Knowledge-Base System for Translators?	3
3 Knowledge-Base Systems Versus Traditional Translator Tools	5
1 Translator's Traditional Knowledge Sources and Knowledge- Management Tools	5
1 Research for a Particular Translation	7
2 Ongoing Research	9
3 Knowledge Storage and Retrieval	10
2 Superiority of Knowledge-Base Systems	10
4 Plan of Thesis	13
2 Knowledge: An Essential Input to the Translation Process	15
1 The Translation Process	15
1 Meaning	16
2 Understanding	20
3 Re-expression	23
2 The Role of Knowledge in Translation	25
3 Classification of Knowledge	29
1 Linguistic Knowledge	30
2 Extra-Linguistic Knowledge	30
3 Translators' Need to Supplement Knowledge	34
4 Knowledge Maximization as a Translation Strategy	39
1 The Opaqueness Factor	39
2 The Three Parameters of Opaqueness	39
1 Assumed Knowledge Profile of Reader	40
2 Actual Knowledge Profile of Reader/Translator	42
3 Textual Clarity	44
4 Dealing with Opaqueness	44
3 Knowledge and Technical Translation	46

5	Informal Experiment With Fourth-Year Translation Class	48
1	Summary	49
2	Transcoding Versus Interpretative Translation	50
1	Accuracy of Meaning	50
2	Resolution of Ambiguity	52
3	Accuracy of Terminology	53
4	Naturalness of Style	54
3	Other Observations	56
4	Conclusions	57
3	Knowledge and Knowledge Representation from the AI Perspective	59
1	Background	60
2	Knowledge, Knowledge Bases and Knowledge-Base Systems	62
3	Practical Issues in Knowledge Representation	64
4	Philosophical Issues in Knowledge Representation	68
5	Three Ways of Representing Knowledge	69
1	Predicate Logic	69
2	Semantic Networks	71
3	Frames	73
6	Conceptual Analysis	76
1	Concepts	76
2	Conceptual Analysis and Knowledge Acquisition	79
3	Conceptual Analysis and Terminology	81
4	A Knowledge-Management System, a Sample Knowledge Base for Translators and an Informal Experiment	84
1	CODE: A Knowledge-Management System	84
1	Overview	84
2	Advantages of CODE for Translators and Other Users	89
1	Flexible and Efficient Access to Knowledge	89
2	Richness	94
3	Open-Endedness	96
2	Example of a Knowledge Base for Translators: OptionCODE	98
1	Overview	98
2	Methodology	100
3	An Example of Conceptual Analysis: Option Strategies	104
3	Some Methodological Issues in Knowledge-Base Design	107
1	Sources of Knowledge	107
2	Knowledge-Base Verification	108
3	Circumscription of Domain	110

4	Definitions	112
5	Natural Types Versus Roles	113
6	Distinction Between Subconcepts and Kinds	114
7	Conceptual Elegance Versus Utility	115
8	Evolution of Properties	116
4	Informal Translation Experiment with CODE and OptionCODE	117
1	Summary	117
2	Results	119
3	Session Observations	120
4	Follow-Up Interviews	121
5	Additional Observations on the Translations	123
5	Future Directions	126
1	A Tool for a Translator's Workstation	126
1	The Translator's Workstation Concept	126
2	Review of Current Translator's Workstations	131
3	Integration of Knowledge-Base Systems into Translator's Workstations	134
2	Integration with Other Systems	135
1	Machine Translation	135
2	The Document-Production Chain	139
3	Distributed-Knowledge Networks	141
3	Design of Knowledge-Base Systems for Translators	144
1	Basic Structure	144
2	Modelling of Interlingual Equivalence	146
3	Content	154
4	User Interface	156
5	Practical Aspects of Knowledge-Base Design	160
6	Conclusion	163
1	Recapitulation	163
2	Recommendations for CODE	165
1	Structural Considerations	165
2	Interface Considerations	166
3	Implications of Hardware and Software Developments	167
4	Directions for Future Research	169
5	Final Word	171

APPENDIX I - Documents for Fourth-Year Translation Class Experiment	173
1 Original French Text ("Bourse")	173
2 Official Translation of "Bourse" Text	173
APPENDIX II - Documents for CODE Experiment	175
1 Original French Text	175
2 Two Sample Translations of French Text by David R. Miller	176
1 A Respectable But Frequently Transcoded Translation	176
2 A More Knowledge-Based Translation	177
3 Translations by Experiment Subjects (A, B and C)	178
1 Translation A	178
2 Translation B	179
3 Translation C	180
4 Evaluation Sheet (including scores for A, B, and C)	181
APPENDIX III - OptionCODE Printouts	182
APPENDIX IV - Graph of Complete OptionCODE	183
Bibliographic References	184

Figures

Figure 2-1	Nida's Basic Model of Translation	21
Figure 2-2	Seleskovitch's Triangle	21
Figure 2-3	Delisle's Triangle	22
Figure 2-4	Role of Knowledge Profiles in the Translation Process	43
Figure 3-1	A Simple Semantic Network	72
Figure 3-2	A Simple Frame	74
Figure 4-1	CODE Screen Dump Showing Browser (Right) and Selected CD View (Left)	87
Figure 4-2	CODE Screen Dump Showing Portion of Main Graph ("Zoomed") ..	91
Figure 4-3	CODE Screen Dump Showing Detail Graph Opened from Browser ..	92
Figure 4-4	Generic Structure of CDs	101
Figure 4-5	Conceptual Analysis of Complex Option Strategies	105
Figure 4-6	CODE Screen Dump Showing Subtree Graph of Complex Option Strategies (cf. Figure 4-5)	106
Figure 4-7	Inheritance over Different Knowledge-Base "Levels"	111
Figure 5-1	Three Kinds of Bilingualism	150
Figure 5-2	Three Possible Bilingual Knowledge-Base Structures	151
Figure 5-3	A Linguistically Integrated Knowledge Base	153

CHAPTER 1

Introduction

1.1 Genesis and Aim of Thesis

It was, in many ways, a typical translation request. A researcher had written a paper entitled "Vulgariser l'évaluation d'une option" and needed an English translation of it — by early next week, please. I glanced through it, noting the practical details (8 pages single-spaced, lots of equations, etc.), and quickly realized that considerable background research would be necessary, since I was faced with paragraphs like the following:

Le théorème de parité est en fait une condition de non arbitrage où la valeur d'une option d'achat, CALL, sur un titre doit être égal à la somme (1) du prix du titre, P, (2) de la valeur d'une option de vente, PUT, sur ce titre et (3) la valeur d'un emprunt égale à la valeur présente du prix de levée, K, escompté au taux sans risque, r .

With a small sigh, I turned to the paper's bibliography to see if there were any obvious candidates for sources of expert knowledge that might help me understand what the author was talking about.

* * *

The fact that a translator should understand the text he is reading in order to produce a proper translation of it is so obvious to someone in the profession that it is always a shock to encounter a client who, when asked questions about the meaning of the source text, responds "It's really too difficult to explain. Just translate the words, and I'll get someone to look it over." Yet such a simplistic view of the translation process is not uncommon. And what is often even more poorly understood is that what the

translator needs in order to understand (like any reader) and so to translate (unlike any reader) is *knowledge*.

Translators are knowledge users. Although they already possess extensive linguistic knowledge of the two languages involved, they may frequently require help with specialized terminology, including advice on how to use this terminology as an expert would. Moreover, they need background information about the subject matter of the source text; given that they are not the audience whom the author had in mind when composing the original text, they probably do not have the specialized knowledge that was assumed of the intended readership.

The topic of the text described above turned out to be stock-market options, a subject about which I knew next to nothing. The all-too-familiar difficulties I encountered in resolving the main knowledge problems connected with the text — in quickly becoming a mini-expert on the topic — fuelled my growing interest in the role of knowledge in translation. And the topic of options came to mind later when I was searching for a suitable specialized subject to illustrate the ideas presented in this thesis. As part of my coursework for my Master's degree, I had explored machine translation, a topic that had long intrigued me, and had become interested in the involvement of artificial intelligence researchers in this area. I discovered that there were some intriguing parallels between their research directions and the work of certain translation theorists. I explored this connection in a paper comparing and contrasting two such groups: the "conceptual dependency" school led by Roger Schank and the "interpretative" school of translation theory led by Danica Seleskovitch. Perhaps the most striking thing they shared was an insistence on the importance of knowledge in the broadest sense for translation.

The aim of this thesis is twofold. First, on a theoretical level, it aims to examine knowledge from two standpoints: translation theory and artificial intelligence (AI). Second, on a more practical level, it aims to explore the applicability of research in the

branch of AI known as knowledge engineering to the eventual development of knowledge-base systems for translators. To this end, the author built a knowledge base for this specific purpose using CODE, a knowledge-management tool under development at the University of Ottawa.

1.2 What is a Knowledge-Base System for Translators?

The modern world is witnessing an explosion of new knowledge. By some estimates, there are already on the order of 30 million general concepts, with some 10 to 40 thousand new ones entering English every year.¹ Clearly, the importance of accurate technical information, including consistent, logical terminology, is greater than ever before. Yet traditional term banks are not necessarily equal to the task. The information stored in them is static, passive. Whenever information must be retrieved, the burden of deciding where to look, how to look, and (most importantly) how to interpret the results of the search fall upon the user. "The richness of the knowledge they contain ... is neither explored nor exploited.... Consequently the full potential of the data has not been realised" [Ahmad et al. 1989, 40]. If we accept that information only becomes knowledge when it is structured or organized towards a particular end, what is lacking in traditional term banks is knowledge.

A *knowledge-base system* essentially consists of 1) a *knowledge base* and 2) a program to enter information into and extract information from the knowledge base, a *knowledge-management system*. These two components are basically analogous to the database and database management system found in traditional database packages.

1. Galinski [1988b, 498]; Gawn [1990, 454].

A knowledge base, then, is similar to a database. Where a conventional database stores *information*, however, a knowledge base stores *knowledge*. Knowledge is information that is *structured* according to its inherent properties in order to bring out the relationships between meaningful units of information or concepts. Thus a knowledge base is inherently concept-oriented, with the concepts formally ordered by means of basic relationships such as generic/specific and part/whole, as well as systematically linked to other concepts through wide-ranging associative links. The properties of concepts are inherited through these links to their subconcepts. In essence, the concepts are the building blocks and the relations between these concepts the mortar. A knowledge base usually restricts itself to a very limited subdomain.

A knowledge-management system provides an environment for entering, reorganizing and retrieving the information stored in a knowledge base. Obviously, its specific capacities will vary from system to system. For instance, it may offer various methods for searching the knowledge base and for displaying the information it contains; it may include mechanisms to verify the logical consistency of the information entered; it may be able to generate its own reports; and it may be able to export information to (or import information from) other programs. Generally speaking, the underlying architecture of the knowledge base is tailored to its use with a particular knowledge-management system. It is the combination of the two that is referred to here as a *knowledge-base system*. These terms will be discussed in more detail in Chapter 3.

A *knowledge-base system for translators* refers to a knowledge-base system that has been built with the needs of translators in mind. The structure and content of the actual knowledge base will certainly reflect the specific needs of professional translators who need convenient access to a variety of information about a specific subject field. The knowledge-management component of the system, on the other hand, might simply offer a general, possibly customizable environment for the translator's interaction with the

knowledge base. Or it may prove feasible in the future to create knowledge-management systems designed specifically for translators and other writers.

It should be noted that a number of other expressions have been coined to describe what I am calling a knowledge-base system for translators. Term banks that are characterized as "knowledge-based," "conceptually-oriented," "intelligent" or "advanced" basically reflect the same idea. While the systems may differ widely in their implementational details, they share a conceptual approach to knowledge representation and a dedication to providing a richer and more flexible environment for knowledge retrieval than traditional systems.

1.3 Knowledge-Base Systems Versus Traditional Translator Tools

1.3.1 Translator's Traditional Knowledge Sources and Knowledge-Management Tools

An entire thesis could be written on the subject of translators' information sources and techniques of documentation research. The following discussion is only a brief overview.²

"Seule la recherche documentaire," wrote Moskowitz, "permet au traducteur de réduire systématiquement son ignorance, d'acquérir des connaissances nouvelles" [1978, 77]. Efforts to acquire knowledge account for a considerable portion of the typical translator's time. In fact, "gaining access to information is the most time-consuming secondary task of the translator" [Vaumoron 1988, 38]. A survey of European translators

2. More complete discussions of translators' information sources may be found in Bédard [1986], Moskowitz [1978], Ministère des Communications du Québec [1978] and a 1980 special issue of *Meta* on documentation (25:1).

by Smith [1986] found that only about half of the typical translator's time was spent actually translating, with another approximately 10 per cent devoted to pure terminological research and 5 per cent to finding and studying background information.³ One company (SITE in France) estimated that its translators were spending up to 40 per cent of their time looking up terms [Joscelyne 1990, 41]. And the preliminary results of a survey to gauge the needs of translators indicate that "considerable time is spent in consulting domain experts and other sources" [Ahmad et al. 1990, 11].

Generally speaking, translators have two *modes* of knowledge acquisition. The first is what in French is called "ponctuel," and which I will call "ad hoc." In ad hoc mode, the translator wants an answer to a particular problem, usually terminological in nature. Besides the target-language term itself, he may require information on its usage. For this mode, speed and accuracy are paramount. However, contrary to what non-translators are apt to think, translators are not simply interested in *terminological knowledge*, but also in *encyclopedic* or *conceptual knowledge* — information about the major concepts or ideas in the particular subject area. In other words, they are seeking both *linguistic* and *extra-linguistic* knowledge. The second mode, which is conceptually oriented, is what I will call "background browsing." In this case, the translator's primary interest is acquiring some deeper knowledge of a particular subject field. Therefore, what he desires most is depth of information and ease of movement or "navigation" within the knowledge source. These two quite different needs or knowledge-acquisition modes have been pointed out by several authors.⁴

One may also distinguish two different *purposes* in translators' knowledge-acquisition activities: research in response to a particular translation project, and ongoing efforts to build up and keep up-to-date personal knowledge in a selected subject field. In the first

3. The remaining 35 per cent or so was devoted to such activities as terminology file updating, client relations, accounting and administration.

4. For example, Vaumoron [1988, 38].

case, the translator may want to consult his source in both ad hoc and background browsing modes, depending on the nature of the text and time constraints. It is very common for the translator to be handed a text in an unfamiliar subject area, requiring him to become an "instant mini-expert" by rapidly acquiring a working knowledge of the key concepts in the field. Such a task is best accomplished in background browsing mode. Ongoing efforts to stay abreast of developments in a particular field often include browsing activity.

When faced with a problem of lack of knowledge that he has been unable to solve through research and consultation, a translator must take a logical, educated guess after bringing all available knowledge to bear. The latter happens more frequently than might be supposed. If a knowledge source is not able to supply the exact piece of information sought, it ideally should at least improve the final solution (i.e. make it better informed) and provide suggestions on where to look further. Since time is almost always of the essence in translation, knowledge sources must offer efficient service.

1.3.1.1 Research for a Particular Translation

In response to the needs of a particular translation request, the translator's traditional first recourse for knowledge acquisition is dictionaries. These come in many varieties. Bilingual and multilingual dictionaries exist for both general language and for specialized (i.e. technical) language. Glossaries (i.e. bilingual and multilingual listings of language equivalents with minimal explanation) are generally restricted to specific technical fields. Professional translators soon learn to approach bilingual dictionaries and glossaries with a healthy scepticism, since cross-language equivalences established out of context are often misleading. Unilingual dictionaries are another traditional source of information. These may be divided into lexical, encyclopedic and specialty dictionaries. Both source-language and target-language works may prove useful here. Full-blown encyclopedias are often a valuable source of knowledge. Generally speaking, target-language

encyclopedias are the most useful to the translator, although a well-known technique is to compare articles from source- and target-language encyclopedias. All of these reference sources, however, are inevitably inadequate when it comes to supplying up-to-date technical knowledge, because technology and its vocabulary are too complex and evolve too quickly to be captured in formally published works.

Over the past 20 years or so, computerized terminology banks have become an increasingly important tool for translators. As their availability, user-friendliness, performance and quality have improved, such term banks have gradually become "first-recourse" sources for many translators. The two major term banks in Canada are the Secretary of State Department's TERMIUM III, which is now available by subscription on CD-ROM, and the Banque de terminologie du Québec managed by the Office de la langue française. Generally speaking, term banks such as these offer straight terminological information, supplemented by definitions, additional information and examples on a term-by-term basis. Thus they are in many ways extensions of traditional bilingual dictionaries. As we shall see, however, there is a growing interest in developing term banks that belong to an entirely different framework — conceptually based knowledge banks.

Various source- and target-language material represents the translator's next line of attack. Popular science texts and magazines are sometimes helpful in providing explanations of key concepts in a given field. Specialized publications — including books, textbooks, journals, manuals and papers — can be valuable sources of specific technical information. The main problem with these latter sources is determining and then locating the most appropriate items. The bibliography or references contained in the source text can sometimes help narrow down the search, but all too often the translator still finds himself rapidly scanning technical material, hoping to stumble onto a discussion of a relevant point.

There are a number of other kinds of written documents that are used by translators on occasion. These include industrial documentation, commercial catalogues, documents on terminological standards, and client-supplied documentation (perhaps including previous translations). The latter material must be treated with both respect and caution: respect because it reflects the needs and practices of the client, and caution because it may prove idiosyncratic or outright faulty.

In addition to the above written sources of documentation, the translator may have recourse to "oral" documentation. Depending on availability and suitability, such sources might include the original author, the client or the end user of the translation. A resource prized by any translator is an independent domain expert who is willing to take the time to help him understand a text. All of these individuals may be unilingual (in either the source or target languages) or bilingual. Of course, fellow translators may also be a valuable source of practical advice and information.

1.3.1.2 Ongoing Research

Either by design or by accident of experience, technical translators usually develop expertise in one or more subject fields. In addition to research in response to particular translation projects, they develop this expertise through extra background research involving both reading and consultation. And once they have developed this expertise, they cannot afford to let their knowledge become out-of-date.

... la formation du traducteur technique n'est jamais terminée: tous les jours il a de nouvelles difficultés à vaincre, et donc de nouvelles connaissances à acquérir. Il doit constamment se tenir au courant de l'actualité, pour accroître à la fois ses connaissances techniques et linguistiques.... il ne peut se permettre de se laisser dépasser [Horguelin 1966, 19].

The Translation Bureau of the Secretary of State of Canada offers ten suggestions in this regard to its own translators and writers [Secretary of State Department 1987, 17]. Among them: consult all available documentation on clients, including annual reports and budgetary estimates, etc.; consult the clients' libraries; obtain official bilingual publications by clients to become familiar with their operations and in-house terminology; develop a file of resource persons; enhance personal knowledge of relevant subject areas in both languages; take courses and attend conferences; read relevant entries in good encyclopedias and specialized reference works; consult special terminology publications; and ask experts in the field for help with picking suitable background reading.

1.3.1.3 Knowledge Storage and Retrieval

It is essential for translators to have a means for storing the knowledge they acquire for later retrieval. Virtually every translator has experienced the frustration of trying vainly to locate a particular item of information that he knows was jotted down somewhere. Translators tend to develop their own individual methods for information storage. While the traditional method is to maintain a system of personal records (most often in the form of file cards), the format and content of such records are themselves idiosyncratic. Today, many translators use computerized systems to store their personal records, but again the format and content (not to mention the software involved) vary widely. Sometimes a translation firm or service will centralize the creation of translation terminology records, imposing some degree of consistency in presentation and content.

1.3.2 Superiority of Knowledge-Base Systems

Knowledge-base systems are potentially better than the traditional methods of knowledge acquisition and retrieval discussed above for at least three reasons.

First, knowledge-base systems can be *richer* sources of knowledge than traditional tools. They can contain a wider variety of information; in fact, there are virtually no restrictions on the kinds of knowledge that may be represented. (It should be noted, however, that the development of knowledge bases is thus no simple task. In order for the final product to be useful and usable, it must be based on a solid and consistent conceptual framework.) Knowledge-base systems are richer than traditional tools, too, because the information they contain is imbedded in a larger hierarchical structure through which valuable implicit knowledge is propagated along a variety of links. Thus when a particular item of knowledge is retrieved, it is supported and enriched by its underlying conceptual environment, which lies ready to be explored if necessary. Knowledge-base systems are also richer because of the variety of ways in which information can be sought and retrieved. As Sager points out [1990, 169], traditional term banks often do contain the information needed to answer searches conducted with partial information or "fuzzy" searches (i.e. searches where the user-supplied information may not be specific or exact), but that information is "hidden" and difficult to retrieve.

Second, knowledge-base systems are better than traditional tools because they provide *more flexible and efficient access*. They offer a single-point counterpart to the various traditional sources of knowledge for translators noted above. "For the user," Sager explains [1990, 187], "the advantage of the computer-based data collection over the conventional dictionary lies in the fact that a single database can now hold information that was conventionally held in fixed and different formats in a number of separate reference tools." By eliminating the duplication and redundancy that characterize traditional data banks, storage requirements and retrieval time can be improved. In addition, the hypertext-like⁵ interface of such a system can provide a means of navigating the knowledge base, offering a relatively friendly and easy way (at least with

5. "Hypertext" refers to text in which there are non-linear links to other related portions of text. In other words, the user is able to indicate that he would like more information on a particular item (often by "clicking" on it with a mouse) and is then taken directly to the appropriate location.

some practice) to zero in on the desired item of information. Such an interface enhances the ability of the system to respond to non-traditional requests, such as searches based on partial or imprecise information, and requests for information matching novel sets of criteria. It also facilitates browsing activities, where the user can indulge in a relatively undirected consultation of the knowledge base in order to gain an overview of the subject area. Finally, such systems offer the potential for convenient on-line access as part of a translator workstation.

Third, knowledge bases have the advantage of being inherently *open-ended*. There is currently a great deal of interest, particularly in Europe, in the development of large knowledge-based term banks that can be used for a variety of purposes. Galinski [1988a and 1988b], for instance, has argued persuasively in favour of term banks that are *multifunctional* — that have many potential applications and whose knowledge can be shared with larger information networks. Sager [1990, 135] notes that a single such term bank could serve as: a tool for translators, a reference tool for standards experts, a source for lexicons in natural-language processing systems, a source for printed glossaries and thesauri, and a reference database for industrial and individual users. And he also points out that there is a growing interest in sharing and exchanging terminological data [1990, 227]. The rapid development of the so-called language industries has spurred international interest in the application of computer-based solutions to the modern problems of document production and communications. Looking more to the future, knowledge-base systems also have potential as components of machine translation systems and as educational tools. Open-endedness means there are fewer limits to conceptually oriented knowledge-base systems, since, by their very nature, they have a greater potential for new applications: "Only a [terminology data bank] designed without built-in limitations, so that its structure can be further developed in order to meet future requirements, can increase its holdings, adapt to new development and control the dynamics of conceptual development" [Galinski 1988c, 175].

While other characteristics of knowledge-base systems for translators could undoubtedly be singled out for mention, these three — greater richness, more flexible and efficient access, and open-endedness — capture the essential advantages of the approach. Baudot, writing in 1986 [153], set down three requirements for the "term banks of the future": multidimensional structure, varied modes of consultation, and distributed interrogation and updating. These correspond fairly closely to my three points.

It is clear that a readily available and easily accessible source of high-quality terminological and domain-specific knowledge could be of enormous benefit to working translators. It would give them speedier access to more and better-quality information and cut down on the time wasted on fruitless undirected research. Their performance and productivity would improve as a result. Until recently, the basic elements required to construct such a multi-purpose translator tool were not available. Today, some extremely interesting products are starting to appear. The problem remains, however, of how best to bend these powerful new technologies to the needs of the translator.

Exploring that question is the purpose of this thesis.

1.4 Plan of Thesis

The following is a brief overview of this thesis. Each chapter begins with a summary of its main topics.

Chapter 2 is concerned with the importance of knowledge to the translation process. After reviewing the nature of the translation process, it discusses the role of knowledge in that process and proposes a classification of knowledge. It then explores the idea of knowledge maximization as a translation strategy. The final section reports on an informal experiment aimed at exploring the role of knowledge in translation.

Chapter 3 examines what knowledge means to researchers working in the area of artificial intelligence. After examining several salient issues in knowledge representation – both practical and philosophical, it discusses the importance of concepts and conceptual analysis.

Chapter 4 introduces a knowledge-management system called CODE and describes a knowledge base on stock-market options developed by the author. After a discussion of some methodological issues, an informal experiment that was conducted to test the usefulness of this sample knowledge-base system for translators is described and analyzed.

Chapter 5 looks at future directions for knowledge-base systems for translators. Specifically, it speculates on how such systems could be integrated into translator's workstations and other appropriate systems. The final section discusses the main factors to be considered when designing knowledge-base systems for translators.

Chapter 6, the conclusion, begins with a recapitulation of the main ideas and conclusions presented in the thesis. It then offers some recommendations for improving CODE and examines the implications of technological developments.

CHAPTER 2

Knowledge: An Essential Input to the Translation Process

The purpose of this chapter is to establish the importance of knowledge for translation and to examine the role played by knowledge in the translation process. It begins with a survey of the translation studies literature on the process of translation, focussing on some key elements that appear frequently in translation models: meaning, understanding and re-expression. The next section, 2.2, examines the role of knowledge in this process and argues that it has not been fully appreciated by translation theorists. Section 2.3 takes a closer look at the kinds of knowledge needed and used by translators. Section 2.4 explores the idea that knowledge maximization is a good translation strategy by examining three parameters that collectively determine the knowledge needs of a translator with respect to a given text. The final section reports on an informal experiment to explore the role of knowledge in translation, which was carried out with a fourth-year class in specialized translation.

2.1 The Translation Process

There is little consensus in the field of translation theory on what translation is and how it is accomplished. Moreover, like many interdisciplinary subjects, the field suffers from wide variation in the terminology used. While generalization is difficult, there are at least three elements that are relatively constant in the various models of the translation process that have been proposed — meaning, understanding and re-expression. I propose to examine these three aspects with a view to investigating the role of knowledge in the translation process. While reference will be made to a number of writers, there will be an emphasis on two major translation models — that of Eugene Nida and that of the so-called "interpretative" school.

2.1.1 Meaning

Most writers on translation have identified meaning as the focal point of the translation process, describing translation as the "transfer" or "invariance" of meaning between languages. House [1977, 25] offers a typical definition: "The essence of translation lies in the preservation of 'meaning' across two different languages." Nida stressed that "translating means translating meaning" [1985, 119]. Other authors, however, have insisted that meaning is language-specific and language-determined, not something that is subject to transfer across linguistic borders: "Meaning is language-internal and critically bound to the particular language" [King 1987, 284]. Translation models based on this viewpoint specifically avoid postulating a "disembodied entity meaning" [Mossop 1983, 19 (footnote)].

The question of whether meaning is language-bound or at least partially language-independent is not a new one. In this century alone, one of the major debates in linguistics has been over the so-called "Sapir-Whorf" hypothesis.⁶ In its strong form, this theory proposed that each language, through its grammatical and semantic systems, delineated a particular view of reality that permanently conditioned how its speakers interacted with the world. The logical corollary was that translation was at best difficult, if not impossible. Eventually arrayed against this theory were the proponents of universal grammar, who argued that languages were more similar than different and that there were some basic similarities or "universals" among languages.⁷

Neither view predominated. The position of hardline Whorfians ran counter to common sense, since ordinary people, not realizing that translating and foreign language

6. This view, more generally known as relativism, was expounded by Van Humboldt in the 19th century. Its main proponents in this century were Benjamin Whorf and Edward Sapir. See, for example, Whorf [1956] and Sapir [1949].

7. This viewpoint is most closely associated with the theories of Noam Chomsky (see Chomsky [1965]), but has been more fully developed by others, notably Greenberg [1978].

learning were next to impossible, went right on doing exactly that. The universalists, on the other hand, failed to find the clear-cut universals they had been looking for and had to be content with lists of extremely general examples of features common to the majority of languages (e.g. the order of colour terms, the presence of proper nouns and interrogatives). Mounin, in a work essentially devoted to the relativist/universalist debate from a translation standpoint, concluded that "la traduction peut toujours commencer" [1963, 279] and tried to put the matter to rest:

Au lieu de dire, comme les anciens praticiens de la traduction, que la traduction est toujours possible ou toujours impossible, toujours totale ou toujours incomplète, la linguistique contemporaine aboutit à définir la traduction comme une opération, relative dans son succès, variable dans les niveaux de la communication qu'elle atteint [1963, 278].

However, other writers on translation continued to find it frustrating that universals did not seem to be immediately applicable to translation. For example, William Frawley contends that research on universals has been irrelevant to translation theory for the most part, since the only universals commonly recognized are too general to be of any use. "Universal grammar and the identity it entails are aspects of linguistic competence," he writes, "whereas translation is a matter of linguistic performance" [1984, 167].

The debate over whether meaning is (or is to some degree) language-independent is actually misleading for translation, since it is clear that translation involves, in some fashion, a transfer of meaning between languages. Yet, as has been stressed by Newmark [1981, 98] and others, it is equally obvious that the language forms used in the source text influence the language forms selected by the translator for the target text. More to the point for translation is that there are two *levels* of meaning in language that need to be distinguished.

The two levels of meaning are that pertaining to *langue* and that pertaining to *parole*.⁸ At the level of *langue*, the meanings of individual lexical items of a given language are defined by their particular function within the lexical system of that language, i.e. by their oppositions to related items. For instance, the English word "cup" may be regarded as essentially defined by its boundaries with its neighbours in the lexical system, such as "mug," "vase" and "amphora." Nida (among others) analyzed the same phenomenon in terms of "semantic fields," each with a "central" meaning and any number of "satellite" meanings following a variety of orbits around it [1975, 123f]. This level of meaning has been variously termed "structural," "linguistic," "lexical," or "formal" meaning – the kind of meaning that dictionary definitions commonly try to capture. Languages are indeed very different at this level. The first attempts to rectify structural linguistics' neglect of semantics (for example, componential analysis and generative semantics) tended to focus on this level.

The "interpretative" school⁹ of translation theory calls this level of meaning "signification." Pergnier [1980, 190ff] offers the following example. The English word "record" can be used in a variety of lexical contexts (i.e. "on the record," "service record," "break a record," "listen to a record," etc.). The common thread of meaning is "a permanent representation." The French word "disque," on the other hand, is found in a

8. The terms *langue* and *parole* were proposed by de Saussure in *Cours de linguistique générale* (first published in 1916). Briefly, given a particular language community, individual utterances by members of that community constitute instances of *parole*. They exhibit considerable variation on several levels (e.g. phonetic, grammatical, lexical). On the other hand, the language system shared by the community, which allows it to communicate and which can be studied indirectly through *parole*, is termed *langue*. While English expressions have been proposed for these notions (Malinowski proposed "language" and "speech" in 1937), the French terms are still generally used. Note, however, that such expressions as "speech act" and "text act" refer to *parole*, and that Chomsky's distinction between "performance" and "competence" reflects the same basic dichotomy.

9. The "interpretative" school refers to a theory of interpretation and translation ("la théorie du sens") espoused by authors associated with the École Supérieure d'Interprètes et de Traducteurs (ESIT) of the Université de Paris III. Its founder was Danica Seleskovitch; other prominent authors associated with this school include Marianne Lederer and Maurice Pergnier. For a more detailed overview, see Garcia-Landa [1981].

completely different set of lexical environments (i.e. "disque de musique," "disque de stationnement," "disque de frein," etc.). The common thread here is "something that is round and flat." Out of context, then, there is no inherent equivalence of meaning between English "record" and French "disque" in their core meanings.

However, for the translator (and for any casual reader, in fact), the meaning of words in actual texts – or in *parole* – is of another level. As Mossop puts it: "The important question for translators is what the writer meant by using a certain word in a particular passage. What words mean, and what people mean *by* them on particular occasions, are two different things" [1983, 271]. To return to Pergnier's example, words such as the French "disque" and the English "record" take on a "désignation" when actually used. Only at the *parole* level of meaning may a cross-language equivalence be established between these and any other groups of lexical items in two different languages in a particular instance in a particular text.

The meaning contained in a text goes far beyond that conveyed by the words alone, however. Words linked in progressively larger units (i.e., phrases, sentences, paragraphs, texts) convey meaning linked to their relationship with other such units. Yet, as House noted, "the importance of the textual aspect of meaning has often been neglected in the practice and theory of translation" [1977, 29]. Besides semantic (referential) meaning and pragmatic meaning (covering connotative, stylistic and collocative meaning), Roberts [1986, 56] includes propositional meaning (covering intentions and groupings of ideas), and discourse meaning (referring to links between propositions at the level of the entire text). Textual meaning — as the sum total of all components of meaning — is the true object of translation. Words as such are only the medium of the meaning.

When a text is created by a specific writer for a specific purpose and aimed at a particular audience, the meaning in it can be regarded as a *message*.¹⁰ As several authors have pointed out, therefore, "the relevant unit of meaning for the translator is not the word, but the message" [Nida 1972, 190]. The communication of the message depends on the reader *understanding* what the writer intended him to understand. Since the translator begins his task as a reader, *understanding* is crucial to the translation process if the message is to be fully conveyed.

2.1.2 Understanding

Understanding takes place through the interaction of the textual meaning with an individual's *knowledge* of the world and of the particular communication situation. A translator, too, first acts as a unilingual receptor to understand the message before he can re-express it in another language. Thus "la traduction n'opère pas sur les langues mais sur les messages" [Pergnier 1980, 26]. Like the receptor of any message, the translator may encounter understanding problems of varying severity. As I shall argue in more detail below, his success in resolving such problems by supplementing his relevant knowledge is a major factor in the quality of the result.

The understanding stage has not always been labelled as such by translation theorists. Nida refers to it as "analysis" ([1974], for example). The process is construed as a search for "kernels" of meaning at the deep-structure level, where the subsequent "transfer" of meaning may more easily take place (see Figure 2-1). For Nida, transfer essentially involves the search for "dynamic equivalences" between the source and target languages at this kernel level. This idea echoes the view of Catford, according to whose structuralist framework translation involves finding "translation equivalents" that "are interchangeable in a given situation" and "relatable to (at least some of) the same

10. The interpretative school calls this transmission the "vouloir-dire."

features of substance" [Catford 1965, 49f]. As noted earlier, the entire notion of formal equivalence is repudiated by the interpretative school. They argue instead that translation equivalents are *created* by the act of translation, but cannot be removed from context and studied as though they existed outside this act.¹¹

The understanding stage is explicitly at the centre of the process in the model of interpretation (i.e. oral translation) advanced

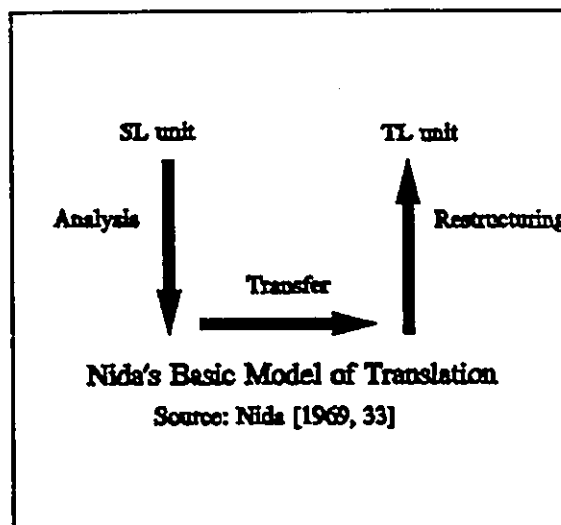


Figure 2-1 Nida's Basic Model of Translation

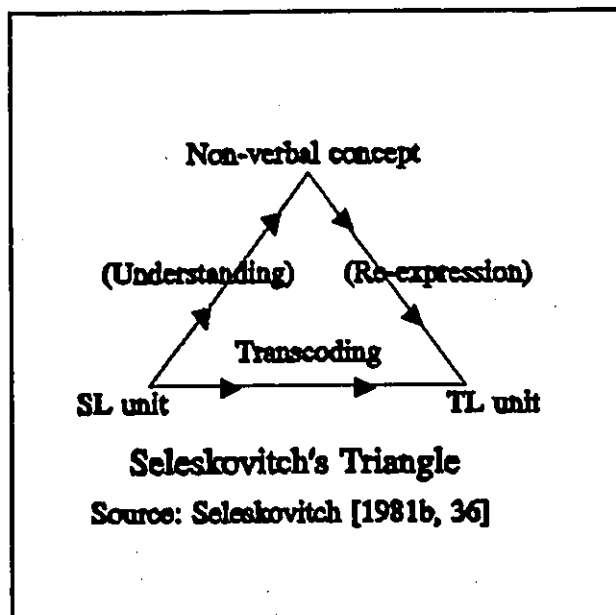


Figure 2-2 Seleskovitch's Triangle

by the interpretative school.¹² As diagrammed in "Seleskovitch's triangle" (Figure 2-2), words enter the listener's "immediate" memory in chunks where they come into contact with the "cognitive complement" (*compléments cognitifs*). This represents the sum total of relevant knowledge, which includes general extra-linguistic knowledge (*bagage cognitif*), specialized knowledge about the subject matter, knowledge of the parameters of the communication situation, memory of previous portions of the text ("cognitive

11. Nida would not disagree entirely with this point of view, since he acknowledges that "once we have isolated a word from its living context, we no longer possess the insight necessary to appreciate fully its real function" [1964, 40].

12. Which is also explicitly offered as a model of written translation, a fact that has been sharply criticized (see, for example, Newmark [1981, 98f]).

context"), knowledge of linguistic probabilities ("verbal context"), etc. The interaction of the words stored in immediate memory and the cognitive complement yields "units of meaning" (*unités de sens*), which are transferred to "working" memory as non-verbal concepts, leaving immediate memory free to process another chunk. Thus the message is built up as the text unfolds, with only the pure language-independent ideas disengaged from language being retained in memory to be re-expressed in the target language.¹³

In Delisle's model [1980, 85], essentially a re-working and adaptation for written translation of the interpretative model, this process is explicitly called "compréhension" (understanding). Delisle divides the process into two steps: "saisie des signifiés" (i.e., of words as lexical items or in *langue*) and "saisie du sens" (i.e., the meaning associated with the words in the particular context). Delisle notes that the distinction is essentially artificial, since the two processes actually work cooperatively and simultaneously. The cognitive

complement (i.e., the reader's knowledge) is viewed as a sort of prism ("le prisme du savoir non linguistique" [1980, 84]) through which the analysis passes in order for understanding to take place. Like the interpretative school, Delisle sees the outcome of this process as a series of non-verbal concepts ("prise en charge des concepts par les mécanismes cérébraux non linguistiques" [1980, 85]).

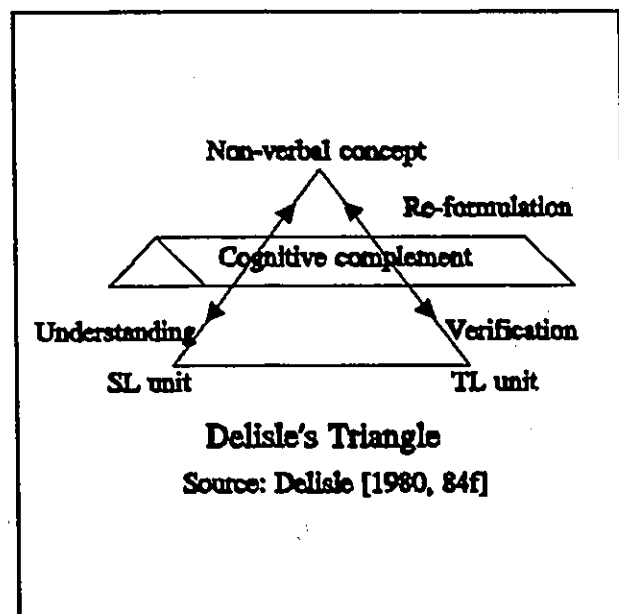


Figure 2-3 Delisle's Triangle

13. The contention that interpretation and translation operate on language-independent meaning is one of the more controversial tenets of the interpretative model.

In the interpretative models (Seleskovitch and Delisle), what exactly happens at the apex of the triangle — when the source-language message turns into a target-language message — is left undescribed, presumably because at that point the message is in completely non-verbal form and so requires no interlingual jump. While Nida at least provides a label for this stage, the discussions of "transfer" commonly included in his works usually focus on a handful of examples that actually relate to problems of analysis or re-expression. In fact, as Ladmiral points out, translation theorists have little choice but to gloss over this aspect of the translation act, because no one really has any idea what is going on during this "alchimie' psychologique ou psycholinguistique un peu mystérieuse" [1992, 331].

It should be stressed that understanding is, of course, relative; while the original author can be assumed to have full understanding of the message he wishes to convey, the reader's understanding may well fall short to some degree. What is important is that a level of understanding is achieved that is adequate to the transmittal of the message. A less than adequate understanding implies that the message will be less than adequately conveyed.

Once the message has been understood by the translator, his role diverges from that of the ordinary reader; he must now endeavour to reformulate the message in another language.

2.1.3 Re-expression

In the Nida model (Figure 2-1), analysis and transfer are followed by the "re-structuring" stage. Re-structuring is construed as a mirror-image process to analysis during which deep or "kernel" meaning in the target language is reconstructed to surface structures. Nida once likened his model to crossing a river [1975, 80]: you must move down the riverbank (analysis) to find the best place to ford (transfer), then work your

way back up the opposite river bank (re-structuring) to get as near as possible to your starting point. Nida does not explicitly acknowledge the central role of knowledge of various kinds in this process.

According to the interpretative school, the whole process of re-expressing the message once it is understood is completely spontaneous and virtually effortless: "On ne répétera jamais assez que le maniement du langage est réflexe" [Seleskovitch and Lederer 1984, 305]. As noted earlier, their theory is based on their experience with interpretation, but is offered up as a theory of translation as well. However, their claim of effortless translation runs so counter to experience — translation can be damned difficult — as to constitute one of the theory's weaker links.

Delisle tries to remedy this shortcoming in his model. Notably, he construes the process of re-expression as passing through the same prism of cognitive complements as the understanding stage, recognizing that the translator's knowledge plays an important role here as well. And he adds another stage, verification, to his model, acknowledging that the translator must check and perhaps adapt or change his initial form of re-expression in accordance with the various variables involved. Delisle stresses the dynamic and intelligent nature of the process, describing it as a "va-et-vient incessant entre le sens 'immatérialisé' ... et les formes linguistiques disponibles propres à les manifester" [1980, 78]. Moskowitz echoes this view, calling the process of understanding and re-expression "[un] intense effort de réflexion intellectuelle" [1978, 84]. Although Nida has been criticized for compartmentalizing the stages of translation in his model as though they were sequential and separate,¹⁴ he too explicitly states that "in the actual process of translating, the translator will constantly swing back and forth between the analytical and the restructuring process by way of the transfer" [Nida and Taber 1969, 104].

14. See, for example, Lörscher [1991].

The interpretative school makes an important distinction between what they regard as "true" translation (or interpretation) and "transcoding." Transcoding relies on the linguistic meaning of words rather than on textual meaning and message. In true or "interpretative" translation, on the other hand, the linguistic forms of the source text play no role in the formulation of the target text. The translator's job is simply to express spontaneously in the target language what it is that he has understood.

It is admitted that certain types of information (proper names, numbers, etc.), as well as words "qui désignent des universaux de connaissances" (i.e. well-defined technical terms) [Seleskovitch and Lederer 1984, 286], must be transcoded from the source text. However, the basic thrust of translation is never transcoding, but understanding and re-expression of the meaning: "traduire n'est pas transcoder mais comprendre et exprimer" [ibid, 19]. According to this point of view, whereas a transcoding approach runs inevitably into problems of polysemy, ambiguity and "translatability,"¹⁵ interpretative translation rarely suffers from these problems, since the translator effortlessly transforms the linguistic meanings of the words ("significations" or "virtualités de sens") into non-verbal "meaning" thanks to his "cognitive complement" or knowledge.

2.2 The Role of Knowledge in Translation

From the foregoing discussion I would submit that, despite some wide differences in approach and terminology, there exists a general consensus on the basic translation process. It can be summed up as follows:

15. Polysemy being the phenomenon of a word having two or more distinct meanings (e.g., "bank"), and ambiguity meaning that a particular passage may be interpreted in more than one way because of a variety of factors, including polysemy, unclear antecedents, and confused syntax (e.g. "Translators hate translating Grateful Dead songs because they are on drugs"). "Translatability" here refers to the relative difficulty experienced in finding idiomatic solutions to passages that resist literal or word-for-word rendering because of the structural differences between the languages involved.

The goal of translation is to re-express in one language a textual *message* originally formulated in another language. The message is conveyed to the translator (or any other reader) by being *understood*. If the understanding is only partial, the message will not be fully conveyed. Understanding takes place through the interaction of the *meaning* of the text with the reader's *knowledge* of the world and of the communicative context. Its *re-expression* by the translator in another language also relies, in large part, on that individual's knowledge.

Both linguistic and extra-linguistic knowledge, then, are involved in the process of understanding the message and reformulating it in the target language. Without knowledge, the translator would be unable to translate.

Although the pivotal role of extra-linguistic knowledge in translation has long been recognized,¹⁶ generally speaking it has been given short shrift by writers on translation. Even a writer with as practical a bent as Newmark makes no mention of the importance of the translator verifying or supplementing his knowledge. His translation model concentrates squarely on the meaning carried by the words of the text, although he does admit that some meaning is implicit [1991, 30]. The closest Newmark comes to acknowledging the role of extra-linguistic knowledge is to include, as one factor in the translation process, "the moral and factual truth," which is defined as "what is being described or reported, ascertained or verified [...] where possible independently of the SL text and the expectations of the readership" [1991, 31].

As noted earlier, Nida too makes little mention of extra-linguistic knowledge in his model. Occasionally he notes in passing that knowledge of the "subject matter" is

16. Machine translation researchers, for instance, had hit the "semantic wall" by the early 1960s. Bar-Hillel's well-known example of "the box is in the pen" was intended to point out that understanding (and so translation) was only possible with some real-world knowledge of boxes and pens, i.e. that the "pen" can only be some sort of enclosure such a play pen or pig pen that is capable of accommodating a box. The literature on machine translation is full of discussions of the importance of world knowledge.

important,¹⁷ but fails to develop it further. However, Nida is certainly sensitive to the importance of extra-linguistic factors, primarily those cultural and situational in nature. This is captured in his notion of "coherence," which is defined as "the manner in which a text fits or adheres to the world view of the participants in the communication, the setting [...] and the existence of similar or diverse texts" [Nida and de Waard 1986, 81f]. In other words, a translation must be developed in accordance with the translator's extra-linguistic knowledge of the audience and the larger communicative situation. Although Nida does not explicitly say so, it is clear that his analysis stage relies on the translator's extra-linguistic knowledge, since the translator will "anticipate" the approach needed for the subsequent restructuring in the particular target language involved [Nida and Taber 1969, 104]. And, as Mayorcas-Cohen points out, "[Nida's] transfer phase which occurs between the reading and understanding stages, and composition in the final target language (TL) results from the bringing together of knowledge and information about the source (SL) and target languages. The translator [also] adds or makes use of information on the nature and subject matter of the text" [1986, 62]. In short, linguistic and extra-linguistic knowledge are an implicit part of Nida's translation model, even though they are not fleshed out.

As discussed earlier, the models developed by the interpretative school (primarily, Seleskovitch and Delisle) explicitly incorporate knowledge in the form of the so-called cognitive complement. This term is a catch-all for the sum total of the receptor's knowledge of the world, the subject and the situation, and is not very fully developed. In some of her later writings (e.g. Seleskovitch [1988]), Seleskovitch has placed more emphasis on the role of knowledge, but has linked it specifically to technical translation. In this she is not alone; Ladmiral, for example, states that "dans le domaine technique, la traduction ne pourra souvent être faite efficacement qu'en faisant fond sur les *realia*

17. A typical quote: "In addition to a knowledge of the two or more languages involved in the translational process, the translator must have a thorough acquaintance with the subject matter" [Nida 1964, 150].

auxquels réfère le texte" [1992, 330], as though this were not true of translation in general.

As noted earlier, however, Delisle incorporates the cognitive complement into his model in the form of a prism through which both understanding and re-expression take place. This intriguing analogy suggests that the higher the quality of the metaphorical prism or, more properly, lens (i.e., the more complete the translator's knowledge), the clearer will be the image produced (i.e., the translator's understanding of the original text *and* his translation of it). Seleskovitch and Lederer themselves acknowledged this factor: "plus les connaissances sont étendues, plus le sens de l'énoncé prend de précision" [1984, 21]. Thus, improvements in the translator's stock of relevant knowledge should lead to improvements in the quality of the translation product. This is an important point to which we will return later.

One translation theorist associated with the interpretative school who explicitly recognizes the importance of knowledge in a translation model is Garcia-Landa. A "knowledge base component (K)" is included in the symbolic representation of the translation process, intended to encompass "everything that is necessary to know to be able to understand what speakers mean" including general encyclopedic knowledge and knowledge of the subject under discussion [1990, 482]. However, Garcia-Landa reserves knowledge of the situation and of the previous conversation for another parameter called "social context of the talk act" (G), and also apparently fails to recognize that the knowledge base of the author may be different from that of the listener/translator, a factor that will be discussed in Section 2.4.2.2.

Another fact worth considering is that re-expressing information is commonly regarded as one of the best ways to test understanding. Asking students to re-express in their own words what they have learned is an ages-old teaching method. It is no coincidence that *précis* and *paraphrase* writing has become a valued technique in

translator training.¹⁸ After all, Mossop [1983] argues that the basic function of a translator is to be a "rapporteur" — someone who reports in one language what was written in another. Clearly, if re-expression is used to *test* knowledge and understanding, successful re-expression must *require* knowledge and understanding.

From the foregoing discussion, it is clear that knowledge in a very broad sense is essential to translation. This is trivially true in that the translator must obviously possess linguistic knowledge of both the source and target languages. Yet any text, from the most literary to the most technical, also assumes an enormous amount of non-linguistic knowledge on the part of the reader. Adequate understanding is only possible when the reader's overall stock of knowledge is equal to the task. Further, the depth of the understanding depends on the comprehensiveness of the relevant knowledge. Most readers will have had the experience of addressing a text in a new and unfamiliar subject area, where the words, although mostly familiar, seem strange and unfriendly. As knowledge of the subject area is gleaned from the text and outside sources, however, things seem to fall into place until, finally, the meaning of the text comes shining through. Only once the lens of relevant knowledge is adequately formed may adequate understanding (and satisfactory translation) follow.

2.3 Classification of Knowledge

A closer look at and more careful classification of the components of knowledge involved in translation will now be attempted. The proposed categories are not necessarily distinct, and hence in some cases the line of demarcation between them may be difficult to draw precisely.

18. See Russell [1989].

1. Linguistic knowledge
 - a) Grammatical knowledge
 - b) Lexical knowledge
 - i) Knowledge of the general lexicon
 - ii) Knowledge of specialized vocabulary
2. Extra-linguistic knowledge
 - a) General world knowledge
 - i) Common-sense general knowledge
 - ii) Current general knowledge
 - b) Situational knowledge
 - c) Specialized (domain-specific) extra-linguistic knowledge

2.3.1 Linguistic Knowledge

Linguistic knowledge refers to the translator's knowledge of the two languages involved. *Grammatical knowledge* is used here in a very broad sense to mean competence in the mechanics of the languages, including phonology, morphology, syntax and discourse structure. As has often been said, a translator, who usually translates into his mother tongue from another language,¹⁹ must possess a thorough grammatical knowledge of the source language and an educated native speaker's grammatical knowledge of the target language. *Lexical knowledge* refers to the translator's knowledge of the vocabulary of the languages involved. This category may be further divided into *knowledge of the general lexicon* and *knowledge of specialized terminology*, the latter of which is not commonly shared by all speakers of the language. Translators typically develop special expertise in the terminology of fields in which they most frequently work.

2.3.2 Extra-Linguistic Knowledge

The second major category of knowledge is *extra-linguistic knowledge* — the translator's knowledge about the world (which includes knowledge of the subject matter of the text).

19. At least in Canada; the practice elsewhere is sometimes different.

It is sometimes termed "world" or "real-world" knowledge.²⁰ Its first component, *general world knowledge*, comprises first what I have termed *common-sense general knowledge*. By this I mean a vast array of basic information about the world and the characteristics and behaviours of the things in it, a kind of knowledge that is gained primarily through experience. People tend to take common-sense knowledge for granted, yet it is what enables them to effortlessly understand natural language most of the time despite its awesome ambiguity when removed from context. It also allows them to translate. For example, human readers have no trouble assigning the proper antecedent to "he" in the two following sentences:

Sigmund told his father he was leaving.
Sigmund told his father he was boring.

It is this common-sense knowledge that has proven so difficult to simulate in any real depth on computers. When machine translation systems profess to have a "semantic" component, it is generally an attempt to simulate some degree of common-sense general knowledge. Case grammar, for instance, which has been widely applied in machine translation, tries to assign roles to the various parts of a sentence on the basis of coded semantic information, so that the machine can distinguish the difference in meaning between, for example:

Sigmund opened the door (Sigmund is "actor" in case grammar terms).
The key opened the door (key is "agent," and the "actor" is unknown or not expressed).

Human beings come to texts equipped with a highly developed semantic competence. As Mossop put it:

20. An important distinction is sometimes made between these two expressions, however. "World" knowledge might concern a possible or idealized world or subworld. "Real-world" knowledge is specifically about the universe we inhabit.

Unlike the programmers of a translating machine, a human translator has no need to acquire any explicit rules for dealing with the world outside language or for dealing with linguistic structures. As a participant in a human culture, he or she already has access to both these types of knowledge and already knows how to combine them to interpret a text [Mossop 1983, 277].

In an adult, the largest part of this pool of knowledge remains unchanging.

In contrast, the second component of world knowledge, *current general knowledge*, comprises information about current events and trends that is constantly being added and updated. It, too, is frequently critical to proper understanding and so to accurate translation. For instance, a newspaper headline of about ten years ago read: "Government Relies On 'Scare Crow' Tactics." The message conveyed in this headline would escape anyone unfamiliar with the name of the Governor of the Bank of Canada (John Crow) and with the fact that the decisions of the Governor largely determine prevailing interest rates in this country. And the full message is only revealed when the headline is placed in the full historical context: a time of record high interest rates that many felt were pushing the economy towards recession. The "scare crow" tactics referred to Crow's warning that rates would likely remain high as long as wage demands continued to fuel inflation. Current general knowledge may be important even to the translation of historical texts. In translating Shakespeare, to use an obvious example, information about the "current" general knowledge of that author's intended audience (i.e., the general public of the day) may be essential to the understanding of an allusion or joke.

The distinction between general world knowledge and *situational knowledge*, the second component of extra-linguistic knowledge, is commonly recognized by translation theorists: "les aspects de la situation que la compréhension ... met en jeu sont divers: certains tiennent à une connaissance préalable d'ordre général; d'autres naissent des conditions même de la communication" [Pergnier 1980, 402]. Situational knowledge

encompasses knowledge of the various factors involved in the particular circumstances of the text act – the communication situation, as it is sometimes called in translation theory. In linguistics, a new appreciation of the importance of this factor led to the contemporary study of *pragmatics*, including discourse analysis and text semantics. The communication situation is one of the most complex factors involved in the translation process, because translation actually involves two distinct communication situations: that connected with the creation of the original text and that connected with the translation of that text. Components of the communication situation, therefore, include the characteristics of the original author and the translator, the intended audience for the original text and for the translation (sometimes quite different), the respective cultural contexts (possibly very different because of temporal or cultural distance), etc. It is also important to realize that the text itself is part of the communication situation and may provide a wide variety of pertinent information. For instance, a vital piece of knowledge that is not supplied in one section may turn up in a later section, as when the titles of bibliographic references provide precious target-language vocabulary.

Specialized extra-linguistic knowledge is the third component of extra-linguistic knowledge. It can also be referred to as "encyclopedic" or "conceptual" knowledge.²¹ Since the translator's task is to understand the message as fully as possible, and since he is normally not a bonafide expert in the field in which he is working, he very often needs to acquire domain-specific background knowledge in order to translate. "Extra-linguistic knowledge has to be available to enable the translator to transmit the factual content of technical texts" [Seleskovitch 1988, 85]. The fact that the translator cannot translate even dry scientific prose in the absence of understanding is well-known to translators but little appreciated outside the profession. In fact, according to Cormier:

21. In truth, both the latter terms are somewhat inaccurate for the purpose. "Encyclopedic" evokes the kind of knowledge found in encyclopedias and other reference works, whereas specialized extra-linguistic knowledge is a broader notion. Calling specialized extra-linguistic knowledge "conceptual," meanwhile, is potentially misleading because *all* knowledge is arguably stored in the human mind in the form of concepts. However, these two expressions will be used when appropriate in this thesis to refer to specialized extra-linguistic knowledge.

Le texte spécialisé présente des difficultés de traduction plus grandes sur le plan de l'appréhension et la compréhension que tout autre texte. Comme on ne peut traduire ce que l'on ne comprend pas, il faut bien sûr acquérir d'abord une certaine connaissance technique, ... mais aussi être en mesure de faire des raisonnements techniques et de suivre une argumentation à caractère spécialisé. On verra une fois de plus que la terminologie n'est pas sans poser de problèmes, mais que la compréhension reste au coeur de la problématique [1991, 440].

Thus technical translators commonly make a considerable effort to keep their specialized knowledge up-to-date. It is important to realize, however, that specialized extra-linguistic knowledge includes far more than simply technical knowledge. It is also required for literary translation; for example, translators of the great classics of literature would be remiss if they did not research their authors thoroughly. Bible translators need in-depth knowledge about the culture of their intended readership. And translators of poetry must possess a vast range of specialized knowledge in order to effectively render the many allusions and special nuances of their source texts.

2.3.3 Translators' Need to Supplement Knowledge

When the competent translator realizes that his knowledge is deficient in any of the above areas, he endeavours to supplement his stock of it. "[Le traducteur] devra donc combler l'écart entre ses connaissances linguistiques (et les connaissances factuelles qui leur sont liées) avant la traduction (c'est-à-dire son propre idiolecte) et celles nécessitées par la traduction du message. D'où l'importance considérable que prend *la documentation* dans l'exercice professionnel de la traduction" [Pergnier 1980, 397]. Turning first to *linguistic* knowledge, professional translators normally only occasionally need to supplement their *grammatical knowledge*, although this is more common than generally supposed. No one ever knows a language totally; it is not unusual for a translator to check a grammatical or stylistic point concerning the source language or the target language (the latter usually in order to respect prescriptive standards). *Lexical*

knowledge, on the other hand, is one of the translator's constant concerns – so much so that an entire professional field, terminology, has developed to serve this crucial need. Again, while the professional translator only occasionally has to check general lexical points (in unilingual and bilingual general dictionaries), the quest for specialized lexical knowledge is never-ending. The fact that most translators find it important to maintain their own personal terminology files (for both general and specialized lexical knowledge) testifies to the importance of this aspect.

Something often overlooked is that translators need not only raw terminological information but also what might be termed *strategic* information on how properly to *use* specialized terminology.²² "[The translator] needs to have access to information which enables him or her to use the appropriate term in an expert-like way" [Ahmad et al. 1989, 15]. Strategic information for translators might include specific suggestions on how to handle common translation problems in the field, as well as warnings and advice about prevailing terminological or conceptual confusion. Thus it may spill over into extra-linguistic knowledge. One of the simplest and most useful types of strategic information that can be added to a knowledge base is collocational information, which can be considered a subset of linguistic knowledge. Collocations are simply combinations of words that, according to the conventions of a given language, are habitually associated.²³ Part of the reason a particular text is judged to have a natural style appropriate to its subject field is that the collocations are correct.

22. It should be noted that some interesting research has been done on "translation strategies" in a more restrictive sense. Wilss [1983] distinguishes between translation "methods" (tried and true techniques) and translation "strategies" (individual, personalized techniques). Lörcher undertakes an empirical study of translation "strategies," which he defines as "potentially conscious procedure[s] for the solution of a problem which an individual is faced with when translating a text segment from one language to another" [1991, 76]. He draws conclusions about translators' mental thought processes from the strategies he identifies.

23. For instance, if someone has a responsibility, he can be said to "carry out" or "fulfil" that responsibility, but not to "complete" it. If he fails to carry it out, it can be said that he "shirked" it, but not that he "missed" it. The word "responsibility" collates, essentially arbitrarily, with certain verbs and not with others.

Turning to extra-linguistic knowledge, we take it for granted that translators possess the basic, *common-sense knowledge* about the world and its inhabitants that all human beings share. Even in this category, however, shades of differences exist between cultures; for instance, between what constitutes a normal reaction to a particular set of circumstances within the bounds of particular times and cultures. It is even more likely that in the category of *current knowledge*, the translator may find that he is lacking some piece of knowledge vital to his understanding of the text. The "scare crow" example cited above indicates how deficiencies in this area may impinge upon understanding. It is part of the translator's skill to recognize and remedy such situations, and this explains translators' concern with remaining abreast of current events, not only in their particular area of activity, but also in spheres of general interest.

In the category of *situational knowledge*, the translator must again rely on his judgment to ascertain whether he needs to supplement his existing knowledge. The importance of knowing the intended audience of a translation and the purpose of having the translation done has been recognized by numerous authors. For example, Newmark, discussing the parameters of textual analysis, distinguishes among others the intention of the text, the intention of the translator, and the readership [1988, 12f]. If this information is not supplied or is not complete enough, the professional translator will endeavour to find out more. Mossop, recommending an interview with the translation client, notes the importance for the translator of "a sense of audience" [1983, 265]. Situational parameters can take on particular importance in literary translation, particularly in the case of older literature, since an appreciation of the circumstances surrounding the creation of the original text requires background research.

The importance of being able to supplement one's stock of *specialized extra-linguistic knowledge* is crucial to the translator. The fact that the translator may be called upon to demonstrate expert-like facility in a field of knowledge at virtually a moment's notice means that he relies heavily on the quality of his reference sources. As noted earlier,

technical translators are constantly on the lookout for new technical dictionaries and reference works. However, knowledge is always expanding by leaps and bounds, with language struggling to keep up. Galinski points out that:

Undoubtedly the development of knowledge is much faster and its changes are more fundamental than that of language. In spite of all "creativity of language", de Saussure's "langue" at all its grammatical levels is very much governed by the linguistic norm, which is probably not so true for specialized languages [as] for the common language. The formation of new terms is impeded first of all by the very limited number of term elements to name new concepts. The number of concepts, on the other hand, increases more or less at the same pace as knowledge grows [1988, 168].

It is for that reason that technical translators like to consult human experts whenever they can. The knowledge of such individuals is far more likely to be up-to-date than that contained in a book, and translators are able to directly put to the expert any relevant question that comes to mind rather than having to depend on the typical book "interface" (i.e. table of contents, index, etc.).

Clearly, the translator is, to use the words of Mayorcas-Cohen, "an information user." As this author indicates [1986, 61], the translator requires information in order to translate at all, to understand the source text, to find the proper terms, to find sources of background knowledge, to keep abreast of changes in techniques, to find out about the evolution and benefits of information technology, and to find out more about their profession and their professional status. With the exception of the last two points (which lie outside the translation process per se), Mayorcas-Cohen's list corresponds more or less with the preceding discussion.

Knowledge, in fact, may be regarded as the providing the hooks to which understanding can attach itself. As long as the reader/translator's stock of knowledge is adequate, the message of the text may be reconstructed in his mind. When the hooks

are missing, however, certain parts of the message may hang loosely (i.e., may not be properly understood) or may fall off completely (incomprehension). Seleskovitch noted this phenomenon for interpreters:

... s'il y a identité entre l'information reçue et la connaissance antérieure, le rattachement est direct et la compréhension équivaut à une *re-connaissance*. S'il y a un trop grand écart entre l'information et les connaissances antérieures, le rattachement n'est pas possible, et il y a incompréhension.... La compréhension rattache l'information à des connaissances voisines; si celles-ci sont absentes, l'information passe inaperçue [1968, 93].

While this aspect of the translator's work has often not been adequately recognized by theoreticians, it is well understood by practitioners. As knowledge continues to grow exponentially and communication needs become ever more complex, it becomes clear that translators, as information users, need to recognize more consciously the importance of increasing their stock of relevant knowledge.

Those involved in translator training have long recognized the problem of inadequate knowledge. Translation teaching programs usually offer courses on documentation and terminology (the University of Ottawa is an example). For instance, Gémard, identifying the main deficiencies of novice translators and translation students, singles out for mention:

la faiblesse de la 'culture générale' du traducteur.... Par culture générale je n'entends pas seulement la culture 'littéraire', mais aussi et *surtout* la connaissance des faits, événements, progrès et autres réalités, indispensable chez le traducteur, à défaut de quoi il ne saisira pas dans toute son étendue le sens réel, profond, des textes qu'il doit traduire [1990, 657].

And Cormier notes the same problem in relation to technical translation:

le problème principal auquel l'étudiant fait face dans l'apprentissage de la traduction technique de premier niveau est la compréhension, élément pivot qui ouvre la voie au message, puis à sa réexpression....
L'incompréhension est le résultat d'une faille dans la connaissance linguistique, dans la connaissance thématique, ou dans les deux [1991, 446].

If, as Cormier says, lack of understanding stems from deficient knowledge, it follows that the most useful strategy at the translator's disposal to combat inadequate understanding is to *maximize* his relevant knowledge.

2.4 Knowledge Maximization as a Translation Strategy

2.4.1 The Opaqueness Factor

One puzzling claim made by the interpretative school is that, while polysemy and ambiguity are characteristic of any word or group of words cited out of context, language loses its ambiguity in actual use. "N'existant pas dans le discours, la polysémie ne pose pas de problème à la traduction" [Seleskovitch and Lederer 1984, 44]. Seleskovitch goes so far as to claim that any ambiguity perceived in the text must therefore be due to a deliberate choice on the part of the speaker. "Le sens qui est attribué [à chaque segment du discours] est parfois faux, mais il est toujours unique. Si, par hasard, il est double, c'est que l'auteur l'a voulu tel; auquel cas il utilise les moyens stylistiques nécessaires pour le faire entendre ainsi" [Seleskovitch 1981, 12]. This assertion is puzzling to the practising translator, who frequently encounters ambiguity in texts.

In fact, ambiguity and consequent lack of understanding on the part of the reader/translator are characteristic of virtually any text. Depending on a variety of factors, a text to be translated presents an inherent degree of "opaqueness" to its reader.

2.4.2 The Three Parameters of Opaqueness

2.4.2.1 Assumed Knowledge Profile of Reader

One characteristic of natural language is *ellipsis*. The vast continuum of reality and experience means that all languages select certain aspects of this continuum as significant and omit others. "Le *continu* n'existe pas dès qu'il s'agit de langue," writes Pergnier [1980, 211]. "Toute langue introduit nécessairement du *discontinu* dans la représentation." As a result of this characteristic of natural language, nothing that is expressed orally or in writing is ever fully explained; in fact, the time involved in explaining any thought totally and comprehensively would be almost infinite. The shared stock of human experience, at the very least, is always taken for granted. Ellipsis can thus be considered a language universal.²³ "À tous les niveaux de langage, on retrouve l'ellipse," observes Lederer. "L'ellipse est constante; un non-dit supposé su accompagne toujours le dit et est rétabli par l'auditeur.... Rien n'est jamais intégralement expliqué" [1981, 66].

The information left implicit in a text by ellipsis may be of many kinds. Larson points out [1984, 39], using her categories of meaning, that implicit meaning may be referential, organizational (i.e. related to discourse structure) or situational in nature. Such ellipsis may be due to the source-language structure (e.g. gender information), or it may reflect the fact that information is available elsewhere in the text or is part of the shared communication situation. The translator may then choose to make explicit any elided information that he feels is necessary for the reader's understanding.

Given that ellipsis is omnipresent, communication between human beings is only possible because of the knowledge they have in common. "All communication is based

23. As noted, for instance, by Weinreich [1963, 176].

on shared information" [Larson 1984, 38]. The role of mutual knowledge and beliefs in communication has been extensively studied by linguists concerned with pragmatics.²⁴ From the linguistic point of view, verbal communication is a highly coded form of social interaction in which "the interactants are continuously supplying the information that is missing in the text" [Halliday 1977, 207]. The author of any text, therefore, besides taking for granted shared human experience, must assume that the receiver shares with him a certain level of specific linguistic and extra-linguistic knowledge – a particular *knowledge profile*. This point was included as a category in Tudor's proposed framework for analyzing texts. He noted that:

the assumed background profile ... is potentially the most wide-ranging category as it can include virtually any element which the writer assumed to be present in readers' minds in the writing of the ST. At the same time, it is a crucial element in comprehending the full communicative force of texts in the sense of what they meant to their target readers in the SL [1987, 81].

As pointed out by the interpretative school, if there were no shared knowledge between author and receiver, communication would be impossible (since there could be no ellipsis), and if their shared knowledge were identical, there would be no need of communication at all [Seleskovitch and Lederer 1984, 38]. These two hypothetical extremes can be considered the opposite poles of what might be termed the *explicit/implicit parameter*. The author of a text makes an assumption about the knowledge profile of his idealized intended reader, then adjusts his text along the explicit/implicit scale according to this assumption. As Moskowitz points out [1978, 74], it is the combination of the explicit information in the text and the presumed knowledge of the reader that grants the reader access to the text's implicit information. Thus the greater the assumed knowledge, the more elliptic can be the formulation. It is also possible, of course, that the author tailors his work for someone with *less* knowledge of the subject

24. See, for example, Sperber and Wilson [1980].

area than he, as in the case of a popular science text. Dressler, among others, points out that the receptor is constantly formulating hypotheses about the direction of the communication. According to him, the ideal is for the author to provide just enough information for the reader to be able to orient and adjust his hypotheses in order to draw correct conclusions about the author's intentions [1981, 71].

2.4.2.2 Actual Knowledge Profile of Reader/Translator

Theoretically speaking, if the degree of explicitness chosen by the author and the actual degree of knowledge shared with the reader were to match exactly (and the text were perfectly written), the reader would enjoy perfect comprehension. This, of course, is an unrealizable ideal, particularly since no two people have exactly the same knowledge and experiences, not to mention the same idiolect. But translatability will be high as long as the particular reader's knowledge falls somewhere close to the knowledge profile assumed by the author (which is generally the case).

Moskowitz makes an important distinction between the "destinataire" or intended receiver (the party whom the author had in mind in writing the text) and the "récepteur" or actual receiver (the party who actually reads the text). The translator is necessarily a "récepteur," but never the "destinataire" [Moskowitz 1978, 73]. Mossop puts the point even more bluntly: "The translator is, generally speaking, not a member of [the] intended audience but an outsider, who so to speak overhears ... the transaction" [1983, 246]. Thus there is little chance that the translator will possess the exact degree of knowledge assumed by the author, because he is not the intended receiver. Indeed, Gile notes that:

le Traducteur a rarement des connaissances égales à celles de l'Émetteur et des Récepteurs dans le sujet de l'énoncé. Cette INFÉRIORITÉ COGNITIVE lui rend plus difficile la compréhension de celui-ci et l'oblige

à déployer des efforts considérables en vue de compléter les connaissances manquantes [1986, 135].

Thus the *actual knowledge profile* of the translator is another factor in translation opaqueness. The more the knowledge profile of the translator falls short of that of the intended reader (i.e. the less the translator knows about the subject matter and its associated vocabulary, compared with a typical reader of the text), the more opaque the text will be to him and the more he will need to supplement his knowledge through research. Ideally, he should acquire sufficient additional knowledge to bridge this "knowledge gap" and thereby bring his knowledge profile more into line with that of the intended reader.

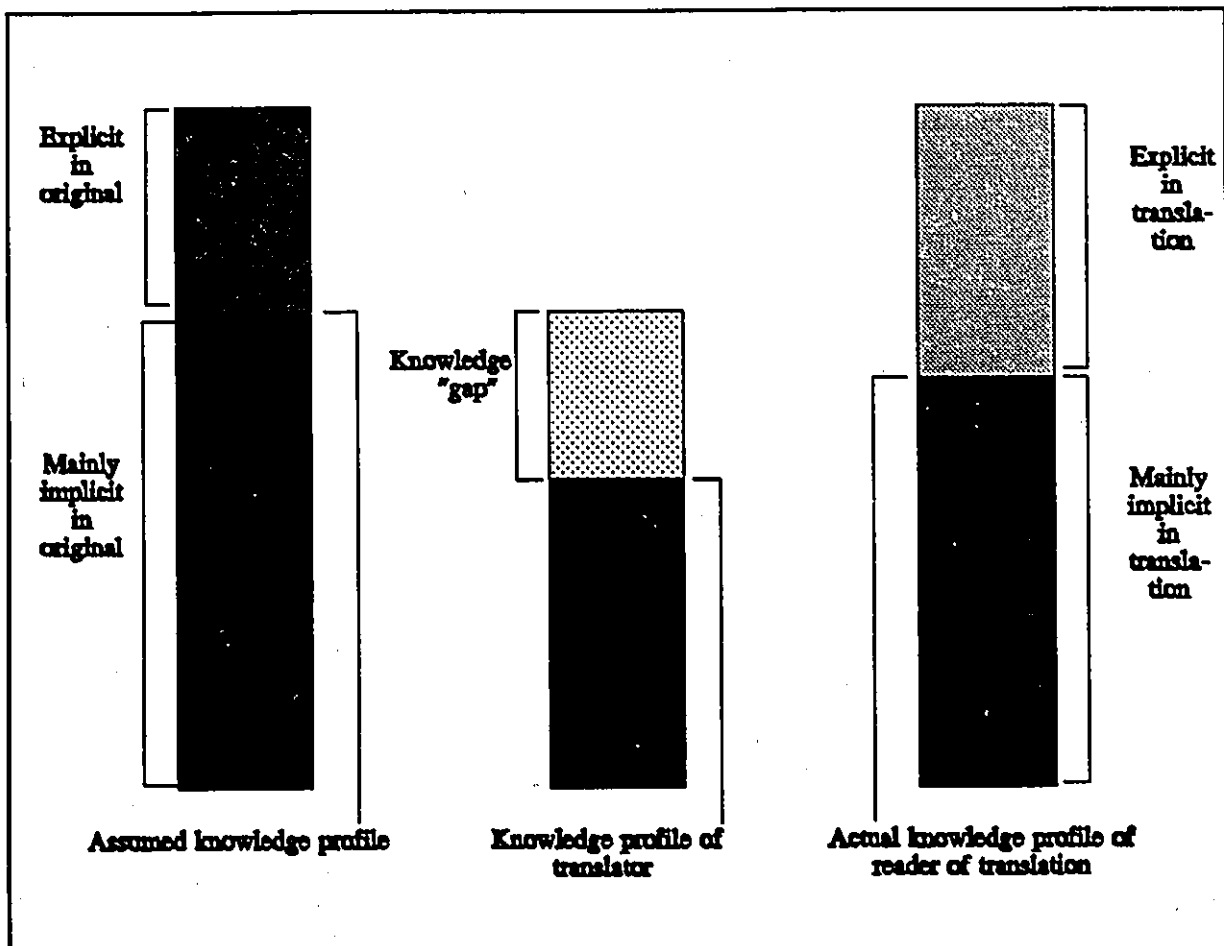


Figure 2-4 Role of Knowledge Profiles in the Translation Process

Moreover, the intended reader for the *translation* may be somewhat different from the intended reader of the original text. As a result, the former's knowledge profile may be different as well. The translator may take this factor into consideration by making explicit some information that may have been left implicit in the original, or vice versa. Nida notes that the restructuring stage of translating may involve "syntactic and lexical expansions" to "fill out ellipses" [Nida and Taber 1969, 67]. The interaction of the assumed knowledge profile, the knowledge profile of the translator, and the knowledge profile of the ultimate reader is diagrammed in Figure 2-4.

2.4.2.3 *Textual Clarity*

Another unfortunate fact about texts that is conveniently ignored by the interpretative school is that not everyone who writes a text or gives a speech is good at it. Translators are often faced with texts which are deficient in both style and logic, vitiated by unwarranted assumptions, illogical connections, and outright contradictions. Mossop uses as a hypothetical example a source text that is "an unedited piece of psycho-sociological prose replete with empty jargon" [1983, 260]. Once again, the perfectly written text is an unrealizable ideal. Part of the talent of a top-notch translator is his ability to transcend such difficulties. *Textual clarity*, therefore, can be considered a third parameter of opaqueness.

2.4.2.4 *Dealing with Opaqueness*

It can thus be argued that there are three parameters influencing the degree of opaqueness of a given text: the assumed knowledge profile of the reader (the degree of ellipsis or explicit/implicit selection made by the author), the actual shared knowledge between author and a specific reader/translator, and textual clarity (how well the text is composed). The interaction of these three factors determines the *overall opaqueness* of

the text and to what extent the translator needs to compensate for his lack of understanding.

As is probably already clear, the opaqueness factor is a continuum bounded by two hypothetical and impossible extremes. A text in which every item of information that is needed for communication (and no more) is made explicit, for which the reader and translator possess the exact knowledge profile assumed by the original author, and which is perfectly written would receive instant and perfect understanding and could even be translated perfectly. It is towards this unrealizable ideal that the translator strives. Faced with texts that inevitably fall far short, the translator's main strategy is to *increase his stock of relevant knowledge*. He judges what kind of knowledge he most needs for the translation job at hand and acts accordingly. "[Le traducteur] devra donc combler l'écart entre ses connaissances linguistiques (et les connaissances factuelles qui leur sont liées) avant la traduction (c'est-à-dire son propre idiolecte) et celles nécessitées par la traduction du message" [Pergnier 1980, 399].

An additional factor in textual opaqueness, sometimes playing a major role, is cultural dislocation. The more the translator is separated culturally and/or temporally from the world of the original author, the more difficult is his job. Given his primary interest in Bible translating, it is not surprising that Nida has always stressed the importance of culture. "The larger cultural context," he wrote, "is of utmost importance in understanding the meaning of any message, for words have meanings only in terms of the total cultural setting, and a discourse must be related to the wider sphere of human action or thought" [1964, 62]. In translation, the text may be removed to a greater or larger extent from the cultural context that gave it birth; thus, "compensation" may be necessary to aid the reader's understanding. Once again, the key is enhancing the translator's stock of extra-linguistic knowledge: in this case, about the particular cultures and the associated cultural differences.

2.4.3 Knowledge and Technical Translation

Technical language is often perceived as being distinct from everyday language. Certainly, there are some notable differences, although these differences are more of degree than absolute. For instance, the statistical frequencies of lexical and grammatical items in technical language are not the same as in everyday language; technical terminology includes many words not found in everyday language or used in a different sense; and there is a constant flow of neologisms into the technical vocabulary.²⁵

However, there are many common misconceptions about technical language. Bédard takes pains to rebut these myths at length [1986, 9ff]. He argues persuasively that technical language is neither rigorous and unambiguous, nor well-established and complete, nor used in a consistent and logical way. Through copious examples, he shows that technical language is just as prone to the problems of ambiguity and polysemy as everyday language.

Similar misconceptions abound regarding the technical translator. It is often assumed that a technical translator is a specialist first and a translator second, with the implication that his extensive expertise in a particular subject area compensates for possibly weaker language skills. But as Moskowitz points out, experts have "une connaissance de conception et d'exécution," while translators have "une connaissance de compréhension" [1978, 71f]. Certainly, technical translators should have some grounding in science. But, as Newman emphatically states, "that does not mean that a translator must have the same education as the scientist for whom he is translating; it only means that the translator must know the vocabulary, understand the basics, and realize when to stop translating and start researching" [1992, 271]. In other words, technical translators marry a broad technical background (i.e. a knowledge of basic scientific concepts and

25. For a detailed discussion of the differences between technical and non-technical language, see Ministère des Communications du Québec 1978.

principles) with a special ability to *understand* (to some degree, at least) new technical subject matter and to re-express it in another language. In fact, most "technical" translators are actually "generalists," in the sense that they can translate in any number of technical fields, thanks to an ability to distil and integrate into their existing knowledge the essential concepts and vocabulary of a new subject area. As Folkart puts it:

Knowledge ... is the *sine qua non* of the technical translator; the technical translator's stock in trade is some sort of background in the actual domain under consideration, or, lacking this, the ability to build up artificially — and rapidly — a competency approaching that of the specialist reader for whom the source-text was written in the first place [1984, 230].

For the technical translator, like any translator, understanding is needed to guide his choice of vocabulary, to allow him to verify that the meaning is accurately transferred, and to ensure the success of the translation in its communicative function [Bédard 1986, 76]. But since the message being conveyed in technical translation is essentially a factual one, understanding is even more critical to technical translation than it is to non-technical translation. In fact, understanding is often the most difficult part of the task.

Le problème de la compréhension se pose de façon particulièrement aiguë en traduction technique. En effet, contrairement au texte dit «général», dont le message renvoie à un fonds de connaissances commun à tous, le texte technique porte sur des réalités étrangères au monde de tous les jours et utilise des notions inconnues des non-initiés [Bédard 1986, 76].

And later Bédard notes that "une compréhension adéquate ... est le premier devoir du traducteur technique" [1986, 121] (original emphasis).

This is an important point. Folkart [1984] stresses that good technical translation is *necessarily* conceptually based or, in her words, "thing-bound." Knowledge of the subject matter frees the technical translator to apply his full range of skills to the linguistic job at hand. In fact, Bédard argues [1986, 89] that even superficial understanding makes for a

better translation; in other words, a little knowledge is better than none. Inadequate understanding, on the other hand, forces the translator to adopt a "word-for-word" or "transcoding" approach, which most likely will produce a stiff, unidiomatic translation plagued by major and minor errors of content. And even if his linguistic skills allow him to reduce the stylistic deficiencies, the content errors will remain. "A translator lacking subject knowledge may produce an elegant translation," Newman says bluntly, "but he is unlikely to produce an accurate one" [1992, 270f].

Relevant specialized knowledge²⁶ is thus a precious commodity in technical translation. At one time it was felt that, for the typical "generalist" technical translator, "la collaboration entre traducteur et expert est indispensable" [Horguelin 1966, 18]. But today knowledge is expanding too quickly for an expert to know intimately more than a narrow area of it. Moreover, access to domain experts is usually problematic. Among the problems: these people are not always available (since they have their own work to do), they have no real motivation to help the translator, they are likely to get annoyed if they are consulted too frequently, and they are not available for what I have termed "background browsing" (i.e. general consultation unrelated to a particular text). That is why the idea of a computerized version of the domain expert in the form of a knowledge-base system for translators is so attractive.

26. It will be recalled from Section 2.3 of this chapter that specialized knowledge is considered to have two components: specialized linguistic knowledge (i.e. knowledge of technical vocabulary and its proper use) and specialized extra-linguistic knowledge (i.e. encyclopedic knowledge about the specific domain).

2.5 Informal Experiment With Fourth-Year Translation Class

2.5.1 Summary

To investigate the importance of knowledge for translators working in an unfamiliar technical domain, an informal experiment was set up with the class of a fourth-year course in French-English technical translation for 27 Francophone students taught by Professor Brian Harris.²⁷ The experiment proceeded as follows. I attended one of the classes in order to present the students with a translation assignment, a text on stock-market options (see Appendix I). After allowing them time to read it over in class, I asked for questions about the text and the subject matter. I replied to these questions as simply as I could, trying to give them a basic overview of the subject. The students then had a week to complete a draft translation. At the next week's session, after the draft assignments had been handed in, I once again asked for questions and this time endeavoured to supply as complete answers as possible. After this session, the students' task was to revise their draft translations in light of the new knowledge I had supplied them and to hand in their final versions the following week. I read and analyzed both their draft and final translations.

The aim of the exercise was to find out what kinds of information generalist translators need when they are faced with the task of translating a text in an unfamiliar field. The first session was intended to assess these needs in a general way and to offer the students just enough knowledge to enable them to complete a draft. The aim of the second session was to supply them with whatever expert knowledge they might request for the revision of their drafts, with the obvious hope that there would be a noticeable

27. I am careful to call this exercise an "informal" experiment because there are obvious flaws in both methodological and practical terms. In particular, the use of Francophones biases the results, since their inevitably poorer target-language skills cannot be factored out. The major design flaw, meanwhile, is the lack of a control group.

improvement in the final assignments which might be at least partially attributed to possession of more expert knowledge.

The results were not all I had hoped, however. In particular, the quality of the final assignments was generally only marginally better than the drafts, certainly no greater an improvement than one would normally expect when a text is reworked. Nevertheless, some valuable insights were gained from the nature of the questions posed by the students in the classroom and from a detailed analysis of their efforts. Among the observations that emerged most strongly was that *a lack of knowledge obliges the translator to adopt a transcoding approach*.

2.5.2 Transcoding Versus Interpretative Translation

It will be recalled from Section 2.1.3 of this chapter that transcoding refers to an approach to translating that is word-bound, in the sense that the units of translation are single words or small groupings of words. When a translator is able to switch from a transcoding to an "interpretative" mode, his units of translation become based on meaning, not words. Interpretative translation can therefore be considered meaning-oriented and knowledge-based, in that it relies on a process of prior understanding.

In general, there are four aspects of translation that are adversely affected by a transcoding approach: accuracy of meaning, resolution of ambiguity, accuracy of terminology, and naturalness of style. These four potential problems are explored below with the help of examples from the experiment with Professor Harris's class.

2.5.2.1 Accuracy of Meaning

Accuracy of meaning refers to how precisely and completely the meaning of the source-language text is transferred to the target-language text. The problem with

transcoding is that the translator who does not understand or only inadequately understands what the text is about has no way of knowing when the meaning is *not* being properly conveyed. It is true that an experienced translator can often turn out a minimally acceptable product armed with only his general knowledge and a cursory understanding of the text. However, many serious pitfalls lie in wait for the translator who dares to work in this way, since the smallest oversight or misinterpretation can seriously skew the meaning. A transcoded translation is far more likely to suddenly stray into grievous error by distorting the meaning, saying the opposite of what was meant, or plunging into utter nonsense — all without the translator knowing.

In the following example, clearly a case of transcoded translation, the passage starts off basically on track, then veers off into meaninglessness, possibly because the student missed the significance of the final comma in the passage.

Quant a l'opération mixte haussière dans le contexte d'options de vente, elle consiste en la vente d'options de vente et en l'achat simultané, sur les mêmes actions, d'options de vente ayant un prix de levée inférieur à celui des options vendues.

As for the bull spread in a put option context, it is the sale of put options, and at the same time, the purchase of the same stock of put options with an exercise price inferior to the options with an exercise price inferior to the option sold.

In particular, a lack of specialized knowledge forces translators to rely on their *default* knowledge, i.e. the set of assumptions judged to be most applicable to a given situation until contradicted or refined by additional information. This is perfectly normal; in fact, it is part of the way human beings deal with the real world. Faced with a new situation, people rely on their default assumptions, established on the basis of current knowledge and past experience, until such time as new information or experience indicates otherwise. A translator working in an unfamiliar subject field does the same. For instance, it became apparent during the classroom discussion at the first session that

several students, knowing that the general subject field was the stock market, assumed that what was being bought and sold was *shares*, since that is the default general knowledge for stock-market trading. Of course, it was the trading of *options* that was under discussion.

2.5.2.2 *Resolution of Ambiguity*

Another area in which lack of knowledge (and the associated transcoding approach) causes meaning distortion is in the resolution of apparent ambiguity. It is very possible that a passage that poses some ambiguity for a translator would not be ambiguous to a domain expert, since the latter can follow the train of thought of the text. For instance, a transcoding translator who finds himself facing syntactic ambiguity may make the wrong decision because of his incomplete knowledge.

In the following example, the referent or head noun of the participle "ayant" is wrongly attributed.

... le vendeur d'une option d'achat ayant un prix de levée de 20 \$...

... the call option seller who has an exercise price of \$20 ...

Another kind of ambiguity that may trouble the transcoding translator but not a domain expert is lexical ambiguity. In the assigned translation, the French word "valeur" was a potential source of difficulty, since it could correspond more or less to the English abstract "value" or could refer to an actual security (such as a stock, a bond or, for that matter, an option). Another pair of words that caused considerable trouble for several students was "achat" and "vente." The French equivalents of "call" and "put" are, respectively, "option d'achat" and "option de vente." Calls and puts, of course, may be bought and sold. Thus the text has passages such as "l'achat ... d'options de vente"

(found in the passage cited in Section 2.5.2.1), which can be very confusing to translate in the absence of expert knowledge.

As discussed earlier in this chapter (Section 2.4.2.3), the original text itself may contain ambiguity. Again, while such ambiguity may pose no problem to the field expert (in fact, may not even be noticed), it can be a major stumbling block for the transcoding translator, who, with no expert knowledge to guide him, may interpret the author's meaning incorrectly. What is worse, incorrect decisions made early in the translation become compounded later on, because every decision about meaning that the translator makes is incorporated into the knowledge base he is mentally building about the text (or the "cognitive context"; cf. Section 2.1.3).

2.5.2.3 Accuracy of Terminology

Describing a particular translation or section of a translation as "transcoded" does not mean that the translator has failed to look up the words in specialized dictionaries or glossaries. In fact, a novice translator who has to translate a technical text typically follows the strategy "first look up all the terms." The target-language equivalents of these terms simply become part of his transcoded, word-bound translation. Aside from the quality of the terminological research, at least two other terminology-related problems can arise with a transcoding approach.

First, the mere fact that the transcoding translator has to rely so heavily on specialized glossaries may deceive him. While professionally prepared terminologies always take into account the overall subject field, often the subdomain distinctions are not fine enough for highly specialized texts. There were several good examples of this phenomenon in the students' texts. The following incorrect equivalences for terms that appeared in the assigned text were taken from specialized glossaries on the *stock market*:

- 1) "opération mixte" → "joint venture" ("opération mixte" actually referred in this text to

an options investment strategy known as a "spread"; 2) "prime de risque" → "risk bonus" (here the correct reference was to the option's "time value" or "time premium"); 3) "rachat" → "redemption" (correct for bonds, which *are* "redeemed," but here simply referring to an investor "buying back" some shares).

Second, there may be "hidden" terminology in the form of general-language words that are actually technical terms, which the translator does not recognize as such because of his lack of specialized knowledge. A transcoding translator would probably never think to look up the translation of "vendeur," for instance, since it seems all too obviously "seller"; yet in the field of options trading the proper term is "writer."

2.5.2.4 *Naturalness of Style*

The naturalness of style of a translation is compromised by transcoding. A transcoding approach is straitjacketed by the words; an interpretative approach has more freedom. A translator working in interpretative mode is able to make a wide range of major and minor adjustments to his work that a transcoding translator would not dare to, including minor additions and deletions of material, greater or less precision, etc.²⁸ The result is a target text that is more natural and easier to understand, and one that may well represent an improvement over the original text. A translator working in transcoding mode has no choice but to adhere slavishly to each and every word of the original, since he has no way of knowing what adjustments he could make to improve the style of the target text. As a result, a transcoded translation usually falls well short of the quality of the original text.

As the following rigidly word-for-word excerpt illustrates, a transcoded translation can be understandable, but it is rarely idiomatic.

28. Such adjustments have been extensively explored by "comparativist" linguists, most notably Vinay and Darbelnet [1958].

Le terme «opérations mixtes» s'explique par le fait qu'il s'agit de stratégies consistant, en règle générale, en l'achat et en la vente simultanés d'options d'achat ou d'options de vente (mais non pas des deux types ensemble) portant sur les mêmes valeurs sous options.

The term "spread" is used because it involves strategies consisting, as a general rule, in the simultaneous purchasing and selling of put or call options (but not of the two types together) on the same underlying stocks.³⁰

A more idiomatic translation of this passage might read as follows:

The term "spread" reflects the fact that spread strategies generally involve the simultaneous buying and writing of call or put options (but not both) on the same underlying stock.

As noted in Section 2.3.3, one type of knowledge that is often hard to come by, since it is rarely provided even in specialized reference works, is collocational information. In order to properly translate the phrase "options ... portant sur valeurs" in the assigned text, for instance, one has to know the proper preposition or prepositional phrase that links options to their underlying securities. Are options "with" securities or are they "for," "on" or "of" securities? Besides these four prepositions, students also suggested "applies to," "based on," "related to," involving," "pertaining to," and "with regards to." Clearly, they had no way of knowing the preposition that properly collates with "options" and "securities," which is "on." Another example is the French expression "effectuer une opération mixte." What does one *do* with an investment strategy? (Answer: "follow" or "implement".)

To sum up, a transcoding approach to translation is fraught with potential problems. There is a greater likelihood that the target text will stray into inaccuracy of meaning,

30. In the interests of accuracy, it must be conceded that the stylistic shortcomings of this and other passages from the subjects' texts could well be attributed, at least partially, to their poorer English skills, since they were Francophone. Still, this is clearly a case of a word-bound, transcoding approach to translating.

reference or terminology without the translator being aware, and the text's naturalness of style will also be adversely affected. On the other hand, knowledge-based or interpretative translation allows the translator to make a variety of adjustments to the text. The result is a clearer, more easily understood target text that more closely approaches the ideal of the spontaneously written text. In the present author's experience, this mode of translation also tends to be faster and more efficient, because the translator can work with confidence, unencumbered by the urge to double-check everything for fear of going wrong.

2.5.3 Other Observations

A few other interesting points emerged from the analysis of students' in-class questions and assignments. One is the difficulty involved in identifying the proper subdomain of a text received "cold" with little or no explanation. In the first classroom session, the students had some difficulty in correctly identifying the general domain of the text (the stock market) and even more difficulty in identifying the specific topic of the text (options).

Another interesting observation is that a little knowledge can sometimes be a dangerous thing. It was already noted that specialized glossaries can sometimes lead a transcoding translator astray. One student "corrected" the expression "stock prices" in her draft to "stock premiums" because I had pointed out in the second session that the price of an *option* is called a premium.³¹

Another important point that emerged strongly from the experiment was that a translator facing a new text in an unfamiliar field usually has little idea of what

31. At first glance, this might seem to contradict the previous assertion (in Section 2.4.3) that the quality of translation increases in proportion to knowledge. However, the current observation involves a translator who apparently did not possess enough background knowledge even to provide a "hook" for the new knowledge I had provided and so used a newly learned term inappropriately.

knowledge he is lacking. In trying to find that knowledge, the most obvious avenue to pursue is to investigate the unfamiliar vocabulary. Thus, at least in the early stages of knowledge acquisition, *the gateway to knowledge is lexical*.³² Even in interrogating an expert knowledge base, the first impulse of a translator will be to look up some words. Only once he has identified his primary area of interest will he be free to start browsing in a more conceptually oriented fashion. This recalls Newmark's remark that "translation and interpretation have to be based on words, sentences, linguistic meaning, language — because [...] they have no material foundation. Meaning does not exist without words" [Newmark 1981, 98].

2.5.4 Conclusions

As noted previously in Section 2.3.3, translation teachers have long recognized that instilling the basic technical knowledge required for understanding is the biggest challenge of teaching technical translation. "Parce qu'il s'adresse à des spécialistes et que l'étudiant est manifestement pas un, le texte spécialisé présente des difficultés de traduction plus grandes sur le plan de l'appréhension et de la compréhension que tout autre texte," states Cormier [1991, 440]. "La terminologie n'est pas sans poser de problèmes, mais ... la compréhension reste au coeur de la problématique." In an experiment designed to see whether a minimal grounding in scientific concepts improved translation skills, Bathgate concludes [1984, 228] that bridging the gap between students' general knowledge and specialized knowledge represents one of the main problems facing translation teachers today.

That is because, as we have seen, the repercussions of a lack of expert knowledge are far-reaching. The most important effect is that the translator with inadequate knowledge has little choice but to adopt a word-bound "transcoding" approach to his work rather

32. This observation was confirmed in the second informal experiment described in this thesis; see Section 4.4.4.

than a meaning-bound "interpretative" approach. Transcoding adversely affects translation quality in several areas: accuracy of meaning, resolution of ambiguity, accuracy of terminology, and naturalness of style.

It was argued in Section 2.4.2 that there are three parameters influencing the "opaqueness" of any given text: the knowledge profile assumed by the author for the reader, the actual knowledge profile of a specific reader/translator, and textual clarity (how well the text is composed). It was further suggested that maximizing knowledge could therefore be considered as an important translation strategy. We are now in a position to conclude that the goal of a knowledge-base system for translators is to provide the translator-user with enough knowledge to enable him to transcend a "transcoding" approach and engage his "interpretative" or "knowledge-based" translating mode.

On a more practical level, the experiment also signalled the great importance of access to collocational information for the translator, including not only nouns but also verbs (processes) and prepositions (relations). Finally, the experiment underscored that, at least for translators, words are the main gateway to knowledge in an unfamiliar field. Both of these last observations have clear implications for knowledge-base design, a topic that will be pursued in Section 5.3.

CHAPTER 3

Knowledge and Knowledge Representation from the AI Perspective

Knowledge is also a central issue in the field of Artificial Intelligence (AI). Like translation studies, AI is inherently an interdisciplinary field, drawing from computer science, linguistics, psychology, philosophy and other disciplines. It is also a relatively young field that has suffered from overhype in its infancy and that some authors feel is now facing a crossroads, if not a crisis (see, for example, Ryan [1991]). Like most interdisciplinary pursuits, it is difficult to settle on a satisfactory definition of AI; however, the basic thrust is to make computers do things that are presumed to require intelligence in humans and possibly to learn something about computers or people from the experience.

The purpose of this chapter is to investigate what knowledge means from the AI perspective. It starts (Section 3.1) with some background on the philosophical roots of AI, and then (Section 3.2) investigates knowledge, knowledge bases and knowledge-base systems. Section 3.3 focusses on some practical issues in knowledge representation and Section 3.4 on some philosophical issues in knowledge representation. Three important schemes for representing knowledge — logic, semantic networks and frames — are examined in Section 3.5. Finally, Section 3.6 takes a closer look at concepts and conceptual analysis, the latter of which is the process of identifying parcels of related knowledge that form concepts, and then examines the relevance of conceptual analysis both to knowledge acquisition in AI and to terminology.

3.1 Background

Humankind's desire to explain and classify knowledge presumably dates back to whenever people first started thinking about thinking. Socrates stressed the importance of carefully defining all the concepts involved in a debate as a prerequisite for rational argument. Aristotle's work on "categories" was the first recorded attempt to break down mental concepts into an irreducible set of elemental units (what today might be termed "primitives"). In the seventeenth century, Leibniz wrote his *Universal Characteristic*, which represented primitive concepts by prime numbers and compound concepts by products of primes. He also proposed that logic could represent a universal and unambiguous scientific language. The latter quest was vigorously pursued in the 19th century, notably by George Boole, who in *An Investigation of the Laws of Thought* (1855) tried to develop "the mathematics of human intellect." Boole's algebra was refined by others, culminating in Alfred Whitehead and Bertrand Russell's *Principia Mathematica*, published in 1910-13, which aimed to show and catalogue the relationship of mathematics to logic and human thought. Also in the early part of this century, Ludwig Wittgenstein explored the relationship between thought and language, and argued in his *Tractatus Logico-Philosophicus* (1921) that any statement which could not be analyzed into atomic propositions stating facts was meaningless. In his later philosophy, however, as presented in *Philosophical Investigations* (1953), Wittgenstein repudiated this extreme position, and offered instead a view of language as a "game" where the meaning of a word is determined by its use, so that a particular word does not have a single fixed meaning or even a fixed set of meanings. In 1958, John McCarthy, who is credited with inventing the expression "artificial intelligence," suggested that human knowledge could be represented using a computer language based on first-order predicate calculus.

Philosophical debate about the nature of knowledge assumed a new character with the advent of computers, since they could be used to automatically process symbols meaningful to human beings. Still, defining a universally acceptable set of primitives for

representing knowledge proved an elusive task. As Wittgenstein had come to realize, all languages contain a variety of close synonyms and complex networks of relationships that cannot be easily represented in terms of either-or distinctions. Early experiments in computational linguistics, especially machine translation, deflated early optimism about the computability of human knowledge and revealed the importance of exploring how people establish and manipulate conceptual structures in everyday life and how they use common-sense reasoning.

The importance of knowledge for computer systems capable of acting in intelligent ways was recognized from the earliest days of AI. Still, in the 1960s, much of AI research focussed on techniques such as search and theorem-proving that were "knowledge-poor." As heuristics were added to restrict the combinatorial explosion of possibilities that arose in systems with broader goals, AI researchers found it necessary to add more and more knowledge to their systems. By the mid-1970s, a fundamental change had occurred in the thrust of AI [Waldrop 1984, 1280]. "The essence of intelligence was no longer seen to be reasoning ability alone. More important was having lots of highly specific knowledge about lots of things - a notion inevitably stated as 'Knowledge is power.'" To a large extent, this observation remains true more than a decade later. In an interview, Feigenbaum stated this clearly:

Knowledge is power. When you know enough and you know the right things, the answers come virtually immediately.... The "knowledge is power" hypothesis has received so much confirmation over the last decade and a half, or two decades, that I simply don't call it a hypothesis any more. I call it the "knowledge principle," that the key source of power for AI programs is in their knowledge.... One could almost extend this to another hypothesis, the "knowledge is all there is" hypothesis. Stop looking anywhere else. How to represent knowledge and how to use knowledge, how to acquire knowledge, that's all there is [Canadian Broadcasting Corporation 1988, 6].

Thus, it may still be argued (in fact, Delgrande and Mylopoulos [1986, 3] call it a "cliché") that knowledge and how to represent it together constitute the central issue in artificial intelligence research. However, knowledge has turned out to be a slippery commodity, and current research tends to be more circumscribed and less ambitious in scope.³¹

3.2 Knowledge, Knowledge Bases and Knowledge-Base Systems

A commonly made distinction is that between *data*, which consists of raw, unprocessed input, and *information*, which is data that has been refined and selected for its meaningfulness. From the AI perspective, there is a third important distinction to be made – between these two concepts and *knowledge*. "Knowledge may be regarded as information in context, as information organized so that it can be readily applied to solving problems, perception and learning" [Tanimoto 1990, 89]. Essentially, unlike information, knowledge has a complex structure that brings out and exploits the relationships among the pieces of data.

While there are theoretically many ways information could be organized in order to transform it into knowledge, the simplest and arguably most useful way is to bundle it into units that are placed according to their characteristics or "properties" into a larger framework within which certain kinds of relationships hold. Such a structure is often called a *knowledge base*, and the branch of AI dealing specifically with inputting knowledge into a computer is generally referred to as *knowledge acquisition*.³² This subfield can also be seen as merely one aspect of the broader problems involved in

31. There are exceptions, of course; Douglas Lenat's "Cyc" project is one of the most ambitious knowledge-based AI endeavours ever launched (see Section 3.3).

32. It should be noted that the term "knowledge acquisition" has expanded through use to sometimes cover more than acquisition per se.

identifying, capturing, retrieving and using knowledge, an area of research known as *knowledge engineering*.

A *knowledge base* is simply a collection of facts about the world [Kramer and Mylopoulos 1987, 883]. "World" here does not mean the real world, however, but an idealized slice of it – a subworld or microworld, or a universe of discourse. A knowledge base is essentially a structured collection of propositions about a selected subworld. These propositions are modelled as a set of "objects" with links to other objects denoting the relationships that are believed to hold among them. A knowledge base is constructed according to some underlying knowledge representation scheme; three such schemes will be discussed later (Section 3.5).

As Sowa points out, however, "knowledge is more than a static encoding of facts; it also includes the ability to use those facts in interacting with the world" [1984, 2]. Thus there must also be instructions for using the knowledge in a meaningful way. Moreover, some kind of meta-knowledge (knowledge about knowledge) is required to govern when and how these instructions are applied. In the 1970s, there was considerable debate for a time over the relative importance of declarative knowledge (facts about objects and events) and procedural knowledge (rules for interpreting and manipulating facts). It is now generally acknowledged that both types of information are necessary, and most current systems employ a mixed strategy.

As discussed in Section 1.2, a knowledge-base system results from the combination of a knowledge base and a knowledge-base management system. A knowledge base's acquisition and retrieval capacity relies on inference. This means that not all information that can be derived from the knowledge needs to be explicitly modelled; thanks to the relationships specified among the various concepts, a variety of implicit information can be retrieved. The most common type of inference used in knowledge bases is *inheritance*. A hierarchical ordering of the objects in the knowledge base allows

concepts lower in the hierarchy to inherit the properties of any higher concepts to which they are related (in other words, for subconcepts to inherit the properties of their superconcepts), subject to various constraints. Inference-based knowledge retrieval is deductive in nature. Implicit coding of information through inheritance is more powerful and more efficient than the explicit inclusion of potentially redundant information (as in a conventional database), although the time required to retrieve it may be longer.

Inheritance is what most distinguishes a knowledge base from a conventional database. A conventional database must rely on the user to provide even the most basic interpretation of the data it contains. A knowledge base, on the other hand, contains explicitly declared rules about its own workings and the information it contains; thus it might be considered to have a *semantic capacity*. The line between a knowledge-based system and an expert system is somewhat harder to draw than that between a knowledge base and a database. Certainly, a knowledge base can be used by, or even form an integral part of, an expert system. In general, however, an expert system is task-specific, whereas a knowledge-based system, while it is typically filled with information about a specific subworld, may be used for a variety of tasks and, in many cases, may be used with little or no adaptation for other subworlds.

3.3 Practical Issues in Knowledge Representation

While the range of theoretical and philosophical issues in knowledge representation is broad, there is also no shortage of practical issues. One is how external objects and events are mapped onto their computer representations. Facts related to the real world (or a selected subset of the real world) must be mapped onto some kind of symbolic formalism to be manipulated. Natural language is itself a symbolic system for talking about the real world. Mapping from natural language to computer representations and vice versa, however, is enormously complex, since such mappings are almost never one-

to-one. Moreover, there are exceptions to any rule, as well as the constant possibility of unanticipated (and unpredictable) instances and applications. As Sowa points out [1991a, 220], it is the complexity of the real world that causes these problems, not any failing of natural language. Consequently, knowledge bases must be able to deal with *imprecision* (because the world itself is heterogeneous and inexact) and with *uncertainty* (when something cannot be either proved or disproved on the basis of current knowledge). Uncertainty poses special problems for formalization, since the truth or falsity of a particular piece of information may never be definitely known. However, special techniques have been developed, notably in expert systems, to assign probabilistic or confidence ratings to knowledge.

There are also a number of issues related to the scope of the knowledge representation – the world it purports to describe. Because of the incredible complexity of reality, to this point knowledge-based systems generally remain confined to limited subworlds or micro-worlds if they are to perform adequately. The precise definition or specification of the subworld is thus crucial – and far from easy. Spillover or crosstalk from the outside world is always possible. As a result, knowledge bases must be equipped with ways of dealing with *inadequacy* (when questions lying outside their world intrude) and *incompleteness* (when knowledge lies within their subworld but is not represented). In earlier logic-based systems, it was tacitly assumed that the knowledge base was complete for its world; anything that could not be proven on the basis of current knowledge was taken to be false. This "closed-world assumption" has now been generally abandoned in favour of a recognition that any but the most trivial knowledge base is necessarily incomplete in some way.

It has proven very difficult to "scale up" successful knowledge-based systems to large domains.³³ As knowledge bases grow in size, a number of inherent problems become increasingly troublesome. *Inconsistency* is one: as Delgrande and Mylopoulos note [1986, 20], "it is a fact of life that large knowledge bases are inherently inconsistent, in the same way as large programs are inherently buggy." *Inaccuracy* is another, because the likelihood of both minor and major errors entered by human operators, as well as the likelihood of inaccuracy caused by inadequate updating, grows with the knowledge base's size. In the jargon of the field, systems tend to grow more "brittle" with size, in that they are increasingly prone to unpredictable, possibly catastrophic failure.

An ambitious attempt is currently under way to solve the "scaling up" problem of large knowledge bases. Since 1984, Douglas Lenat and his associates have been inputting large amounts of basic, common-sense knowledge into a knowledge base known as Cyc. It is hoped that a critical mass of knowledge will be eventually be built up which will allow the system to transcend the "brittleness" barrier that has so far limited the extension of large knowledge-base systems. There are currently some 10 million concepts entered in Cyc.³⁴

The above issues are mostly related to the scope of the knowledge base. The converse issue is the "grain size" of the system; in other words, the degree of detail used in representing objects and relationships. To a large extent, the answer depends on the intended application and desired performance. Certainly, global consistency in grain size is imperative. This issue is closely tied to the selection of primitives – the basic

33. In fact, it has proven difficult in general to scale up AI systems from successful experiments in highly restricted domains to less restricted domains approaching anything like the real world. System performance tends to degrade exponentially. Shastri, among others, contends that it is a fallacy to consider this simply a problem of scale, i.e., that more facts and rules are needed: "The problem is that the *solutions* do not add up" [Shastri 1988: 2 (*italics added*)]. In the field of natural-language processing, Sowa contends that "the successes of language processors on small domains and their failure on unrestricted domains result from the fundamental nature of language" [quoted in Pacquin 1991, 328].

34. For more information on Cyc, see Lenat and Guha [1990].

elements that will be used to encode the system's knowledge. As Barr and Feigenbaum remark [1981, 149], "the selection of primitive elements ... is a basic problem in all representation schemes, whether the primitives are represented as nodes in a *semantic net*, predicates in *logic* formulas, or slots in a *frame*." These three knowledge representation formalisms are discussed in more detail in Section 3.5.

Knowledge-based systems must also strive for adequacy on several practical levels: representational adequacy (the ability to efficiently represent everything in the appropriate subworld), inferential adequacy (the ability to efficiently derive implicit knowledge from explicit), and acquisitional adequacy (the ability to easily incorporate new knowledge). Ideally they should also offer retrieval adequacy (the ability to efficiently interpret user requests and deliver the knowledge needed). In practice, however, the "front end" or user interface for a knowledge-based system is frequently a separately developed component.

Now that knowledge-based systems are evolving from "toys" to genuinely useful applications, another issue of increasing importance is compatibility. The enormous amount of work that goes into the development and construction of a knowledge base can go to waste if this knowledge cannot in some way be exported to another system. Moreover, in order for the limited knowledge bases of today to develop and grow into the multifunctional knowledge bases of tomorrow, they must be able, at the very least, to talk to each other and, ideally, to merge their knowledge into a larger whole. Recognition of this necessity has led to the formation of knowledge-base standards groups.

3.4 Philosophical Issues in Knowledge Representation

It has been pointed by several authors [e.g., Johnson 1986, 239] that there are two fundamental approaches in AI research. In the late 1970s, the two schools were sometimes identified by the labels "neats" and "scruffies." The former group believed that a strong and elegant theoretical component was an essential prerequisite for an AI system. The scruffies, on the other hand, argued that the job of AI was to build interesting programs that worked; theory, if any, could come later. These two approaches might be better characterized as a theoretical approach and an engineering approach. While many AI experts have taken somewhat extreme positions, the distinction is really not cut and dried. For instance, one of the most prominent "scruffies" was (and still is) Roger Schank; yet he is not at all reluctant to claim that a general theory of cognition underlies his work.

Related to this debate is the question of the relationship between AI systems and human thinking. AI researchers belonging to what has been called the cognitive-modelling school believe that AI programs should draw inspiration from and try to emulate how human minds work. The most striking example of this philosophical approach is research into parallel processing or "connectionism," where systems with multiple microprocessors establish networks of links among themselves in imitation of networks of biological neurons. Other researchers with a more epistemological orientation are primarily concerned with the nature of knowledge itself rather than the nature of the mind. Still other researchers eschew any comparisons with human cognition and feel that computer programs should be written on their own merits, relying on those strategies to which computers are best suited.

The extent and intensity of debate in the field of AI over philosophical issues has already been alluded to. Proponents of "strong AI" believe that the ultimate goal is to produce systems that literally "think" in the way human beings do. "I draw no boundary

between a theory of human thinking and a scheme for making an intelligent machine," wrote Minsky [1981, 247], and on another occasion he termed brains "meat machines." At the other extreme is a long string of critics of AI; notably, Searle, Dreyfus, Weizenbaum and Penrose. The thrust and intensity of their attacks vary, but the common thread is that there is a fundamental difference between computer processing and human cognition, a gap that will never be bridged. Between these two extreme positions, of course, lie some voices of moderation (e.g. Winograd), who are careful to avoid excessive claims for their research and to adopt a reasoned, critical perspective on their successes. Still, the debate over the fundamental goals and the very nature of AI continues and, from an informed outsider's point of view, makes great reading.

As noted earlier, the question that stands at the centre of the philosophical debate is knowledge – what it is, how it is acquired, how it is stored, how it is used. Even Schank has admitted [1977, 1007f], from his cognitive-modelling perspective, that "researchers are starting to understand that tour-de-forces in programming are interesting but non-extendable.... The AI people recognize that how people use and represent knowledge is the key issue in the field."

3.5 Three Ways of Representing Knowledge

3.5.1 Predicate Logic

First-order predicate logic (or calculus) has been widely used as a means of representing knowledge. Basically such a knowledge base consists of a series of axioms, which consist of facts and rules about those facts. Formal inference rules allow deductions to be made. For example, given the rule that everything mom cooks is delicious, i.e. " $\forall x. \text{mom's_cooking}(x) \rightarrow \text{delicious}(x)$," and the fact that mom cooks a dish known as "gunk," i.e. " $\text{mom's_cooking}(\text{gunk})$," it can be inferred that " $\text{delicious}(\text{gunk})$."

The inference rule that allows this conclusion is technically known as "modus ponens" and can be summarized as: if $A \rightarrow B$ (A implies B) and A is true, then B is true. The propositions that result from the application of inference rules are called theorems.

Because of first-order logic's rigorously formal structure, deductions are necessarily correct insofar as the initial information is correct. First-order logic is both sound (it is impossible to prove a false statement) and complete (any true statement can be proved). Thus the consistency of a knowledge base implemented in first-order logic is, theoretically at least, guaranteed. Another reason that logic-based representations have been so popular in AI research is that the derivation of new facts from old can be mechanized. Thus the validity of any new statement can be checked by attempting to prove it from existing statements. The computer language Prolog, which is based on the principles of predicate logic, has proven a flexible tool for a wide variety of knowledge representation situations. In Prolog, each step consists of a "goal" which the system tries to prove.

However, certain types of knowledge are difficult to represent using first-order logic; for example, "fuzzy" concepts characterized by relative degrees (e.g., "hot"), uncertainty and probabilities, and default values in the absence of evidence to the contrary. As Reiter points out [1978, 406], the number of negative facts about the world, even a restricted world, vastly outnumber the positive facts (e.g., penguins don't play piano, at least in the "real" world) and it is "totally unfeasible" to explicitly represent these negative facts. In addition, systems based on first-order logic are "monotonic," in that statements can either be true or false, not unknown, or true until otherwise indicated, and statements once proved to be true cannot later be retracted. A variety of nonmonotonic logics, with extensions to admit such features as default reasoning and the addition and retraction of statements in accordance with changing conditions, have been proposed by AI researchers as better frameworks for representing human-like knowledge. Unfortunately, other problems arise with such nonstandard logics, notably that the

guarantee of consistency and completeness may be lost. Systems based on nonstandard logics also tend to be less efficient, and a large knowledge base can become "an amorphous, unstructured listing of statements" [Waldrop 1984, 1280]. While some researchers have continued to believe that "a number of important features of commonsense reasoning can be implemented *only* within a logical framework" [Moore 1982, 337],³⁵ a variety of other formalisms for representing knowledge have been investigated.

3.5.2 Semantic Networks

One of the most important knowledge representation techniques in AI is the semantic net or network. A semantic network is a structure for representing knowledge as a pattern of interconnected nodes and directed arcs. The nodes represent concepts, usually entities, attributes, events and states. The arcs in the net indicate associations or relations (such as properties) between these concepts (see Figure 3-1). It is important to realize that the graphs themselves are simply a convenient way to diagram the network; the actual computer-level implementation depends on the language used. The roots of this formalism can be traced to "existential graphs" of Charles Peirce and the "assertional networks" of Gottlob Frege in the late 1800s, as well as to the graphs developed by Otto Selz, a psychologist, in the early part of this century. The modern concept for computer applications is generally credited to Ross Quillian with his Ph.D. thesis *Semantic Memory* [1966], although machine translation researchers were using network-like structures in the early 1960s.³⁶

Because of their less rigid and less formal structure, semantic networks are perhaps better suited than predicate calculus to the task of representing the kind of knowledge

35. And see Lapalme [1991] for more recent arguments along the same lines.

36. A detailed discussion of the origins of semantic networks may be found in Sowa [1991a].

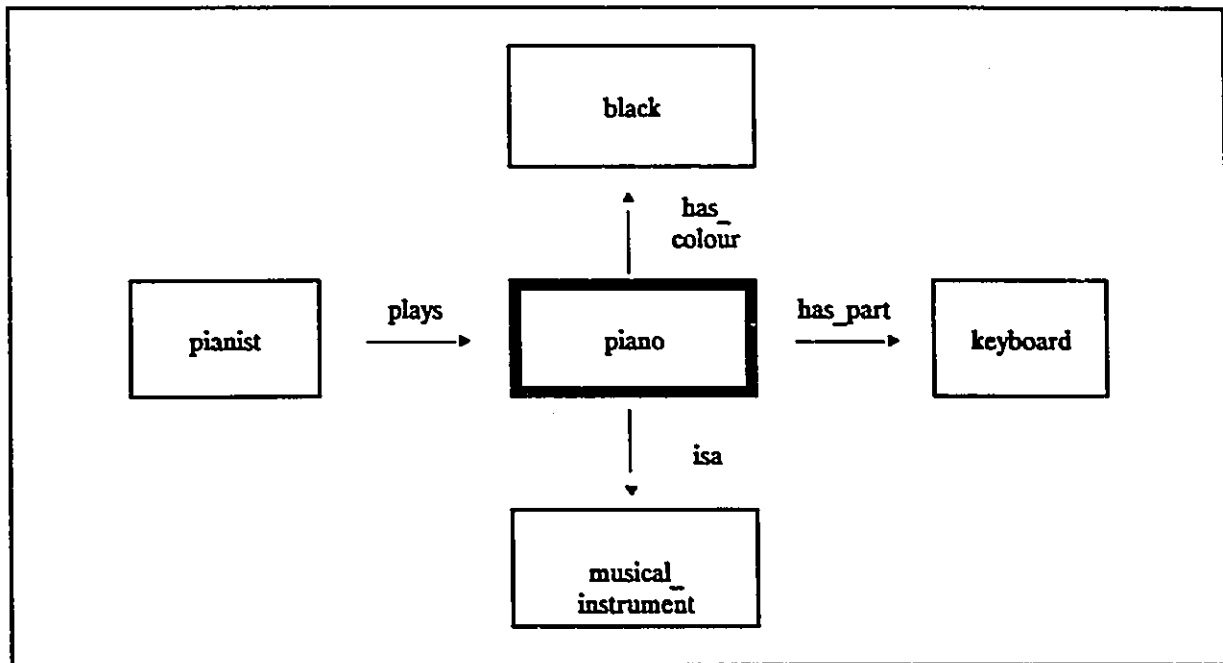


Figure 3-1 A Simple Semantic Network

expressed in natural language. One of the main uses of semantic networks in AI has been to represent the meaning of natural language, after parsing, in a deep-level (some would even claim language-independent) form. The pioneering work of Tesnière in 1930s on dependency grammar, the forerunner of valency grammar and case grammar, was an important influence, primarily on machine translation. Semantic nets were often used in question-answering systems in the late 1960s and early 1970s. In the mid-1970s, with a shift in emphasis towards more semantic-based parsing, semantic networks were used by Schank [1975] and his school to represent "conceptual dependencies" and by Sowa [1984] for his "conceptual graphs."

Semantic networks can also provide the framework for type or inheritance hierarchies. Here the nodes are concepts, and relationships that hold for all concepts of a given type are inherited from the superordinate concept through the hierarchy. In this case, the most important relation is inclusion (or generic-specific), commonly known as the "isa" relation (e.g., an aardvark isa mammal isa animal isa living_entity). A special type of "isa" relation is the instance, i.e. a specific and unique example of a particular

subtype. Another commonly used relation is partitive (part-whole), usually abbreviated as either "is_part" or "has_part." A wide variety of other relations may be allowed, such as composition ("made_of"), use or function.

Semantic networks became a very popular knowledge representation scheme because they were relatively simple to conceptualize and implement and because they were very flexible. They allowed important relations about objects to be shown explicitly and succinctly. This very flexibility was one of their weaknesses, however. An influential article by William Woods [1975] pointed out the weak formal foundations of semantic networks and prompted a re-examination of their theoretical underpinnings. Barr and Feigenbaum were compelled to note in 1981 [180] that "the superficial similarity of ... notation is all that some semantic network systems have in common." More recently, Schubert argued that the term "semantic network" has become almost meaningless [1991, 95]. One of the most troublesome aspects of representing commonsense knowledge in semantic nets remains the handling of exceptional and conflicting information [Shastri 1988, 7]. The inherent problem of semantic nets from a logic point of view is their lack of formal semantics, so that the inferences made through the net are not guaranteed to be valid and consistent the way they would be under a formal logic framework.³⁷ Nevertheless, the influence of semantic nets can be detected in many current systems.

3.5.3 Frames

Another important knowledge representation concept is "frames." First proposed by Minsky in a seminal 1975 paper, a frame is simply "a data structure for representing a stereotypical situation" [1975, 246]. In Minsky's original formulation, the top level of the frame comprises fixed attributes and the lower level has a variety of "slots" that can be filled according to certain criteria. These slots frequently have associated default values

37. On the other hand, Minsky has argued [1975, 260] that consistent systems tend to be weak systems.

Piano frame		
	isa	musical_instrument
	has_part	keyboard (1) legs (4)
	colour	black (default)
	brand	one of: Steinway Yamaha Bösendorfer

Figure 3-2 A Simple Frame

that may be modified. Significant deviations from default values may serve as an attention focuser. In addition, slots may have associated procedures that are triggered as necessary to determine the appropriate values.

For example, a frame for a piano might include slots for its type (upright, baby grand, concert grand, etc.), make (Steinway, Yamaha, Bösendorfer, etc.), colour, condition, location, and so on (see Figure 3-2). The slots are "filled" (i.e. given values) according to available information or through inheritance, with defaults stipulated in the absence of other information (for instance, a concert grand might be assumed to be black unless otherwise specified). Frames model the intuitively attractive idea that people possess repertoires of structures based on their previous experiences. In other words, they do not start from scratch when faced with new situations, but have expectations about what they will encounter. Frames may be related to one another, usually in a hierarchical manner, to form frame systems.

Frames have their problems, too. Perhaps the most important is identifying *which* frames should be used, *when* they should be used, and when they should be *discarded*.

For example, if a piano drops on a cartoon character, is it necessary to "activate" the piano frame? The problem of foreseeing all possible slots and values for all cases has also been frequently pointed out (for instance, a piano could happen to have two keyboards). Consequently, frames cannot hope to truly encapsulate a significant proportion of common-sense knowledge. The frame formalism has also been criticized as incapable of supporting composite descriptions based on the interrelation of parts [Firebaugh 1988, 289].

Nevertheless, frames have proven a useful, flexible and computationally efficient means of knowledge representation. One important spin-off of the frame concept was the "script," basically a frame describing a stereotypical *sequence* of events, like a visit to the doctor. However, scripts suffer from many of the same problems as frames. Other higher-level structures have been proposed and developed, such as "plans," where the emphasis is on constructing and anticipating, and "memory organization packets," which mediate and monitor script activation, as well as lower-level subroutines such as "goals."⁴⁰ In fact, the notion of frames has evolved considerably from Minsky's original concept. As Hayes pointed out [1979, 294], "the *real* force of the frames concept was not at the representation level at all, but rather at the implementational level: a suggestion about how to organize large memories." Today, so-called "frame-based" systems are usually knowledge representation systems using frame-like structures with slots and inherited default values for organizing information about individual concepts that are themselves organized in type hierarchies.

In fact, in many respects the strategies for knowledge representation discussed above – formal logic, semantic networks and frames – have much in common. As Sowa noted [1991a, 45], "During the 1980s, the boundary lines between network, framelike, and linear forms of logic tended to disappear." Modern knowledge representation systems

40. The major figure in these developments is Schank; see, for example, Schank and Abelson [1977].

may well use frame structures for concepts that are themselves organized into a semantic network hierarchy, with the entire system written in a predicate logic language like Prolog or Lisp. Cyc, for example, is a frame-based system written in a representation language based on Common Lisp. In any event, one of the essential preliminary steps in constructing all such systems is conceptual analysis.

3.6 Conceptual Analysis

3.6.1 Concepts

Today it is generally accepted that human beings make sense of the world around them — the continuum of reality — by constructing systems of concepts that focus on specific aspects of that continuum. Beyond that, debate and questions abound: what concepts are, how they develop, their differences and similarities within a single linguistic-cultural community, their necessary and associative criteria, their relationship to language and to the perception of reality, and so on.

In the classical view, as expounded by Aristotle and others, a category was like a bucket containing objects that all shared certain properties; any given object could be placed either inside or outside the bucket on the basis of its properties. In other words, the properties of a concept are singly necessary (every member of the concept must have the property) and jointly sufficient (every entity with those properties must be a member). As mentioned earlier in this chapter, the quest for an ultimate set of defining properties or primitives proved quixotic. Yet the classical view of categories or concepts has dominated Western philosophy and science for 2000 years. Sometimes termed objectivism, it heavily influenced the "strong" version of AI, because if classification is essentially a matter of "yes, it belongs" and "no, it doesn't," it is a process that can be

automated. Any straight hierarchical structuring of concepts implies a degree of objectivism.

But the influence of objectivism has been waning in the face of research and experimentation indicating that human subjectiveness is integral to the process of categorization. "Most human concepts have imprecise, context-dependent meanings," observed Michalski [1987, 50]. "Such concepts enable people to express a great variety of meanings by fewer words, and this serves cognitive economy." Contrary to the classical view, some entities turn out to be "better" examples of a category than others and can be considered as conceptual prototypes: "Categories tend to become defined in terms of prototypes or prototypical instances that contain the attributes most representative of items inside and least representative of items outside the category.... Most, if not all categories do not have clear-cut boundaries" [Rosch 1978, 30ff]. Indeed, the boundaries of even the most apparently precise concept rapidly dissolve into fuzziness upon close observation. Labov's experiments with such allied concepts as "cup" and "mug" are a good illustration [Labov 1973]. Hofstadter's introspective analysis of the concept of the letter "A" led him to observe that there is no specific set of features for "A-ness"; there instead "lurks a concept, a Platonic entity, a *spirit*" [Hofstadter 1985, 279 (original italics)]. If concepts are inventions of the human mind that focus on certain features of the world at the expense of others, then a model of the world (or any piece of it) can never be perfect, since the concepts it employs are abstractions that necessarily gloss over many details and complexities. At best, they are workable approximations.

Yet it should be realized that there is no real incompatibility in dealing with the world using discrete categories when the underlying concepts are fuzzy at the edges. It is a question of what AI people would call the "grain size." In fact, it can be argued that a conceptual view of the world is a prerequisite for survival. People could not begin to cope with the complexity of the real world if they did not make an enormous number of conceptually based default assumptions; otherwise, they would be constantly

overwhelmed by all sorts of theoretically possible but unlikely exceptions to general rules.

The conceptual approach used by human beings is actually an extremely efficient and logical method of dealing with the real world. Concepts offer at least three major advantages [Smith 1989, 501f]: they promote cognitive economy, decreasing the amount of information that must be processed and retained; they allow past experience to be brought to bear on present circumstances, which are never exactly the same as, but at least partially similar to, previous experiences; and they permit inductive inferences in the absence of complete information.

Unsurprisingly, the very nature of human-style concepts renders them difficult to represent in computer form. There is no finite set of definitive properties, no precise template, against which membership in a conceptual category can be judged. Moreover, most complex human concepts are not derived by simply combining more elemental concepts; for example, the concept of a good piano cannot be adequately described as the conjunction of the concepts good and piano. In fact, it is quite possible to say "This piano is a good table" when sharing a post-concert snack of wine and cheese.³⁹

Johnson-Laird notes that "the rules for combining interpretations operate, not on the referents of expressions, but on their meanings — and in ways that are not properly understood" [1988, 337]. Nevertheless, a variety of schemes (such as predicate logic, frames, and semantic networks) have been developed that make it feasible to implement conceptual networks for capturing knowledge in computerized form. The problem becomes deciding the internal structure of the concepts and their organization within the system — in other words, what properties should be represented and how inheritance should operate between the concepts. The process of making such decisions is called conceptual analysis.

39. More precisely, this reflects the fact that the concept good is what George Miller [1978, 306] calls "syncategorematic," in that its meaning depends on what it modifies.

3.6.2 Conceptual Analysis and Knowledge Acquisition

"The ultimate goal [of conceptual analysis] is a precise, formalizable catalog of concepts, relations, facts, and principles" and the product of conceptual analysis is "an ontology for a possible world - a catalog of everything that makes up the world, how it's put together, and how it works" [Sowa 1984, 344]. As noted earlier, efforts to organize and characterize knowledge into categories are at least as old as Aristotle. Sporadically there have been attempts to classify the sum total of human knowledge in this fashion. Roget's famous thesaurus is a good example (and some researchers have recently returned to Roget's system as the basis for a computerized knowledge representation model). Today, as Sowa points out [1984, 294], several disciplines practise conceptual analysis under different names. In particular, conceptual analysis has become a central focus of terminology.

In AI, the process of collecting, analyzing and formalizing the essential pieces of information in a particular subject field is called knowledge acquisition and those who practise it, knowledge engineers. The information is typically collected through direct interaction with a field expert. The end product is a representation of the knowledge in some kind of language, knowledge base or expert system. The knowledge engineer must be skilled in choosing his questions and analyzing the responses in order to extract this information, since the expert usually has little appreciation about the nature and extent of his own knowledge. Whether he calls it such or not, the knowledge engineer is carrying out conceptual analysis throughout this process. This subject, which is often glossed over in books on AI and expert systems, is crucial to the success of knowledge-engineering projects, since the conceptual analysis phase largely determines the internal structure of the knowledge base and is a key factor in its quality.

A good knowledge engineer should be an expert in logic, language and human reasoning. His primary goal is to elicit and identify the essential facts and rules that

express information about the subject area. Facts are statements that are known to be true, false or to some extent probable, and rules, which can often be phrased as "if X then Y," codify general principles that relate these facts. As the knowledge engineer goes about his task, he is alert to at least five kinds of information: hierarchical (e.g., X is a kind of Y, X is a part of Y), defining (e.g., X always has the property Y), constraints (e.g., the value of property Y of X must be Z), associations (e.g., X is made of Y), and behavioural (e.g., X does Y if Z). He also records any useful background information.

Conceptual analysis techniques have tended to be idiosyncratic and ad hoc. Sowa [1984, 297ff] lists a number of strategies and practical tests that can assist the process, including brainstorming, free association, and co-occurrence patterns (i.e. phrases or structures shared by a set of words). Other authors have suggested such techniques as drawing concept maps, imagining one is teaching the concepts, and story telling. Certain more formal techniques have also been developed. For instance, binary features analysis involves constructing the smallest set of true-or-false questions that distinguish between a group of concepts; it is similar to linguistics' "distinctive features" and terminology's "semantic components" or "descripteurs." Another method known as conceptual clustering measures the similarities between objects and groups of objects according to a set of attributes to determine "clusters." A similar technique involves measuring the similarity between an instance of the concept and its prototype. Some of these methods are "top-down" or "deductive" (working from larger conceptual structures and filling in the details), some are "bottom-up" or "inductive" (working from details and building structures to accommodate them), and some combine the two approaches. On a related topic, Lenat and Guha [1991] provide a list of "mistakes commonly made when knowledge is entered."

Once the conceptual analysis is complete (or, more typically, as the conceptual analysis is being done), the knowledge is mapped into a formal knowledge representation language (here "language" means a specialized computer language, or a knowledge-

representation scheme such as a frame-based system). There is often more than one way to represent knowledge, and the knowledge engineer has to make choices on the basis of such considerations as the purpose of the system and efficiency. As Sowa notes [1984, 302], the results may sometimes seem simple, but finding the right set of features may take months or years of analysis.

3.6.3 Conceptual Analysis and Terminology

The modern science of terminology "deals principally with concepts, conceptual relations, systems of concepts and in this connection also with terms and systems of terms, as well as symbols and other systems of symbols for the representation of concepts" [Galinski 1988a, 486]. Thus it too is concerned with conceptual analysis, and it has what Ahmad et al. [1989, 2] have called "a symbiotic relationship" with knowledge engineering. The links between terminology and knowledge engineering have been explored by other authors, notably Meyer [1993], and it has been suggested that terminologists have something to offer designers of expert systems [Parent 1988].

Galinski observed [1988b, 494] that there are currently two basic approaches to terminology. The first focusses on individual terms in a specialty language and is linguistically oriented. The second focusses on concepts and systems of concepts in a subject field and is knowledge-oriented. Such work is the heir of the General Theory of Terminology founded by Eugene Wüster, which is based on the idea that the purpose of a terminological record is to store information about a concept and its relations with other concepts.⁴⁰

As Picht and Draskau [1985, 36] note, while there are differing views on the details, there is widespread agreement on the importance of the concept. The conceptual

40. For a description of the work and contribution of Wüster, see Picht and Draskau [1985, 27ff].

approach in terminology always treats the concept in the context of a network of concepts connected by logical or associative relations. Traditionally, the two main conceptual relations in terminology were generic-specific (corresponding to the "isa" link in AI) and partitive (X is a part of Y). Other relations only occasionally appeared. In the main, this approach has given way to an open-ended set of associative relations dictated by need [Sager 1990, 29]. Generic-specific relations still tend to dominate, however, because they combine inherent intuitiveness and simplicity with considerable expressive power. It is now recognized, however, that in any division of a superconcept into subconcepts there is an explicit or implicit criterion for this division [Sager 1990, 36], and that such a division influences the properties that are assigned to the concept subsystem. What this implies is that there is usually more than one way to classify the members of any given superconcept, a phenomenon that terminologists call "multiple partitioning of reality" or "multidimensionality" [Meyer and Skuce 1992; Bowker 1992].

As in knowledge acquisition, the techniques used for conceptual analysis in terminology are primarily intuitive. Again, the two basic strategies are top-down or "onomasiological" (creating the system of concepts and assigning terms to them) and bottom-up or "semiasological" (starting with the terms and building a conceptual structure for them). In actual practice, a combination of the two strategies is usually employed, in what Parent [1989, 412] calls "un mouvement dialectique." Some general principles for conceptual analysis have been offered. For instance, Picht and Draskau [1985, 64] stress clarity, intelligibility and transparency. On a more practical level, Auger and Rousseau [1978, 30f] list a number of general objects that frequently correspond to "unités terminologiques" in technical fields (phenomena, devices, procedures, etc.).

What is striking in all these efforts in both AI and terminology is how heavily the process of conceptual analysis relies on "paper-and-pencil" and "do-it-all-in-the-head" approaches. Sager [1990, 9] observes that "in this area in particular [conceptual systems], conventional thinking is still largely dominated by pen and paper processing techniques."

Sowa himself proclaims [1991, 178], "before any technique can be automated, the tool developers themselves must know how to use it with paper and pencil." What is clearly needed is a computer tool for conceptual analysis, a spreadsheet for ideas, a system for managing knowledge.

Such a tool exists, and it is the focus of the next chapter.

CHAPTER 4

A Knowledge-Management System, a Sample Knowledge Base for Translators and an Informal Experiment

The purpose of this chapter is to introduce a knowledge-management system known as CODE (Section 4.1) and to describe a knowledge base developed by the author using CODE (Section 4.2). The knowledge base, called optionCODE, is intended to illustrate the potential offered by such systems for translators. Some methodological issues are discussed in Section 4.3. The final section of the chapter reports on an informal experiment that was conducted to test how useful CODE and optionCODE could be to translators in their present form.

4.1 CODE: A Knowledge-Management System

4.1.1 Overview

CODE (for *Conceptually Oriented Description Environment*) is a general-purpose knowledge-management system developed by Douglas Skuce at the Artificial Intelligence Laboratory of the University of Ottawa. Written in Smalltalk-80 (an object-oriented language), it can run on a UNIX workstation, or on a sufficiently powerful microcomputer (a 386-based IBM compatible or a high-end Macintosh). CODE has been under development for several years, and Version 4 of the system is currently undergoing testing. Version 2 was used for the purposes of this paper, however, because

it is a thoroughly tested and fairly stable system. This is the same version used for some current research into the terminological applications of knowledge-based systems.⁴¹

The following description of CODE is necessarily very cursory, touching only upon the main features of the system. Further information may be found in Skuce, Wang and Beauvillé [1989] (Version 2) and Skuce [1993] (Version 4).

The central notion of CODE is the *concept*; in fact, CODE has been described as a "spreadsheet for concepts" [Skuce and Meyer 1990, 188]. The concepts correspond to the labelled nodes of a semantic network. The system imposes no preconceived theory on the user and offers a great deal of flexibility in the internal structure of concepts and their inter-relationships. Concepts in a CODE knowledge base are generally arranged in an inheritance hierarchy of superconcepts and subconcepts. For instance, a knowledge base about penguins could have a concept *penguin* with a number of subconcepts representing the various species of penguin (*emperor penguin*, *king penguin*, *little blue penguin*, etc.). Concepts are defined according to their *properties*, which are themselves grouped into meaningful property *categories*. The penguin knowledge base, for instance, might include a property category "Physical appearance," which could include such properties as "colour," "size" and "number of feet." A property consists of two main parts: its name and its body (the value or expression specifying it to some level of detail). For instance, a possible body (value) for the property "size" could be "small." Explanatory comments can also be attached to properties. The properties of concepts are inherited through the hierarchical network from superconcept to subconcept according to defined rules (controlled by the inheritance *flags* assigned to each property). Each of the three physical appearance properties of penguins mentioned above, for instance, would probably have different inheritance flags. In the penguin superconcept the property "colour" would be assigned the body "black and white" and that value would be

41. Notably, the "COGNITERM" project. See Meyer et al. [1992a].

automatically inherited by all penguin species unless explicitly overridden (i.e., a default value). This default value would have to be changed in the case of the emperor penguin, which is basically blue-gray and white. The property "size" would be inherited to subconcepts with no body specified, since penguin species vary considerably in size; the value for the property would be indicated separately for each subtype. The property "number of feet," however, would have the value of 2 for all penguins; in fact, this property might have been originally inherited from a superconcept *bird*, of which penguin is a subcategory. Note, however, that there could be an *instance* of a concept (i.e., a particular individual) that violates the defined value for this property; for example Percy, a penguin who earlier lost one foot to frostbite. The inheritance rules governing a property may have to change in the case of instances. CODE allows inheritance to be overridden, forced or blocked if necessary in a particular CD.

The bundle of properties that collectively define a concept is grouped in a *conceptual descriptor* (CD), which is analogous to a frame in artificial intelligence. The term CD is often used by CODE users to refer to the concept and all its associated properties.

The basic purpose of CODE is to serve as a knowledge acquisition and retrieval tool. It allows a user to store and manipulate knowledge, to analyze and organize it for specific purposes, and to retrieve it in a hypertext-like, graphical environment. Multiple windows may be opened, closed, or "collapsed" (temporarily closed), and the user can jump from one to another with a simple mouse click. Changes made in one window that directly affect the knowledge base are immediately reflected in all other windows.

There are two basic modes of interacting with CODE. The first is through the *browser*. In this mode, all concept names or a selected subset are listed in either hierarchical or alphabetical order. Selecting a particular concept causes the browser to display that concept's property categories and their associated properties. Selecting a particular property calls up a variety of information about that property, including its

Objectworks [read] [edit] [del] [copy] [paste] [undo] [redo] [print] [quit] [help] [version] 2.5 [file] [edit] [view] [print] [quit] [help] [version] 2.5

synthetic strategy in "pp" format [read] [edit] [del] [copy] [paste] [undo] [redo] [print] [quit] [help] [version] 2.5

cdName: option
super: derivative security
hasPrepOf: SPECIAL
kinds:
(1) put, call
(2) European option, American option
(3) exchange-traded option, over-the-counter option, subConcepts: stock option, American option

creation date (option) : 14 August 1991
done by (option) : dave
last update date (option) : 19 August 1992
level (option) : 4
status (option) : nh

DEFINITION
definition (option) : An option is a formal contract which grants the holder of the contract the right but not the obligation to buy or to sell, as the case may be, a certain quantity of a specific underlying interest (i.e., a security, currency, commodity or stock index) at a stipulated price for a stated period of time. There are two types of options: a call which gives the buyer the right to purchase an underlying interest and a put which gives the buyer the right to sell an underlying interest. The option holder (buyer) is not obligated to exercise the option and will not do so unless it is advantageous. If the option is not exercised or sold by the option holder

ENCYCLOPEDIA INFORMATION
class (option) : All options on a particular underlying interest: French class (fr) (Lavaursur, p. 16)
duration (option) : a span of time [q.v.]; for exchange-traded options, a multiple of 3 months. [Sometimes referred to as the life of the option or expiration period; french words (fr) (de vie) (August, 63); période (fr) d'exercice (Madsen, 98)
exchange-traded? (option) : yes
exercise date (option) : an exercise date [q.v.]
for a European option [q.v.], the exercise date = the expiry date [q.v.].

UNDERSTANDING INFORMATION
English collocations (option) : Options are sold to be (on) an underlying interest.
Options may be exercised up to their expiry date.
If not exercised, options expire on their expiry date (not mature).
Options are traded -- i.e., sold or written and bought or purchased -- on stock exchanges, and thus may be referred to as exchange-traded option or listed options.
English synonym (option) : option contract [but see property "underlying interest" in CD "stock option"].

SUPPLEMENTARY INFORMATION
example of option quotes (option) : (1) DOME JA 32 1/2
(2) COMWCO FE 24 P
(1) Call option on Dome Petroleum with an exercise month of January and an exercise price of 32 1/2.
(2) Put option (indicated by "P") on Combe with an exercise month of February and an exercise price of 24.
history of options trading (option) : Organized call option trading began in the United States with the creation of the Chicago Board Options Exchange [v. Options Clearing Corporation] in April 1973. This led to trading of call

cd > category > property > category > property > cd

AI
Definition
Encyclopedic information
English synonym
French collocations
French collocations
Supplementary information
observations

concrete entity
security
tangible security
equity security
common stock
preferred stock
debt security
capital market security
bond
secured bond
unsecured bond
note
money market security
treasury bill
short-term government
bankers acceptance
certificate of deposit
commercial paper
securities contract
derivative security

put
call
European option
American option
exchange-traded option
over-the-counter option
stock option
fixed-income option
currency option
stock-index option
gold or silver option
futures option
forward contract
commodity futures contract
financial futures contract
swap
interest-rate swap
warrant
subscription right
convertible security
convertible bond
callable convertible bond
convertible preferred stock

actor
agent
broker
investor
buyer
seller
institution
options clearing house
Trans Canada Options Inc

restrict

alpha

main | location | view

formats: option
super: derivative security
kinds:
(1) put, call
(2) European option, American option
(3) exchange-traded option, over-the-counter option
subConcepts: stock option, fixed-income option, currency option, stock-index option, gold or silver option
[Source: T
[Source: option
Comments:
PrepDesc:
User: dave
Status: in progress
Category: Linguistic information
Name: french equivalent
Number: 2
option (fr), contract (en) d'option (TERMINA)

sternLogin

Figure 4-1 CODE Screen Dump Showing Browser (Right) and Selected CD View (Left)

value. Most of the conceptual information displayed in the browser can be directly edited, and the results take effect immediately. Individual concepts and individual properties of concepts can be revised, created, destroyed, etc. A *CD view* can be quickly opened in order to display all the system's knowledge about a concept at a glance (see Figure 4-1).

The second main way of using CODE is through a graph. The concepts in the knowledge base may be displayed in a variety of graphical formats: the normal hierarchical graph showing all active concepts, a subtree graph showing a particular concept and its subconcepts, a detailed graph showing a concept with its immediate superconcepts and subconcepts, and a property graph showing the concept and any other concepts associated with it through property links. On the full graph, particular nodes (concepts) or subtrees can be singled out for attention in a separate graph or can be temporarily hidden to allow the user to focus on one part of the network. Graphs can be moved and scrolled at will, and the placement and display of concepts can be easily changed. Graphs are editable; for instance, nodes and subtrees can be moved from one branch to another, and concepts can be created or destroyed. These changes are immediately reflected in the knowledge base. (See Figures 4-2, 4-3 and 4-6 for examples of CODE graphs.)

CODE draws upon the three knowledge-representation formalisms discussed in Section 3.5. As noted earlier, concepts or CDs can be regarded as the nodes of a semantic network. In as much as they are bundles of properties with associated values or bodies, individual CDs also correspond to frames, with slots to be filled. Finally, the system also has a logic component in the form of an associated first-order logic deduction system that can perform logical operations on CDs, such as the imposition and verification of constraints on properties. Property inheritance is also logic-based.

4.1.2 Advantages of CODE for Translators and Other Users

CODE has proved to offer a number of advantages for terminological work.⁴² Among these are the graphical display, which allows terminologists to have a readily customized "picture" of the domain; the ability to partition reality in more than one way; the hypertext-like browsing capability; the flexibility offered in specifying concept properties and relations between concepts; and CODE's automatic signalling of logical inconsistencies in the knowledge base. Several of these features are also of particular advantage to translators using CODE and a CODE knowledge base as a knowledge-retrieval tool.

It was argued in Chapter 1 that knowledge-base systems potentially offer three main advantages to translators: they are richer sources of knowledge, they offer more flexibility and more efficient access to the information they contain, and they are inherently open-ended. The following section will examine how CODE fares generally under each of these categories. Some more detailed observations in this regard will be offered in the analysis of the field experiment.

4.1.2.1 Flexible and Efficient Access to Knowledge

As discussed above, there are two main ways of interacting with CODE: through the graph and through the browser — and each of these modes itself offers considerable flexibility. In the browser environment (see Figure 4-1), concepts may be listed alphabetically or hierarchically. In the latter case, indenting serves to indicate the hierarchical structure. Once a relevant concept is located alphabetically, for instance, a single mouse click displays a hierarchical listing for the selected concept, showing all its superconcepts and subconcepts. At any time, the user may specify that further

42. Discussed in detail in Skuce and Meyer [1990 and 1991].

operations are to be restricted to the current list. A mask can also be applied to the information displayed in the browser to restrict the CDs, property categories, properties listed (among other things). A "CD view" can be opened at will to provide a formatted listing of the information for a particular concept, and the user may specify whether it is to retain its current information or dynamically update as other CDs are selected. A CD view may also have a mask associated with it.

In the graph environment (see Figure 4-2), the user may move easily around the graph, hide parts of the tree not of interest, and display detailed graphs of certain concepts or subtrees in separate windows. A "mask" may be applied to the current graph in order to further restrict the display in accordance with a wide range of criteria. Masks may be easily modified or cancelled at any time.

CODE also makes it easy to coordinate the information displayed in the graph and the browser. While working with the graph, the user may ask for any node (representing a concept) to be automatically selected in the browser. Conversely, in the browser environment, the user can ask for the currently selected concept to be centred in the open graph. Any of the various types of graphs may also be opened for the currently selected concept directly from the browser (see Figure 4-3).

It is important to realize that any number of windows can be opened at any time, each displaying a particular part of the network in a particular format. It is quite possible to have multiple graphs, browsers and CD views open simultaneously. Each of these windows may be quickly moved, reshaped or closed, and may also be "zoomed" to fill the entire screen (and "unzoomed") or "collapsed" to an icon to be reopened at will. In addition, windows may be opened from the operating system, so that, for example, a text editor window can be available for the translator to type his work in. Any open window (including one "collapsed") becomes instantly active with a mouse click anywhere in it. The system also allows text to be cut, copied and pasted between windows.

All of these facilities add up to an extremely flexible knowledge-retrieval environment for the translator. He has a variety of strategies available to enable him to zero in on the area of the knowledge base in which he is interested, and, once he has located it, he has an extremely wide range of ways to tailor what information is displayed and how it is displayed. He can jump easily from one window to another and, as discussed above, can coordinate the display in one window to reflect that of another.

Because of this flexible, hypertext-like environment, a CODE knowledge base is ideally suited to browsing. Once the translator has located a concept that is relevant to his information needs, he can explore the particular subtree or follow property links leading to other parts of the network. As Galinski and others have pointed out, a conceptually based knowledge base is an excellent environment for knowledge retrieval:

Via the concept and its conceptual relationships to other concepts the user is guided to the individual units of knowledge and information which he is looking for. Thus information loss is kept a low level. At the same time "noise" (i.e. false or ambiguous information resulting in irrelevant or even misleading responses) can be kept at a controllable level [Galinski 1988c, 177f].

As it now stands, however, CODE is far from ideal on certain points. The most obvious shortcoming is the lack of an index or "global lexicon" that can efficiently locate all occurrences of a particular word or phrase. A translator faced with the task of translating a text in a relatively unfamiliar subject area will typically want to look up terms in the source language, not only to find their target-language equivalents, but also as a logical pointer to the relevant concepts in the knowledge base. Version 2 of CODE does not make such a request easy to carry out. The best existing method appears to be masking the browser to restrict the display to those CDs matching the appropriate property name and body. Once this is done, cancelling the mask condition allows the selected CD to be seen in its proper listing relative to other CDs. Unfortunately, this operation is tricky to carry out repeatedly, prone to errors, and slow. At a minimum, the

information contained in whatever property is designated to hold the source-language equivalent should be indexable.

Another shortcoming of CODE is that its current interface has a steep learning curve. It takes several intensive sessions with the system before a novice user can hope to gain any real facility in its architecture and use. As an experimental version, it still has unpredictable bugs and problems that can be vexing to the user.

4.1.2.2 *Richness*

The extensive flexibility offered by CODE for searching and displaying information adds to the richness of its knowledge bases. The ease of browsing means that any piece of relevant information located by the user is supported by a knowledge network of related properties within the CD and hierarchically related concepts outside it. The knowledge stored in CODE actually has what might be regarded as three dimensions; the first dimension is the knowledge local to the concept; the second the knowledge related to the hierarchical structure of generic-specific ("isa") relations, primarily the information inherited from hierarchically related concepts; and the third the knowledge provided through the various associative relations with other concepts via property links. These properties and associative relations are very flexible and can be defined by the knowledge engineer/terminologist developing the knowledge base to be as useful as possible for the field and the purpose of the knowledge base.

Another consequence of CODE's repertoire of ways to access and view information is that the same knowledge can be viewed in several different ways and different levels of detail simultaneously (e.g. in a browser, a CD view and a graph). This adds depth to the information.

Multiple partitioning of reality has been a central problem of terminological methodology. The fact that CODE allows knowledge to be partitioned in alternative ways also adds depth. Besides the standard generic-specific link (shown as an "s" link on graphs), CODE has another built-in subconcept relation called "kinds" (represented on a graph as a "k" link). This link represents a "partition" in the strict mathematical sense of the term: a division into subconcepts that covers all possibilities, so that each member of the class is necessarily one of the specified "kinds."⁴³ In our penguin knowledge base, for instance, the concept **penguin**, besides having a number of subconcepts representing the various species of penguins, could also be partitioned into the concepts **male penguin** and **female penguin**, because each penguin, no matter what its species, is either male or female. Multiple partitions can be most informative for the translator (and other users), since they can capture potentially important information that may be only implicit in a text.

The fact that the translator is using the same system to interrogate the knowledge base as was used to develop it is significant. Using CODE in the development process forces the developer to carefully consider and clarify the conceptual framework. CODE has a number of built-in consistency checks that prompt the developer when inconsistencies are detected. As a result, the user (the translator) has a better guarantee of conceptual clarity and consistency in the knowledge base than may be typically found in expert reference sources.

On the negative side, the implementation of associative links in CODE needs improvement in order to fully exploit the "third dimension" mentioned in the first paragraph of this section. At present, the system only explicitly recognizes associative links when a property body corresponds *exactly* to another concept name. For instance, if the concept **penguin** had the property "main diet," which had the value "fish," an

43. In more technical terms, a partition is a mutually disjoint and exhaustive division.

associative link with a concept **fish** would be automatically established by the system. If, however, the value entered into the "main diet" property were "fish, crustaceans," no link would be made with either **fish** or **crustacean**. Ideally, a true hypertext approach could be taken to associative relations: the existence of an associative link could be signalled by some convention (such as the concept name in boldface) and a mouse click on this "button" would suffice to take the user directly to the corresponding CD. These links could be displayed on a property graph in a semantic network format: a series of arcs labelled with the property name.

To fully exploit the inherent richness of a CODE knowledge base, the user interface is again a crucial consideration. A front-end query system could perhaps eventually be developed to interpret user requests. It could prompt for clarification when necessary and also interpret system responses for the user. A sophisticated query module would ideally allow the user to carry out searches with incomplete information, permitting questions of the kind "what do you call a thing that does X and has Y and Z as properties?"

4.1.2.3 *Open-Endedness*

As discussed in Chapter 1, open-endedness refers to the possibility of easily expanding and modifying the knowledge base, of using the knowledge it contains in novel ways, and of sharing that knowledge.

CODE's flexible knowledge structure makes it applicable to a wide variety of applications. The fact that it has been successfully used for terminological research and for systems design testifies to its adaptability. Once a knowledge base is created, it is easy to add new concepts and properties to it. In our penguin knowledge base, if a new species of penguin were suddenly discovered, it could be easily added as another subconcept of the concept **penguin** and would automatically inherit all the system's

knowledge about penguins; only the exceptional or distinguishing information would have to be added to the new CD. If it became necessary to add another property or even an entire property category to the penguin knowledge base, it would suffice to add the new information to the penguin CD and it would be automatically propagated throughout the appropriate subtrees.

CODE stores its CD information in the form of standard ASCII files, so the possibility exists of using that information for other purposes. For example, another database or knowledge base that can read ASCII files could import the information in CODE files for its own purposes. In addition, selected CD information can be written out as ASCII files via an option in the browser. With the appropriate selection of CDs and browser masking, it is possible to produce ASCII listings in a terminological record format or in a glossary style with definitions and other-language equivalents (see Appendix III). While this feature is not fully developed, CODE can create a kind of natural-language description for a concept on the basis of its hierarchical and property information.

Two subsystems are included in CODE but not fully developed. The first, the ClearTalk parser, can check CD information for syntactical and lexical compatibility with ClearTalk, a simple English-like language developed by Douglas Skuce,⁴⁴ and with a user-defined lexicon. This illustrates the possibility that a CODE knowledge base could be used to help ensure terminological and grammatical consistency in a documentation project. The second subsystem is a full first-order logical deduction system written in Prolog that can parse and debug sets of rules and facts. Such a module could be further developed in order to expand CODE's logical consistency checks to take into account constraints on the values assigned to properties. This could be of benefit in a large systems design project, for instance, where logical consistency is paramount.

44. See Skuce [1992].

The utility of CODE for terminological work has been extensively explored by Meyer and Skuce [1992]. From a translator's point of view, there are a number of advantages connected with using a knowledge base developed by terminological experts using CODE, including a better guarantee of consistency of structure and terminology. This makes a CODE knowledge base a potentially ideal resource for other kinds of writing as well, such as technical writing and system documentation. There is also the potential for machine use of the knowledge base. A CODE-like system would be ideally suited to capturing a variety of commonsense knowledge, such as might be required by a machine-translation system or a database query system.

The foregoing discussion illustrates that a knowledge-management system like CODE has the potential to be adapted to a variety of other applications. Unfortunately, this potential remains mostly latent for the moment. Part of the problem is the general lack of standards for knowledge-base design, so that there is little possibility of sharing information between different knowledge-based systems. It is to be hoped that this is a situation that will change.

4.2 Example of a Knowledge Base for Translators: OptionCODE

4.2.1 Overview

OptionCODE is a CODE knowledge base developed by the author to explore the principles and problems of designing knowledge bases for translators. It contains about 200 concepts (or CDs) of relevance to stock-market options. More specifically, while its full domain extends to stock exchanges in general, its *focus* is on options. The purpose of this knowledge base is to provide a generalist or even a technical translator with the knowledge he would need to translate from French into English texts concerning options and options trading. It contains definitions, linguistic information (terminology,

synonyms and collocations in both English and French, as well as observations about usage), encyclopedic or conceptual information (the salient characteristics of concepts and their links to other concepts), and occasional background information that might be of use to the translator.

Options are a type of security or investment that can be bought and sold on several major North American stock exchanges. They are formal contracts granting the holder the right but not the obligation to buy or sell a certain quantity of an underlying "interest" — typically a security such as a stock or bond, but also a currency, commodity, or stock index — at a stipulated price for a stipulated period of time. A *call* option gives the buyer the right to purchase and a *put* option the right to sell. Option transactions are often combined in a variety of ways to create investment strategies aimed at maximizing profits while limiting risks.

This subdomain was selected as suitable for modelling in a knowledge base for a number of reasons. First, the author had already had first-hand experience with translating texts in the field and with the task of acquiring expert knowledge from traditional sources in a short period of time. Second, it appeared to be a fairly well-circumscribed domain that, because of its degree of specialization, constituted a relatively autonomous subworld. Third, it was a domain that had some substance, in the sense that options played a significant role in investment trading domestically and internationally and were a subject of some interest in the technical and theoretical literature on financial products.

A number of caveats should be made in offering optionCODE as an example of a knowledge base for translators. First, a translator would have to know how to use CODE efficiently, a skill that would be gained only through instruction and considerable experience. Second, optionCODE conforms to the subworld assumption; i.e. that the world of the stock market is the universe and nothing is known about what lies outside.

This division is not watertight, however, since a few higher-level (i.e. general-knowledge) concepts are also included in the knowledge base (this issue is discussed in more detail in Section 4.3.3). Third, no claim is made to completeness and consistency. Not only are there undoubtedly many important concepts (and, in fact, whole conceptual subtrees) missing, but the level of detail throughout the knowledge base is not consistent. (The question of knowledge-base verification is discussed in Section 4.3.2.)

4.2.2 Methodology

OptionCODE was developed on the basis of a number of texts, articles, encyclopedias and dictionaries collected by the author, primarily from the library of the Bank of Canada. Only two sources were bilingual, with the vast majority unilingual. Since the knowledge base was developed in English, English texts were the main sources of knowledge. These were selected on the basis of an intuitive assessment of their clarity and usefulness. The French terminology included in the CDs was gathered from the French-language sources, and no attempt was made to rank the quality of these equivalents. In general, then, a prescriptive, selective approach was followed with regard to English sources and a descriptive, non-selective approach with regard to French sources. This issue is discussed further in Section 4.2.3.

The basic methodology was systematic scanning and detailed note-taking of the information contained in the selected reference works, followed by analysis and extensive cross-referencing to identify the essential concepts of the field and their important properties. After the basic structure of the knowledge base had been sketched out on paper (about a dozen concepts), most of the conceptual analysis was done interactively using CODE.

All CDs in optionCODE share a common structure defined by property categories (see Figure 4-4). These categories — "Definition," "Encyclopedic Information," "Linguistic

Information" — are inherited by all CDs from the top-level concept T. The "Encyclopedic Information" category contains properties that collectively define the concept and distinguish it from other concepts. These properties are generally inherited from superconcept to subconcept, subject to their associated inheritance flags. The properties in the other two categories — "Definition" (containing the single property "definition") and "Linguistic Information" (comprising the

Property Categories	Properties
System properties	cdName super hasPropsOf kinds subConcepts inheritPropsTo instanceOf instances
Definition	definition
Encyclopedic information	[specific to each CD]
Linguistic information	English collocations English synonym French collocations French equivalent observations
Background information [optional]	[as required]

Figure 4-4 Generic Structure of CDs

properties "English synonyms," "English collocations," "French equivalents," "French collocations," and "observations") — are automatically inherited with *empty bodies* that are filled in separately for each CD. An additional property category, "Supplementary Information" is included in CDs on an ad hoc basis whenever it was judged that some additional information (primarily historical) might prove useful. For example, the option CD has the following properties grouped under "Supplementary Information": "examples of option quotes," "history of options trading" and "process of options trading."

A number of arbitrary conventions were adopted in the knowledge base. Parentheses were used not only when material was parenthetical in the original, but also to indicate grammatical information in the "French equivalent" property. Square brackets were used to indicate additions by the author. The formula "[q.v.]" was used to indicate that the

indicated expression is a concept in its own right elsewhere in the base. Information sources were also indicated in square brackets in the form of a code based on the author's name and a page number. A dummy concept **REFERENCES** was created as a subconcept of **T** to hold reference information; each author code corresponds to a property in that CD with full bibliographic information contained in the property body. Note that this means that the knowledge base user can complement his search with relevant printed materials as required.

Other conventions were adopted for properties. It was decided that the first appearance of a property in CD would be matched by an entry under "PropDescr" (property description) explaining the meaning of the property (although this was not followed consistently). The property "Comment" was sometimes used to specify the possible values (bodies) of the property when there was a limited set. The "Comment" was also used occasionally to provide other information, including linguistic, when there was no appropriate CD in the knowledge base.

Another decision was to use the "k" ("kinds") link only in cases where there was a true partition (see Section 4.1.1 on **CODE** for a fuller description). Thus the majority of subordinating relations are simple "s" ("subconcept") links. (This issue is discussed further in Section 4.3.6.)

The definitions provided for most concepts in option**CODE** deserve some explanation. As discussed in the following section, definitions should ideally be constructed by an expert in definitions in consultation with a domain expert. In option**CODE**, the author simply selected for each CD the definition or defining context from the references that appeared to be the most complete and most lucid. Thus all definitions are taken verbatim from the indicated reference source. In many cases, it is obvious that better definitions could be formulated, particularly in terms of providing knowledge to the intended user — a translator.

A couple of observations based on the author's experience are perhaps in order. First, the references provided vast amounts of information about the subject, but, generally speaking, that information was rarely presented in a conceptually logical manner. For example, a fundamental distinction is made in optionCODE between **tangible security** and **derivative security**. The former includes "regular" securities such as stocks and bonds, and the latter securities such as options and futures that are actually contracts to buy or sell tangible securities at a later date. This distinction, which is crucial to an understanding of how option markets operate, had to be inferred from the documentation; virtually nowhere was it explicitly spelled out. Another example concerned the division of the concept **debt security** into its numerous subconcepts (bonds, notes, bills of exchange, certificates of deposit, etc.). It took me some time to realize that the fundamental distinction between this group of securities was the *duration of their terms*.

Second, the initial stage of development — when I was attempting to identify the key concepts and the essential properties distinguishing each concept from its immediate relatives — was quite laborious. Once the basic "ontology" or basic conceptual framework was completed and some basic decisions taken regarding CD format and conventions, progress became much smoother. New pieces of information could be inserted fairly quickly in their proper locations and subtrees could be developed more rapidly. A non-expert striving to learn something in a new subject field using traditional reference sources likely goes through essentially the same process. At first he finds it arduous as he peruses the unfamiliar material, but gradually a basic conceptual framework crystallizes in his mind, and new information becomes easier and easier to assimilate. Many people have had the experience of deriving much more from a book on second reading, because they now possess the basic conceptual "skeleton" on which to hang the new knowledge.

4.2.3 An Example of Conceptual Analysis: Option Strategies

As mentioned earlier, options can be used and combined in many different ways to effect a variety of investment strategies tailored to an investor's particular situation, goals and tolerance for risk. In fact, it is this flexibility that has made them an increasingly popular vehicle for professional investors. The conceptual modelling of the part of the knowledge base dealing with these strategies turned out to be an interesting challenge.

The extremely large number of ways that options can be combined — in other words, their very flexibility — means that there is considerable variation in the terminology and its usage. Eventually I compiled a list of the most important option strategies, culled from a wide range of sources. I then tried to identify what was different in each of them and how they could be organized into a conceptual framework.

Eventually, I came to the realization that there was an underlying conceptual coherence and consistency to these strategies that was not identified clearly in any reference source. First, there were the basic option strategies: writing (i.e. selling) or buying a call or a put. Next there were the "hedge" strategies, which involved combining one of these basic option strategies with the sale or purchase of the same stock (or other interest) underlying the option. Lastly, there were what I characterized as "complex" option strategies, all of which involved two different transactions involving options. These complex strategies could be divided conceptually according to the nature of the two transactions, which I captured in the four properties "transaction1," "transaction1 object," "transaction2" and "transaction2 object." This conceptual division is summarized in Figure 4-5 and a screen dump showing a graphical representation of the option strategy subtree in Figure 4-6.

Developing the knowledge base optionCODE was an illuminating experience. Not only did I learn probably more about options than I would wish on anybody, but I gained

Figure 4-5 Conceptual Analysis of Complex Option Strategies

Concepts	Properties	transaction1			transaction2			transaction2 object			
		no.	kind	series	no.	kind	series	no.	kind	series	
				ex. price	ex. price	ex. month				ex. price	ex. month
complex option strategy	a transaction		an option				a transaction		an option		
straddle	= t2		call (f)	=		=	= t1		put (f)	=	=
long straddle	P		call	=		=	P		put	=	=
strap	P	2	calls	=		=	P	1	put	=	=
strip	P	1	call	=		=	P	2	puts	=	=
short straddle	S		call	=		=	S		put	=	=
spread	S (f t2)		= t2	f		f	P (f t1)		= t1		f
simple spread	S	1	= t2	f		f	P	1	= t1		f
variable spread	S	f	= t2	f		f	P	f	= t1		f
vertical spread	S		= t2	f t2		=	P		= t1		f t1
v.s. using calls	S		call	f t2		=	P		call		f t1
v.s. using puts	S		put	f t2		=	P		put		f t1
bear spread	S	1	(call)	< t2		=	P	1	(call)		> t1
bull spread	S	1	= t2	> t2		=	P	1	= t1		< t1
butterfly spread	S	2	(calls)	< & > t2		=	P	2	(calls)		t1
sandwich spread	S	2	= t2	= & t2 ¹		=	P	2	= t1		< & > t1 ¹
strangle	S	2	= t2	f & t2 ²		=	P	2	= t1		< & > t1 ²
tabletop spread	S	1	= t2	f t2 ³		=	P	2	= t1		> or < t1 ³
calendar spread	S		= t2			f t2	P		= t1		f t1
horizontal spread	S	1	= t2	= t2		f t2	P	1	= t1		f t1
diagonal spread	S	1	= t2	f t2		f t2	P	1	= t1		f t1
combination	= t2		= t2	f t2		=	= t1		= t1		f t1
long combination	P		= t2	f t2		=	P		= t1		f t1
short combination	S		= t2	f t2		=	S		= t1		f t1

Legend: t1 - transaction1, t2 - transaction2, P - purchase, S - sale, = - equal/same, f - not equal/different, > - greater than, < - less than, > - greater than, | - between; (call) means typically, but not necessarily, a call; shading indicates defining characteristic.

1 Sandwich spread: exercise price of both t1 options the same; exercise price of t2 option on either side of t1 exercise price.
 2 Strangle: exercise price of t1 options not the same; exercise price of t2 option on either side of t1 exercise prices.
 3 Tabletop spread: in case of calls, t2 exercise price higher than t1 exercise price and ascending; in case of puts, lower and descending.

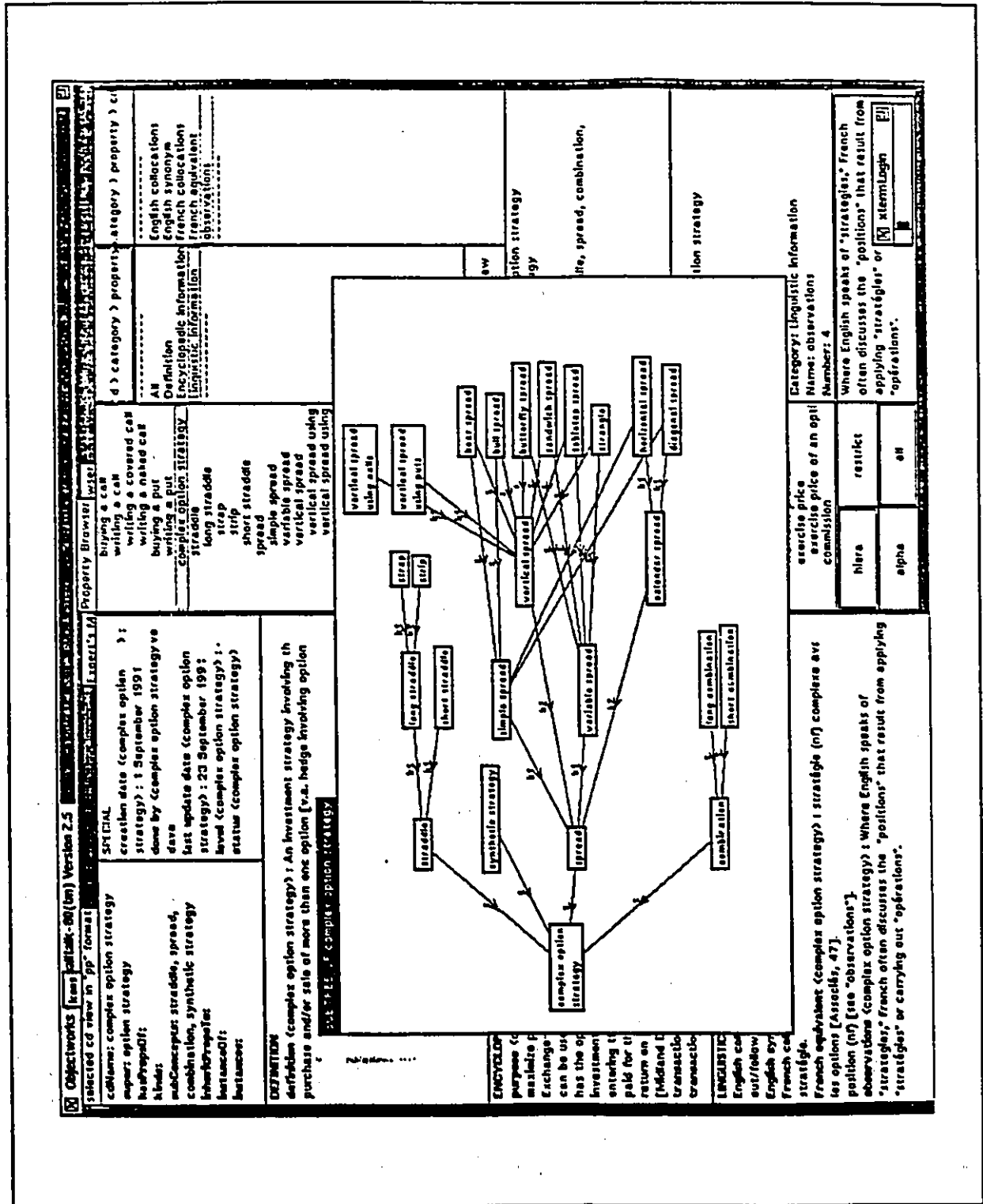


Figure 4-6 CODE Screen Dump Showing Subtree Graph of Complex Option Strategies (cf. Figure 4-5)

a keen appreciation for the task facing a knowledge engineer or terminologist attempting to formalize the essential information about a given subject field in a shareable form. A number of interesting methodological issues cropped up in the course of the exercise, some of which have already been alluded to in this section. Some of them are discussed in more depth in the following section.

4.3 Some Methodological Issues in Knowledge-Base Design

4.3.1 Sources of Knowledge

In AI, knowledge engineers developing a knowledge base or expert system for a particular purpose typically rely on verbal exchange with a domain expert or experts. Their main problem is extracting the right information and *all* the right information from the knowledge that the expert is carrying in his head. Essentially, this means asking the right questions and correctly interpreting the answers.

Developers of term banks (and eventually of knowledge-base systems for translators), on the other hand, usually rely primarily on written sources for their information, even though they endeavour to consult with a domain expert as extensively as possible. Their main problem is similar to that faced by knowledge engineers — to extract the right information and *all* the right information from the works they have available. Not being domain experts themselves, however, they do not necessarily know which works are more authoritative than others. Ideally, their references should be rated by an expert. In the absence of this assistance, they have to make informed guesses. The question becomes crucial when they are faced with conflicting or contradictory information. In this case, they have two options: either choose one source over the other, ideally in consultation with an expert; or else include the information from both sources.

In developing optionCODE, I did not have the services of a domain expert to assess the quality of my reference materials. I essentially relied on my own (admittedly limited) expertise as a translator specialized in economics and finance to judge which material to use and which to discard. When faced with conflicting information, I tried to find the common denominator of knowledge and to distil it into a conceptual form. I frequently found that descriptions that appeared to be incompatible on the surface often simply reflected terminological differences. In fact, having different perspectives on the same subject was often valuable.

There is another aspect to this question. A knowledge-base system for translators (KBST) may have a descriptive or prescriptive focus. If the former, it theoretically should include whatever its developers find, since there is a possibility that a translator may encounter that very concept or expression in a text he has to translate. If exclusively the latter (i.e. the term bank has a rigidly *normative* function), there is a risk that it may not be of maximum use to a translator who has to work with real-world texts, which are inevitably incompletely standardized. One compromise is to apply a descriptive approach to the source language and a prescriptive approach to the target language. OptionCODE basically subscribes to the latter philosophy, although not rigorously. The main disadvantage of this solution is that the term bank then becomes one-way — useful, for example, to a French-English translator but not to an English-French one. It should also be recognized that there is an implicit prescriptive weight carried by the concept label relative to any synonyms offered in a concept property. This issue will be further explored in Section 5.3.2 on bilingual knowledge-base design.

4.3.2 Knowledge-Base Verification

I was fortunate to have a recognized expert in the field of options (in fact, the author of one of my source books) to review my knowledge base: Claude Henin, the Assistant Dean of the Undergraduate Office of the Faculty of Administration at the University of

Ottawa. Dr. Henin had two approximately 1½-hour sessions with optionCODE, the first to become familiar with its nature and overall functioning, and the second to review part of its content. This amount of time is clearly far from sufficient to thoroughly verify a knowledge base of even optionCODE's modest proportions. Still, Dr. Henin made a number of excellent observations and suggestions, and the experience impressed upon me the importance of expert verification and suggested many of the points discussed below.

Ideally, any knowledge base must be verified by a domain expert. The very structure of a knowledge base poses some problems for systematic verification, however, since there is no easy way to proceed logically start to finish. On the other hand, the fact that corrections made at one level are immediately reflected throughout the knowledge base via inheritance mechanisms is a definite advantage. There is a wide scope for further research into knowledge-base verification. One possibility would be to draw up a standard questionnaire that the domain expert would be asked to complete for each concept in the knowledge base.

Among the aspects of the knowledge base that the domain expert should check are the following:

- the ontology — that the basic conceptual framework of the knowledge base is logically and intuitively sound;
- the properties — that the properties selected for the concepts are logically motivated and of practical value, and that the appropriate distinctions have been made between defining properties (necessary and sufficient) and non-defining (of interest only);
- completeness — that there are no key concepts missing;
- consistency and coherence — that choices and conventions adhered to in one part of the knowledge base are compatible with those in other parts, and that the depth of detail is relatively uniform throughout the knowledge base; and

- the quality of linguistic information — obviously, in a knowledge base for translators the quality of the linguistic information is very important: the definitions, synonyms, other-language equivalents and all other linguistic information should be thoroughly vetted by the expert, and he may have some opinion on the selection and quality of the sources (although this should properly have been done during development).

The expert may also have some helpful ideas to offer on how the knowledge-base system itself (e.g. its interface and features) could be improved.

4.3.3 Circumscription of Domain

As noted earlier, a KBST necessarily describes a subworld, a fairly well-defined (usually technical) subdomain of knowledge. The importance of the subworld concept was discussed earlier in the section on AI. In building a knowledge base, however, the developer may find it convenient to represent "high-level" concepts belonging to the realm of more general knowledge. In principle, this violates the subworld assumption, but it often enriches and broadens the usefulness of the KBST. For instance, in the optionCODE knowledge base, a CD was created for the high-level concept of value. This allowed the representation of price (and various types of prices) and of valuation (the process of determining the value of a security). Without the high-level concept, the lower-level concepts would be "orphans" without superconcepts.

An alternative approach was followed in the COGNITERM project (see Meyer et al. [1992a]). A special type of link — the "r" (for "related") link — was created to associate "orphan" concepts with the main hierarchical tree. For instance, the CD price of an option would have been connected to the CD option using the "r" link to indicate that it is related to the latter without entering into a specific hierarchical or associative relation with it. The original idea was to have a means of adding such concepts quickly during

the development phase, leaving decisions regarding their most appropriate placement or treatment to a later date. It evolved into a "see-also" link, similar to that found in technical thesauri.

The purpose of a KBST is to provide the translator with knowledge on concepts and terms that he might

encounter in a text in a specific domain. But in real-world texts, he will inevitably encounter some concepts and terms that lie outside that domain. The problem for knowledge-base development is where to draw the line. At some future point in the evolution of KBSTs, it

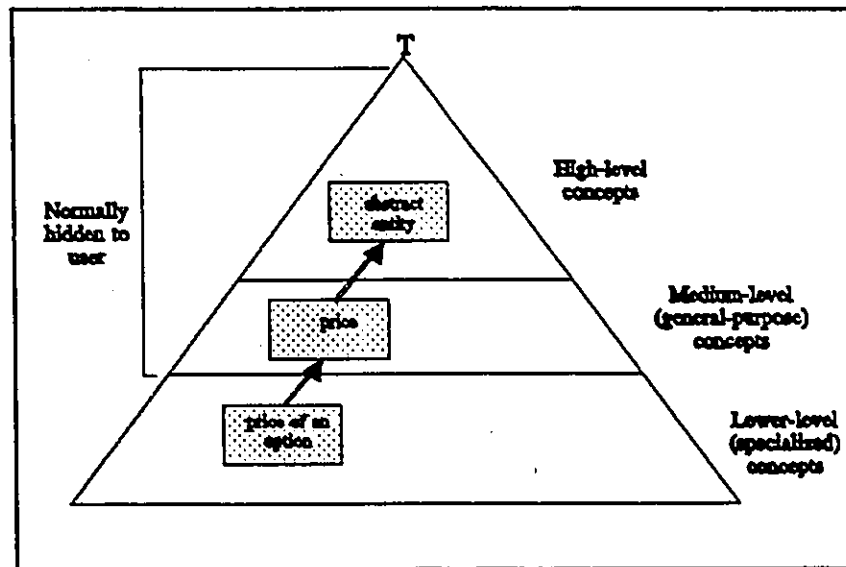


Figure 4-7 Inheritance over Different Knowledge-Base "Levels"

may be possible to have a general knowledge base of higher-level general-purpose concepts that invisibly underpins any number of specialized knowledge bases, so that, for example, declaring the concept price of an option to be a price (a general-purpose concept) would suffice for it to inherit an array of information about prices (e.g. expressed in a unit of currency, may change according to certain conditions, must be paid to purchase an associated item). (See Figure 4-7.)

Other authors have suggested modelling knowledge at different levels of abstraction. Raittikorn and Hayes-Roth [1991], for example, describe an approach where the abstract processes and structures underlying physical systems are represented as "prime models," and actual examples of these processes and structures are represented as "domain

models" that draw upon the prime models. For instance, a *flow* model at the prime level can serve as the basis for domain models of the human excretion system or a water-delivery system.

My solution for the purposes of optionCODE was to declare that the *general domain* of my knowledge base was the stock market, but the *focus* was options. In other words, the concept option and concepts "close" to it were modelled in some detail, while other relevant but more distant areas of the knowledge base were sketched out as required. This is not a perfect solution, since there are still concepts that lie outside even the general domain, but this is inevitable. The problem is not one of size or scope; no matter how vast the knowledge base, there will always be linkages to concepts lying outside the knowledge base's subworld.

In fact, having high-level concepts modelled either explicitly or through some underlying general-knowledge base is a prerequisite for machine use of the knowledge base, a subject that will be explored further in the section on machine translation (5.2.1).

4.3.4 Definitions

One of the most important pieces of information attached to each concept in a KBST is its definition. Definition formulation is a delicate and specialized task.⁴⁵ Definitions in the knowledge base should be constructed expressly for the purpose by a terminologist on the basis of the conceptual information modelled by the KBST developer. Ideally, it should be possible to match each key word in the definition with a property, and perhaps to group these properties in a special "definitional" category, with all the other properties being considered "encyclopedic." This corresponds more or less with what terminologists

45. A subject explored in depth in Eck [1993].

call the "descripteurs." Another possibility is have the system generate draft definitions from the conceptual information. CODE already has a rudimentary capacity to do this.

4.3.5 Natural Types Versus Roles

Certain concepts, upon analysis, turn out to describe *roles*, which are performed by some entity, rather than *natural types*, which are concepts related directly to the essence of the entity. Many common concepts like *pet* and *food*, for example, are actually role types. Both could actually be describing a penguin, for instance (the latter if the knowledge base concerned killer whales). Modelling both natural types and role types as concepts in a knowledge base is usually necessary and not at all inconsistent. Sometimes, however, problems arise in conceptual analysis that turn out to ultimately stem from the interaction of these two fundamentally different types of concepts.

AI researchers recognize that the natural type/role type distinction can be problematic. Concepts such as *father* can sometimes be analyzed in either way. Brachman [1991, 430ff] offers a practical guideline for making the distinction: a natural type has an independent existence, whereas a role depends on another concept or concepts for its existence. The supporting concept for a role may be implicit: for example, *landlord* implies that the person is a landlord *of* something. In fact, one rule of thumb for determining whether a concept is a role is to see whether it accepts an "of" phrase.

A number of CDs in optionCODE represent role-type concepts. For instance, the CD *underlying security* refers to the security that is the subject of an option or of some other derivative security. It can be a stock, a bond, or some other "tangible" security. In other words, some *tangible security* (a *natural type*) is playing the *role* of the *underlying security*. Trouble may arise, therefore, if the KBST developer starts to describe the

conceptual properties of the CD underlying security in isolation, since most of these are derived from whatever tangible security is acting in this role for a particular option.

4.3.6 Distinction Between Subconcepts and Kinds

As discussed in Section 4.1.1, CODE offers two types of subordinating links: the "s" (for "subconcept") link and the "k" (for "kind") link. The latter is intended to indicate "partitions": a particular subdivision into subconcepts that are disjoint, in that any subconcept of that concept must belong to one and only one of the "kinds." In optionCODE, for instance, the CD option is partitioned into European option and American option. The former can only be exercised *on* its specified exercise date, whereas the latter can be exercised *up to* its exercise date. The same CD also has a second partition: into *call* and *put*, calls being options to buy and puts options to sell. The twofold division into kinds means that every example of an option, whatever else it might be (and there are many different types of options) is necessarily either a European or an American option and either a put or a call.

Sometimes the choice of whether to use the "k" or "s" subordinating link is problematic. For instance, what of a division where every member of a concept A is necessarily an a, a b or something else (i.e. not a nor b)? The situation where there is implicitly a concept representing the logical complement of its sibling concepts occurs quite often. For instance, if one creates a CD as a subconcept of typesetting called digital typesetting (a perfectly acceptable "s" link), should one also create a CD non-digital typesetting, since any method of typesetting is necessarily one or the other? If so, then this division becomes a partition, and the "k" link should be used.

4.3.7 Conceptual Elegance Versus Utility

As described above, the KBST developer may find himself creating concepts and properties that may be logically necessary for conceptual consistency and elegance, but which ultimately turn out to be confusing for the end user of the knowledge base.

The optionCODE knowledge base, for example, contains the concept **derivative security**. This is a well-motivated concept, in the sense that both the term and the concept are well-established in the literature. It has a number of subconcepts (among them **option**) that have in common the fact that they are contracts for the possible purchase or sale of "real" securities. Derivative securities *are* securities themselves, however, since they are traded on exchanges at prevailing market prices. In contradistinction to the CD **derivative security**, the CD **tangible security** was also created. Its members are the "real" securities like stocks, bonds and treasury bills. However, **tangible security** is not so well-motivated a concept as **derivative security**; although the expression is occasionally found in the literature, it is not a term in the formal sense, rather a description.⁴⁶ Nevertheless, its conceptual properties are clear, and its presence in the knowledge base enhances conceptual elegance. Because it has been elevated to the same status as **derivative security**, however, the user may gain an inaccurate impression of its importance as a concept or accuracy as a term.

Properties, too, may fall prey to the same phenomenon. Sometimes a set of properties that makes good conceptual sense can end up simply confusing the user. For instance, the CD **derivative security** has two superconcepts: **security** and **contract**. This double hypernymic relationship reflects the fact that, besides being securities, derivative securities such as options and futures are legal contracts between two parties. **Contract**,

46. As discussed in Section 4.3.5, the real distinction between **tangible security** and **derivative security** is that the former is a natural type and the latter a role type. As well, **tangible security** could be regarded as the logical complement of **derivative security**, a situation discussed in Section 4.3.6.

meanwhile, has among its properties "party1" and party2" to reflect the fact that contracts are necessarily reciprocal. Consequently, option, a subconcept of derivative security, inherits these two properties. Conceptually, this makes perfect sense, allowing us to specify that "party1" of an option is known as the "writer" and "party2" as the "buyer." However, a user of the knowledge base seeking a quick explanation of what an option is may find properties with names such as "party1" and "party2" simply perplexing.

4.3.8 Evolution of Properties

Sometimes a property introduced at one level acquires a new "name" when it propagates to a concept at a lower level. For instance, the property "price," which was introduced at the concept security, inherits down the hierarchical tree to the concept option. The price of an option is known as its "premium." How should this change be represented?

The solution adopted in optionCODE was to *block* the property in the CD where its name changed, indicating its new name in the property comment. A new property with that name was introduced at the same level. This solution has a number of disadvantages. First, it breaks the continuity of inheritance. Second, it is yet another example of a practice that, while it may make sense conceptually, is clearly confusing to the user. A useful trick in CODE was found to be applying a mask to the browser, CD view or both to filter out blocked properties, so that only the property with the new name was shown. It should be noted that the most recent version of CODE (Version 4) includes a facility for locally renaming properties.

A related problem with some properties was that there was additional information that could be noted about them once they were inherited to certain CDs; e.g. a French equivalent, some additional conceptual information. For instance, option inherits the property "duration" from its superconcept contract. There are some interesting things

that could be said about the duration of an option. The only way to do this is to specify that the body of the "duration" property in option is "duration of an option" and then create a new CD with that name, thereby establishing an associative link between that CD and contract.

There are at least two difficulties with this approach. First, it is confusing and inconvenient for the user, who looks up the property "duration" in the CD option hoping to find some information and discovers only that it is the "duration of an option." He then has to jump to the CD with that name for further information. This would perhaps be acceptable if it involved an easy, hypertext operation. Second, it proliferates the "orphan concept" problem discussed in Section 4.3.3. Creating duration of an option logically demands the creation of its superconcept duration (which was in fact done in optionCODE).

4.4 Informal Translation Experiment with CODE and OptionCODE

4.4.1 Summary

To test how well CODE and optionCODE functioned as a knowledge-base system for translators, an informal experiment was devised. Three French-English translators of similar background (all 1991 graduates of the School of Translation and Interpretation, none with any particular expertise or experience in finance) were asked to translate a text on options. Two of the test subjects (referred to as Subjects A and B) were Master's students who had been using the same version of CODE as part of their own thesis work. Thus they had a good understanding of the system and were comfortable with its use. The third subject was working as a terminologist's assistant in the translation section of the Bank of Canada, where she had access to all the documentation and research resources of that institution.

The initial part of the experiment was conducted in the Artificial Intelligence Laboratory at the University of Ottawa. Subjects A and B, working at separate workstations, were given three hours to translate the text. Aside from a French-English set of Harrap's (two-volume "standard" edition) and a Gage Canadian Dictionary, their sole reference source was optionCODE. The subjects typed their translations in an "editor" window offering the standard set of Smalltalk editing functions (e.g. backspace and delete, cut and paste). Before they started, the author provided them with a few very general tips on using CODE as a translator reference tool. They were given no prior information about the text or its subject matter. The author was present during the experiment in order to observe their methods of interaction with CODE and to help in the event of technical problems (there were none).

The second part of the experiment was conducted two days later on the premises of the Bank of Canada. Subject C was given the same three-hour time limit to translate the same text. The author was not present during this time. Again, Subject C was given no prior information about the text, and she was instructed to use whatever reference sources she desired. At the Bank of Canada, these included Termium on CD-ROM, the Banque de terminologie de Québec (BTQ) on-line, Edibase (a textual database),⁴⁷ a specialized translation reference collection, and the complete library of the Bank of Canada.

The French-language source text was a 300-word excerpt from a much longer article by Bernard Jacquillat entitled "Les options négociables sur action et sur indice : de la théorie à la création d'un nouveau produit financier." It had appeared in the journal *Banque*, which is published in Paris. At the time the article appeared (1980), there was a proposal to set up specific markets for stock options and index options, and the purpose

47. *Edibase* was developed by Inform II-Microfor (based in Westmount, Quebec). English and French versions of the major Bank of Canada publications are regularly stored in Edibase, where they may be retrieved by searching for keywords using a variety of logical operators.

of Jacquillat's article was to explain stock and index options and to analyze the ways they could be combined into investment strategies. The particular passage was selected because it offered a complete (though cursory) description of a particular aspect of options (how their value is determined), and because it was judged to be typical of texts that professional translators face in terms of difficulty and style.

Appendix II includes the original French text, two sample translations of that text by the author (intending to illustrate the difference between transcoded and interpretative translation), and the three translations by Subjects A, B and C.

4.4.2 Results

The three translations were evaluated by Helen Meibus, the English editor at the Bank of Canada, who is responsible for revising French-English translations. She received an evaluation sheet prepared by the author for the purpose. There were two parts to the evaluation: a count of major faults and minor faults, based on the evaluation system of the Translation Bureau of the Secretary of State [Brisson 1992]; and ratings on a scale of 1 to 10 for the overall accuracy of the translation, the overall naturalness of the translation, how well the translator understood the text, and the overall quality of the translation. A copy of this evaluation sheet showing the scores for Subjects A, B and C is included in Appendix II.

The revisor found no major faults in any of the translations (although, in fact, there was at least one major fault in Translation C; this is discussed below). Translation A was judged to have 9 minor faults; Translation B, 9; and Translation C, 12. The overall quality of the translations was rated on a scale of 1 to 10 as follows: Translation A, 7; Translation B, 8; and Translation C, 6. On each of the other three ratings, Translations A and B scored the same as or better than Translation C.

The results of the evaluation appear to be very favourable to CODE and optionCODE. Given the time constraint, the two translators using optionCODE as their sole source of expert knowledge performed consistently as well as or better than the translator with access to the full range of reference sources of a large translation service.

4.4.3 Session Observations

As mentioned earlier, the author was present during the CODE session and was able to observe the interaction of Subjects A and B with optionCODE. Given the usefulness of graphs for terminological work (see Meyer et al. [1992a and 1992b]) and for knowledge retrieval (see Section 4.1.2.1), the graphical features of CODE were used surprisingly little. Each subject briefly consulted the full graph only once; no use was made of subgraphs, and the features that coordinate the browser and graph displays (see Section 4.1) were not used. The reason may lie with the imposition of tight time constraints, which forced the subjects to work more in "ad hoc" mode (i.e., searching for answers to specific problems) than in "background browsing" mode (see Section 1.3.1). Instead of graphs, the browser was the primary consultation tool for both subjects. Both ended up opening a "selected CD view" window next to the browser in order to have more complete CD information constantly available. The "go to CD" option from the CD view was not used. Cross-referencing (i.e. jumping to another CD from a "q.v." reference in another) was done manually, either by scanning for the CD name in the hierarchy or by alphabetizing the browser listing.

One interesting method the subjects used to interact with CODE was to open a mask on the browser and use the "Property body =" feature (which the present author had originally suggested as a way of accessing French equivalents) to search for a variety of property information with the help of wildcards ("*" for any string, "#" for any single character). It will be recalled that applying a mask serves to restrict the display to the CDs that meet the specified criteria. So it is possible, for instance, to obtain a listing of

all CDs that contain the string "option" anywhere in any property body by stipulating "Property body = *option*" in the mask. While this method is certainly far from speedy in the current implementation of CODE, it is extremely flexible. Both subjects reported that they found this a way to do "fuzzy" searches, in the sense that they were not quite sure what they were looking for, but wanted to locate the area of the knowledge base where they were likely to find the desired information.

On the basis of these observations, it would appear that certain techniques are available for interacting with CODE that work well for translators. Clearly, in addition to training on CODE itself, special training on these techniques must be provided. There were several potentially valuable techniques that the subjects failed to make use of or made only partial use of. At the same time, they also came up with some techniques of their own. Part of the issue is individual "workstyle"; different people work more comfortably and efficiently in certain ways than in others. Thus a KBST should offer a wide variety of interaction possibilities and should be flexible enough to adapt to a translator's particular workstyle. To some extent, too, translators will need to adapt their workstyles to such new tools; treating a KBST as simply a super-dictionary is not necessarily the most effective approach.

4.4.4 Follow-Up Interviews

Follow-up interviews were conducted with all three subjects. For Subjects A and B, the author wanted to explore three basic areas: for CODE — which features were useful and which were not, what was easy to use and what was not; for optionCODE — general impressions, what aspects of the knowledge base were helpful and informative and which were not; and for the text and translation — feelings about the final result, whether the translator felt she understood the text and was able to make adjustments accordingly. Subject C was asked the same questions regarding the text and was also questioned about her work methods and reference sources.

Both Subjects A and B were quite enthusiastic about their experience. In terms of their mode of interaction with CODE, they confirmed the author's observations. Both agreed that they had not used the graph to any real extent, but found that the hierarchical structure of knowledge was clear from the indenting convention used in the browser (see Section 4.1.2.1). Also as noted above, they found the most useful window configuration to be the browser coupled with a selected CD view (see Figure 4-1). Predictably, the subjects found the process of locating French equivalents using the browser mask a frustrating experience and felt strongly that better and more efficient access methods (global lexicon, indexing of language equivalents, etc.) were imperative for translator users. They pointed out that the words of the source text are the translator's primary key to unlocking the information stored in a KBST.⁵⁰ Consequently, a KBST must offer a "lexical gateway" to the knowledge it contains.

On the subject of optionCODE, Subjects A and B were very complimentary. They felt that the structure of the knowledge base and the CDs was good, and that the depth of information presented was appropriate. They found the various "Linguistic Information" properties useful, particularly "English collocations," but felt that more collocational information and observations were needed. Subject B noted that the "q.v." convention to indicate cross-references was useful, but would be better if such a reference was a hypertext link. Subject A commented that optionCODE represented a source of "distilled knowledge."

On the subject of the source text and their translations, both Subjects A and B found the original text almost incomprehensible on first reading, but stated that they ended up with a good understanding of its content through their interaction with the knowledge base. Subject A observed that she felt she was learning something about the topic in the course of the session. Certain points of the text remained obscure to both, however.

50. A point noted in the previous experiment; see Section 2.5.3.

Subject A noted that the process of revising her draft was made easier by the knowledge she had acquired, and that she was able to cut out some transcoded verbiage as a result.

Subject C, the non-CODE translator, found that the three-hour time limit severely restricted her opportunities to research the topic. She simply did not have enough time to start browsing through specialized reference works in order to gain an overview of the field. As a result, her primary reference tools were lexical — Termium, Edibase and Translation Bureau glossaries. She, too, found the text difficult, but felt that by the end she basically understood what it was saying.

The follow-up interviews with the three subjects confirmed a number of points that have been made earlier in this paper; e.g. the need for a global lexicon and better hypertext capabilities. One interesting aspect was that the time limit appeared to be more of a constraint for the non-CODE translator, limiting the breadth and depth of her research efforts. This underlines one of the main advantages of on-line KBSTs: because they are integrated into the translator's working environment, they are an inherently more efficient research tool. In fairness, optionCODE had an edge because of its rigid focus on a small subdomain, compared with the far greater comprehensiveness of traditional knowledge sources. Still, it is important to realize that even though there is considerable room for improvement in the performance of the CODE system, it is, by its very nature, a more efficient way to acquire knowledge.

4.4.5 Additional Observations on the Translations

In reviewing the translations produced by the three subjects, the author noted some aspects that confirm points made earlier in this paper.

It was argued earlier that a transcoding approach to translation is inherently riskier because the translator has no way of knowing when he has strayed from the meaning; a

wrong interpretation of a referent or of some other ambiguity may throw the translation seriously off-track. A related problem is an outright mistake on the part of the translator. For instance, Subject C made a major error in the first paragraph (unfortunately not noted by the independent revisor), three times translating the French "action" by "option." Of course, translators are human and sometimes make inexcusable errors of this kind in their drafts (I speak from extensive experience). Such errors are usually caught at the revision stage. But I suggest that a translator operating in transcoding mode (i.e. without sufficient understanding of the semantic content of the source text) is *less likely to notice such an error*.

Another interesting example concerns the translation of the passage "l'avec, le prix de l'option" in the fifth paragraph. Here the word "avec" is a noun referring to the option premium; it is a term more commonly used in France than in Canada. "Le prix de l'option," set off by commas, is an appositive offering an alternative expression for the same concept. Subjects A and B successfully handled this passage by collapsing it to "the price of an option" and "the option premium," respectively. I interpret this as an instance where the translators were able to make an adjustment⁴⁸ based on their knowledge of the subject matter. Subject C, on the other hand, following more of a transcoding approach, was less free to make this kind of adjustment and felt obliged to find a word to fill the "slot" established by the word "avec." She settled on the incorrect term "cum" (used with a bond to indicate that a dividend is attached). In passing, it is interesting to note that the word "avec" cannot be looked up in Termium, since prepositions are entirely transparent (i.e. not indexed) in that system. In fact, two other expressions in the source text — "dans le cours" and "hors des cours" — cannot be looked up in Termium for the same reason.

48. Specifically, "économique" in the Vinay and Darbelnet classification; see Vinay and Darbelnet [1958].

A final example shows how useful it is to have collocational information readily at hand. The second paragraph of the source text refers to "l'échéance" of an option. In French, this word can be applied to both bonds and options (among other things). In English, however, a bond is said to "mature," but an option "expires." This information is included in optionCODE in the "English collocations" property in the CD option. Subject B used the proper term, while Subjects A and C did not. This is a good illustration of the kind of information that can be made easily accessible in a KBST, but that tends to be rarely included in specialized glossaries and only implicitly found in expert reference works.

CHAPTER 5

Future Directions

This chapter explores the potential of knowledge-base systems from several points of view. Section 5.1 describes the translator's workstation concept and surveys current efforts in this area. The question of how a knowledge-base system such as the one described in this thesis might be integrated into a translator's workstation is then examined. In Section 5.2 the focus is broadened to explore the place of knowledge-base systems in machine translation, the document-production chain (i.e. the entire process of creating documents from initial conception to finished product in various languages), and distributed-knowledge networks. Each of these topics is vast in its own right, and so the discussion is necessarily restricted to some general observations. Lastly, some specific problems in designing knowledge-base systems for translation applications are studied, including translators' special needs.

5.1 A Tool for a Translator's Workstation

5.1.1 The Translator's Workstation Concept

In 1987, Alan Melby declared, referring to the translation workstation, that "its time has come" [Melby 1987b, 7]. A translator's workstation (TW) simply refers to a computer setup, including all associated hardware and peripherals and all software, that has been designed specifically around the needs and workstyles of translators with the goal of optimizing the speed, accuracy and efficiency of their work. Thus as E. Macklovitch noted [1991, 319], the TW is essentially a concept, not a fixed product. Melby, who has been writing about TWs since the early 1980s, has repeatedly sketched out the three "levels" of a TW based around a personal computer (PC). In each case, the software and hardware are off-the-shelf products selected with a view to their

compatibility in format and in ability to function well together. A similar four-level classification for the automation of translation was suggested by Thouin [1988].

A Level One TW consists of a multilingual word processor (possibly with features such as spell checking, word counting, grammatical checking and synonym finding), telecommunications hardware and software, and a multilingual terminology management program. (Thouin's classification distinguishes between a first level with word processing, desktop publishing and communications, and a second level with computerized spelling and grammar checkers, on-line terminology, and documentation aids.) In many cases, several of these functions may be provided by a single software package (for instance, WordPerfect offers, in addition to word processing, spell checking, word counting and a thesaurus). Terminology management can be based on a specially programmed version of a generic database system or can rely on a specialized terminology management package, a number of which are available for PCs, including Melby's own Termex.⁴⁹ Terminology support may include access to government or commercially prepared term banks, term banks created specifically for a particular organization, and personal terminology records. The telecommunications package allows texts to be received and sent electronically and would grant access to remote dictionaries, term banks and other information sources. Many translation services now operate on a local area network (LAN) that makes distributed software available to translators and allows them to communicate with each other. A translator who is "plugged into" the rest of the world via a LAN, modem, fax and electronic mail is able to more easily solicit the advice of translator colleagues and domain experts. A Level One TW might also include a variety of other target-language aids, such as on-line verb conjugators.

49. Bédard points out [1990a, Q8] that, aside from one little-known exception (a program called "Le Terminologue"), there are no such specialized packages available for the Macintosh.

A Level Two TW adds access to source texts and possibly to previously translated texts (source and target languages) in machine-readable form. In some cases, an optical character reader is available to convert a source text. Given the proliferation of facsimile machines, the combination of a modem/fax board installed directly in the computer with an OCR program capable of processing "soft" faxes may be advisable. With the source text in machine-readable form, a variety of specialized text processing tools can be brought to bear, capable of producing word lists, dynamic key-word-in-context displays, and text-related glossaries. The source text can be indexed in order to find all occurrences of word in specific contexts (a concordance), to produce word frequency lists or to identify unknown terms. A number of other ancillary programs can be added to the workstation at this stage as required. Document-management systems, for instance, can assist in document location and retrieval by providing date stamping, retrieval handles, scanning to match given search criteria, and document statistics. Document-comparing software can be used to identify updates and track revisions. Conversion programs may be useful for converting text and other files from one format to another. Textual databases containing previously translated texts can put a wealth of information at the translator's disposal. As Harris has pointed out, the major advantage of such "bitextual" databases is that they preserve the actual context of a previous translation and allow the translator to work with "translation units" (i.e. segments of text larger than words), as he normally does [Harris 1988]. Job-management software can even be included to assist with client tracking and billing, scheduling and deadline management, work analysis, and production statistics. Specialized desktop publishing software may be appropriate when the production unit to which translators belong is responsible for producing camera-ready copy.

Melby's Level Three TW, besides all the components discussed above, gives the translator access to a machine translation (MT) system,⁵⁰ which the translator would be

50. Here we are speaking of a current commercial system, which would *not* be knowledge-based, but dictionary-based. MT is discussed further in Section 5.2.1.

free to use as he judges appropriate. Depending on his workstyle and needs and the nature of the text, he may use the MT system for producing a first draft, for assistance with a particular passage, or simply for dictionary lookup. Ideally, he could invoke the MT system at appropriate moments (say, for a passage that he knows it will handle efficaciously, such as a repetitive list) and then take back control when he wishes. The MT component might also be set to run in the background during microprocessor idle time to improve its performance.

Rather than or in addition to an MT component, I feel that a Level Three TW would benefit from having on-line access to advanced knowledge bases specifically designed for the use of translators and other writers. Properly designed KBSTs could even replace the conventional term banks of a Level One TW, at least in their respective subject fields. As has been repeatedly argued in this thesis, knowledge maximization is an important factor in translation, and accessing expert knowledge can be a very time-consuming process. Since the goal of the TW is to enhance the efficiency and productivity of the translator, it makes sense to provide him with the richest, most flexible and open-ended knowledge retrieval tool possible. Knowledge bases will augment the value of translator's workstations.

What Melby has long argued is that a TW is not a thing of the future, but something that can be constructed at the desired level from currently available off-the-shelf components. The crucial consideration at each level of TW is the seamless integration of all components, so that information can be moved easily and transparently among them. Clearly, however, a complete Level Three TW including a knowledge base on anything but an experimental basis still lies in the future.

Turning to hardware, there was a time when a clear distinction was made between "workstations" built around desktop PCs and "industrial workstations" such as those made by Sun and DEC, usually running some flavour of UNIX. The latter machines were

expensive and had much faster microprocessors, large amounts of RAM by the standards of the day, large high-resolution graphical screens, and extensive windowing capabilities. Recent developments in microcomputer technology, however, are putting increasing power on the desktop, and the distinction between PCs and industrial workstations is starting to blur, even in terms of graphics ability and screen size. The one remaining edge enjoyed by high-end workstations is raw speed, but even this distinction is being erased as increasingly faster and cheaper desktop systems appear.⁵¹ Knowledge-management systems such as CODE will soon be able to be run efficiently on desktop systems. CD-ROM technology, which offers relatively efficient access to large quantities of information, may prove suitable for storing knowledge bases.

It is interesting to draw a parallel, although it is less than perfect, between Melby's three TB "levels" and the three "stages" of new technology identified by Alan Gilchrist [1982]. Gilchrist suggested that new technology is introduced in three stages: At the first stage, it enables you to do the same thing more cheaply, faster and better; at the second stage, it enables you to do things you could not do before; and at the third stage, it changes your behaviour and ways of doing things to match the new capabilities bestowed by that technology. Word processing and associated programs clearly belong to the first stage; they essentially help the translator to do his work faster and better. In a Level Two TW, the ability to process the source text and previously translated texts does indeed allow the translator to do things he has never done before, such as text concordance and bitextual lookup. And the MT component and (as I propose) on-line knowledge base of a Level Three TW would, I believe, eventually change the working habits of a translator as he became more adept at exploiting its full capabilities. For example, a translator might adopt the habit of brushing up on the most important concepts in a subject field before beginning work on a new translation request. The availability of on-line KBSTs would tend to break down the distinction between a

51. Products based on Intel's new "Pentium" microprocessor, with its 64-bit internal data bus, will go a long way towards bridging this gap. They should appear this year (1993).

separate "research" stage and a "translating" stage, since the two activities could readily be carried out simultaneously. Gagnon [1992] suggests that advanced term banks will bring about a shift in the "techno-economic paradigm" of terminology and translation, by changing the means of production, the conditions of production, the form of distribution, and consumption patterns in these areas of activity.

5.1.2 Review of Current Translator's Workstations

Over the years, a number of word-processing packages tailored specifically to translators have made their appearance. Melby described a prototype of such a system as early as 1984 [Melby 1985]. Most of these have had as their centrepiece a split screen allowing the computer to display the source text in one window as the translator worked in another window. Generally speaking, none of these products has caught the fancy of working translators, primarily because commercial word processing packages have steadily increased in quality and power while dropping in price.

Several companies have developed integrated TWs around what originally were machine translation systems. ALPNET, for instance, offers a system of software tools known as TSS (Translation Support System) that runs on an industrial workstation⁵² [Childs 1989]. Tailored to situations where high-volume translation is required, its core is a series of dictionaries, including a "document dictionary" built up for each text. The dictionaries can be used for on-line lookup and insertion (either automatic or human-guided), or for machine translation. The ALPNET system requires a considerable investment in terms of time (for customizing dictionaries) and money. Although it assists with dictionary lookup and terminological consistency, it does not provide translators with expert knowledge.

52. The latest version will apparently operate on a fairly powerful microcomputer running the OS/2 operating system.

The alternative to the integrated approach to TWs is the modular approach. As discussed earlier, it involves carefully combining off-the-shelf components to create a workstation that is specifically designed for translators, even if the individual components are not. Shreve [1991] notes that this approach lends itself to both the "toolbox" metaphor (a tool at hand for every task) and the "desktop" metaphor (an imaginary desk with icons representing functions; e.g., a book to represent dictionary lookup). The main advantages of the modular approach are increased power, since each application is dedicated to its particular job, and greater flexibility, since any component may be changed to suit evolving needs or technologies. The main disadvantage is ensuring that the components work together properly, i.e. that they are compatible in terms of operation and format, and that information can be moved quickly and easily between them. This necessitates an "integrator" or "shell" program whose job is to tie the various components together through a consistent interface and to mediate communication among them. Ideally, the user himself would be able to customize this interface around his needs and preferences.

Large industrial workstations manufactured by such companies as SUN and DEC have long offered many of these capabilities in their software. An international team at the University of Surrey, England is working on the "Translator's Workbench Project" [Ahmad et al. 1990] aimed at developing an integrated set of computerized aids for machine-aided human translation. The Translator's Workbench, which will run on a SUN workstation, includes a custom-built text analyzer program that can generate word frequency lists, statistics, concordances, and examples in context. Terminology, including conceptual knowledge (hierarchical and encyclopedic), is currently stored in a conventional relational database system (called ORACLE), but a frame-based knowledge-representation module is under development.

There is also considerable activity in building a PC-based TW from off-the-shelf components. The Apple Macintosh, with its inherent user-friendliness and consistent

software structure, would seem to be an ideal microcomputer platform for a TW. Bédard described such a Mac-based system, which he called "MacTranslator" [Bédard 1988], and a Secretary of State translator named Michel Thibodeau has developed his own turnkey TW for the Macintosh (described in Bédard 1989b). However, efforts associated with businesses and government organizations have tended to focus on IBM-compatible systems, because these machines dominate corporate and government circles.

Since 1989, the Translation Bureau of the Secretary of State Department has been developing such a workstation in collaboration with the Canadian Workplace Automation Research Centre [Macklovitch, L. 1991]. In its latest version, the workstation is built around a 386 PC equipped with, among other things, a CD-ROM player. Software includes word processing, memory-resident glossaries, a verb conjugator, and a document manager and comparer, all running under a shell and program integrator. The CD-ROM player is used primarily for Termium. While the official evaluation is not yet completed, participants' reaction to the trial workstation has reportedly been positive, with highest accolades reserved for the local availability of Termium. In its next version, the Translation Bureau's workstation will be linked to a network to allow distributed access to Termium and other CD-ROM reference materials. It will likely be based on Windows, an enormously successful shell and multitasking environment for IBM-compatibles offering a graphical environment patterned after the Macintosh's.

It is clear that, as it stands, the Translation Bureau's workstation does not represent a quantum leap forward. In fact, similar and even more powerful setups for translators are becoming more and more common as hardware and software prices tumble. One interesting result of the Translation Bureau's project is the enthusiasm of their translators for ready access to Termium. This underscores a point that has been made several times in this paper: translators' need for fast, efficient retrieval of information.

5.1.3 Integration of Knowledge-Base Systems into Translator's Workstations

As discussed above, current options for terminology management are restricted to traditional-style term banks. Glossary programs such as Termex and TermTracer provide little more than a quick look-up of the "source-language term = target-language term" variety. Even Termium, despite its wealth of information and formal structure, is focussed on terminology and is not a knowledge-based system.

That is where the technology stands today. However, the translator's workstations of the future will be able to draw upon knowledge bases of the kind this paper has been discussing. Translators on powerful workstations (either desktop PCs or high-end industrial workstations) linked to a central computer will be able to browse freely through distributed knowledge bases to locate a wide range of specialized information about their current field of interest. Windows simultaneously running different applications will allow them to jump instantly or to move information easily between their texts and the knowledge-base system. This wealth of resources can pose its own problems, however; as Otman points out, "le problème ne sera plus l'absence de l'information ou la difficulté d'accéder instantément à l'information requise, mais la gestion du flux d'information" [1991, 20].

It seems clear that the hypertext approach offers the best way of navigating knowledge bases, which by their very nature are not linear or sequential as are traditional term banks. It was suggested earlier (Section 4.1.2.2) that the knowledge provided in such a system can be considered to have three dimensions: the immediate information about the concept; the hierarchical information inferred from its superconcept(s), subconcepts and sibling concepts; and the concept's associative links with other concepts. Only a hypertext interface can bring this third facet of knowledge bases into sharp focus by allowing these associative links to be easily and fully explored.

The question of the interface will be discussed further in the section on knowledge-base design (5.3).

5.2 Integration with Other Systems

Knowledge-base systems hold out considerable potential for integration with other systems. In particular, there is currently a great deal of interest in designing special "terminologist's workstations." Given the emphasis placed by contemporary terminology on conceptual analysis, knowledge-management tools designed specifically for this purpose clearly offer some exciting possibilities. In the present thesis, however, the focus will remain on translation and translators. For a discussion of the role of conceptual analysis and knowledge-management systems like CODE in terminology, see Miller et al. [1991] and Skuce and Meyer [1990].

5.2.1 Machine Translation

Machine translation (MT) has passed through three distinct, but overlapping, generations over the past 40 years. The first generation of MT started with high expectations in the 1950s. Computer translation was viewed as a process of mapping the lexicon and syntactic structures of the source language directly to the corresponding structures of the target language. Activity slowed down considerably in the mid-1960s (at least in the English-speaking world) as dissatisfaction grew with the poor quality of results. The second generation, which spanned the early 1970s to the mid-1980s, concentrated more on underlying linguistic theory and relied mainly the "transfer" approach, where the syntactic and, to some extent, semantic structure of the source-language text was analyzed with the specific target-language in mind before being

transferred to that language. Despite some notable successes,⁵³ second-generation systems mainly failed to achieve their promise. In the late 1970s, a third generation of MT emerged from AI. It was largely inspired by the "interlingual" model, where the semantic content of the source text was converted to a theoretically language-independent representation of meaning, which was then used to generate a target-language text. Again, some exaggerated claims on the part of certain researchers and a failure to deliver practical systems led to harsh criticism of the semantic-based approach.

Today, MT systems from all three generations co-exist, although it is primarily the second-generation systems that are used for serious practical work. In fact, though, the lines between the three approaches have blurred as researchers have tried to utilize the best features of each. Many authors have argued that there is one clear lesson from the history of MT to date. "The first 40 years of MT experience teach us that significant long-term progress depends on advances in our basic knowledge of how to model natural language using computers" [Hutchins 1986, 2].

Many current functional MT systems encode some degree of real-word knowledge in their dictionaries. However, they continue to rely mainly on the manipulation of linguistic forms (i.e. words) to perform translation. As a result, human intervention in one form or another is essential for high-quality output. Systems with the most impressive results impose quite severe restrictions on subject domains (and their sublanguages⁵⁴) and/or on text types.

AI researchers became seriously involved in MT in the early 1980s. But their number has always been relatively small, even though Wilks, for example, has repeatedly

53. The most-often cited example being MÉTEO, which translates weather forecasts from English into French.

54. See Kittredge [1983] for a discussion of sublanguages.

argued that MT is an excellent testing ground for AI theories of natural language.⁵⁵ And Tucker [1987, 25] states that MT should, but curiously has not yet, become a major application of AI. Recently, however, some major projects have emerged that are explicitly concerned with *knowledge-based MT* — translation by systems that are coupled with extensive knowledge bases that purport to represent not only specialized knowledge, but also (and more importantly) general common-sense knowledge.⁵⁶ The thrust of current research, therefore, is, first, to determine what kind and what degree of knowledge is required for more successful MT, and, second, how such an inevitably large amount of knowledge can be encoded efficiently.

MT systems can make little use of conventional term banks, since, as noted earlier in this paper, the burden of interpretation remains on the user, a task that itself requires a vast array of supplementary information about the particular subject field and the world in general, information which MT systems simply do not possess.

However, knowledge-representation systems of the kind discussed in this thesis, which encode implicit information through inheritance mechanisms, may be better suited to exploitation by MT systems. The implicit information contained in such a system, much of which is not generally required by humans, can be made explicit when required for machines. In particular, this includes the system's higher-level information: knowledge associated with general language concepts and even with very high-level concepts such as "entity," "event," and so on (see also Section 4.3.3). It should be noted that knowledge bases are potentially useful not only to knowledge-based MT per se, but also to other MT architectures (i.e. direct, transfer and interlingual), since they represent an efficient means for storing and retrieving a variety of information. Hutchins proclaims that "the

55. See, for example, Wilks [1983].

56. The most notable example is a project under way at Carnegie Mellon University involving a team of researchers led by Sergei Nirenburg. Their latest system, known as KBMT (Knowledge-Based Machine Translation), is coupled with a knowledge-acquisition system called Ontos. See Goodman and Nirenburg [1991].

major theoretical issue which faces all MT researchers is the place of artificial intelligence in MT systems... The question is not whether understanding has a role in MT but how large a role it should have" [Hutchins 1988a, 12]. While this statement may be somewhat overblown, since interesting progress continues to be made in MT that is not necessarily knowledge-based, it is probably accurate to say that the next major advance in MT will not arrive until some of the fundamental problems connected with knowledge representation and machine understanding are solved.

The question of what "understanding" means in relation to MT is not easy to answer. Even for human translators, understanding is relative. The level of understanding required of the system conditions not simply *what* knowledge must be represented, but *how deep* that knowledge must be. As discussed earlier, translators do not necessarily have to possess the same knowledge as a field specialist in order to translate a text in that field. In fact, their knowledge and understanding are inevitably inferior to the expert's. However, generally speaking, the greater the translator's degree of understanding, the better the quality of the result. Tsujii sums up the issue for MT as follows [1991, 9]: "From a practical point of view, the question to be asked is how MT systems have to understand texts in order to produce reasonable translations for *average ... sentences* [italics added]."

There are some promising currents evident in current MT research. Some notable successes have been achieved through a narrow focus on appropriate sublanguages.⁵⁷ Another approach is bi/multilingual text generation by machine as an alternative to MT.⁵⁸ Yet another possibility is pooling the complementary knowledge of computers and humans; for instance, coupling a domain expert who is not a translator with an MT system that possess the required linguistic knowledge or, conversely, coupling a human

57. See Kittredge [1987] for a thorough discussion of sublanguages in relation to MT.

58. See, for example, Kittredge [1989].

translator with a system that can provide expert knowledge.⁵⁹ The latter approach is akin to the knowledge-base systems for translators that are the subject of this thesis.

5.2.2 The Document-Production Chain

It was stated in the Chapter 1 that the modern world is witnessing an explosion of new knowledge. And, as Galinski [1988c, 168] notes, language inevitably lags behind the rapid development of knowledge and the new concepts associated with that knowledge because of the limited number of existing term elements to name new concepts. High-technology tools are developing at a similar pace, and companies have come to recognize that the efficient transfer of highly specialized knowledge across national and linguistic boundaries and the provision of high-quality documentation in multiple languages based on consistent, logical terminology are crucial components of their business strategies. In addition, the time between product design and production is constantly shrinking, imposing ever tighter deadlines on the provision of documentation. As communication needs grow increasingly complex and stringent (and expensive), the need increases for systems that can bring order to this terminological and conceptual chaos.

The rising importance of documentation, particularly its commercial aspects, reflects these trends. Some companies have developed their own knowledge bases to handle their documentation management needs (for example, Royal Dutch/Shell, with "Sherpa" [Brace 1990]). In fact, documentation is fast becoming a major industry in its own right. Corporations specializing in documentation management — which includes technical writing, abstracting, translating, graphics, digitizing, production and more — are one of the fastest growing language industries. Another niche that is rapidly expanding is the "language conversion" industry, which refers to the integrated production of foreign-language materials in a variety of media, from original to final product. A leader in this

59. See Tsujii [1991] for a discussion of these approaches.

area is Ovum, a communications consultancy firm based in London, England. In addition to written documentation, the language conversion industry also encompasses such activities as dubbing and subtitling of film and video productions.

Another striking example of the growth of the documentation industry is provided by France's SITE, now the largest technical communication company in Europe [Joscelyne 1990]. It has developed an advanced term bank for its translators, which, while not conceptually based, incorporates some of the features of the new generation of term banks, including a natural language interface and a tree-like structure. In its approach to documentation, it envisages each consumer product it handles as a "vector," along which move a variety of systems and subsystems, each with its own range of documentation and communication needs. This philosophy reflects what has been termed an "object-oriented" approach to the publishing process. Another reflection of this approach is recent efforts aimed at producing a universal "document description language" that could be understood by all manner of document-production systems.⁶⁰

As part of the document-production team, translators too will benefit from knowledge-based systems for managing reference documentation and maintaining terminological consistency. For translators do not simply need knowledge bases to *provide* information in particular subject areas, they need knowledge *about* information: where to find it and how to access it. As Mayorcas-Cohen notes [1986, 80], "Generally, translators are badly informed about the wealth of information sources available for solving many of their queries, and poorly versed on efficient methods for handling the information that does come their way."

60. An ambitious attempt backed by the International Standards Organization is now under way to create such a language, known as the Standard Generalized Markup Language (SGML). See Wright and Melby [1992].

Within such a system, translation will have the opportunity to evolve from a secondary activity that is neglected until the next-to-last moment to an integral step in the document-production chain. "Translation has to be designed into the automated document production and transmission processes and can no longer be accepted as a hiatus in an otherwise paperless flow of information and documentation," writes Sager [1989, 9]. "Translation can no longer be considered an isolated activity at the margin of text production." The translation process itself may benefit from specialized work-management systems; the European Parliament, for instance, has a computerized system running under UNIX for tracking, scheduling and logging translation activity.

Corporations of various kinds facing an "information crisis" may find that advanced knowledge-based term banks may help solve many of their problems. Such knowledge bases have the potential to serve as the interface among a company's various management-information systems and other systems. A knowledge base might provide support for an entire documentation project, providing greater quality and functionality of data, ensuring terminological consistency, and enhancing internal communication through its shareability. It could provide essential support for the company's own communication experts, including writers and translators, as well as for its non-human specialists — expert systems, and automatic text generation and machine-translation systems. In the final stages of a project, the knowledge base may prove to be an effective in-house training tool, as well as an instrument for accessing internal documentation. And logically, better internal communication should translate into improved external communication across national and linguistic boundaries. Knowledge bases used in this way may form the basis for distributed-knowledge networks.

5.2.3 Distributed-Knowledge Networks

In a typical modern corporate environment, there are computers everywhere — mainframes and minicomputers, industrial workstations, microcomputers on virtually

every desk, some connected to wide- and local-area networks and some autonomous. Each of these installations houses vast amounts of data, much of stored in a variety of database or spreadsheet formats. There is a wide array of knowledge that could be derived from combining and analyzing the information contained in these disparate sources, but the technology does not yet exist to integrate and display this information upon request.

Some ambitious efforts are under way to master information proliferation through systems that integrate computer data by providing a common framework for representation and retrieval. Such a system may be termed a *distributed-information network*. At its heart there should lie a knowledge base.

Galinski argues that terminological data banks "of a higher order" offer the solution to problems of knowledge processing and the universal availability of information [1988a, 489]. And he further contends that such systems are ideally suited to serving as the core of comprehensive knowledge networks "not only because they are just good for storage and retrieval purposes, but also because our natural, technical and colloquial languages are poorly suited to the task of ordering knowledge items ... in a systematic way as well as of retrieving stored texts and other representations of complex knowledge" [Galinski 1988c, 174].

Among the most ambitious of these efforts is the *Carnot* project, which is under development at Microelectronics and Computer Technology Corp. The project's ultimate goal is to develop a basic mechanism for heterogeneous database access. "Carnot acts as a corporate intelligence network, consolidating diverse data resources and synthesizing varied information formats into a semantically common whole" [Rasmus 1991, 250]. Its knowledge base is Cyc, which was discussed briefly in Section 3.3. Cyc's job is to assist with interpretation of requests for information through its understanding of corporate information relationships and data sources.

Another research thrust aimed at solving the problems of dispersed information within an organization is knowledge acquisition from natural-language texts. Systems are being developed to sift through masses of written documentation and extract key elements of the enormous body of knowledge stored in that form. One thing that researchers are starting to realize, however, is hardly surprising in light of the analysis presented in this thesis: "les concepts ne sont pas déposés dans les textes mais littéralement construits par le lecteur en interaction avec le texte. Une maîtrise suffisante du domaine, ainsi qu'une connaissance des conditions de production, s'avèrent préalables à la reconstitution des concepts qui sont à l'oeuvre" [Pacquin 1991, 319]. In other words, knowledge is needed to extract knowledge. Consequently knowledge-acquisition systems of this kind will require knowledge bases of their own.

Ultimately, other information processing systems might be plugged into a distributed-knowledge network. Management-information systems and project-management programs are obvious candidates. In addition, modern expert systems, which rely heavily on knowledge bases,⁶¹ could tap into a much richer array of information on which to base their conclusions once they are part of a distributed-information network. As discussed in the previous section, machine-translation systems might also make use of the same knowledge base.

Distributed-information networks offering richer information and better communication among users may fundamentally alter translation practice. As Mayorcas-Cohen [1986, 66] speculates:

I think that the most significant way in which technology will change our behaviour and way of doing things is that we will cease to regard each translation as a new task to be tackled on an *ad hoc* basis.... The whole approach to translation will change: it will be possible to organise

61. To such an extent that the term "knowledge-based system" seems to be becoming near-synonymous with "expert system."

translation on a more systematic basis, and to make far better use of all the know-how that the community of translators has acquired over the years.

5.3 Design of Knowledge-Base Systems for Translators

This section will discuss some basic issues related to the design of knowledge-base systems for translators.

5.3.1 Basic Structure

There seems to be common agreement among terminologists and technically literate translators that the new generation of term banks must be conceptually based. As discussed extensively earlier in this thesis, this means that those term banks will in fact be knowledge bases. While the label "conceptually based" has been rather loosely applied,⁶² this thesis has assumed it refers to systems that conform to the basic semantic network model — a series of nodes representing concepts and links representing relations. The nodes themselves usually employ a frame-like structure to store information relevant to the concept.

Shreve and Vinciguerra [1990, 299] and others have pointed out that it can be useful to consider each node in a knowledge base as an object in an object-oriented system. Each such object names a concept, describes the concept, organizes the concept into a hierarchy, relates the concept to other objects, and constrains the behaviour of the concept.

62. For instance, TERMIUM has been described as a conceptually based system [Pavel 1992]. Some of the inadequacies of TERMIUM from a conceptual point of view are noted in Meyer et al. [1992b].

There are many ways that this basic design can be realized. Attempts have been made to implement this architecture using standard database systems (for example, Ahmad et al. [1990]; Freibott and Heid [1990]). In contrast, the work described in this thesis, as well as that of Meyer and her colleagues, relies on a system developed for artificial intelligence research (viz. CODE). CODE itself has similarities to an object-oriented system [Skuce 1990, 4].

Another area of common agreement regarding the design of knowledge-base term banks is that they must make better provision for the exchange and sharing of their knowledge. This is in accordance with the principle of *open-endedness* discussed earlier. Yet in Canada, for instance, Termium and the BTQ, the country's two major terminological bases, do not yet participate in systematic information exchange. Sager [1990, 227] observes that

There is now a greater will to share, sell and exchange terminologies, and the view of terminology as the private property of its compiler is dying out.... The exchange of terminological data is one of the most frequently recurring topics at international meetings involving term bank organisers, and a major factor in the design of term record formats for new term bank projects is compatibility with existing terminological data banks.

To this end, there has been extensive activity aimed at developing standards for terminology banks. Notably, the International Organization for Standardization (ISO) has published a number of official standards, including a working document entitled *Systèmes de notions, élaboration et représentation*.⁶³ It has also developed a special format for the exchange of terminological information called MATER, as well as a reduced format using only standard ASCII characters called MicroMATER.⁶⁴ While

63. A discussion of the role of conceptual analysis in terminology standards may be found in Miller et al. [1991]. See also Rondeau [1981, 89ff] and Sager [1990, 114ff] for a more complete discussion of standardization issues in terminology.

64. See Melby [1990].

these formats have been used primarily in conventional record-oriented databases, the Translator's Workbench project supports MATER, for example, and Shreve and Vinciguerra [1990] adopted MicroMATER for their translator-oriented knowledge base (basically a hypertext semantic network). A new "markup" language for terminological purposes is being developed by ISO in collaboration with a number of national and international associations. Called TEI-TERM ("TEI" standing for "Text Encoding Initiative"), it is probably destined to replace MATER and MicroMATER. It is based on the Standard Generalized Markup Language (SGML) and aims to facilitate the exchange of formatted information between term bases and, in fact, between all text-based systems.⁶⁸ As Sager [1990, 137] points out, however, the growing trend towards shareable terminological information increases the danger of poorer quality work being more widely disseminated.

Another basic consideration in the architecture of knowledge bases for translators is *flexibility*. The structure of the knowledge base should make it easy to produce output in novel ways and formats, such as customized glossaries and terminological thesauri.

5.3.2 Modelling of Interlingual Equivalence

Obviously, a knowledge base specifically designed for translators should be bilingual or multilingual. The question of how best to represent the conceptual structure of the chosen subdomain in two or more languages is a very interesting question that will only be covered in a superficial manner here. In a more traditional term-oriented bank, the question does not arise so acutely, since the "objects" of the database are the bi/multilingual records themselves. In a conceptually oriented system, however, the objects are the *concepts*, and, from the point of view of certain theorists, concepts are tied to particular languages. As discussed in Chapter 1, in fact, some writers on translation

68. See Wright and Melby [1992].

would argue that any attempt to establish equivalents between languages out of context (i.e. in the absence of an actual "text act") is, at the very least, artificial. However, while acknowledging the artificiality and subjectivity of cross-language equivalence, terminologists and lexicographers consider it worthwhile to study the possibilities, leaving the ultimate choice to the end user.

Theoretically it is possible to distinguish between total equivalence, partial equivalence, and absence of equivalence. However, these categories are of limited validity, since only the extreme points are fixed. Between them, there may exist any degree of equivalence whatsoever; whether, despite this, an equivalence proves acceptable or otherwise depends on many factors which the translator and lexicographer ... must take into account. Without the context it is not possible to judge whether, in a given situation, a conceptual difference is important [Picht and Draskau 1985, 104].

There are two basic ways two languages may differ in the way they partition a given portion of reality:

- Lexical gap, involving the complete absence of a term in one language;
- Partial equivalence, involving either:
 - a more specific/general term in one language that may need qualification, or
 - terms in the two languages that are essentially equivalent, but which present some difference of nuance or usage.

Note that it would inaccurate to say, in the case of a lexical gap, that a *concept* is missing in one language (an extreme Whorfian viewpoint), because the idea can undoubtedly be described in some way; a concept can exist without a standardized term with which to label it.

Fortunately, in technical domains (i.e. languages for special purposes or "sublanguages" in Kittredge's sense), there is a greater degree of correspondence between

languages because of the international scope of most scientific activity and extensive efforts to standardize technical terminology. However, such phenomena as those noted above still do occur. One oft-cited example is French terms such as "informatique," "bureautique," and "terminotique." Even though the language of computing is dominated by English, concise English equivalents for these handy French terms do not exist; one is forced to employ such compounds as, respectively, "data processing" or "computer science," "electronic office technology," and "computerized tools for terminology." In the author's optionCODE knowledge base, there was one clear case of a French concept with no obvious English equivalent: "option synthétique," which refers to the simultaneous purchase of shares and an option to hedge against risk on the same shares. English must use a descriptive phrase to render this idea ("hedge involving options" in optionCODE).

Proposed cross-language equivalents should also carry some indication of reliability. This is commonly done in traditional term banks; Termium employs a system that includes labels such as CORRECT, UNCONFIRMED, TRANSL. SOURCE and AVOID. Reliability information is equally important in knowledge bases for translators, however, because of the conceptual precision. The Terminological Information System of the Council of Ministers of the European Communities, for example, uses a four-level system of "confidence codes"; 0 indicates an equivalent established by the system from an inverted unidirectional link, 1 an equivalent proposed by a translator, 2 an equivalent reviewed and approved by a terminologist, and 3 an equivalent that can confidently be used in final products.⁶⁶

It has been suggested [Meyer 1991] that a conceptual approach is best suited to resolving the problems of the different partitioning of reality by different languages. In fact, Van Campenhoudt [1991, 11] goes so far as to suggest that "seule une approche

66. For more information, see Brøns [1987].

notionnelle permet de résoudre les problèmes traditionnels de la traduction des langues de spécialité (vide notionnelle, intersection partielle, inclusion, etc.)." Van Campenhoudt goes on to say that a conceptual approach also leads to better-quality bi/multilingual glossaries, dictionaries and so on.

Galinski is another strong advocate of the conceptual approach to term banks. He stresses that, among other advantages, a conceptual framework makes it possible to "clarify the equivalency between terms of different languages ... with a higher degree of reliability [and to] find unknown terms for concepts of which only the position within the system of concepts is known" [1988c, 177]. Sager examines both standard database management and information-retrieval packages as the basis for a conceptually based term bank and finds them wanting, then suggests that "semantic networks have considerable potential for a conceptual and cognitive representation in a term bank" [1990, 186].⁶⁷

However, an important issue remains: how are bilingual and multilingual equivalences to be modelled in the knowledge base?

Researchers on bilingualism have long argued about how language and concepts are stored in the minds of bilingual and multilingual individuals. Are the basic concepts common (i.e. not tied to a particular language), or are they language-specific and activated independently? If the latter, how does translation take place? Essentially, these questions are an extension of the relativist/universalist debate discussed in Chapter 2. Whatever model is chosen, a related question is the difference between linguistic and extra-linguistic knowledge.

67. The section in Sager [1990] referred to here (6.3) was drafted by Richard Candeland.

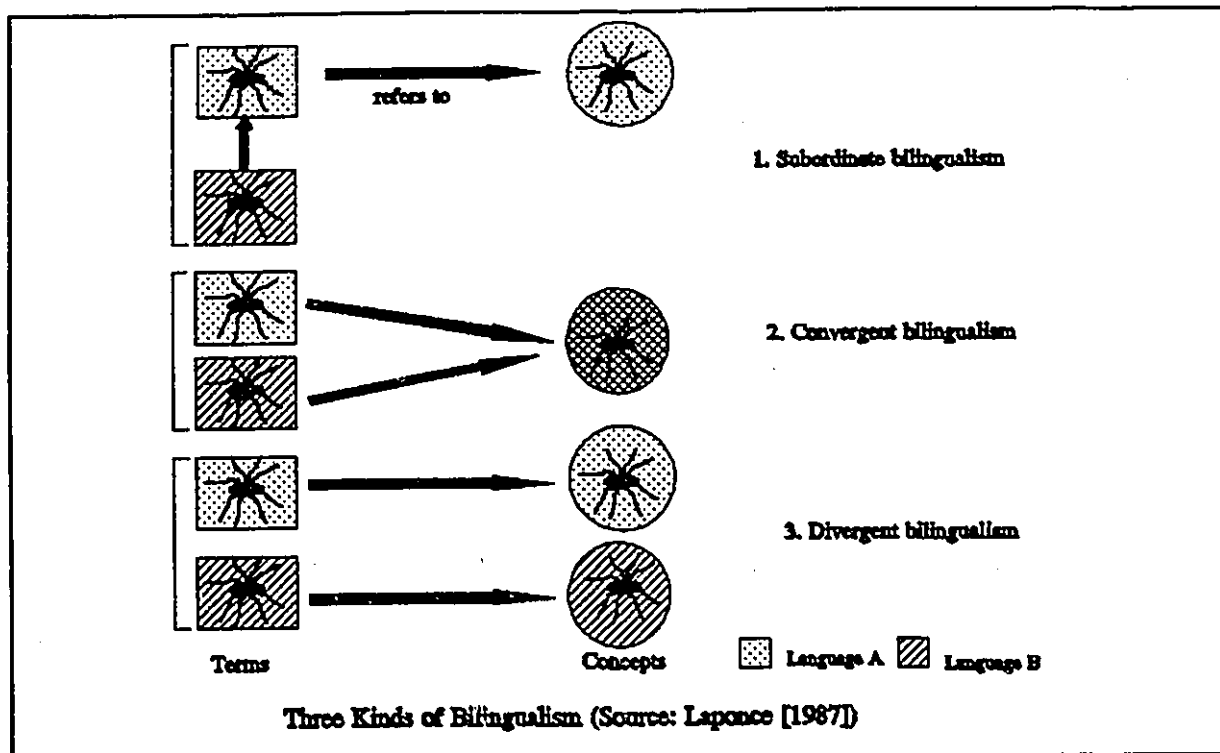


Figure 5-1 Three Kinds of Bilingualism

A general consensus has emerged that there are three basic *kinds* of bilingualism, in accordance with the relationship between the signifier or linguistic sign and the signified or concept (see Figure 5-1). In subordinate bilingualism, the individual has learned an alternative label (signifier) in the second language for the same concept. In convergent or compound bilingualism, the individual is able to apply the second-language label independently and directly to the concept, which remains, however, common to both languages. In divergent or coordinate bilingualism, the individual has independent labels and concepts in the two languages — and the concepts are likely subtly different. Laponce [1987, 32] concludes that "each language has its own association of ideas, each pulls in a different direction — not so much by its signifieds as by their sets of signifiers."

The foregoing discussion suggests that there could be three corresponding models for bi/multilingual term banks (see Figure 5-2). Some traditional term-oriented banks reflect the subordinate bilingualism model, in that the target-language term is simply one

more item of information included in a source-language record with information on the concept in that language (e.g. the BTQ, where English is subordinate to French). Others reflect more the convergent bilingualism model, in that the records themselves are bi/

multilingual, but they describe a single concept assumed to be identical in both languages (e.g. Termium). To my knowledge, no term banks conform to the divergent bilingualism model, which would imply completely independent conceptual knowledge bases for each language involved with some mechanism for establishing possible equivalences between their concepts.

It is clear that, whatever the theoretical attractiveness of independent conceptual ontologies for each language, the latter approach is flawed in practical terms. It would require at least double the development work and would be impossibly unwieldy in implementation. Moreover, it would make one of the main purposes of term banks — the provision of bi/multilingual terminology — a secondary and unnecessarily complicated process.

One solution [Sager 1990, 173ff] might be to establish a language-independent conceptual ontology, then assign records in various languages to the appropriate concept through a code. A specialized concept-term link would be used to indicate the nearest term in another language when the respective conceptual systems did not match. It is

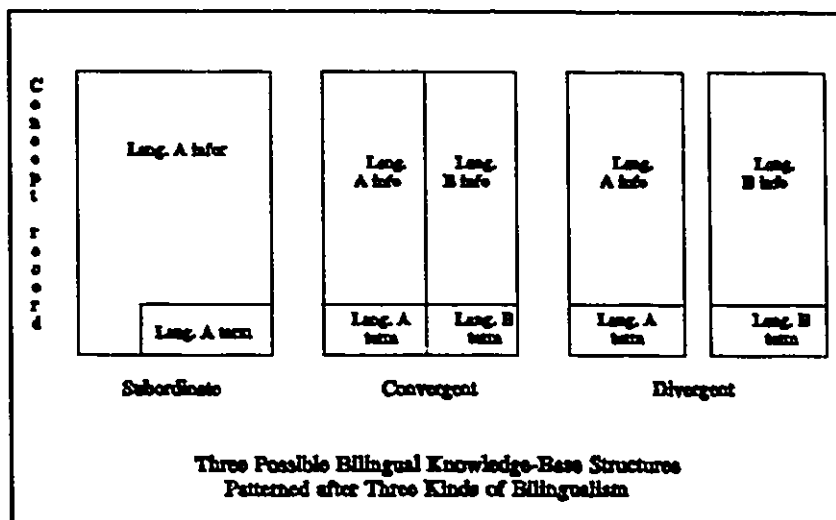


Figure 5-2 Three Possible Bilingual Knowledge-Base Structures

pointed out that the "nearest equivalent" relationship is neglected in most traditional term banks, where equivalence is considered an all-or-nothing phenomenon. In addition, equivalence is considered a unidirectional relationship, unless the term bank is expressly designed to be multidirectional, in which case all concepts are defined in all languages in an identical manner [Sager 1990, 139].

A quite different approach has been proposed by Freibott and Heid [1990]. In their system, based on a commercial database manager, the database "objects" are lexical items that are related one to another through a variety of links such as SYNONYM, QUASI-SYNONYM, EQUIVALENT, QUASI-EQUIVALENT, and VARIANT-OF. The cross-language equivalence links are considered as unidirectional, and the system will propose the "best" candidate when there are no matches satisfying all requirements. However, while the system is ostensibly conceptually based, its reliance on a standard database manager means that it has no inheritance and does not conform to the semantic network model.

One possibility based on the semantic network model is to have a *linguistically integrated* knowledge base, with parallel or divergent concepts for particular languages present as required by their respective ontologies (see Figure 5-3). Concept nodes would be flagged as belonging to a particular language or languages. The user would then be able to specify what language he wanted the knowledge base to reflect during interrogation — in effect, "hiding" the other concepts — but would still be able to access them to determine the best equivalent. He would be able to switch from one language to another and to contrast the conceptual ontologies of different languages on request. It should be noted that such an approach implies that the most limited concept (i.e. that with the smallest extension or the most restricted set of instances) would dictate the "grain size" of division. In addition, the developer would be faced with decisions on *how much* a concept in one language would have to differ from that in another language before it warranted separate representation. It is intended that future versions of CODE

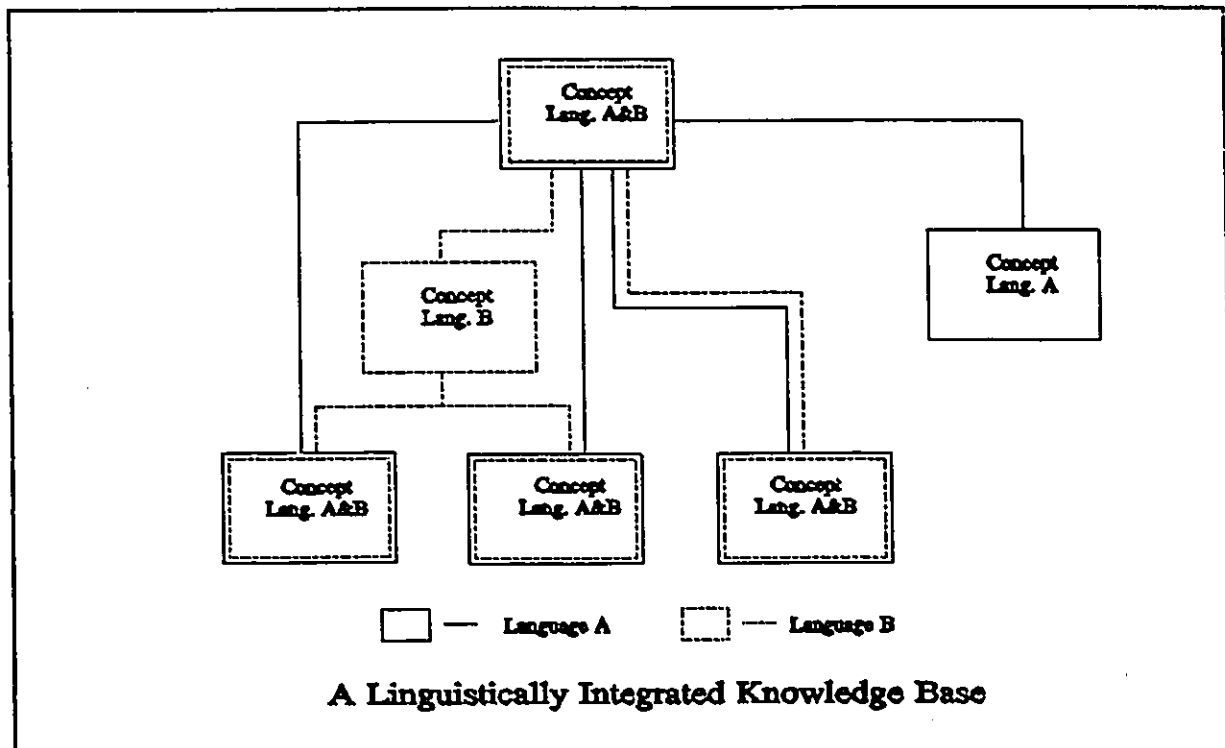


Figure 5-3 A Linguistically Integrated Knowledge Base

would allow the implementation of the integrated approach.⁷¹

Ideally, a knowledge base subscribing to the model outlined above would be based on a "universal" set of categories and high-level concepts. As discussed in Chapter 3, however, the determination of universal "semantic primitives" is a problem that has already occupied philosophers for centuries, and there is little prospect of it being resolved in the near future. Nevertheless, Skuce and Monarch [1990, 15] argue that it is time to start thinking about a universal set of primitive categories, perhaps in conjunction with translators themselves.⁷²

71. It might also be possible to do this with Version 2 of CODE by using the multidimensionality techniques described by Bowker [1992] and treating each language as a separate dimension.

72. See also Skuce [1992].

5.3.3 Content

Terminologists commonly store many kinds of information in their records. Besides administrative information, such as the record number, author and date of entry, it is common to have fields for domain (subject field) and subdomain, sources, reliability codes, and scope notes (i.e. country and level of language). However, a working translator is most interested in four kinds of knowledge when he searches a knowledge base. As discussed in Section 2.3, the first is *specialized lexical knowledge* and the second *specialized extra-linguistic (encyclopedic or conceptual) knowledge*. Recall that "specialized" here means knowledge pertaining to a particular domain or subdomain. As argued in Section 2.3.3, the translator is also interested in *strategic information* — information on how to use the vocabulary in an expert-like way. Fourth, strategic information can be supplemented by *textual information* in the form of actual texts in the languages involved, perhaps even good examples of previous translations or parallel texts. The most efficient way to provide this capability would likely be a separate textual database with index links to the knowledge base.

These four components of the content of knowledge bases for translators — lexical, conceptual, strategic and textual knowledge — have been proposed by several authors. Neubert describes them as the elements of a system that is a software analogue of a translator's competence.⁷⁰ As Shreve and Vinciguerra [1990, 298] note, the kind of knowledge sought by translators varies with the stage of their work — reading, researching, drafting, editing — and with their individual work styles. However, the first two components are clearly the most important. The conceptual information in the knowledge base flows from the conceptual analysis performed by the system's developers and actually determines the overall structure of knowledge in the base. The lexical information (i.e. the actual source- and target-language terms) is, of course, of primary

70. Cited in Shreve and Vinciguerra [1990, 297].

interest to the translator, because, as discussed earlier, the source-language terms often act as a "lexical gateway" to relevant knowledge and because the appropriate target-language terms are critical to the quality of his product.

The optionCODE knowledge base covers essentially only these first two components. While there was some strategic information provided, it was essentially limited to collocational information and a few observations on usage, and the coverage was admittedly incomplete and ad hoc. No textual information was provided, since Version 2 of CODE offers no easy way to link textual passages to the knowledge base (although Version 4 will).

A more complete knowledge base for translators would have to offer a more highly developed strategic-information component. Ideally, collocational information would be broadened to offer "phraseologies" [Harris 1982, H-15] — repertoires of common phrases for a key idea (or concept) that are commonly translated in a certain way. In fact, if the "phrase" corresponds to a fairly well-defined concept, there is no reason why it could not be included in the knowledge base as such. Strategic information could also encompass hints, suggestions or "recipes" for solving common translation problems. In fact, strategic information is really a selected and organized subset of textual information, since the strategies are derived from usage. But, as Harris [1988] has stressed, there is a special perspective gained by having actual texts available for perusal. In my own experience, I have often benefited from consulting the textual database maintained by the Bank of Canada's Translation Section.⁷¹

There are few limits on what might eventually be incorporated into knowledge bases for translators. Next in line might be information in graphical form, such as schematics

71. The software is called *Edibase*, developed by Inform II-Microfor. For further information, refer to Section 4.4.1.

of machinery or other devices, organizational charts and pictures.⁷² The multimedia approach taken by the latest CD-ROM products, some of which incorporate high-quality graphics, animation, and sound, testifies to the wide range of possibilities.

5.3.4 User Interface

As has been stressed earlier in this thesis, a crucial consideration in the design of knowledge-based term banks for translators is the user interface. Since it is the primary channel of communication between the user and the knowledge base, the interface largely conditions the ultimate utility and efficiency of the system.

While it is clearly in the interests of the terminologists involved in computer-assisted concept analysis and the knowledge-base developers themselves to learn the ins and outs of the system they are using, translators as end users should not have to. In most cases, they are interested only in retrieving the knowledge. Ideally they should be able to interact with the system through an interactive natural-language interface designed around their needs. Not only should such a natural-language front end be available to mediate and guide the user's interaction with the system, it should be seamlessly integrated with the other components of the translator's workstation.

A very active area of research is query languages and natural-language interfaces for databases and expert systems. Some systems, for instance, offer the best candidates when there is no perfect match for the search query. Others use transformation algorithms to convert unsuccessful queries into other "semantically close" queries. However, while progress has been made in automating and making more user-friendly the mechanical aspects of searching, there has been less progress in helping users to formulate their *search strategies*. As Hawkins [1988, 33] puts it, "it is as if the

72. For example, the knowledge base described by Freibott and Heid [1990] incorporates pictures linked to database objects.

unsuspecting novice researcher were being guided through the online retrieval jungle, carefully skirting all the pitfalls along the way. Then, when the last and biggest pitfall appears, the guide drops the searcher's hand and runs away." In the case of knowledge-base systems for translators, training will be required before users understand how to utilize the interface optimally. One approach is for the system to develop the search strategy interactively with the user through a series of questions and answers. In effect, an expert system mediates between the user and the knowledge base.⁷³

There is a growing recognition in software development and computer systems design that greater attention must be given to the needs of end users. A user-oriented approach has two dimensions: user friendliness and flexibility.

"User friendliness" has become an over-used term, but it reflects an important idea: that the interface should make it as quick and easy as possible to do the kind of things users want to do and should minimize frustration when things do not work out. As Sager [1990, 194] notes, "the concept of user friendliness has recently been extended to cover not only ease of use but also user-satisfaction with the response given.... 'No Find' is no longer acceptable." In addition, user friendliness means avoiding what Moser-Mercer [1987, 157] sees as the two most common problems in the man-machine interface: the "peephole" effect, where the user feels as though he is looking through a peephole into a large ballroom and so lacking an overview of the system; and the "concrete" effect, where the user feels compelled to conform to the computer's way of doing things rather than using the system the way he wants to.

Flexibility is the necessary counterpoint to user friendliness. First, there should be a variety of ways to interrogate the system. It has been argued in this thesis that translators have two different modes of interacting with knowledge bases (ad hoc and

73. See Sager [1990, 161] and Hartley [1987].

background browsing; see Section 1.3.1), and the interface should facilitate both modes. In terms of the interface, this means that the system should offer what Shreve [1990, 305] calls two ways of navigating the knowledge base: through its formal organization and through a hypertext-like facility. This echoes what Sager [1990, 196] terms the dual "access paths": to the term and to the concept. Access to the term (the "lexical gateway" mentioned earlier) may be enhanced through a conventional database-management system acting as the lexical front end to a knowledge base, keeping track of the location and status of the various lexical items contained in the knowledge base — terms, abbreviations, synonyms, orthographic variants, etc. A browsing capability is equally important, particularly for occasions when the translator or other user is faced with the task of becoming an "instant expert" in a technical field. Yet some systems designed specifically for translators do *not* offer browsing capabilities.⁷⁴ Flexibility should also be reflected in multiple possibilities for hardcopy output [Sager 1990, 189].

Second, provision should be made for the varying needs of different users. According to Shaikevich and Oubine [1988, 10], it is now "common knowledge" that *dictionaries* themselves must address the needs of various categories of users. In this thesis, we have been concerned primarily with how translators' needs can be met by a knowledge-base system, but, if it is to be multifunctional, a knowledge-base system should be able to adapt to a variety of users. A partial list of potential users of a knowledge-base system might include subject specialists, technical writers, professional communicators, translators, terminologist, lexicographers, information and documentation specialists, language planners, language teachers, researchers in applied linguistics, and a variety of generalist occasional users.⁷⁵ Each type of user may have a different need with respect to some aspect of the knowledge base. For instance, Sager [1990, 196] observes that a scientist probably wants a very precise definition of a concept, a translator a less

74. Such as that described in Hartley [1987].

75. This list is based mainly on Sager [1990, 197ff].

technical one, and an undergraduate student an encyclopedia-like entry. As well, a fundamental distinction should be made between monolingual users (e.g. technical writers, standardization specialists) and multilingual users (e.g. translators and terminologists). In the latter category, terminologists require both input and retrieval facilities, while translators are usually concerned with retrieval. In short, because there may be a wide range of needs, ideally the system interface should be able to adjust to a number of basic user profiles. In addition, these profiles should be individually customizable, so that they can conform to the particular likes and dislikes and workstyles of individual end users of the system.

Finally, KBST developers can also benefit from more general research into the optimal design of user interfaces. For instance, an excellent book by Shneidermann [1992] lays out a number of basic principles, such as the need to distinguish clearly between system-engineering goals and human factors. Shneidermann proposes and develops "eight golden rules" of interface design [72ff]:

1. strive for consistency
2. enable frequent users to use shortcuts
3. offer informative feedback
4. design dialogues to use closure (in other words, sequences of actions should have a clear beginning, middle and end)
5. offer simple error handling
6. present easy reversal of actions
7. support internal locus of control (in other words, let users feel in control of the system, rather than vice versa)
8. reduce short-term memory load.

A KBST that respects these rules will be easier and more enjoyable to use, and so *will* be used.

5.3.5 Practical Aspects of Knowledge-Base Design

Obviously, fast response time is vital. No matter how sophisticated it is, a system that takes too long to provide an answer or is sluggish to use will be shunned by users. For the ad hoc search mode of translators,⁷⁶ the system must offer the translator speed and ease of access. He must be able to retrieve the desired item of information at least as easily and quickly as he could by turning to a conventional dictionary. "The most important characteristic of an information retrieval system, after the quality of the information," stresses Vaumoron [1988, 38] "is its access time; this is the time elapsed between the moment the translator stops typing the next word in order to seek information, and the moment he can resume typing after having brought to the screen the information needed." That is why a knowledge base for translators must offer a comprehensive and efficient indexing and search mechanism, or what was earlier termed a "global lexicon."⁷⁷ Measuring system response as the time it takes from the decision to consult the knowledge base to the return to the home application (presumably some kind of word-processing program) underscores the need for seamless integration of the knowledge-base system into the user's working environment. For instance, the translator should be able to signal that the word on which his word processor's cursor is sitting should be used as a search key. But integration means not only that entry and exit from the system is quick and easy, but also that there is an efficient mechanism for copying information from the knowledge-base into the home application.

76. See Section 1.3.1.

77. See Section 4.1.2.1.

Another desirable feature would be automatic lemmization based on morphological and syntactic analysis to derive the root form of the search key.⁷⁸ In order to guarantee acceptable performance, it might be feasible to have a background parser constantly at work during CPU idle time, so that the lemma of the current word would be instantly available should the user decide to use it as a search key. A background analyzer might also help improve the performance of grammatical assistance and machine translation modules, should these be incorporated into the workstation environment.

For the second mode of translator search — the background browsing mode — response time within the system would be the critical factor. Operations such as opening and switching between graphs and browsers should be speedy, and screen updates should be fast. If the system is busy doing something, it is important that the user is so informed, so he does not worry that the system has "hung." Such individually minor but cumulatively important details have often been overlooked by system developers. Speaking primarily of machine-assisted translation systems, Moser-Mercer notes that designers have concentrated their efforts on the linguistic aspects of systems, relegating psychological factors for later. "After 30 years of research in this area," she exclaims, "the psychology of the user ... cannot be further neglected" [Moser-Mercer, 1987, 161].

While not a design factor, there is another important consideration that has been alluded to earlier in this thesis: users of knowledge-base systems will need training to learn how to use them. Because of the way they usually work, translators will inevitably start off by using a knowledge-base system as a "super-dictionary"; in other words, they will want to look up terms. Using it efficiently as an online source of expert technical knowledge will require a shift in mindset. Translators will not instinctively know the

78. Lemmization is the conversion of an inflected word to its "lemma" or standard dictionary form. For example, plural words would be made singular, modifiers made singular and masculine (in most cases), and verbs converted to infinitives. Lemmization may also involve reduction to a stem; e.g., the conversion of "translator," "translating" and "translation" to "translate."

kinds of things they can look up in such a knowledge base, nor will they automatically recognize opportunities for enhancing their knowledge. Hence they will need training not only in how to operate the system, but also in strategies for ferreting out what they are looking for. In addition, they will need encouragement (i.e. both incentives and evidence of usefulness) to convert from their traditional time-tested techniques for acquiring expert knowledge to the world of conceptual browsing. As discussed in Section 5.1.1, knowledge bases for translators represent a "third-stage" technological innovation; a change in behaviour on the part of users is required before their full power can be exploited.

CHAPTER 6

Conclusion

6.1 Recapitulation

Chapter 1 laid the groundwork for this thesis, setting out a theoretical aim — to study the role of knowledge in the translation process — and a practical aim — to explore the applicability of research in knowledge representation associated with artificial intelligence to the development of knowledge-base systems for translators (KBSTs). It noted that translators have two main modes of knowledge acquisition: ad hoc searches to find answers to specific problems, and background browsing to broaden their personal knowledge in a particular field. These two knowledge acquisition activities may be linked to a particular translation job or may be related to the translator's ongoing efforts to keep abreast of the latest developments in a particular field. After examining translators' traditional sources of knowledge, the chapter suggested that knowledge-base systems designed specifically with the needs of translators in mind could represent a superior source of knowledge in specific subdomains for three main reasons: they are potentially richer sources of knowledge, they can offer more flexible and efficient access, and they are inherently open-ended.

Chapter 2 explored the importance of knowledge for the translation process. After examining the treatment of knowledge in the literature on translation, it concluded that adequate knowledge is extremely important to good translation. The reason is that understanding, which is a critical stage of the translation process, takes place through the interaction of the meaning of the text with the reader's knowledge. Consequently, knowledge maximization can be considered a viable translation strategy for countering the "opaqueness" of a text, or its resistance to understanding. From the viewpoint of knowledge, opaqueness can be considered to have three parameters: the knowledge profile assumed by the writer for the reader, the translator's actual knowledge profile,

and the "clarity" or quality of writing of the text. The chapter also suggested that knowledge and understanding are, if anything, more important in technical translation than in general translation. An informal experiment with a fourth-year translation class showed, among other things, that in the absence of adequate knowledge, a translator is forced to adopt more of a "transcoding" or word-bound approach to translation. The goal of a knowledge-base system for translators, therefore, must be to provide enough knowledge to allow the translator to engage his "interpretative" or meaning-based mode of translating.

Chapter 3 investigated knowledge and knowledge representation from the viewpoint of Artificial Intelligence. A number of practical and philosophical issues related to knowledge representation were discussed, and three ways of representing knowledge — predicate logic, semantic networks, and frames — were described. It was noted that current research tends to combine these three approaches. The chapter ended with a discussion of conceptual analysis, noting that it underlies not only the work of knowledge engineers, but also terminologists and future KBST developers.

Chapter 4 introduced a knowledge-management system called CODE and described a knowledge base on stock-market options called optionCODE, which was developed by the author to explore the principles and problems of designing and using KBSTs. A number of methodological issues reflecting the author's experience during this exercise were discussed. Lastly, another informal experiment was described where the same text was translated by three different translators, two using optionCODE as their sole source of knowledge, and the third using standard reference tools. The experiment indicated that, whatever its shortcomings as a KBST, optionCODE appeared to provide sufficient knowledge-retrieval capacities to perform as well as, if not better than, standard reference sources.

Chapter 5 explored the concept of the translator workstation, surveyed current developments in this area, and discussed how KBSTs might be integrated into the translator workstations of the future. The potential for integrating KBSTs with machine translation systems, document-production systems and distributed-knowledge networks was also discussed. A number of considerations in the design of KBSTs were investigated, including their basic structure, their content and their interface. Lastly, some ideas on how best to model bi/multilingualism in such a system were offered.

6.2 Recommendations for CODE

In the course of this thesis, a number of ways CODE might be improved to better serve as a KBST have been suggested. The main ones are summarized below.

6.2.1 Structural Considerations

- **Improve the implementation of associative links between CDs (conceptual descriptors) — i.e. links other than supertype to subtype or "isa" — so that these relations a) are readily apparent to the user, b) may be easily explored by the user, and c) may be graphed in a variety of ways by the user, including in a semantic network format.**
- **Develop the hypertext capacities of the system, so that, for example, a user may quickly jump to a relevant area of the knowledge base by clicking on a clearly indicated hypertext link. In addition, ensure easy hypertext access between all system displays (graphs, browsers and views).**

- Support local renaming of properties in a CD, so that a property's name may be altered as appropriate when it inherits to that CD, while maintaining continuity of the link (discussed in Section 4.3.8).
- Provide the possibility of multiple levels of representation in the knowledge base, levels that could be specified by the KBST developer and changed as required by the user. This capability could be used to implement various levels of abstraction in the knowledge base (see Section 4.3.3) or to model the different partitioning of reality by different languages (see Section 5.3.2).
- Provide an elegant way to include bibliographic references in the knowledge base.
- Improve CODE's support of accented foreign-language characters. In particular, ensure that they can be easily entered, that they are properly indexed, and that they will appear on hardcopy.⁸²
- In accordance with the principles of open-endedness, ensure that the knowledge contained in a knowledge base can be exported in a variety of forms and formats. In particular, the MATER, MicroMATER and new TEI-TERM standards should be supported (see Section 5.3.1).

6.2.2 Interface Considerations

- In accordance with the principle of the "lexical gateway" (see Sections 2.5.3 and 4.4.4), provide a global lexicon with multiple indexes to allow the user to quickly locate all occurrences of a particular word or phrase in the knowledge base in a specific context (e.g. in the body of a particular property or category of

⁸² This is not a small request, since the handling of non-English characters is an incredibly confused area. For an overview of the problems and possible solutions, see Melby [1992].

properties). Support the use of wildcards in such queries. Provide a hypertext link to allow the user to jump directly from the result of such a request to the corresponding location in the knowledge base.

- Improve the flexibility and adaptability of the user interface by making it more customizable, both by the developer (in accordance with client needs) and by the user (in accordance with differences in individual workstyles).
- Consider adding a front-end query system to CODE that would provide a more sophisticated interface for users (particularly novice users). Investigate how such a query system could permit users to make queries with partial or imprecise information. In general, make the system more friendly to novice users.
- In light of translators' two different modes of knowledge acquisition (see Section 1.3.1), facilitate both ad hoc searches for particular items of information and background browsing for conceptual learning. For example, both ad hoc searches and background browsing would benefit from improved hypertext links, including clear indications that such links exist (i.e. hypertext "buttons") and efficient means of navigation between links. KB developers must keep the needs of their target clients in mind and must try to *anticipate* the things they will want to do with the KBST (without *limiting* the possibilities, of course!).

6.3 Implications of Hardware and Software Developments

It was mentioned in Section 5.1.1, that the gap between high-end industrial workstations and personal computers (PCs) is narrowing rapidly in terms of power and capabilities. At the same time, the prices of PCs are declining precipitously. Indications

are that these trends will not only continue, but accelerate.⁸⁰ It will not be too many years before the distinction between industrial workstations and PCs is completely erased.

What this means for KBSTs is that the typical translator working at his computer will eventually have the power and speed to run the kind of software described in this thesis. Undoubtedly there will be a wide range of other sophisticated software tools at his disposal, some of them belonging to the types discussed in Section 5.1, others of types yet to be conceived.

Another area where exciting developments are taking place is the organization and presentation of information on the computer. This is a vast topic, but it is worth pointing out some interesting initiatives. The Natural Language Theory and Technology Group of Xerox PARC is working on the concept of the "information theatre," which allows a user to view his information as three-dimensional structures that stretch, slide and spin.⁸¹ This approach has clear relevance to the display of hierarchical structures in knowledge bases. There are some innovative ideas being explored in the area of information retrieval: for example, intelligent agents and even "friendly" viruses that work behind-the-scenes to search out and organize a user's information. Research is also under way on how to improve responses to queries for text-based information: vector and clustering analysis (retrieving documents that most closely match queries), pattern matching (using algorithms to automatically search for variants of the query), and "snippet browsers" that show the user fragments of text and update the request on the fly as the user indicates what is interesting and what is not. This research is complemented by work on the ergonomics of user interfaces, reflecting a shift in focus from what *programs* can do to what *people* want them to do and *how* they want them to do it. All

80. To cite just one example, the X Window graphics operating system is now available to PC users.

81. Described in BYTE magazine, November 1992.

of these areas of research should lead to more efficient tools for all information users, including translators.

6.4 Directions for Future Research

Several possible areas for future research are suggested by the topics explored in this thesis.

- **Translators' knowledge needs**

It was argued in Section 5.3.3 that translators have two modes of knowledge acquisition and two purposes in knowledge acquisition. While translators require several different kinds of knowledge, it was also suggested that they will expect four types of knowledge from a KBST: specialized lexical knowledge, specialized extra-linguistic (i.e. encyclopedic or conceptual) knowledge, strategic information on terminology and translation, and textual information in the form of actual texts. All these topics are well deserving of investigation. This research would be useful not only to KBST developers, but also to terminologists, lexicographers, and translation theorists.

- **Translators' workstyles**

This issue is related to translators' knowledge needs. It was suggested that KBSTs should take into account translators' particular working habits and, moreover, that KBST interfaces should be able to adapt to *differences* in translators' workstyles. However, what is distinct about translators' workstyles and how they vary among translators are little-explored questions.

- Methodology for Developing KBSTs

A number of methodological issues related to the design of knowledge bases for translators were discussed in Section 4.3. This is a fertile area for future research, because a solid framework for this methodology remains to be developed. It will certainly involve cross-fertilization between terminologists and knowledge engineers, a "natural symbiosis" that is being explored by several researchers. Among the most important methodological issues are definition construction (see Eck [1993]), modelling of multidimensionality (see Bowker [1992]), and incorporating two or more languages into the conceptual ontology (see Section 5.3.2).

- Training

It was pointed out in Section 5.3.5 that translators will have to be trained to use KBSTs profitably. There are two aspects to this training. First, on a technical level, they will need training in the mechanics of the system: how to start and end a session, the various search procedures available, getting around within the system, producing the various displays and graphs, adjusting the screen layout, etc. On a conceptual level, they will need to be trained to use KBSTs efficiently as an online source of expert technical knowledge. This will require a shift in mindset and workstyle, since translators will not instinctively know the kinds of things they can look up in such a knowledge base, nor will they automatically recognize opportunities presented by the particular text for enhancing their knowledge. In other words, they will need guidance in knowledge acquisition strategies. It is clear that research on how best to train translators (and other users) to take full advantage of such systems would be very useful.

6.5 Final Word

Sager suggested [1988, 5] that a generation will pass before translators fully embrace the new tools being developed for them by terminologists and knowledge engineers, or, as Otman [1991, 20] put it, before "traducteurs" become "traductiens." The first computer-technology revolution — word processing — has transformed translating forever. The appearance of sophisticated tools intended specifically for translators and other wordworkers — including online dictionaries, thesauri and spellcheckers, automatic text-processing software, and CD-ROM-based references, to mention only a few — may be regarded as a second major advance, allowing translators to do things that they could not do before. But computer-based knowledge acquisition represents an even more radical change; it is a "third-stage" technological innovation, one that requires a change in behaviour on the part of users before its full power can be exploited (see Section 5.1.1.1).

One positive aspect of this revolution is that it may bring about a reconciliation or convergence of interests between terminologists and translators. Terminologists have long been concerned with the conceptual analysis of their subject fields, but translators have often displayed impatience with their methodical approach ("Enough of the conceptual stuff; just give me the target-language equivalent!"). As this thesis has demonstrated, however, translators *do* need conceptual knowledge. Now the means are at hand to provide it to them in a relatively convenient, efficient form.

Translators must seize the day, however. They have every reason to take an active part in the development of these knowledge tools. Terminologists and other knowledge-base developers should welcome their input, since the result will be to make the systems produced by their efforts more useful to the people who will actually use them. "The incorporation of data processing into translating is inexorable," wrote Moser-Mercer

[1987, 158]. "It is in part up to the translator ... to shape this development rather than be shaped by it."

APPENDIX I - Documents for
Fourth-Year Translation Class Experiment
(See Chapter 2, Section 2.5)

I.1 Original French Text ("Bourse")

Opérations mixtes haussières et baissières

Il existe plusieurs types de stratégies "d'opérations mixtes" répondant aux objectifs et prévisions variés de l'investisseur. Le terme "opérations mixtes" s'explique par le fait qu'il s'agit de stratégies consistant, en règle générale, en l'achat et en la vente simultanés d'options d'achat ou d'options de vente (mais non pas des deux types ensemble) portant sur les mêmes valeurs sous options. Les *positions mixtes* sont des stratégies qu'adopte le vendeur désireux de limiter ses risques.

Le but principal poursuivi par l'investisseur qui effectue des opérations mixtes est de percevoir le revenu de la prime de risque à titre de vendeur d'options tout en limitant les risques inhérents à la vente découverte d'options. Par exemple, le vendeur d'une option d'achat ayant un prix de levée de 20 \$ peut décider d'acheter simultanément une option sur la même action et ayant un prix de levée plus élevé, 22½ \$ ou 25 \$, par exemple. Ainsi, s'il se trompe et que le cours de l'action sous option de l'option d'achat qu'il a vendue à découverte augmente, cet investisseur détiendrait une position « stop » lui permettant de limiter ses pertes : il détient en effet une autre option d'achat pour le "rachat" des mêmes actions. Il s'agit là d'une *opération mixte baissière* puisque le vendeur adopte une attitude baissière sur le cours des actions sous option de sa position. En d'autres termes, il s'agit d'une stratégie qui s'avère fructueuse lorsque le cours de la valeur sous option chute modérément.

Quant à l'*opération mixte haussière* dans le contexte d'options de vente, elle consiste en la vente d'options de vente et en l'achat simultané, sur les mêmes actions, d'options de vente ayant un prix de levée inférieur à celui des options vendues.

I.2 Official Translation of "Bourse" Text

Spreads: Bulls and Bears

There are several types of *spread* strategies depending upon the investor's goals and forecasts. The term "spread" arises from the fact that spread strategies generally consist of the simultaneous buying and writing of call options or put options (but not puts and calls together) on the same underlying stock. Spreads are strategies for the option writer who wants to limit his total risk.

The basic goal of the options spread trader is to collect time premium revenue as an option writer but reduce the risks associated with "naked" option writing. For example, the writer of a call option with a strike price of \$20 may decide to simultaneously buy an option on the same stock but with a higher exercise price of either \$22½ or \$25. Thus, if he is wrong and the price of the stock underlying his short call option should rise, the investor will have a "stop loss" position since he holds another call option to "buy back" the same stock. This is known as a *bear spread* since the "spreader" is by definition bearish on the value of the underlying stock. That is, it is a profitable strategy if the price of the underlying stock drops moderately.

A *bull spread* consists of writing put options and simultaneously buying put options on the same stock with a lower exercise price.

APPENDIX II - Documents for CODE Experiment
(See Chapter 4, Section 4.4)

II.1 Original French Text (Excerpt from Bertrand Jacquillat, « Les options négociables sur action et sur indice: de la théorie à la création d'un nouveau produit financier », *Banque* N° 397 (juillet-août 1980)).

La valeur intrinsèque d'une option d'achat

La valeur intrinsèque d'une option d'achat mesure la valeur pour l'investisseur d'une option qu'il achèterait et exercerait immédiatement. Si le cours de l'action est supérieur au prix d'exercice, cette opération suivie de la vente de l'action produit un revenu instantané. Dans ce cas l'option a une valeur intrinsèque positive et se trouve « dans le cours ». Si le prix d'exercice de l'option est supérieur au cours de l'action, l'option a une valeur intrinsèque nulle et se trouve « hors des cours ».

La valeur d'un option non arrivé à l'échéance doit toujours être supérieure ou égale à sa valeur intrinsèque. Si la valeur de l'option est inférieure à sa valeur intrinsèque, des arbitragistes s'engageront simultanément dans les deux opérations suivantes:

- achat de l'option et exercice instantané de cette option;
- vente de l'action à un prix supérieur au cours de l'option plus son prix d'exercice.

Ils utiliseront les actions obtenues lors de l'exercice de l'option pour livrer celles qu'ils ont simultanément vendues, ce qui aboutit à un profit sans risque, de telles possibilités d'arbitrage sans risque empêchent l'avec, le prix de l'option, de baisser au-dessous de sa valeur intrinsèque.

La valeur temporelle d'une option

Les opérateurs évalueront une option avec une surcote par rapport au revenu qu'ils pourraient obtenir immédiatement en l'exerçant, dans la mesure où ils estiment qu'il y a une chance que l'exercice de l'option soit encore plus profitable s'ils intervient à une date ultérieure. Lorsque l'avec d'une option est supérieur à sa valeur intrinsèque, on dit que l'option a une valeur temporelle positive.

La valeur totale d'une option

La valeur totale d'une option est la somme de sa valeur intrinsèque et de sa valeur temporelle. Elle tend donc vers sa valeur intrinsèque au fur et à mesure que l'on se rapproche de la date d'exercice du contrat.

II.2 Two Sample Translations of French Text by David R. Miller

II.2.1 A Respectable But Frequently Transcoded Translation

Intrinsic Value of a Call Option

The intrinsic value of a call option measures the value it offers to the investor who might buy and exercise it immediately. If the price of the stock is higher than the exercise price, such an operation followed by the sale of the stock would produce an immediate income. In this case, the option has a positive intrinsic value and is "in the money." If the exercise price of the option is higher than the price of the stock, the option has zero intrinsic value and is "out of the money."

The value of an option before maturity must always be higher than its intrinsic value. If the value of an option is lower than its intrinsic value, speculators will simultaneously carry out the two following operations:

- purchase and immediate exercise of the option;
- sale of the stock at a higher price than the option price plus its exercise price.

They will use the shares obtained from exercising the option to deliver the shares they simultaneously sold, producing a risk-free profit. The possibility of such arbitrage prevents the premium, the price of the option, from falling below its intrinsic value.

Time Value of an Option

Investors will evaluate an option with a premium relative to the income that they could obtain immediately by exercising the option in accordance with what they feel the chances are that exercising the option will prove even more profitable if done later. When the premium of an option is higher than its intrinsic value, the option is said to have a positive time value.

Total Value of an Option

The total value of an option is the sum of its intrinsic value and its time value. Thus the total value will gradually approach the intrinsic value as the exercise date of the contract draws nearer.

II.2.2 A More Knowledge-Based Translation⁸²

Intrinsic Value of a Call Option

The intrinsic value of a call option refers to the value it represents to an investor were he to buy it and exercise it immediately. If the market value of the underlying stock is higher than the exercise price, an immediate income is produced by doing that — buying and exercising the option — and immediately selling the shares. In this case, the option has a positive intrinsic value and is said to be "in the money." An option with an exercise price higher than the market price of the stock has zero intrinsic value and is said to be "out of the money."

However, the total value of option necessarily remains above its intrinsic value prior to its expiry. If the value of an option were to fall below its intrinsic value, speculators would simultaneously:

- purchase and immediately exercise the option;
- sell the shares involved at a price higher than the option price plus its exercise price.

The shares obtained from exercising the option would be used to make good on the simultaneous sale, producing a risk-free profit. The possibility of this kind of arbitrage makes it impossible for the option premium (the price of the option) to fall below its intrinsic value.

Time Value of an Option

In evaluating an option, investors will factor in an amount above the income that could be obtained by exercising the option immediately; this additional value reflects the perceived likelihood that exercising the option will be even more profitable in the future. An option whose premium is higher than its intrinsic value is said to have a positive time value.

Total Value of an Option

The total value of an option is its intrinsic value plus its time value. The total value will gradually decline to reach the intrinsic value by the expiry date of the contract.

82. "More knowledge-based" in that the present author has made a number of adjustments (such as expansion and suppression of information, reordering of information, shifts of perspective, etc.) that are possible only because he possesses some degree of expert knowledge and experience in the subject field. See Section 2.5.2.

II.3 Translations by Experiment Subjects (A, B and C)

II.3.1 Translation A

Intrinsic Value of an Option

The intrinsic value of an option indicates the value that the option would have for an investor if he were to buy and exercise that option immediately. If the share price is higher than the exercise price, this operation, followed by the sale of the stock, produces an immediate yield. In such a case, the option has a positive intrinsic value and is "in the money". If the exercise price of the option is higher than the share price, the option has an intrinsic value of nil and is "out of the money".

The value of an option that has not reached maturity is always higher than or equal to its intrinsic value. If the value of the option is lower than its intrinsic value, the following two operations are simultaneously carried out through arbitrage:

- the option is purchased and immediately exercised;
- the stock is sold at a price which is higher than the combined price of the option and its exercise price.

The shares obtained at the time of exercising the stock are used to deliver those which had simultaneously been sold. The result is a non-risk profit, as such possibilities of non-risk arbitrage prevent the price of an option from dropping below its intrinsic value.

Time Value of an Option

An option with a [surcote] is priced in relation to the immediate yield that can be obtained by exercising it, insofar as it is considered that there is a chance that exercising the option could be more profitable at a later date. When the price of an option is higher than its intrinsic value, the option is said to have a positive time value.

Total Value of an Option

The total value of an option is the sum of its intrinsic value and its time value. Therefore, as the exercise date of the contract approaches, the total value tends to become increasingly closer to the intrinsic value.

II.3.2 Translation B

The Intrinsic Value of a Call

The intrinsic value of a call measures the value of an option that an investor is able to buy and exercise immediately. If the share price [cours de l'action] is higher than the exercise price, this operation, followed by the sale of the shares, produces instant revenue. In this case, the option has a positive intrinsic value and is considered "in the money." If the exercise price is higher than the share price, the option has no intrinsic value and is considered "out of the money."

The value of an option that has not yet expired must always be higher or equal to the option's intrinsic value. If the option's value is lower than its intrinsic value, arbitragists will simultaneously carry out the following two strategies:

- purchase the option and immediately exercise it;
- sell the share at a price higher than the option price plus its exercise price.

The arbitragists will use the shares obtained when the option is exercised to deliver those that they have simultaneously sold, which results in a non-risk profit. Such non-risk arbitrage strategies prevent the option premium from going lower than its intrinsic value.

The Time Value of an Option

Operators will evaluate the time value of an option in relation to the revenue that they can immediately obtain by exercising it, if they believe that exercising the option at a later date will be even more profitable. When the premium of an option is higher than its intrinsic value, the option is said to have a positive time value.

The Value of an Option

The value of an option is the sum of its intrinsic value and its time value. It thus tends to move closer in value to its intrinsic value as the exercise date of the contract approaches.

II.3.3 Translation C

The intrinsic value of a call option

The intrinsic value of a call option indicates to the investor the current value of an option if it was bought or exercised. If the option price is higher than the exercise price, buying the option and subsequently selling it, produces instantaneous income. In this case, the option has a positive intrinsic value and is "in-the-money". If the exercise price of the option is above the option price, the option has an intrinsic value of zero and is "out-of-the-money".

The value of an option which has not yet matured should always be higher than or equal to its intrinsic value. If the value of the option is less than its intrinsic value, arbitragers will simultaneously perform these two procedures:

- buying the option and immediately exercising it;
- selling the stock at a rate higher than the option price plus its exercise price.

They will use the stocks obtained outside of the exercise of the option to deliver those options which they have simultaneously sold, which results in a no-risk profit. These no-risk arbitrage alternatives stop the cum, or call price, from falling below its intrinsic value.

The time value of an option

Hedgers will rate an option with a premium, based on the income they could gain by immediately exercising it, if they feel they could earn a greater profit if they intervened at a future time. When the cum of an option is higher than its intrinsic value, the option is said to have a positive time value.

The aggregate value of an option

The aggregate value of an option is the sum total of its intrinsic value and its time value. As the exercise date of the option contract approaches, this total gradually comes into line with the intrinsic value.

II.4 Evaluation Sheet (including scores for A, B, and C)Criteria for Translation EvaluationTranslation: A B C**A. Faults****1. Major Faults**

A major fault represents:

- a. a mistake in meaning, i.e.
 - a change in essential elements of meaning
 - the opposite said of what was meant or nonsense
- b. a mistake in language, i.e.
 - an inexcusable lexical or syntactic error

Number of major faults A:0 B:0 C:0**2. Minor Faults**

A minor fault represents:

- a. a mistake in meaning, i.e.
 - a change in meaning that may cause confusion but does not alter the essential elements of meaning: e.g. inaccuracy, ambiguity, unwarranted addition or omission
- b. a mistake in language, i.e.
 - an error of syntax, spelling, punctuation
 - wrong level of language
 - incorrect term
- c. a mistake in style, i.e.
 - inappropriate style to type of text
 - redundancy
 - excessive literalness

Number of minor faults A:9 B:9 C:12**B. Overall Ratings (using scale of 1 to 10)**

1. Overall accuracy of translation A:6 B:7 C:6
2. Overall naturalness of translation A:6 B:7 C:5
3. How well do you feel the translator understood the text? A:7 B:8 C:7
4. Overall quality of translation A:7 B:8 C:6

APPENDIX III - OptionCODE Printouts

The printout on the following pages illustrates how CODE can produce a glossary-style listing. It was produced from the property browser by restricting the display to CDs hierarchically related to option and then limiting the information by a mask to those properties belonging to the categories "Definition" and "Linguistic information." The CDs were then written to an ASCII file, which was subsequently printed without additional editing. Note that accented French characters were filtered out.

Cd Name: T
SubConcepts: concrete entity, abstract entity, time, event, REFERENCES

DEFINITION

definition : An ontology of stock market concepts relevant to options..

Cd Name: concrete entity
SuperConcepts: T
SubConcepts: security, securities contract, actor, institution, portfolio

Cd Name: security
SuperConcepts: concrete entity
SubConcepts: tangible security, derivative security

DEFINITION

definition : [securities] A generic term for any financial instrument traded on a stock exchange [Valentine]..

LINGUISTIC INFORMATION

English synonym : financial asset, financial instrument.
investment instrument..
French equivalent : titre (nm) (de placement), valeur (nf) (mobili re)
[OLF]..

Cd Name: securities contract
SuperConcepts: concrete entity
SubConcepts: derivative security

Cd Name: derivative security
SuperConcepts: security, securities contract
SubConcepts: option, forward contract, futures contract, swap, warrant,
subscription right, convertible security

DEFINITION

definition : A derivative security is a security whose value depends on the values of other more basic underlying variables.... Derivative securities are also known as contingent claims. Very often the variables underlying derivative securities are the prices of traded securities [Hull, 1]..

LINGUISTIC INFORMATION

English synonym : derivative instrument [TERMIUM]..
French equivalent : [derivative instrument] effet (nm) accessoire [TERMIUM].
[derivative product] produit d riv , produit secondaire, produit accessoire [TERMIUM]..

Cd Name: option
SuperConcepts: derivative security
SubConcepts: stock option, fixed-income option, currency option,
stock-index option, gold or silver option, futures option

Kinds: (1) put, call (2) European option, American option (3) exchange-traded option, over-the-counter option

DEFINITION

definition : An option is a formal contract which grants the holder of the contract the right but not the obligation to buy or to sell, as the case may be, a certain quantity of a specific underlying interest (i.e., a security, currency, commodity or stock index) at a stipulated price for a stated period of time. There are two types of options: a call which gives the buyer the right to purchase an underlying interest and a put which gives the buyer the right to sell an underlying interest.... The option holder (buyer) is not obligated to exercise the option and will not do so unless it is advantageous. If the option is not exercised or sold by the option holder before expiration, the option privilege expires and the contract becomes void. The person who grants the option (i.e., writer) must be prepared to honour the terms of the contract if the contract is exercised during the lifetime of the option [CSI, 158]..

LINGUISTIC INFORMATION

English collocations : Options are said to be <on> an underlying interest.

Options may be exercised up to their expiry date.

If not exercised, options expire on their expiry date (not mature).

Options are traded -- i.e., sold or written and bought or purchased -- on stock exchanges, and thus may be referred to as exchange-traded options or listed options..

English synonym : option contract [but see property "underlying interest" in CD "stock option"].

financial asset option [Binhammer, 93]..

French collocations : Options sont exerc es/lev es ou alors abandonn es. Dans ce cas, elles arrivent l' ch ance/expirent..

French equivalent : option (nf), contrat (nm) d'option [TERMIUM]..

Cd Name: put

SuperConcepts: option

DEFINITION

definition : An option with a finite life that gives the owner (buyer) the right to sell a specified number of shares in the underlying security ... at a specified price and within the space of time until expiration date. Upon expiration the unexercised put option becomes worthless. [Thomsett, 222].

LINGUISTIC INFORMATION

English synonym : put option.

French equivalent : option (nf) de vente [OLF].

put (nm)..

Cd Name: call

SuperConcepts: option

DEFINITION

definition : An option which has a finite life and gives the owner

(buyer) the right to buy a specified number of shares ... at a specified price and within the space of time up until expiration date. Upon expiration, the call option becomes worthless. An individual buys a call in the hope that the stock will increase in value. If it does so, the value of the call increases accordingly. [Thomsett, 40].

LINGUISTIC INFORMATION

English synonym : call option.

French equivalent : option (nf) d'achat [OLF].
call (nm)..

Cd Name: European option

SuperConcepts: option

DEFINITION

definition : European options can only be exercised on the expiration date itself.... Note that the terms "American" and "European" do not refer to the location of the option or the exchange. Some options trading on North American exchanges are European [Hull, 9 and footnote]. N.B. In North America, options are generally American. All stock options are American..

LINGUISTIC INFORMATION

English synonym : European-style option.

French equivalent : option (nf) europ enne [Augros, 26].
option de type europ en [Jacquillat89, 277]..

Cd Name: American option

SuperConcepts: option

SubConcepts: stock option

DEFINITION

definition : American options can be exercised at any time up to the expiration date.... Most of the options that are traded on exchanges are American.... Note that the terms "American" and "European" do not refer to the location of the option or the exchange [Hull, 9 and footnote]. N.B. In North America, options are generally American. All stock options are American..

LINGUISTIC INFORMATION

English synonym : American-style option.

French equivalent : option (nf) am ricaine [Augros, 26].
option de type am ricain [Jacquillat89, 277].
option l'am ricaine [Pilverdier, 256]..

Cd Name: stock option

SuperConcepts: option, American option

DEFINITION

definition : A stock option is a transferable right to transact before a future date at a predetermined price in a particular class of common stock.... A stock option does not represent ownership in the

company underlying the option.... A stock option merely gives the owner the right to acquire or sell the shares on or before a future date (MSE, 8). N.B. a stock option is necessarily an American option [q.v.]..

LINGUISTIC INFORMATION

English synonym : equity option, share option.
French equivalent : option (nf) d'achat d'actions [OLF].
option sur action [Jacquillat80, 825].
option sur actions [Jacquillat89, 340].
option de souscription [TERMIUM]..

Cd Name: exchange-traded option
SuperConcepts: option

DEFINITION

definition : [listed option] An option traded on the public exchanges [Thomsett, 156]. They have standardized terms set by the exchange and are administered by an options clearing house..

LINGUISTIC INFORMATION

English synonym : listed option.
French equivalent : option (nf) n gociable [Jacquillat81, 237].
option cot e, option n goci e en bourse [CSI, 235].
option n gociable en bourse [TERMIUM]..

Cd Name: over-the-counter option
SuperConcepts: option

DEFINITION

definition : A contracted option for which the premium and expiration are negotiated between the buyer and the seller. [Thomsett, 199].
[Conventional option] An unlisted option.... It may have strike price or expiration terms not in standing with those of listed options [Thomsett, 64]..

LINGUISTIC INFORMATION

English synonym : OTC option, unlisted option, off-the-board option.
French equivalent : option (nf) n goci e hors cote, option n goci e hors bourse [Collier].
option hors bourse [TERMIUM].
option conventionnelle [Augros, 47].
option traditionnelle [Jacquillat81, 237].
option non n gociable [Associ s, 13]..

Cd Name: stock option
SuperConcepts: option, American option

DEFINITION

definition : A stock option is a transferable right to transact before a future date at a predetermined price in a particular class of common stock.... A stock option does not represent ownership in the company underlying the option.... A stock option merely gives the

owner the right to acquire or sell the shares on or before a future date (MSE, 8). N.B. a stock option is necessarily an American option [q.v.]..

LINGUISTIC INFORMATION

English synonym : equity option, share option.
French equivalent : option (nf) d'achat d'actions [OLF].
option sur action [Jacquillat80, 825].
option sur actions [Jacquillat89, 340].
option de souscription [TERMIUM]..

Cd Name: fixed-income option
SuperConcepts: option

LINGUISTIC INFORMATION

English synonym : debt instrument option, interest-rate option [CSI, 161]..
French equivalent : option (nf) sur dette(s) [Roure, 65].
option sur titre d'emprunt [CSI, 242]..

Cd Name: currency option
SuperConcepts: option

LINGUISTIC INFORMATION

English synonym : foreign currency option, currency option contract [TERMIUM]..
French equivalent : option (nf) de change, option sur devises [Roure, 73].
option sur monnaies trang res, option sur devises trang res, contrat d'option sur devises [TERMIUM]..

Cd Name: stock-index option
SuperConcepts: option

LINGUISTIC INFORMATION

French equivalent : option (nf) sur indice boursier [Roure, 83]..

Cd Name: gold or silver option
SuperConcepts: option

LINGUISTIC INFORMATION

English synonym : gold or silver option contract.
French equivalent : option (nf) sur l'or ou l'argent [CSI, 241]..

Cd Name: futures option
SuperConcepts: option

DEFINITION

definition : Options on futures contracts or futures options are now traded on many different exchanges. They require delivery of an underlying futures contract when exercised.... Futures options are

written on both financial futures and commodity futures [Hull, 143]..

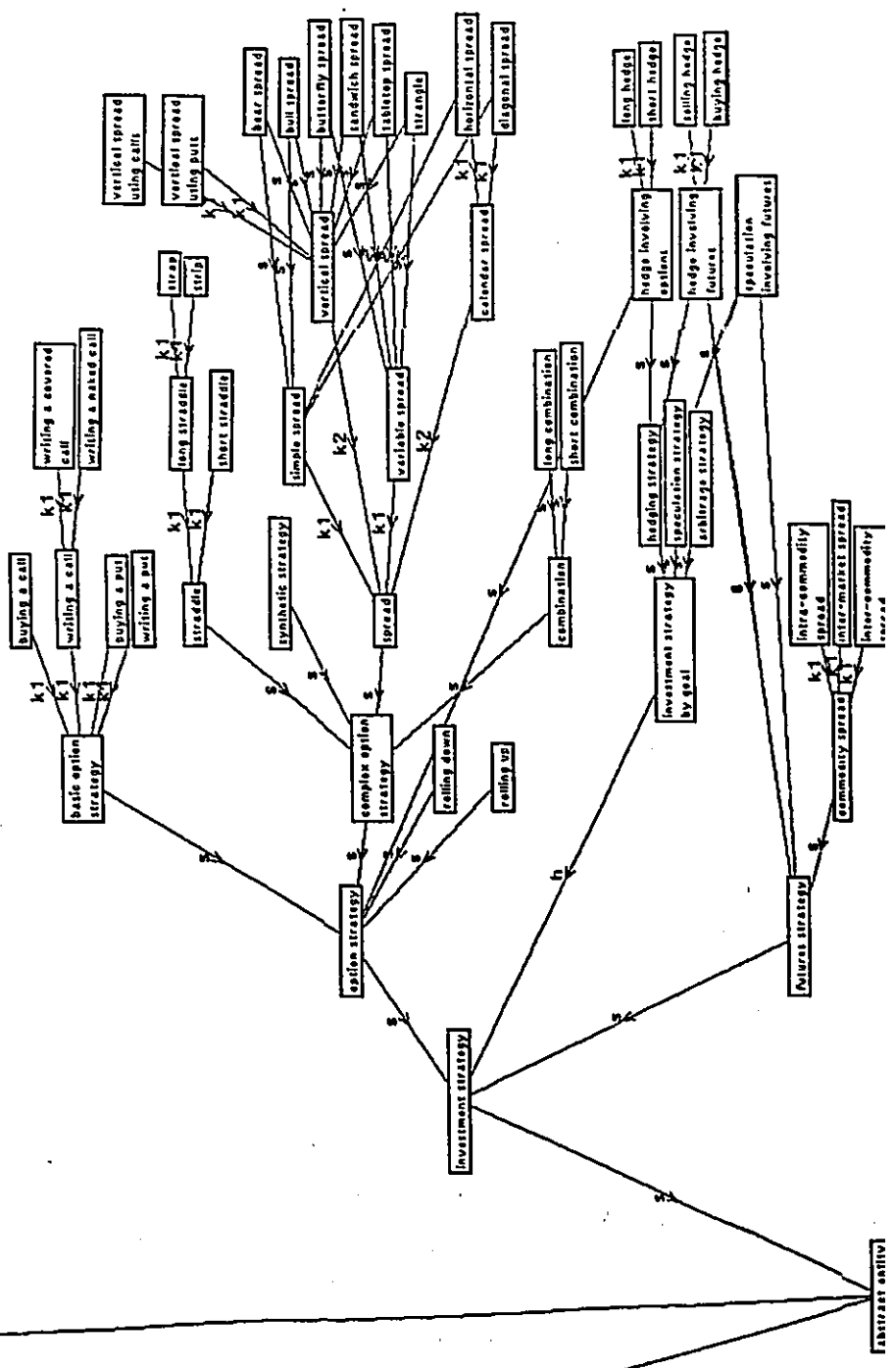
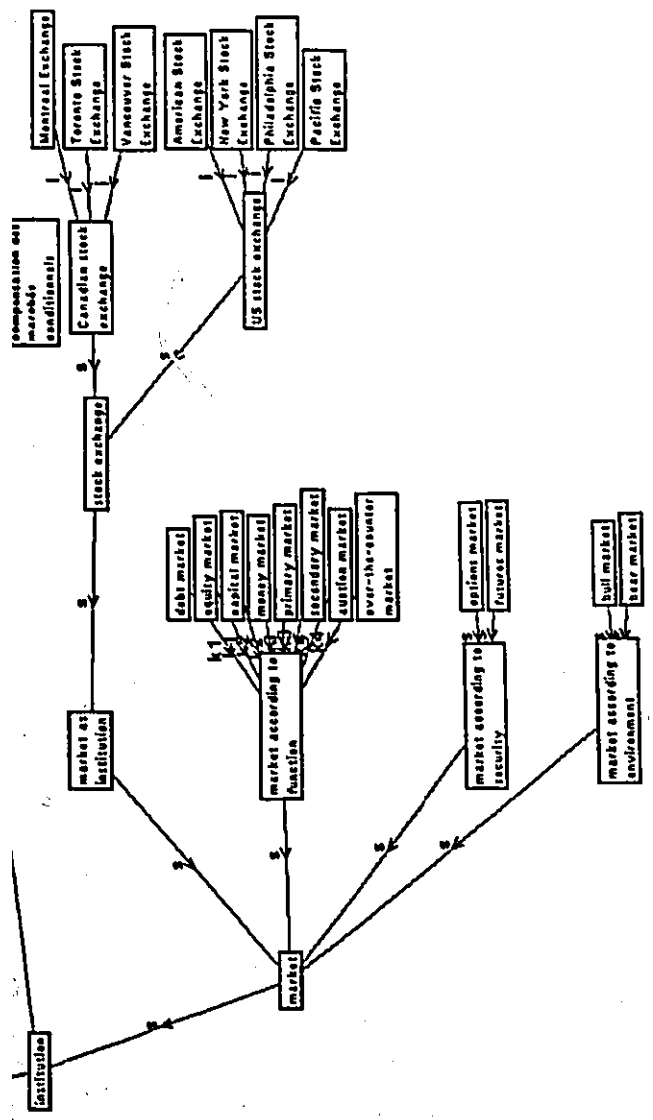
LINGUISTIC INFORMATION

English synonym : option on futures, option on a futures contract.

French equivalent : option (nf) sur contrat(s) terme [TERMIUM].
option sur futures..

APPENDIX IV - Graph of Complete OptionCODE Knowledge Base

The following fold-out page presents a complete hierarchical graph of the entire optionCODE knowledge base.



Bibliographic References

- Ahmad, Khurshid, Heribert Picht, Margaret Rodgers and Patricia Thomas (1989). *Terminology and Knowledge Engineering: A Symbiotic Relationship Explained*. Technical Report, KITES Project. Guildford, England: University of Surrey.
- Ahmad, Khurshid, Heather Fulford, Paul Holmes-Higgin, Margaret Rodgers and Patricia Thomas (1990). "The Translator's Workbench Project" in *Translating and the Computer 11: Preparing for the Next Decade*. Ed. Catriona Picken. London: Aslib, 9-19.
- Allen, James (1987). *Natural Language Processing*. Menlo Park, CA: Benjamin/Cummings.
- Auger, Pierre and Louis-Jean Rousseau (1978). *Méthodologie de la recherche terminologique*. Québec: Office de la langue française.
- Barr, Avron and Edward A. Feigenbaum (1981). *The Handbook of Artificial Intelligence Volume I*. Los Altos, CA: William Kaufmann.
- Barr, Avron and Edward A. Feigenbaum (1982). *The Handbook of Artificial Intelligence Volume II*. Los Altos, CA: William Kaufmann.
- Bathgate, R.H. (1980). "Studies of Translation Models." *The Incorporated Linguist* 19:4, 113-114.
- Bathgate, R.H. (1984). "Teaching translation students to think about science." *The Incorporated Linguist* 23:4, 228-232.
- Baudot, Jean (1986). "Les banques de terminologie de l'avenir." *Meta* 31:2, 153-158.
- Bédard, Claude (1986). *La traduction technique: principes et pratique*. Montreal: Linguatech.
- Bédard, Claude (1988). "MacTranslator - a translator workstation on a Mac." *Language Technology* 10, 32-34.
- Bédard, Claude (1989a). "The BKB (bilingual knowledge bank): a promising new approach to machine translation" in *Coming of Age (Proceedings of the 30th Annual Conference of the American Translators Association)*. Ed. Deanna Lindberg Hammond. Medford, N.J.: Learned Information, 437-441.
- Bédard, Claude (1989b). "Technologie: les services linguistiques sauront-ils prendre l'initiative ?" *Circuit* 27, 33-34.

- Bédard, Claude (1990a). "Le PTT: au CCRIT et ailleurs." *Circuit* Hors-série No. 1, Q8-Q11.
- Bédard, Claude (1990b). "Les produits « infolinguistiques »: un tour d'horizon." *Terminogramme* 55, 7-9.
- Boole, George (1854). *An Investigation of the Laws of Thought*. New York: Dover.
- Bowker, Lynne (1992). *Guidelines for Handling Multidimensionality in a Terminological Knowledge Base*. M.A. thesis, School of Translation and Interpretation, University of Ottawa.
- Brace, Colin (1990). "The Sherpa project." *Electric Word* 18, 26-30.
- Brachman, Ronald J. (1979). "On the epistemological status of semantic networks" in *Readings in Knowledge Representation*. Ed. Ronald J. Brachman and Hector J. Levesque. Los Altos: Morgan Kaufman, 192-215.
- Brachman, Ronald J. and Hector J. Levesque (1985). "Introduction" in *Readings in Knowledge Representation*. Ed. Ronald J. Brachman and Hector J. Levesque. Los Altos: Morgan Kaufman, xiii-ix.
- Brachman, Ronald J. et al. (1991). "Living with CLASSIC: when and how to use a KL-ONE-like language" in *Principles of Semantic Networks*. Ed. John F. Sowa. San Mateo, CA: Morgan Kaufmann, 401-456.
- Brisson, Hélène (1992). "L'évaluation: un second regard." *Terminology Update* 25:1, 9-12.
- Brøns, Uffe (1987). "Aspects of translation equivalence and their impact on the use of confidence codes in a multi-bilingual terminological database." *Terminologie et traduction* 3, 37-43.
- Brown, Gillian and George Yule (1983). *Discourse Analysis*. Cambridge, England: Cambridge University Press.
- Bruzzone, Catherine (1988). "Software for translators — the terminological database." *The Linguist*, 27:4, 160-161.
- Canadian Broadcasting Corporation (1988). *Reflections in the Mind of the Machine* (transcript of program *Ideas*, broadcast January 18 & 25, 1988). Montreal: CBC Transcripts.

- Casagrande, Joseph B. (1954). "The ends of translation." *International Journal of American Linguistics*, 20:4, 335-340.
- Catford, John (1965). *A Linguistic Theory of Translation*. London: Oxford University Press.
- Childs, Larry G. (1989). "ALPNET terminology tools" in *Coming of Age* (Proceedings of the 30th Annual Conference of the American Translators Association). Ed. Deanna Lindberg Hammond. Medford, N.J.: Learned Information, 455-460.
- Childs, Larry G. and Joseph Pentheroudakis (1991). "ATA automation results" in *Horizons* (Proceedings of the 32nd Annual Conference of the American Translators Association). Ed. A. Leslie Willson. Medford, N.J.: Learned Information, 317-324.
- Chomsky, Noam (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Cormier, Monique C. (1991). "Traduction et textes destinés à des spécialistes: approche pédagogique." *Meta* 36:2/3, 440-447.
- Coughlin, Josette M. (1988). "Artificial intelligence and machine translation: present developments and future prospects." *Babel* 34:1, 3-9.
- Delgrande, James P. and John Mylopoulos (1986). "Knowledge representation: features of knowledge" in *Fundamentals of Artificial Intelligence*. Ed. W. Bibel and Ph. Jorrand. Berlin: Springer-Verlag, 3-36.
- Delisle, Jean (1980). *L'Analyse du discours comme méthode de la traduction*. Cahiers de traductologie No. 2. Ottawa: University of Ottawa Press.
- Delisle, Jean (1981a). "De la théorie à la pédagogie: réflexions méthodologiques" in *L'enseignement de l'interprétation et de la traduction: de la théorie à la pédagogie*. Ed. Jean Delisle. Cahiers de traductologie No. 4. Ottawa: University of Ottawa Press, 135-151.
- Delisle, Jean (1981b). "La compréhension des textes et le processus cognitif de la traduction" in *Comprendre le langage*. Ed. J. Barbizet et al. Paris: Didier, 68-70.
- Dressler, M.V. (1981). "La compréhension des textes et des discours comme base d'un modèle linguistique" in *Comprendre le langage*. Ed. J. Barbizet et al. Paris: Didier, 71-73.

- Eck, Karen (1993). *An Knowledge-Based Approach to the Construction of Intensional Terminological Definitions*. M.A. thesis, School of Translation and Interpretation, University of Ottawa (in progress).
- Felber, Helmut (1982). "Computerized terminology in TermNet: the role of terminological data banks" in *Translating and the Computer 4: Term Banks for Tomorrow's World*. Ed. Catriona Picken. London: Aslib, 8-19.
- Firebaugh, Morris W. (1988). *Artificial Intelligence: A Knowledge-Based Approach*. Boston: Boyd and Fraser.
- Folkart, Barbara (1984). "A thing-based approach to the practice and teaching of technical translation." *Meta* 29:3, 229-246.
- Frawley, William (1984). "Prolegomenon to a theory of translation" in *Translation: Literary, Linguistic and Philosophical Perspectives*. Ed. W. Frawley. Newark: University of Delaware Press, 159-175.
- Freibott, Gerhard and Ulrich Heid (1990). "Terminological and lexical knowledge for computer-aided translation and technical writing" in *TKE'90: Terminology and Knowledge Engineering*. Frankfurt: INDEKS Verlag.
- Frost, Richard (1986). *Introduction to Knowledge Base Systems*. New York: MacMillan.
- Gagnon, René (1992). "La place des banques de terminologie dans les technologies de l'information." *Terminogramme* 66, 5-9.
- Galinski, Christian (1988a). "Terminology and information/knowledge management" in *Technology as Translation Strategy* (Proceedings of the 29th Annual Conference of the American Translators Association). Binghamton, NY: State University of New York at Binghamton, 485-492.
- Galinski, Christian (1988b). "Recent developments in knowledge engineering and the international division of labor within TermNet" in *Technology as Translation Strategy* (Proceedings of the 29th Annual Conference of the American Translators Association). State University of New York at Binghamton, 493-499.
- Galinski, Christian (1988c). "Advanced terminology banks supporting knowledge-based MT" in *New Directions in Machine Translation*. Ed. Dan Maxwell et al. Dordrecht, Holland: Forbis, 167-181.
- Galinski, Christian and Gerhard Budin (1989). "Terminology, knowledge theory and language industries." *Terminology Update* 22:3, 3-4.

Garcia-Landa, Mariano (1981). "La «théorie du sens», théorie de la traduction et base de son enseignement" in *L'enseignement de l'interprétation et de la traduction: de la théorie à la pédagogie*. Ed. Jean Delisle. Cahiers de traductologie 4. Ottawa: University of Ottawa Press, 113-132.

Garcia-Landa, Mariano (1990). "A general theory of translation (and of language)." *Meta* 35:3, 476-488.

Gawn, Peter (1990). "Tools for terminology." *Meta* 35:2, 452-455.

Gémar, Jean-Claude (1990). "Pour une méthode générale de traduction: traduire par l'interprétation du texte." *Meta* 35:4, 657-668.

Gilchrist, Alan (1982). "Patterns of information transfer, storage and retrieval in various types of organisation revealed by recent research. Part 2" in *Information Management and Organisational Change* (Proceedings of an Aslib conference April 6-8, 1981). London: Aslib.

Gile, Daniel (1986). "Traduction et interprétation: deux facettes d'une même fonction." *The Linguist* 25:3, 135-138.

Goodman, Kenneth and Sergei Nirenburg (1991). *The KBMT Project: A Case Study in Knowledge-Based Translation*. San Mateo, CA: Morgan Kaufmann.

Green, Georgia (1989). *Pragmatics and Natural Language Understanding*. Hillsdale, N.J.: Lawrence Erlbaum.

Greenberg, Joseph H. (ed.) (1978). *Universals of Human Language*. Stanford, CA: Stanford University Press.

Halliday, M.A.K. (1977). "Text as semantic choice in social context" in *Grammars and Descriptions*. Ed. T.A. van Dijk and J.S. Petöfi. Berlin, New York: de Gruyter, 176-225.

Harris, Brian (1982). "The need for definitions on term records: a translator's view" in *Colloque international de terminologie: les problèmes de la définition et de la synonymie en terminologie*. Québec: Université Laval.

Harris, Brian (1988). "Bitext, a new concept in translation theory." *Language Monthly* 54, 8-10.

Hartley, Anthony F. and Emmanuel J. Yannakondakis (1987). "An intelligent computational environment for terminology and text handling." *Meta* 32:2, 139-148.

- Hawkins, Donald T. (1987). "Artificial intelligence (AI) and expert systems for information professionals." *Online* 11:5, 91-98.
- Hawkins, Donald T. (1988). "Applications of artificial intelligence (AI) and expert systems for online searching." *Online* 12:1, 31-43.
- Hayes, Patrick J. (1979). "The logic of frames" in *Readings in Knowledge Representation*. Ed. Ronald J. Brachman and Hector J. Levesque. Los Altos: Morgan Kaufman, 287-295.
- Hewett, Rattikorn and Barbara Hayes-Roth (1991). "Representing and reasoning about physical systems using prime methods" in *Principles of Semantic Networks*. Ed. John F. Sowa. San Mateo, CA: Morgan Kaufmann, 507-525.
- Hofstadter, Douglas R. (1985). *Metamagical Themas: Questing for the Essence of Mind and Pattern*. New York: Basic Books.
- Horguelin, Paul A. (1966). "La traduction technique." *Meta* 11:1, 15-25.
- House, Juliane (1977). *A Model for Translation Quality Assessment*. TBL. Tübingen: Verlag Günther Narr.
- Hugo, Charles (1988). "Building your own personal workstation." *Language Technology* 7, 11.
- Hunt, Morton (1982). *The Universe Within: A New Science Explores the Human Mind*. New York: Simon and Schuster.
- Hutchins, John (1986). *Machine Translation: Past, Present and Future*. Chichester, U.K.: Ellis Horwood.
- Hutchins, John (1988a). "Recent developments in machine translation: A review of the last five years" in *New Directions in Machine Translation*. Ed. Don Maxwell et al. Dordrecht, Holland: Forbis, 7-63.
- Hutchins, John (1988b). "Future Perspectives in Translation Technologies" in *Technology as Translation Strategy*, Vol. II. Ed. Muriel Vasconcellos. Binghamton, NY: State University of New York at Binghamton, 223-240.
- Isabelle, Pierre (1988). "Terminologie et traduction automatique." Interview by Jean Perron. *Terminogramme* 46, 16-19.
- Johnson, George (1986). *Machinery of the Mind*. Redmond, WA: Tempus.

- Johnson-Laird P.N. (1988). *The Computer and the Mind: An Introduction to Cognitive Science*. Cambridge, MA: Harvard University Press.
- Joscelyne, Andrew (1990). "On Site." *Electric Word* 18, 38-43.
- King, Margaret (1987). "On the proper place of semantics in machine translation" in *Language and Artificial Intelligence: Proceedings of an International Symposium on Language and Artificial Intelligence*. Ed. Makoto Nagao. Amsterdam: Elsevier, 283-202.
- Kittredge, Richard I. (1983). "Semantic processing of texts in restricted sublanguages" in *Computational Linguistics*. Ed. Nick Cercone. Oxford: Pergamon Press, 45-58.
- Kittredge, Richard I. (1987). "The importance of sublanguage for automatic translation" in *Machine Translation: Theoretical and Methodological Issues*. Ed. Sergei Nirenburg. Cambridge, England: Cambridge University Press, 59-67.
- Kittredge, Richard I. (1989). "Multilingual text generation as an alternative to machine translation" in *Coming of Age* (Proceedings of the 30th Annual Conference of the American Translators Association). Ed. Deanna Lindberg Hammond. Medford, N.J.: Learned Information, 465-469.
- Kramer, B.M. and J. Mylopoulos (1987). "Knowledge representation" in *Encyclopedia of Artificial Intelligence*. Ed. Stuart Shapiro. New York: Wiley, 882-890.
- Labov, William (1973). "The boundaries of words and their meanings" in *New Ways of Analyzing Variation in English*. Ed. Charles-James Baily and Roger W. Shuy. Washington, D.C.: Georgetown University Press.
- Ladmiral, Jean-René (1992). "L'ordinateur est une vraie machine à écrire" in *L'environnement traductionnel: le station de travail du traducteur de l'an 2001*. Ed. André Clas and Safar Hayssam. Sillery, Quebec: Presses de l'Université du Québec, 329-340.
- Lapalme, Guy (1991). "Utilisation de la logique dans le traitement des langues naturelles" in *Les industries de la langue: Perspectives des années 1990*, Actes du colloque, Tome I. Government of Quebec, 335-347.
- Laponce, J.A. (1987). *Languages and their Territories*. Translated by Anthony Martin-Sperry. Toronto: University of Toronto Press.
- Larson, Mildred L. (1984). *Meaning-Based Translation*. Lanham, MA: University Press of America.

- Lederer, Marianne. "La compréhension des textes" in *Comprendre le langage*. Ed. J. Barbizet et al. Paris: Didier, 65-67.
- Leibniz, Gottfried Wilhelm (1679). "Elementa characteristica universalis" in *Opuscules et fragments inédits de Leibniz* (1903). Ed. L. Couturat. Paris: Ancienne Librairie Germer Baillière, 42-92.
- Lenat, Douglas B. and R.V. Guha (1990). *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project*. Reading, MA: Addison-Wesley.
- Lörscher, Wolfgang (1991). *Translation Performance, Translation Process, and Translation Strategies: A Psycholinguistic Investigation*. Tübingen: Verlag Günter Narr.
- Lyons, John (1968). *Introduction to Theoretical Linguistics*. Cambridge, England: Cambridge University Press.
- Macklovitch, Elliott (1991). "The translator's workstation ... in plain prose" in *Horizons* (Proceedings of the 32nd Annual Conference of the American Translators Association). Ed. A. Leslie Willson. Medford, N.J.: Learned Information, 317-324.
- Macklovitch, Lison (1991). "The trial of the translator's workstation at the Canadian Translation Bureau" in *Horizons* (Proceedings of the 32nd Annual Conference of the American Translators Association). Ed. A. Leslie Willson. Medford, N.J.: Learned Information, 301-306.
- Malinowski, B. (1937). "The dilemma of contemporary linguistics." *Nature* 140, 172-173.
- Mayorcas-Cohen, Pamela (1986). "The translator as information user" in *Translating and the Computer 7*. Ed. Catriona Picken. London: Aslib, 61-88.
- McCorduck, Pamela (1979). *Machines Who Think*. San Francisco: W.H. Freeman.
- Melby, Alan (1985). "The translator's work station of tomorrow" in *Proceedings of the National Symposium on Linguistic Services* (October 9-12, 1984). Ottawa: Supply and Services Canada.
- Melby, Alan (1987a). "On human-machine interaction in translation" in *Machine Translation: Theoretical and Methodological Issues*. Ed. Sergei Nirenburg. Cambridge: Cambridge University Press, 145-154.
- Melby, Alan (1987b). "The translator workstation." *Language Technology* 2, 7.

Melby, Alan (1990). "MicroMATER: Its importance to translators in the 1990s" in *Looking Ahead* (Proceedings of the 31st Annual Conference of the American Translators Association). Medford, N.J.: Learned Information, 281-287.

Melby, Alan K. (1992). "Interchanging documents containing non-English characters" in *Proceedings from the International Symposium on Terminology and Documentation in Specialized Communication*. Ottawa: Supply and Services Canada, 104-132.

Meyer, Ingrid and Douglas Skuce (1992). "Computer-assisted concept analysis for terminology: A framework for technological and methodological research" in *Proceedings of the Euralex Fourth International Congress*. Barcelona, Spain: Bibliograf, 129-138.

Meyer, Ingrid, Lynne Bowker and Karen Eck (1992a). "COGNITERM: an experiment in building a knowledge-based term bank" in *Proceedings of the Euralex Fifth International Congress*. Barcelona, Spain: Bibliograf, 159-172.

Meyer, Ingrid, Lynne Bowker and Karen Eck (1992b). "Constructing a knowledge-based term bank: fundamentals and implications" in *Proceedings from the International Symposium on Terminology and Documentation in Specialized Communication*. Ottawa: Supply and Services Canada, 232-256.

Meyer, Ingrid (1993). "Concept management for terminology: a knowledge management approach" in *Standardizing Terminology for Better Communication: Practice, Applied Theory, and Results*. Ed. R.A. Strehlow and S.E. Wright. Philadelphia: American Society for Testing and Materials.

Michalski, Ryszard S. (1987). "How to learn imprecise concepts: a method employing a two-tiered knowledge representation in learning" in *Proceedings of the Fourth International Workshop on Machine Learning*. Ed. Pat Langley. Los Altos: Morgan Kaufmann, 50-58.

Miller, David R., Ingrid Meyer and Diane Michaud (1991). "Terminologie et analyse notionnelle assistée par ordinateur" in *Les industries de la langue: Perspectives des années 1990*, Actes du colloque, Tome II. Government of Quebec, 781-800.

Miller, George A. (1978). "Practical and lexical knowledge" in *Cognition and Categorization*. Ed. Eleanor Rosch and Barbara B. Lloyd. Hillsdale, NJ: L. Erlbaum, 305-319.

Ministère des Communications du Québec (1978). *Guide du traducteur*. 3rd edition. Québec: Éditeur officiel du Québec.

Minsky, Marvin (1975). "A framework for representing knowledge" in *Readings in Knowledge Representation* (1981). Ed. Ronald J. Brachman and Hector J. Levesque. Los Altos: Morgan Kaufman, 248-262.

Moore, Robert C. (1982). "The role of logic in knowledge representation and common-sense reasoning" in *Readings in Knowledge Representation*. Ed. Ronald J. Brachman and Hector J. Levesque. Los Altos: Morgan Kaufman, 336-341.

Moser-Mercer, Barbara (1987). "Man/machine interaction in translation and terminology." *Meta* 32:2, 156-163.

Moskowitz, D. (1978). "Le traducteur: récepteur et destinataire du message" in *Études de linguistique appliquée* 12. Paris: Didier, 71-84.

Mossop, Brian (1983). "The translator as rapporteur: a concept for training and self-improvement." *Meta* 28:3, 244-278.

Mounin, George (1963). *Les problèmes théoriques de la traduction*. Paris: Gallimard.

Nagao, Makoto (1989). *MT: How Far Can It Go?* Translated by Norman D. Cook. New York: Oxford University Press.

Newman, Patricia (1992). "Transmission of scientific content" in *Translating in North America — A Community of Interests* (Proceedings of the Third Congress of the FIT Regional Center for North America). Ed. Michèle Valiquette. Montreal: Trandek.

Newmark, Peter (1981). *Approaches to Translation*. Oxford: Pergamon.

Newmark, Peter (1988). *A Textbook of Translation*. London: Prentice-Hall.

Newmark, Peter (1991). "Translation: an introductory survey" in *About Translation*. Avon, England: Clevedon, 14-41.

Nida, Eugene (1964). *Toward a Science of Translating*. Leiden: Brill.

Nida, Eugene (1975). *Language Structure and Translation*. Stanford, California: Stanford University Press.

Nida, Eugene (1985). "Translating means translating meaning. A sociosemiotic approach to translating" in *Translators and Their Position in Society* (Proceedings of Tenth World Congress of the International Federation of Translators, Vienna 1984). Ed. H. Bühler. Vienna: Wilhelm Braumüller, 119-124.

- Nida, Eugene and Charles Taber (1969). *The Theory and Practice of Translation*. Leiden: E.J. Brill.
- Nida, Eugene and Jan de Waard (1986). *From One Language to Another: Functional Equivalence in Bible Translating*. Nashville: Thomas Nelson.
- Nirenberg, Sergei (1987). "Knowledge and choices in machine translation" in *Machine Translation: Theoretical and Methodological Issues*. Ed. Sergei Nirenburg. Cambridge, England: Cambridge University Press, 1-16.
- Ogden, C.K. and I.A. Richards (1969). *The Meaning of Meaning*. 10th edition. London: Ridgewood.
- Otman, Gabriel (1989). "Terminologie et intelligence artificielle." *Banque des mots*, special edition (December 1989), 63-95.
- Otman, Gabriel (1991). "Aspects de l'informatisation des activités terminologiques et traductionnelles." *Terminologies nouvelles* 5, 15-20.
- Pacquin, Louis-Claude (1991). "Le passage des termes aux concepts" in *Les industries de la langue: Perspectives des années 1990*, Actes du colloque, Tome I. Government of Quebec, 313-333.
- Parent, Richard (1988). "Terminologie et modélisation des connaissances." *Terminogramme* 47-48, 5-6.
- Parent, Richard (1989). "Recherche d'une synergie entre développement linguistique informatisé et systèmes experts: importance de la terminologie." *Meta* 34:1, 611-614.
- Park, William M. (1988). "The ideal workstation: perspective of the independent translator" in *Technology as Translation Strategy* (Proceedings of the 29th Annual Conference of the American Translators Association). State University of New York at Binghamton, 27-35.
- Pavel, Silvia (1992). "TERMIUM: le traitement informatisé des difficultés de traduction" in *Proceedings from the International Symposium on Terminology and Documentation in Specialized Communication*. Ottawa: Supply and Services Canada, 22-43.
- Pergnier, Maurice (1980). *Fondements sociolinguistiques de la traduction*. Paris: Atelier Reproduction des Thèses.
- Pergnier, Maurice (1981). "Théorie linguistique et théorie de la traduction." *Meta* 26:3, 255-262.

Picht, Heribert and Jennifer Draskau (1985). *Terminology: An Introduction*. Department of Linguistic and International Studies Monographs. Guildford, England: University of Surrey.

Quillian, M. Ross (1966). *Semantic Memory*. Report AD-641671. Clearinghouse for Federal Scientific and Technical Information.

Rasmus, Daniel W. (1991). "Integrating distributed information." *Byte* 16:12, 247-253.

Reiter, Raymond (1978). "On reasoning by default" in *Readings in Knowledge Representation*. Ed. Ronald J. Brachman and Hector J. Levesque. Los Altos: Morgan Kaufman, 402-410.

Rich, Elaine and Kevin Knight (1991). *Artificial Intelligence*. 2nd edition. New York: McGraw-Hill.

Roberts, Roda P. (1986). "Translation theory for practitioners" in *Building Bridges* (Proceedings of the 27th Annual Conference of the American Translators Association). Ed. Karl Kummer. Medford, N.J.: Learned Information, 51-62.

Rondeau, Guy (1981). *Introduction à la terminologie*. Montreal: Centre éducatif et culturel.

Rosch, Eleanor (1978). "Principles of categorization" in *Cognition and Categorization*. Ed. Eleanor Rosch and Barbara B. Lloyd. Hillsdale, NJ: L. Erlbaum, 27-48.

Russell, Pamela R. (1989). *How to Write a Précis*. Ottawa: University of Ottawa Press.

Ryan, Bob (1991). "AI's identity crisis." *Byte* 16:1, 236-246.

Sabourin, Pascal (1992). *Sondage sur la tarification professionnelle: rapport final*. Ottawa: Association of Translators and Interpreters of Ontario.

Sager, Juan C. (1989). "Ten years of machine translation design and application: from FAHQT to realism" in *Translating and the Computer 10: The Translation Environment 10 Years On*. Ed. Pamela Mayorcas. London: Aslib, 3-10.

Sager, Juan C. (1990). *A Practical Course in Terminology Processing*. Amsterdam: John Benjamins.

Sampson, Geoffrey (1987). "MT: A nonconformist's view of the state of the art" in *Machine Translation Today: The State of the Art*. Ed. Margaret King. Edinburgh University Press.

Sapir, Edward (1949). *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace.

Schank, Roger (1975). *Conceptual Information Processing*. Amsterdam: North Holland.

Schank, Roger and R. P. Abelson (1977). *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*. Hillsdale, NJ: Lawrence Erlbaum.

Schubert, Lenhart K. (1991). "Semantic nets are in the eye of the beholder" in *Principles of Semantic Networks*. Ed. John F. Sowa. San Mateo, CA: Morgan Kaufmann, 95-107.

Scown, Sloan J. (1985). *The Artificial Intelligence Experience: An Introduction*. Digital Equipment Corporation.

Secretary of State Department (1987). *Aide-mémoire d'autoperfectionnement à l'intention des traducteurs et des rédacteurs*. Ottawa: Secretary of State.

Seleskovitch, Danica (1981a). "Introduction. Pourquoi un colloque sur la compréhension du langage ?" in *Comprendre le langage*. Ed. J. Barbizet et al. Paris: Didier, 9-15.

Seleskovitch, Danica (1981b). "L'enseignement de l'interprétation" in *L'enseignement de l'interprétation et de la traduction: de la théorie à la pédagogie*. Cahiers de traductologie 4. Ed. Jean Delisle. Ottawa: University of Ottawa Press, 23-46.

Seleskovitch, Danica (1988). "Technical and literary translation: a unifying view" in *ITI Conference 2: Translators and Interpreters Mean Business*. Ed. Catrione Picken. London: Aslib, 83-88.

Seleskovitch, Danica and Marianne Lederer (1984). *Interpréter pour traduire*. Paris: Didier.

Shaikovich A. and I. Oubine (1988). "Translators and researchers look at bilingual terminological dictionaries." *Babel* 34:2, 10-16.

Shapiro, S.C. (1987) (Ed.). *Encyclopedia of Artificial Intelligence*. New York: John Wiley & Sons.

Shastri, Lokendra (1988). *Semantic Networks: An Evidential Formalization and its Connectivist Realization*. London: Pitman.

Shneidermann, Ben (1992). *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Reading, MA: Addison-Wesley.

- Shreve, Gregory M. (1990). "Hypertext knowledge-bases for computer-assisted translation: organization and structure" in *Looking Ahead* (Proceedings of the 31st Annual Conference of the American Translators Association). Medford, N.J.: Learned Information, 295-305.
- Shreve, Gregory M. (1991). "Evolution of the translator's workstation as virtual desktop: software services and user environment" in *Horizons* (Proceedings of the 32nd Annual Conference of the American Translators Association). Ed. A. Leslie Willson. Medford, N.J.: Learned Information, 307-316.
- Skuce, Douglas (1990). "Conceptual analysis and knowledge acquisition for software engineering." University of Ottawa. Unpublished.
- Skuce, Douglas (1992). "Notes on ClearTalk." Ottawa: Artificial Intelligence Laboratory, Department of Computing Science, University of Ottawa.
- Skuce, Douglas (1993). "A multifunctional knowledge management system." *Knowledge Acquisition* (forthcoming).
- Skuce, Douglas, S. Wang and Y. Beauvillé (1989). "A generic knowledge acquisition environment for conceptual and ontological analysis" in *Proceedings of the 4th Knowledge Acquisition for Knowledge-Based Systems Workshop*. Banff, Alberta.
- Skuce, Douglas and Ira Monarch (1990). "Ontological issues in knowledge base design: some problems and suggestions" in *Proceedings of the 5th Knowledge Acquisition for Knowledge-Based Systems Workshop*. Banff, Alberta.
- Skuce, Douglas and Ingrid Meyer (1990). "Computer-assisted concept analysis: an essential component of a terminologist's workstation" in *TKE'90: Terminology and Knowledge Engineering*, Vol. 1. Frankfurt: INDEKS Verlag, 187-199.
- Skuce, Douglas and Ingrid Meyer (1991). "Terminology and knowledge engineering: exploring a symbiotic relationship." *Proceedings of the 6th Knowledge Acquisition for Knowledge-Based Systems Workshop*. Banff, Alberta, 29-1 to 29-21.
- Smith, David (1986). "Translation practice in Europe" in *Translating and the Computer 8: A Profession on the Move*. Ed. Catriona Picken. London: Aslib, 74-82.
- Smith, Edward E. (1989). "Concepts and induction" in *Foundations of Cognitive Science*. Ed. Michael I. Posner. Cambridge, MA: MIT Press, 501-521.
- Sotiropoulos, Megaklis T. (1989). "Conceptual-mathematical models as a bridge from the general theory of terminology to informatics." *Terminologie et traduction* 3, 135-148.

- Sowa, John F. (1984). *Conceptual Structures: Information Processing in Mind and Machine*. Reading, MA: Addison-Wesley.
- Sowa, John F. (1991a). "Semantic networks" in *Selected Papers*. Unpublished [an earlier version appeared under the same title in Shapiro (1987)].
- Sowa, John F. (1991b). "Conceptual analysis as a basis for knowledge acquisition" in *Selected Papers*. Unpublished.
- Sowa, John F. (1991c). "Issues in knowledge representation" in *Principles of Semantic Networks*. Ed. John F. Sowa. San Mateo, CA: Morgan Kaufmann, 1-11.
- Sparck Jones, Karen and Martin Kay (1973). *Linguistics and Information Science*. New York: Academic.
- Sperber, Dan and Deidre Wilson (1982). "Mutual knowledge and relevance in theories of comprehension" in *Mutual Knowledge*. Ed. N.V. Smith. London: Academic Press, 61-87.
- Steiner, George (1975). *After Babel: Aspects of Language and Translation*. London: Oxford University Press.
- Tanimoto, Steven L. (1990). *The Elements of Artificial Intelligence*. New York: Computer Science Press.
- Thompson, Bill and Bev Thompson (1991). "Overturning the category bucket." *Byte* 16:1, 249-256.
- Thouin, Benoît (1988). "L'industrialisation de la traduction and la montée des multinationales." Interview by Solange Lapierre. *Circuit* 22, 16-21.
- Touretzky, D.S. (1987). "Inheritance hierarchy" in *Encyclopedia of Artificial Intelligence*. Ed. Stuart Shapiro. New York: Wiley, 422-431.
- Tsujii Jun-icki (1991). "Machine translation and machine-aided translation — what's going on" in *Translating and the Computer 12: Applying Technology to the Translation Process*. Ed. Catriona Picken. London: Aslib, 3-24.
- Tucker, Allen B. (1987). "Current strategies in machine translation research and development" in *Machine Translation: Theoretical and Methodological Issues*. Ed. Sergei Nirenburg. Cambridge, England: Cambridge University Press, 22-41.

- Tudor, Ian (1987). "A framework for the translational analysis of texts." *The Linguist* 26:2, 80-82.
- Van Campenhoudt, Marc (1991). "TI, le logiciel d'expérimentation notionnelle de Termisti." *Terminologies nouvelles* 5, 11-14.
- Vaumoron, Jean (1987). "Two-level data banks for translators." *Computers and Translation* 2, 275-280.
- Vaumoron, Jean (1988). "An ideal workstation? Perspective of the in-house translator" in *Technology as Translation Strategy*, Vol. II. Ed. Muriel Vasconcellos. Binghamton, NY: State University of New York at Binghamton, 36-41.
- Vinay, Jean-Paul and Jean Darbelnet (1958). *Stylistique comparée du français et de l'anglais: Méthode de traduction*. Bibliothèque de Stylistique comparée 1. Paris: Didier and Montreal: Beauchemin.
- Waldrop, M. Mitchell (1984). "The necessity of knowledge." *Science* 223, 1279-1282.
- Weinreich, Uriel (1963). "On the semantic structure of language" in *Universals of Language*. Ed. Joseph H. Greenberg. Cambridge, MA: MIT Press, 142-216.
- Whorf, Benjamin Lee (1956). *Language, Thought and Reality: Selected Writings of Benjamin Lee Whorf*. Ed. John B. Carroll. Cambridge, MA: Technology Press, MIT.
- Wilks, Yorick (1983). "Machine translation and the artificial intelligence paradigm of language" in *Computers in Language Research* 2. Ed. Walter and Sally Sedelow. Berlin: Mouton.
- Wilss, W. (1993). "Translation strategy, translation method and translation technique: a clarification of three translational concepts." *Revue de phonétique appliquée* 66/68, 143-152.
- Winston, Patrick Henry (1977). *Artificial Intelligence*. Reading, MA: Addison-Wesley.
- Wittgenstein, Ludwig (1921). *Tractatus Logico-Philosophicus*. London, England: Routledge & Kegan Paul (1961).
- Wittgenstein, Ludwig (1953). *Philosophical Investigations*. Oxford: Basil Blackwell.
- Woods, William (1975). "What's in a link: foundations for semantic networks" in *Readings in Knowledge Representation*. Ed. Ronald J. Brachman and Hector J. Levesque. Los Altos: Morgan Kaufman, 217-241.

Wright, Sue Ellen and Alan K. Melby (1992). "TEI-TERM: an SGML-based interchange format for terminological data within the framework of the text encoding initiative" in *Proceedings from the International Symposium on Terminology and Documentation in Specialized Communication*. Ottawa: Supply and Services Canada.