A Corpus-driven Approach to Source Domain Determination

by

Siaw-Fong Chung

Institute of Linguistics, Academia Sinica, Taipei, Taiwan
2009
Acknowledgements

For the completion of this book, I would like to thank the following people who have invested tremendous time and energy into my work.

First of all, I would like to thank my supervisor, Professor Kathleen Ahrens, for her support through all of these years, not only in academics but also in all aspects of my life. For this book in particular, I will never forget the time we spent discussing its fine details—from its research questions and methodology to its argumentation, as well as the organization of its contents. I will never forget the look of satisfaction mingled with fatigue on her face as she saw this book getting thicker and thicker. I would also like to thank Professor Ahrens for keeping her expectations high for my work, making me work harder and encouraging me to excel. Even after my oral defense, she told me that I should not take the revisions lightly, and as a result, I re-wrote several chapters of this book. I am grateful because without her high expectations, I would not have striven to be better.

The second person I would like to thank is Professor Chu-Ren Huang. Professor Huang works differently from Professor Ahrens but has many innovative ideas, and I have greatly benefited from discussions with him. Although Professor Huang was often occupied with other commitments, he still took the time to listen to my problems with this book. From the selection of target domains and cut-off points to the psycholinguistic experiments, I could see the amount of attention Professor Huang gave to my work, in spite of his limited time. In addition, I would like to express my gratitude to him for allowing me to work with his research team at Academia Sinica. Without the help of Professor Huang and his team, things could never have gone as smoothly with this book as they did.

In addition to Professor Ahrens and Professor Huang, I would also like to thank my other book committee members—Professor I-wen Su, Professor Yung-O Biq, Professor Hintat Cheung and Professor Sue J. Ker. I thank Professor Su for her critical questions, which I realize were crucial to the heart of the problem. I also thank her because she made me realize problems I did not notice before. I thank Professor Biq for the time she spent with me on my work. I also thank her for the questions she raised regarding this book. I thank Professor Cheung for his kind advice (always) and his wisdom in different aspects of academic life. I also thank him for his questions, which triggered me to think outside of the context of my research. Finally, I thank Professor Ker for being available whenever I sought her opinion. I also thank her for her suggestions regarding problems
I faced during the writing of this book. I am very grateful to my entire committee for the help they gave me. They were the best possible team to have for my research topic.

I would also like to thank the entire faculty at the Graduate Institute of Linguistics, National Taiwan University, for their teaching all through these last five years that I have been at the institute. I would like to thank Professor Shuanfan Huang for his teaching and insightful comments on my research in his class. My gratitude also goes to Professor Li-may Sung, who is always available when we have questions; Professor Wenyu Chiang, who has always been concerned about my progress; and Professor Janice Fon, who always encourages me with a kind word.

In addition, I would like to thank Professor Chung-Ping Cheng for his advice regarding statistics as well as his patience in explaining all of my statistics questions. I also thank him for being encouraging—because of his help, and thanks to discussions with him, I have become better at statistics.

I would also like to thank Professor Shu-Chuan Tseng for discussing the problems in the book with me. Although we only communicated for a short period of time, her suggestion regarding the cut-off points has been useful, as it really shed light onto a problem that had me stuck at a critical point in my research.

I would also like to thank Dr. Laurent Prévot, who has helped me tremendously with the calculation of conceptual distance, without which I could not possibly have completed that section of the book. I also thank him for the time he spent discussing problems and issues related to the calculation methods. Working together with Dr. Prévot is Ming-Wei Hsu, who helped me with all sorts of programming problems in this book. I thank him for his patience and his continuous help with this book. In addition, Petr Šimon has also helped me greatly with this book.

I would also like to thank the Academia Sinica Fellowships for Doctoral Candidates in the Humanities and Social Sciences. I thank the Institute of Linguistics, Academia Sinica, for providing the facilities and the financial support for me to complete the writing of this book. My special thanks go to the director, Professor Dah-an Ho, who kindly provided a comfortable environment so that I could concentrate on my work. My gratitude also goes to the research fellows at the institute who have advised me on my research. My special thanks also to those who have encouraged me in research as well as in many aspects of my life. The comments from the research fellows during my stay at Academia Sinica were also invaluable. Finally, I would like to thank the 2006 Taiwan Summer Institute of Linguistics (TSIL) Research Award for supporting the research of this book.
I would also like to take this opportunity to thank my in-laws for being so supportive of my research. The trust my family has shown me has been a great source of encouragement. I would also like to thank my daughters, Shih-Yuan Lee and Shih-Ning Lee, for being a driving force for me to work better, and in addition, I would like to thank my family in Malaysia who misses me and who has been tirelessly encouraging me over the phone in their many long-distance calls. Without my family (in both Taiwan and Malaysia), I would not have been able to get through this difficult time in my life. I thank them for being supportive.

I am also very fortunate because my husband, Eric Chia-Hwa Lee, has been so supportive and encouraging through the entire graduate school process. I am very grateful for the patience and understanding he has shown throughout these past five years, which has made this journey all the more rewarding and worthwhile.
English Abstract

The present book provides a detailed investigation of metaphors in Mandarin Chinese based on the premise that different knowledge domains are mapped in the formulation of metaphors. In conceptual metaphors (Lakoff & Johnson 1980, Lakoff 1993), the two mapped domains include the target domain, which is usually more abstract, and the source domain, which is usually more concrete, providing key information to understand the concepts of the target domain. Due to the importance of source domains, studies attempting to define operationally how source domains are involved with conceptual metaphors have been undertaken in the past. However, these studies have encountered difficulties due to two main reasons: (a) the variation in the specificity of the source domains; and (b) the subjectivity of individuals when determining source domains.

This book overcomes these difficulties by employing both top-down and bottom-up approaches to determine the source domain of a conceptual metaphor. A top-down approach to source domain determination usually sorts metaphorical instances according to pre-determined source domains that come from a conceptual knowledge system (such as an ontology). Alternatively, a bottom-up approach accumulates knowledge about the source domain through language use (or collocation). Both approaches, top-down and bottom-up, have been used to identify source domains in metaphor studies; however, the precision of these two approaches has not previously been compared. The present book intends to fill the gap between these two approaches by comparing the approaches using computational tools, such as WordNet, SUMO (Suggested Upper Merged Ontology), and Sketch Engine.

The outcome of this study provides a comprehensive comparison of the top-down and bottom-up approaches regarding the determination of source domains. Based on the combined results of the two approaches, this study proposes a hierarchical model for source domain definition. This model facilitates automatic identification of metaphors and explains why certain domains overlap in conceptualization. By using computational tools with a linguistic framework of domain identification that contribute to computational linguistics, this study finds further support through psycholinguistic experiments, thus enhancing the ability to define source domains. This, in turn, makes a substantial contribution to understanding the human categorization of concepts through linguistic evidence.

Key words: source domains, conceptual metaphors, top-down approach, bottom-up approach, ontology, collocation
一般來說，領域訊息在隱喻形成的時候會產生互換的情形，而在概念隱喻裡 (Lakoff & Johnson 1980, Lakoff 1993)，互換的兩個領域訊息分別為源域與目標域。源域通常比目標域較為具體，所以源域提供的訊息對於理解目標域的概念來說是十分關鍵的要素。因此，在過去曾經有相關的研究嘗試定義源域，可是這些研究通常會遇到以下兩個問題：第一，對源域定義的差異、第二，對源域界定的主觀程度不一致。

本研究為了克服這些困難，使用了兩個面向的方式將隱喻之源域做系統化的界定：一是由上而下、二則為由下而上。一般來說，由上而下的方式通常會假設一個上層的概念，而源域就是這些概念裡的一部分。唯有當源域已決定之後，才會把各種的隱喻例子歸類到不同的源域。而隱喻概念這個理論 (Lakoff & Johnson 1980, Lakoff 1993) 用的就是這一種方式，例如：在 LOVE IS JOURNEY 的隱喻概念裡，當決定 JOURNEY 為源域之後，一些相關的例子才會接著分類到該源域內。

相反地，由下而上的方式則會先累積人類的用語，而這些資訊會構成較上層的知識。此種方式通常會以頻率為基礎，如：Labov (1973)、Rosch & Mervis (1975) 所舉出的典型模範 (prototype) 指的就是共現度高的搭配詞組。在隱喻相關研究中，這兩種方式都曾有學者使用過，但是，在文獻中卻少有研究對這兩種方式做全面性的比較。本研究使用了 WordNet、SUMO (Suggested Upper Merged Ontology) (知識本體) 與 Chinese Sketch Engine (中文詞彙特性速描系統) 等計算語言學的工具來界定源域，是第一個整合從由上而下及由下而上兩種方式來比較領域訊息的研究。並且，我們透過計算語言學工具來驗證領域界定的理論架構，得到的結果有助於我們進一步了解人類對事物的歸類。而更重要的是，本研究結合了計算語言學、語言學理論及心理語言學的知識來界定領域概念，因此其成果將具有跨領域的貢獻。

關鍵詞：源域，概念隱喻，由上而下的方式，由下而上的方式，知識本體，搭配詞組
# Table of Contents

Acknowledgements..................................................................................................................... i
English Abstract ......................................................................................................................... v
中文摘要 ............................................................................................................................... vii

**Chapter 1: Introduction** ......................................................................................................... 1

**Chapter 2: What Constitutes a Source Domain?** ................................................................. 9

2.1 How to Define General versus Specific Source Domains .............................................. 11
2.2 How to Define Concrete Domains Based on Human Experience ................................. 13
2.3 Terminology .................................................................................................................. 15
2.4 Summary of Chapter .................................................................................................... 16

**Chapter 3: Metaphor Identification versus Source Domain Determination** ............ 17

3.1 Top-down Approach to Metaphor Identification ......................................................... 18
3.2 Bottom-up Approach to Metaphor Identification ....................................................... 23
3.3 Summary of Chapter .................................................................................................... 28

**Chapter 4: Lexical Knowledgebases and Corpus** .............................................................. 29

4.1 WordNet......................................................................................................................... 29
4.2 SUMO (Suggested Upper Merged Ontology) ............................................................... 31
4.3 SinicaBow (Academia Sinica Bilingual Ontological Wordnet) .................................. 34
4.4 Chinese-English Merged Word List .......................................................................... 36
4.5 Sketch Engine ................................................................................................................ 37
4.6 Chinese Gigaword Corpus Version 1.0 ..................................................................... 42
4.7 Use of Lexical Resources in the Book ...................................................................... 43
4.8 Summary of Chapter .................................................................................................... 48

**Chapter 5: Selection of Corpus, Target Domains and Sub-Corpora for Target Domains** ................................................................................................................................. 49

5.1 Selection of Corpus: Chinese Gigaword Corpus 1.0 .................................................... 51
5.2 Selection of Target Domains......................................................................................... 52
5.3 Corpora Instances for Different Target Domains ....................................................... 60
5.3.1 Single Target Domains ......................................................................................... 61
5.3.2 Coordinated Target Domains ............................................................................. 63
5.4 Total CNA and XIN Data for the Different Target Domains ....................................... 67
5.5 Summary of Chapter .................................................................................................... 71
Chapter 6: Metaphorical Expression Identification Using Target Domains

6.1 Global Hypotheses versus Local Hypothesis of Linguistic Behaviors
6.2 Procedures for Metaphorical Expression Identification
6.3 Results of Metaphorical Expression Identification
   6.3.1 Percentages of Overall Tokens of Metaphorical Expressions
   6.3.2 Tokens-per-Type Ratios for Overall Metaphorical Expressions
   6.3.3 Comparisons between Target Domains
      6.3.3.1 Comparisons between Single Target Domains
      6.3.3.2 Comparisons between Coordinated Target Domains
6.4 Summary of Chapter

Chapter 7: Top-down Approach to Source Domain Determination

7.1 A Top-down Approach to Source Domain Determination
7.2 Mappings to WordNet and SUMO
   7.2.1 Mappings through SinicaBow
   7.2.2 Mappings through the Chinese-English Merged Word List
7.2.3 Reasons for Types of Metaphorical Expressions That Are Not Found
7.3 Concreteness of Source Domains
   7.3.1 Methodology of Selecting Concrete Senses through Measuring Concreteness of SUMO Nodes
      7.3.1.1 Measuring Conceptual Distance
      7.3.1.2 Human Rating of Concreteness for SUMO Nodes
    7.3.2 Results of Measuring Concreteness of SUMO Nodes
7.4 Determining Source Domain Names
7.5 Grouping of Types of Metaphorical Expressions Based on Source Domain Names
7.6 Summary of Chapter

Chapter 8: Results of the Top-down Approach

8.1 Performance of the Top-down Approach
8.2 Source Domains in the Top-down Approach
   8.2.1 Source Domains for the Single Target Domains
   8.2.2 Most Frequently Appearing Types of Metaphorical Expressions in Different Source Domains for the Single Target Domains
   8.2.3 Source Domains for the Coordinated Target Domains
   8.2.4 Most Frequently Appearing Types of Metaphorical Expressions in Different Source Domains for the Coordinated Target Domains
   8.2.5 Top Ten Source Domains for the Top-down Approach
8.3 Strengths and Weaknesses of the Top-down Approach

8.3.1 Being Economic versus Regional Differences

8.3.2 Measurement of Conceptual Similarity versus ‘Tree-independence’

8.3.3 Replicability versus Lack of Lexical Knowledgebases

8.3.4 Evaluation of Resources versus Noise

8.4 Summary of Chapter

Chapter 9: Bottom-up Approach to Source Domain Determination

9.1 A Bottom-up Approach to Source Domain Determination

9.2 Extraction of Collocates from the Chinese Sketch Engine

9.2.1 Types of Metaphorical Expressions Found in the Chinese Sketch Engine

9.2.2 Reasons for Types of Metaphorical Expressions That Are Not Found

9.2.3 Selected Grammatical Relations from the Chinese Sketch Engine

9.3 Computing Cut-off Points

9.3.1 Methods Based on Distributional Data of Saliency

9.3.1.1 Transformation of Data

9.3.1.2 Results of Methods One and Two

9.3.2 Methods Based on ‘Mean of Means’

9.4 Clustering

9.4.1 Mappings to WordNet through SinicaBow and the Chinese-English Merged Word List

9.4.2 Clustering of Found Collocates above ‘Mean of Means’

9.5 Analyses of Source Domain Names

9.6 Summary of Chapter

Chapter 10: Results of the Bottom-up Approach

10.1 Performance of the Bottom-up Approach

10.2 Source Domains in the Bottom-up Approach

10.2.1 Source Domains for the Single Target Domains

10.2.2 Most Frequently Appearing Types of Metaphorical Expressions in Different Source Domains for the Single Target Domains

10.2.3 Source Domains for the Coordinated Target Domains

10.2.4 Most Frequently Appearing Types of Metaphorical Expressions in Different Source Domains for the Coordinated Target Domains

10.2.5 Top Ten Source Domains for the Bottom-up Approach

10.3 Strengths and Weaknesses of the Bottom-up Approach
Appendices .................................................................................................................................................. 307
Appendix A4.1: Procedures for Running the WordNet Database .......................................................... 307
Appendix A4.2: Procedures for Running the KSMSA Database ........................................................... 311
Appendix A4.3: Procedures for Running the SinicaBow Database ...................................................... 314
Appendix A4.4: Procedures for Running the Chinese Sketch Engine .................................................. 316
Appendix A5.1: Percentages of Coordinated Instances Deleted from the First 4 Per cent of the Single Target Domains (CNA) .................................................................................... 320
Appendix A5.2: Percentages of Coordinated Instances Deleted from the First 4 Per cent of the Single Target Domains (XIN) .................................................................................... 320
Appendix A5.3: Instance Number per Year from 1991 through 2002 in CNA and XIN in Each Target Domain ................................................................................................................................. 321
Appendix A5.4: Chi-Square Analyses of Instances in Each Target Domain and in CNA and XIN .................................................................................................................................................. 321
Appendix A6.1: Types of Metaphorical Expressions, Tokens of Metaphorical Expressions and Tokens-per-Type Ratios in CNA, XIN, CNA+XIN and the Overlapped in CNA+XIN in the Single Target Domains and Coordinated Target Domains ........................................................................................................ 322
Appendix A6.2: Types of Metaphorical Expressions, Tokens of Metaphorical Expressions and Tokens-per-Type Ratios in CNA, XIN, CNA+XIN and the Overlapped in CNA+XIN in Different Target Domains ........................................................................................................ 323
Appendix A6.3: Comparisons of 政治 zhèngzhì ‘politics’ and 經濟 jīngjí ‘economy’ to other Single Target Domains through Kruskal-Wallis Tests ........................................................................................................ 324
Appendix A6.4: Comparisons of Single Target Domains versus Coordinated Target Domains through Kruskal-Wallis Tests ........................................................................................................... 325
Appendix A7.1: The Percentages of Found Types and Tokens of Metaphorical Expressions in SinicaBow When Searched in the Merged Word List ........................................................................ 326
Appendix A7.2: Questionnaire for the Concreteness Rating of the SUMO Nodes .................................. 327
Appendix A8.1: List of General Keywords in the Top-down Approach .................................................. 329
Appendix A9.1: Formula for the Calculation of Shortest Distance ......................................................... 330
Appendix A10.1: List of General Keywords in the Bottom-up Approach ................................................ 332
Appendix A11.1: Questionnaire for the Evaluation of Source Domains ................................................ 333
Index ............................................................................................................................................................ 337
List of Figures

Figure 3.1: Finding Related Metaphor for ‘Enter-Lisp’ (Martin 1990:69) ...................... 19
Figure 3.2: The Complete Hierarchy of Chinese Concept Dictionary Used by Wang (2006:46) .................................................................................................................. 20
Figure 3.3: Top-down Approach to Metaphor Identification ........................................... 22
Figure 3.4: The Search for 起飛 qǐfēi ‘take off’ in Central News Agency (CNA) in the Chinese Sketch Engine ........................................................................................................... 25
Figure 3.5: Bottom-up Approach to Metaphor Identification ........................................... 27
Figure 4.1: Search Result for ‘transport’ in WordNet 1.7.1 Browser ................................ 30
Figure 4.2: Upper Hierarchy for ‘transport’ in SUMO (Left) and WordNet (Right) ........ 32
Figure 4.3: The Merged WordNet-SUMO Taxonomy for ‘Finance’ in KSMSA ............... 33
Figure 4.4: Sense 1 for 政治 zhèngzhì ‘politics’ in SinicaBow ........................................ 35
Figure 4.5: The Chinese-English Merged Word List ......................................................... 36
Figure 4.6: Collocates for the Query Word 經濟 jīngjì ‘economy’ in the Chinese Sketch Engine ............................................................................................................................ 38
Figure 4.7: The Search Results for 經濟 jīngjì ‘economy’ Using the Thesaurus Function in the Chinese Sketch Engine .............................................................................................. 39
Figure 4.8: Sketch-Difference for 經濟 jīngjì ‘economy’ and 政治 zhèngzhì ‘politics’ ................................................................................................................................. 41
Figure 4.9: Top-down and Bottom-up Approaches to Source Domain Determination Used in Present Study ........................................................................................................ 45
Figure 5.1: Collocates for the Query Word 經濟 jīngjì ‘economy’ in the Chinese Sketch Engine ............................................................................................................................ 54
Figure 5.2: ‘And/Or’ Relation of 經濟 jīngjì ‘economy’ ...................................................... 55
Figure 5.3: Related SUMO Nodes for 經濟 jīngjì ‘economy,’ 政治 zhèngzhì ‘politics,’ 外交 wàijiāo ‘foreign affairs,’ 教育 jiàoyù ‘education’ and 財政 cáizhèng ‘finance’ ...................................................................................... 58
Figure 5.4: The Percentages of Corpora Instances for Coordinate Target Domains with 經濟 jīngjì ‘economy’ in Both CNA and XIN .......................................................... 64
Figure 5.5: Percentages of Coordination of 政治 zhèngzhì ‘politics,’ 外交 wàijiāo ‘foreign affairs,’ 教育 jiàoyù ‘education’ and 財政 cáizhèng ‘finance’ with 經濟 jīngjì ‘economy’ in Coordinated Target Domains in CNA .............................................. 66
Figure 5.6: Percentages of Coordination of 政治 zhèngzhì ‘politics,’ 外交 wàijiāo ‘foreign affairs,’ 教育 jiàoyù ‘education’ and 財政 cáizhèng ‘finance’ with 經濟 jīngjì ‘economy’ in Coordinated Target Domains in XIN .............................................. 66
Figure 6.1: Percentages of Tokens of Metaphorical Expressions in CNA and XIN in the Single Target Domains and Coordinated Target Domains .................................................. 77
Figure 6.2: Tokens-per-Type Ratios of Metaphorical Expressions for CNA, XIN and CNA+XIN in the Single Target Domains and the Coordinated Target Domains ................................................................. 81
Figure 6.3: Tokens-per-Type Ratios of Metaphorical Expressions for Overlapped and Non-overlapped Metaphorical Expressions in the Single Target Domains and the Coordinated Target Domains ............................................................................................................................ 85
Figure 6.4: Percentages of Tokens of Metaphorical Expressions in Five Target Domains with Different Datasets .................................................................................. 87
Figure 6.5: Tokens-per-Type Ratios in the Single Target Domains ................................................................................................................................. 90
Figure 6.6: Tokens-per-Type Ratios in the Coordinated Target Domains ................................................................................................................................. 94
Figure 7.1: Steps Involved in Top-down Approach to Source Domain Determination ................................................................................................................. 102
Figure 7.2: Combined Types of Metaphorical Expressions in Single+Coordinated Target Domains for the Purpose of Computation ............................................................................................................................ 103
Figure 7.3: Figure Used in Wu & Palmer (1994:136) for Measurement of Conceptual Similarity ................................................................................................................................. 116
Figure 7.4: Measuring Conceptual Similarity between SUMO Nodes ................................................................................................................................. 117
Figure 7.5: The Top Hierarchy of SUMO ................................................................................................................................................................................................. 120
Figure 8.1: Steps Involved in Top-down Approach to Source Domain Determination ................................................................................................................. 139
Figure 8.2: Percentages of Types of Metaphorical Expressions (Found in the Overall SinicaBow and the Merged Word List, Respectively) with Source Domains Determined in the Top-down Approach ................................................................................................................................. 141
Figure 8.3: Percentages of Tokens of Metaphorical Expressions (Found in the Overall SinicaBow and the Merged Word List, Respectively) with Source Domains Determined in the Top-down Approach ................................................................................................................................. 142
Figure 9.1: Steps Involved in the Bottom-up Approach to Source Domain Determination ................................................................................................................................. 167
Figure 9.2: Collocates of 成長 chéngzhǎng ‘grow/growth’ in the Chinese Sketch Engine ................................................................................................................................................................................................. 169
Figure 9.3: Patterns of Distributional Data ................................................................................................................................................................................................. 174
Figure 9.4: Collocates of ‘Subjects’ of 起飛 qǐfēi ‘take off’ in CNA in the Chinese Sketch Engine ................................................................................................................................................................................................. 175
Figure 9.5: Distributional Pattern for ‘Subjects’ of 起飛 qǐfēi ‘take off’ in CNA in the Chinese Sketch Engine ................................................................................................................................................................................................. 176
Figure 9.6: Ways to Find Cut-off Points ................................................................................................................................................................................................. 178
List of Tables

Table 4.1: Senses for 政治 zhèngzhì ‘politics’ Provided by SinicaBow ....................... 34
Table 4.2: Sub-Corpora in the Chinese Sketch Engine ............................................. 43
Table 5.1: Selected Target Domains in the ‘And/Or’ Relation with 經濟 jīngjì ‘economy’ in Both CNA and XIN Datasets Provided by Chinese Sketch Engine ... 55
Table 5.2: Selected Senses from SinicaBow for Metaphor Analysis .......................... 57
Table 5.3: Five Target Domains x CNA/XIN x Coordinated/Single Target Domains ............................................................................................................................ 60
Table 5.4: First 4 Per cent of Corpora Instances in the Chinese Gigaword Corpus for Single Target Domains in CNA and XIN .............................................................. 62
Table 5.5: Total Corpora Instances for Single Target Domains in CNA and XIN (after Removing All Coordinated Constructions) ....................................................... 62
Table 5.6: Total Instances for Coordinate Target Domains with 經濟 jīngjì ‘economy’ in Both CNA and XIN ................................................................................ 63
Table 5.7: Number of Corpora Instances in CNA and XIN ......................................... 67
Table 5.8: Period of Time for the Corpora Instances in CNA and XIN ........................ 69
Table 6.1: Mean Ranks for the Single Target Domains in CNA and XIN ...................... 92
Table 7.1: Tabulated WordNet and SUMO Information for 成長 chéngzhǎng ‘grow/growth,’ 起飛 qǐfēi ‘take off’ and 瘫瘓 tānhuàn ‘paralytic’ in SinicaBow ....................................................... 105
Table 7.2: Types of Metaphorical Expressions and Tokens of Metaphorical Expressions That Are Found in SinicaBow .......................................................... 106
Table 7.3: Tabulated WordNet and SUMO Information from 成長 chéngzhǎng ‘grow/growth’ in the Chinese-English Merged Word List ..................................................... 108
Table 7.4: Overall Types of Metaphorical Expressions and Tokens of Metaphorical Expressions That Have Been Found in the Chinese-English Merged Word List ................................................................. 110
Table 7.5: Metaphorical and Literal Meanings of 成長 chéngzhǎng ‘grow/growth’ ............................................................................................................................... 114
Table 7.6: Selected Prototypical Concrete Nodes and Their Mean Ratings .................. 122
Table 7.7: Selected SUMO Nodes with Highest Concreteness Scores from SinicaBow .................................................................................................................. 124
Table 7.8: Selected SUMO Nodes for 成長 chéngzhǎng ‘grow/growth’ with Highest Concreteness Scores from the Chinese-English Merged Word List .... 126
Table 7.9: Examples of Possible Source Domain Names for the Top-down Approach . . . . ............................................................................................................. 130
Table 7.10: Source Domains for 成長 ‘grow/growth,’ 起飛 qīfēi ‘take off’ and 麻痺 tānhuàn ‘paralytic’ ................................................................. 133
Table 7.11: Sorting Types of Metaphorical Expressions According to Source Domains (SinicaBow) ........................................................................................................ 136
Table 8.1: Most Frequently Appearing Source Domains for Each Dataset in the Single Target Domains in the Top-down Approach .......................................................... 145
Table 8.2: The Top Ten Source Domains Most Frequently Used for the Sum of the Single Target Domains in the Top-down Approach ........................................ 146
Table 8.3: The Top Two Types of Metaphorical Expressions in the Single Target Domains in the Top-down Approach ................................................................................ 150
Table 8.4: Most Frequently Appearing Source Domains in the Coordinated Target Domains (with 經濟 jīngjì ‘economy’) in the Top-down Approach .................................... 152
Table 8.5: The Top Ten Source Domains Most Frequently Used for the Sum of the Coordinated Target Domains ..................................................................................... 154
Table 8.6: The Top Two Types of Metaphorical Expressions in the Coordinated Target Domains (with 經濟 jīngjì ‘economy’) in the Top-down Approach ................. 156
Table 8.7: The Top Ten Source Domains Most Frequently Used for the Sum of the Single+Coordinated Target Domains in the Top-down Approach .................... 158
Table 8.8: Strengths and Weaknesses of the Top-down Approach ................................................................................................................................. 159
Table 8.9: Overall Types of Metaphorical Expressions and Tokens of Metaphorical Expressions That Are Found in the Chinese Sketch Engine .................... 171
Table 9.1: Grammatical Relations in Each Type of Metaphorical Expression That Contain the Five Target Domains .......................................................................................... 173
Table 9.2: Calculation of Cut-off Points Using Methods One and Two (CNA) ............... 180
Table 9.3: Method One (Short Distance): ‘Subject’ 起飛 qīfēi ‘take off’ (CNA) .............. 181
Table 9.5: Mean of Means: ‘Subject’ 起飛 qīfēi ‘take off’ (CNA) .................................... 184
Table 9.6: Calculation of Cut-off Points Using Mean of Means ..................................... 185
Table 9.7: Collocates That Are Successfully Mapped to WordNet through SinicaBow and the Chinese-English Merged Word List .................................................... 188
Table 9.8: Tabulated WordNet Information from SinicaBow ........................................... 190
Table 9.9: Percentages of Successfully Clustered Types of Metaphorical Expressions (Based on Their Found Collocates) in CNA and XIN .................................................. 191
Table 9.10: Examples of Two Hierarchies with the Highest Conceptual Density ............ 195
Table 9.11: Collocates of ‘Subjects’ of 復甦 fùsū ‘recover/recovery’ in CNA in the Chinese Sketch Engine .................................................. 196
Table 9.12: Examples of Selected Three Top Levels with High Conceptual Density ........ 200
Table 9.13: Selection of Top Levels from WordNet for Source Domain Names .......... 202
Table 10.1: Types and Tokens of Metaphorical Expressions with Found Source Domains in the Bottom-up Approach ............................................................................................................... 206
Table 10.2: Collocates for 建設 jiàngsè ‘construct/construction’ in CNA and XIN .... 208
Table 10.3: Most Frequently Appearing Source Domains in the Single Target Domains in the Bottom-up Approach ........................................................................................................ 209
Table 10.4: The Top Ten Source Domains Most Frequently Used for the Sum of the Single Target Domains in the Bottom-up Approach ...................................................... 211
Table 10.5: The Top Two Types of Metaphorical Expressions in the Single Target Domains in the Bottom-up Approach .............................................................................. 213
Table 10.6: Most Frequently Appearing Source Domains in the Coordinated Target Domains in the Bottom-up Approach (with 經濟 jìngjì ‘economy’) .................. 215
Table 10.7: The Top Ten Source Domains Most Frequently Used for the Sum of the Coordinated Target Domains in the Bottom-up Approach ........................................ 217
Table 10.8: The Top Two Types of Metaphorical Expressions in the Coordinated Target Domains in the Bottom-up Approach (with 經濟 jìngjì ‘economy’) ..... 218
Table 10.9: The Top Ten Source Domains Most Frequently Used for the Sum of the Single+Coordinated Target Domains in the Bottom-up Approach ............... 221
Table 10.10: Strengths and Weaknesses of the Bottom-up Approach ................................................................. 222
Table 10.11: Collocates of ‘Subjects’ of 成長 chéngzhǎng ‘grow/growth’ in CNA in the Chinese Sketch Engine ............................................................................................................. 225
Table 11.1: Examples of Stimuli for Human Rating Task ................................................................. 230
Table 11.2: Types of Metaphorical Expressions That Overlap in CNA and XIN ............... 233
Table 11.3: Final List of Stimuli for Human Ratings ................................................................. 234
Table 11.4: Total Number of Stimuli from the Top-down Approach ........................................ 235
Table 11.5: Total Number of Stimuli from the Bottom-up Approach ........................................ 236
Table 11.6: Top Three WordNet Noun Hierarchy Selected and Their Bottom Nodes... 237
Table 11.7: Means for Overall Percentages of ‘Yes’ ................................................................. 242
Table 11.8: Means for Overall Percentages of ‘Yes’ According to Conditions.............. 243
Table 11.9: Comparisons between SinicaBow and the Merged Word List in the Top-down Approach by the Non-CWN Group ................................................................. 244
Table 11.10: Examples of Source Domains Obtained through SinicaBow in the Top-down Approach ............................................................................................................. 245
Table 11.11: Examples of Source Domains Obtained through the Chinese-English Merged Word List in the Top-down Approach ................................................................. 246
Table 11.12: Mean Percentages of ‘Yes’ (Divided by the Number of Non-CWN Subjects=8) for All Stimuli with Source Domains with Saliency ‘Rank 1’ to ‘Rank 6,’ Respectively ................................................................. 249
Table 11.13: Kruskal-Wallis Tests Comparing between ‘Rank 1’ and ‘Rank 6’............ 250
Table 11.14: Means for the Percentages of ‘Yes’ for Non-CWN in CNA versus XIN ............................................................................................................................... 251
Table 11.15: Means of the Percentages of ‘Yes’ for the Non-CWN Group in Overlapped versus Non-Overlapped........................................................................... 252
Table 11.16: Source Domains with the Highest Ratings for Different Types of Metaphorical Expressions in the Top-down Stimuli ................................................ 254
Table 11.17: Source Domains with the Highest Ratings for Different Types of Metaphorical Expressions in the Bottom-up Stimuli ........................................ 257
Table 13.1: Possible Source Domains for 成長 chêngzhâng ‘grow/growth’ (CNA) Obtained through the Bottom-up Approach ............................................................. 290
Chapter 1: Introduction

Different people define source domains according to various levels of specificity. Consequently, there is little agreement between one another regarding whether or not a metaphor such as ‘building economy’ should belong to the source domain of HOUSE, BUILDING or CONSTRUCTION. Lakoff (1993:212) attempts to answer this question by suggesting that the mapping between source and target should occur at the superordinate level. However, he does not attempt to explain whether the source domain of BUILDING is more superordinate than CONSTRUCTION or vice versa. He also does not explain what belongs to the basic level. He goes on to state that VEHICLE is found at the superordinate level, while CAR and BOAT both reside at the basic level (thus, VEHICLE is a more suitable source domain than CAR or BOAT). In other instances, however, it is difficult to determine the superordinate category. For example, Lakoff has labeled WAR as a source domain, although it remains unclear as to why he is certain that WAR should be located at the superordinate level.

It is, therefore, obvious that the issues involved in source domain determination range from the varied definitions of source domains to the varied concreteness of source domains as well as encompass the inconsistent quality of human intuition inherent in the process of defining source domains. The impossibility in defining a closed set of items for source domains can also be due to the nature of the Conceptual Metaphor Theory (Lakoff & Johnson 1980, Lakoff 1993). For example, the Conceptual Metaphor Theory states only ‘what source domains do’ but not ‘what source domains are.’ Lakoff (1987: 219) defines conceptual metaphors as “metaphors that map complex conceptual structures in a source domain onto conceptual structures in a target domain.” This definition is not helpful in clarifying what constitutes a source domain. Croft (2003:175) makes the following statement regarding the Conceptual Metaphor Theory:

...if one accepts Lakoff and Johnson’s theory of Metaphor, as I do, one must be more specific as to what domain or domains are involved in a metaphor. I argue that the two domains being compared are base domains, that is, the bases of the profiled predication (Croft 2003:175).

Kövecses (2002:4), too, has tried to define source domains more clearly. He defines the source domain as “[t]he conceptual domain from which we draw metaphorical expressions to understand another conceptual domain."

1 In this excerpt by Croft (2003:175), ‘profiled’ is used in the sense of Langacker (1987).
Lakoff’s (1987, 1993) view of conceptual metaphors adopts a top-down approach, which suggests that conceptual metaphors appear “at the level of thought” (Deignan 1999: 180). In a book review regarding figurative language from cross-cultural and cross-linguistic perspectives, Kövecses (2006:191) expresses his personal opinion of top-down and bottom-up approaches. He states that top-down approaches first “find certain data, make particular generalizations given that data, and suggest global cognitive structures” underlying the data. The Conceptual Metaphor Theory (or Contemporary Theory of Metaphor) (Lakoff & Johnson 1980, Lakoff 1993) proposes a scenario approach to conceptual metaphors, whereby prior knowledge of the mapped domains (target domains and source domains) is assumed to already exist before conceptual metaphors are created. This view of conceptual metaphors embodies the top-down approach, where all occurrences of conceptual metaphors will be attributed to embodied conceptualization, which can be proved cross-linguistically. Due to this embodied conceptualization, conceptual metaphors in different languages have been found in the works of Yu (1995) for English and Chinese; Marim-Arrese (1996) for English and Spanish; and Özçalışkan (2003) for English and Turkish. These studies support the assertion that conceptual metaphors are psychologically valid, and that moreover, they are found abundantly in a number of different languages.

Since the top-down approach only looks for global conceptualization, it usually does not clearly state just what goes on between the two mapped domains. The bottom-up approach, however, takes into consideration specific details in the creation of conceptual metaphors. Therefore, bottom-up approaches usually “begin with an extensive (not selective) set of data, make minimal generalizations about the data, and are much less in the business of suggesting global cognitive structures that account for the data” (Kövecses 2006:191). For example, the Conceptual Metaphor Theory (Lakoff & Johnson 1980, Lakoff 1993) does not specify the criteria for source domains (with the exceptions of ‘concreteness’ and ‘at superordinate level’). Many studies that follow this model also have different ways of determining the names of source domains. The variety of methods employed in the determination of source domains is also due to the limitations of the model regarding the provision of quantitative data to build a logical statistical analysis. In addition, no specific principles can be said to exist between the source-target domains’ mappings. For instance, in the Conceptual Metaphor Theory, there is no mention of what governs the mapping within domains in a conceptual metaphor, such as ARGUMENT IS WAR (as also argued in Ahrens 2002). Lakoff (1993:206) offers the following thought regarding the general principle of Conceptual Metaphor Theory:

Is there a general principle governing how these linguistic expressions about journeys are used to characterize love?
Is there a general principle governing how our patterns of inference about journeys are used to reason about love when expressions such as these are used?

The answer to both is yes. Indeed, there is a single general principle that answers both questions, but it is a general principle that is neither part of the grammar of English, nor the English lexicon. Rather it is part of the conceptual system underlying English (Lakoff 1993:206; bold added).

In fact, the Conceptual Metaphor Theory operates through Idealized Cognitive Models (ICMs) (Lakoff 1993), which assumes the existence of a cluster of concepts from which a conceptual category can be derived. These “cognitive models structure thought and are used in forming categories” (Lakoff 1993:13). Lakoff emphasizes the variation in “classical” categories using ICMs. He claims that “classical” categories are based on “folk theory” (Lakoff 1993:5). He goes on to state that how laymen categorize things in their daily lives may not be true scientifically. For example, a person will naturally categorize colors according to the visual experiences encountered in his or her daily life but has no way of knowing whether or not these categories are the same as those of any other community (Lakoff 1993:24-26). When Lakoff suggests that ICMs are based on subconscious conceptual knowledge, he does not attempt to differentiate between various communities.

The bottom-up approach is evident in studies such as Charteris-Black & Ennis (2001) and Chung, Ahrens & Sung (2003). These studies do not try to claim global regularities but, rather, draw conclusions from the results of the (sampled) data analyzed. Nevertheless, the source domains identified by these studies are divergent. We will now discuss the source domains used in these studies in detail.

One example of divergent source domains is encountered in Charteris-Black & Ennis’ (2001) examination of metaphor in Spanish and English financial reports. In this particular study, the authors collected financial reports from newspapers published during the October 1997 stock market crash. In terms of similarities, both Spanish and English are shown to have used the conceptual metaphors ECONOMY IS AN ORGANISM and MARKET MOVEMENTS ARE PHYSICAL MOVEMENTS. Both languages also describe the downward spiral of the market as NATURAL DISASTERS. The differences between the two languages become clear when contrasting the more frequent use of psychological metaphors in the Spanish data (pánico ‘panic,’ tranquilizador ‘calming,’ desconfianza ‘distrust,’ etc.) to the more nautical movement of English metaphors (‘plunge,’ ‘weather the storm,’ ‘haven,’ etc.). In a separate study, Charteris-Black & Musolff (2003) compare metaphors for euro trading in British and German reports and find that both languages have metaphors that describe euro trading as an up-down
movement and as being healthy. They note that the British use more combat metaphors in their reports compared to those of the Germans, owing to the fact that the German reports perceive euro trading as beneficial action. Another study of interest is that of O’Connor (1998), which examines FINANCE metaphors in Spanish and suggests that MONEY and FINANCE metaphors can be categorized into three different types, i.e., SOLID, LIQUID and GAS. It is seldom questioned, however, whether or not SOLID, LIQUID and GAS are at the superordinate level (cf. Lakoff’s 1993 definition of source domain mentioned in page 1 of this book) or whether they are specific enough to be source domains. In the examination of euro metaphors in the British and Italian newspapers, Semino (2002) goes as far as to have the source domain of BIRTH, which is unclear in terms of its level of abstraction.

Although these studies provide a comparative examination of conceptual metaphors in different languages, these papers do not clearly define the scope of a source domain. For instance, the source domain can be as general as PHYSICAL MOVEMENTS in MARKET MOVEMENTS ARE PHYSICAL MOVEMENTS (Charteris-Black & Ennis 2001) or as specific as COMBAT in EURO TRADING IS COMBAT (Charteris-Black & Musolff 2003), in which COMBAT is also a type of PHYSICAL MOVEMENTS.

The problems associated with identifying source domains are acknowledged by Chung, Ahrens & Sung (2003), as well as by Chung, Ahrens & Huang (2005). In particular, Chung, Ahrens & Sung (2003) re-analyze the metaphor MARKET MOVEMENTS ARE NAUTICAL OR ARE WAYS OF MOVING IN THE WATER in Charteris-Black & Ennis (2001).2 They proceed to re-categorize the items into two source domains, i.e., BOAT (with the linguistic items of ‘plunge,’ ‘ripples,’ ‘floating,’ ‘bale out,’ ‘dive,’ ‘anchor’ and ‘flagship’) and OCEAN WATER (‘haven,’ ‘turn tide’ and ‘calm’). Their re-analysis also illustrates that the division of a general source domain into more specific domains, such as BOAT and OCEAN WATER, allows for different linguistic patterns to surface for further study.

The observations made from the studies mentioned previously show that (a) the definitions for source domain vary; (b) the concreteness for the source domains also varies; and (c) as all of these variations rely on the individual author’s judgment to make a final decision, all of the aforementioned studies have notably different criteria for source domain determination. The main factor at the root of this inconsistency is the fact that what constitutes a source domain has always been based on intuition.

---

2 This conceptual metaphor is a specific example of MARKET MOVEMENTS ARE PHYSICAL MOVEMENTS in Charteris-Black & Ennis (2001). Therefore, Charteris-Black & Ennis (2001) suggest several conceptual metaphors that are arranged in several levels of abstraction. See Chapter Thirteen of this book for Charteris-Black’s (2004) Critical Metaphor Analysis.
Chapter 1: Introduction

The problem regarding the use of different source domains also stems from the original idea of the Conceptual Metaphor Theory (Lakoff & Johnson 1980, Lakoff 1993), which does not provide clear criteria for defining source domains. As a result, all other studies (both top-down and bottom-up) use different criteria for naming the source domains, which is the heart of the problem when examining conceptual metaphors.

This book approaches the issues of source domain determination from an empirical point of view. We suggest that the criteria used in determining source domains can indeed be proposed and subsequently evaluated. This book also aims to compare both the top-down and bottom-up approaches in determining source domains. These two approaches are used to determine source domains with minimal human interference. The ideas behind these two approaches are similar to the characteristics of these two approaches mentioned previously: The top-down approach is based on a global conceptualization, whereas the bottom-up approach is based on a generalization from the data analyzed (cf. the distinction made by Kövecses (2006) mentioned in pp. 2-3 of this book). A comparison of these two approaches constitutes a great contribution to metaphor research, not only because source domains have not been examined previously, but also because the use of two approaches to determine source domains has not been explored in any other studies to date.

Based on the issues of interests brought forth previously, regarding source domains and different approaches, two main research questions may be postulated:

(1) a. How can lexical and computational methods reduce human subjectivity in determining source domains?
   b. What are the various advantages and disadvantages of the top-down and bottom-up approaches in determining the source domains of conceptual metaphors?

The hypotheses for these two research questions are:

(2) a. Lexical and computational methods are able to reduce human subjectivity in determining source domains through both top-down and bottom-up approaches. Several lexical resources (such as SinicaBow, WordNet, SUMO (Suggested Merged Upper Ontology) and Sketch Engine) will be used to determine source domains so that human intuition can be reduced.
   b. The top-down and the bottom-up approaches each have their inherent advantages and disadvantages. As a result, both the top-down and bottom-up approaches will perform differently in the determination of source domains: The top-down approach will return general source
domains, while the bottom-up approach will return specific source domains.

For the first research question in (1a), this book proposes to use lexical and computational methods to determine source domains for the reason that through the use of principled methods such as these, one is able to see how much human intuition is needed (or not needed) in order to generate source domains for different metaphorical expressions. For the second research question (1b), we first hypothesize that the two approaches will perform differently in source domain determination. We then further hypothesize that the top-down approach will return general source domains (since the top-down approach assumes a global conceptualization), while the bottom-up approaches will return specific source domains (since the bottom-up approach is based on generalization from data analyzed). This second hypothesis is made in accordance to the ideas of the two approaches outlined by Kövecses (2006:191).

This book will be organized according to the following chapters: In Chapter Two, we first discuss the lack of attention historically given to source domain determination. We show in this chapter that most previous studies in metaphor research tend to ignore the importance of source domains. Chapter Three will review several studies that are related to metaphor identification. These studies can also be seen in terms of using top-down and bottom-up approaches (but with the purpose of “identifying metaphors” rather than “determining source domains”). Chapter Four will introduce the lexical knowledgebases and corpus used in this book. Chapter Five will point out how the choice of corpus and corpora data will affect the analyses. This chapter will also explain how target domains are selected based on one of the knowledgebases introduced in the previous chapter. As well, the selection of corpora data based on the different target domains will be discussed in this chapter. Chapter Six will discuss the steps for metaphor identification, which is a step a priori to source domain determination (since metaphors must be collected in order to carry out source domain analyses). In this chapter, the selection of datasets and the steps for manual metaphor extraction will also be discussed. Chapter Seven will explain the determination of source domains using a top-down approach aided by the use of computational tools, such as WordNet (Miller, Beckwith, Fellbaum, Gross & Miller 1990, Fellbaum 1998) and SUMO (Niles & Pease 2001). These tools are used to define a concrete sense belonging to each metaphor type collected in Chapter Six. This concrete sense will, in turn, be used as a clue to determine source domains. Chapter Eight will discuss the results of the top-down approach, i.e., the source domains selected for different datasets. Chapter Nine will use another approach, i.e., a

---

bottom-up approach, to determine source domains for the same set of metaphor types used in the top-down approach. The bottom-up approach utilizes a collocation system called the Chinese Sketch Engine (Kilgarriff, Huang, Rychly, Smith & Tugwell 2005). Chapter Ten will discuss the results of the bottom-up approach. Chapter Eleven will evaluate the results of the source domain determination, relying on human judgment to decide whether or not the source domains are correct. Chapter Twelve will discuss the linguistic findings regarding source domains, and a hierarchical model for source domain definition will be proposed. Chapter Thirteen will conclude this book. The conclusion will address the two research questions posed and suggest possible future research directions.

---

4 The Chinese Sketch Engine is available at http://corpora.fi.muni.cz/chinese_all/ and http://wordsketch.ling.sinica.edu.tw/. See Chapter Four for the introduction of these resources.
Chapter 2: What Constitutes a Source Domain?

Although studies in various languages supporting the Conceptual Metaphor Theory (Lakoff 1993) have been undertaken, there has been minimal agreement among researchers as to what constitutes a source domain, owing to the fact that most previous studies (such as Yu 1995, Marim-Arrese 1996, Charteris-Black & Ennis 2001, and Özçalişkan 2003) considered metaphor identification to be the most important issue in metaphor research. The majority of previous researchers have adopted Lakoff’s (1993) definition of source domain (i.e., to be more concrete than target domain). For example, Charteris-Black (2004:15) says that source domains are “[b]asic domains [which] originate in human experience and are not derived from other more fundamental domains; they include, for example, space, matter and temperature.”¹ This definition holds the same notion of applying human experiences to conceptual metaphors that has been suggested by Lakoff & Johnson (1980). However, whether or not one knows how to define what constitutes human experience is another issue entirely. For instance, when Lakoff & Johnson (1980) suggest that JOURNEY is a source domain, they also suggest that JOURNEY is based on human experience (which is true). However, the question of whether or not it is appropriate for the term JOURNEY to constitute a source domain (that reflects human experiences) is rarely taken into account by most previous studies, including those of Lakoff & Johnson (1980).

In point of fact, many studies have come to the realization that it is difficult to identify certain source domains. In a prime example, ‘growth’ has been grouped under the divergent source domains of PLANT (Kövecses 2002, Charteris-Black 2004), ORGANISM (Stefanowitsch 2005, Charteris-Black & Ennis 2001) and PERSON (Chung, Ahrens & Huang 2005). In particular, Charteris-Black & Ennis (2001) have suggested that ‘growth’ be grouped under ORGANISM only to have the same term grouped under the source domain of PLANT in Charteris-Black (2004). While the vexing inconsistency in naming and determining source domains has become a major problem in metaphor analysis, it must be reiterated that none of the above mentioned studies have placed any emphasis on the importance of determining source domains, most likely due to the fact that most studies on metaphor analysis are still concentrating on metaphor identification. The Conceptual Metaphor Theory (Lakoff 1993), in which the source domains are always intuitively determined, does not provide a satisfactory solution to this question.

¹ Note that “they” in this quotation is ambiguous for being either “basic domains” or “fundamental domains.”
Most of the problems encountered in source domain determination can be categorized from one of two aspects. The first is concerned with whether or not source domains should be defined as general source domains (e.g., ORGANISM) or specific source domains (e.g., PLANT/PERSO). The second aspect is related to whether or not there is a unified definition for the concreteness of source domains and how human experiences converge and combine to agree on a similar definition of concreteness. In some cases, human experiences are in direct contradiction with the notion of concreteness. For example, DREAM and IDEA are both part of the human experience, but neither is concrete. If all human experiences are the basis for source domains, why then does IDEA require the additional source domain of BUILDING in the conceptual metaphor of IDEA IS BUILDING (Lakoff & Johnson 1980)? It may, therefore, be concluded that the definition for what constitutes a source domain is not clearly laid out, and this exacerbates the difficulty of determining what constitutes a source domain. The twin criteria of ‘human experiences’ and ‘concreteness’ alone are not sufficient to enable humans to determine source domains. Since human judgments are sometimes inconsistent in determining source domains, this book provides an alternative way of looking at the problem via its use of computational and lexical methods to objectively decide upon source domains.

As previously stated, the vast majority of metaphor studies have focused on how to identify metaphors. This book provides a different way to look at the issues involved in metaphor identification. Specifically, it probes the question of whether source domains can be determined automatically. If this is possible, then it is possible that metaphor identification can likewise become automatic. For example, there have been studies (Fass & Wilks 1983, Martin 1990) that tried to identify metaphors automatically by first examining whether or not a word violates the literal meaning; if so, the program proceeds to search as to whether this word is being used metaphorically. These studies, however, require two levels of processing (by first rejecting that a word is literal) to arrive at the conclusion that a word is metaphorical. This follows the traditional pragmatic interpretation of Searle (1979), which has been falsified by many psycholinguistic studies that found that metaphorical expressions are processed as fast as literal expressions (Glucksberg, Gildea & Bookin 1982, Gibbs, Bogdanovich, Sykes & Barr 1997, Ahrens 2002). Therefore, one question regarding the two-level metaphor identification programs is whether they are counter-intuitive and, if so, what problems this causes.

Unlike the other corpus-based and computational studies, this book focuses first on determining source domains and then turns to the question of how they can be used to identify metaphors in corpora. Therefore, determining the source domains are of utmost importance, as this constitutes the starting point for the identification of mapped domain information. Linguists such as Clausner & Croft (1997) explain the formation of
Chapter 2: What Constitutes a Source Domain?

domain mappings through schematicity and entrenchment (Langacker 1987). Yet if the
domains are not clearly stated at the outset, then how can the domain mappings be
found? Therefore, the most important issue in metaphor analysis is not merely the
identification of metaphor per se. Rather, it is imperative that the starting point must be
revisited in order to discover what constitutes a domain and, in this way, how domains
are mapped. This chapter will focus on the first issue of ‘what constitutes a domain’ by
looking specifically at source domains. Chapter Three will discuss this issue in greater
detail by examining previous works, focusing, in particular, on how these works define
general versus specific source domains first as well as how they define concrete domains
based on experience.

2.1 How to Define General versus Specific Source Domains

Despite the importance of source domains, most people cannot agree on a single
definition for source domains. Researchers have a wide range of interpretations as to
what constitutes a source domain. Some interpret source domains as specific ‘scenes’
(Grady 1999, Heywood & Semino 2005), and some use the term ‘scenarios’ (Musolff
2004). Heywood & Semino’s (2005:14) definition of ‘scene’ comprises “an action, the
participant(s) involved in the action, and a goal” and Musolff (2004:13) places his
‘scenarios’ as an “intermediate analytical category between the level of the conceptual
domain as a whole and its individual elements.” Both ‘scenes’ and ‘scenarios’ contain
individual elements (i.e., event/action, participants and goal) within themselves but
differ in their level of specificity. For example, Heywood & Semino (2005:14) propose
that some primary ‘scenes’ such as ‘support’ are used in a variety of different contexts,
such as “financial support, moral support and technical support” (italics added). They
consider Musolff’s (2004) ‘scenarios’ as non-primary scenes because of examples such as
‘to bolster support,’ which is a specific example of the primary scene of ‘support,’ where
“an extra pillow is added to prop up someone in bed” (Heywood & Semino 2005:16).
Therefore, they argue that Musolff’s definition of ‘scenario’ is more complex than their
primary scenes.

In actual fact, the distinction made by Heywood & Semino (2005) regarding
primary and non-primary scenes is directly related to the conventionality defining the
uses of source domains. As mentioned in the prototypical model (Labov 1973, Rosch &
Mervis 1975), the more prototypical a notion is, the more commonly it is found. Ahrens,
Chung & Huang (2003) have applied the prototypical paradigm to the analyses of
metaphors, whereby the most frequently occurring metaphor types become the most
prototypical and, therefore, the conventional concept. If the above case of Heywood &
Semino (2005) were to be classified according to the Ahrens, Chung & Huang (2003),
the primary scenes would form conventional metaphors (such as ‘support’), while the non-primary scenes would form ‘novel metaphors’ (such as ‘to bolster support’). While novel metaphors are not prototypical, they can be understood nonetheless.

The distinction between novel and conventional metaphors is important in source domain analyses. Without this distinction, any element could be mapped to form metaphors. For example, we can create a novel metaphor by saying ‘He is a microphone,’ but this is only one instance, and its occurrence is not entrenched enough (cf. Langacker 1987) to establish a sufficiently strong relationship. For these kinds of one-shot metaphors, the information in the domains cannot be generated systematically because there are not enough co-occurrences of MALE PERSON and MICROPHONE to make a conclusion regarding a source domain. A similar situation also occurs in poetic metaphors: Even though various poetic metaphors are extended throughout a poem, their rate of occurrence is not frequent enough to form a domain. For instance, the poem The Sick Rose by William Blake (1794) is an example of creative word usage that has not been adopted in ordinary speech. For this reason, even though the use of ‘sick rose’ is coherent throughout the poem, ROSE IS A PATIENT/PERSON is still not a conventional conceptual metaphor. Distinction between conventional and creative metaphors can also be found in Knowles & Moon (2006:5-6).

In Ahrens, Chung & Huang (2005), the distinction between conventional and novel source domains is made by comparing the frequency of a particular source domain among various other source domains. For example, the source domain of WAR has many metaphorical expressions among which are ‘attack,’ ‘at war’ and ‘fire.’ In addition, these metaphorical expressions are also frequent in tokens (i.e., each one appears frequently). The sum of these tokens constitutes the frequency of the source domain of WAR. Comparatively, the source domain of ANIMAL has fewer metaphorical expressions and a correspondingly lower token frequency. The overall frequency for the source domain of ANIMAL will be far lower than WAR, which results in the source domain of WAR being more prototypical than that of ANIMAL.

However, as mentioned, how specific a source domain is, in the end, depends on the approaches taken as well as on the various decisions that different humans make. In the previous examples, WAR and ANIMAL are still determined manually. Different individuals may have named them differently. Therefore, the inconsistency of human judgment is a serious problem in metaphor analysis, as different interpretations of a source domain may cause problems when making comparisons across studies in conceptual metaphors.

In fact, some believe that the boundaries for source domains are impossible to specify. An example of this belief, held by Lemmens (2001:38), is shown in the following paragraph:
Within this [ICM] framework, the identification of a particular usage as metaphorical hinges then on whether or not the domain boundary is crossed and this at least seems to give us a handle on how to define metaphor. However, the obvious problem is that all depends on the demarcation of conceptual domains, which is in principle not really possible. As a matter of fact, the delineation of domain boundaries is very often a post hoc operation…(Lemmens 2001:39; term in brackets added)

Lemmens believes that the “demarcation of conceptual domains” is often “a post hoc operation” and that the identification of domain boundaries is not possible. In fact, most linguistic categorizations are post hoc operations, as they occur after observations of linguistic data. The question that remains then is this: If humans are inconsistent in naming source domains, is automatic determination of source domains a possible, viable solution? Another way of looking at this problem is to probe whether there is a non-subjective way of defining source domains, despite their ambiguity. When humans find it hard to choose between source domains when attempting to find the most suitable one, computational tools can help decide which of the domains is better, thereby reducing the need for human judgment.

Some of the previous attempts to use computational approaches to identify metaphors will be shown in Chapter Three. The next section will further discuss what, in fact, constitutes a source domain by focusing in-depth on the problems encountered in source domain determination.

2.2 How to Define Concrete Domains Based on Human Experience

In addition to the specificity of source domain, the second issue at hand is related to how concrete domains are defined in metaphor analysis, as it has been suggested that source domains are more concrete than target domains. Several conceptual metaphors are shown in (1). These conceptual metaphors are taken from the Conceptual Metaphor Homepage.²

(1) a. Emotional Stability Is Balance
   b. Coherent Is Aligned
   c. Creation Is Cultivation
   d. Love Is a Journey
   e. Time Is a Resource

² The Conceptual Metaphor Homepage is available at http://cogsci.berkeley.edu/lakoff/.
The source domains of these conceptual metaphors (underlined) are supposedly more concrete than the target domains (‘emotional stability,’ ‘coherent,’ etc.). However, we can clearly see in (1) that these source domains are not necessarily concrete. In fact, most of the decisions regarding concreteness of domains are based on individual experience. There has been no criteria outlined attempting to specify this concreteness because human experiences are not necessarily concrete, so it is therefore impossible to have the same measure of concreteness remain invariable across a group of individuals. Based on the prototype theory (Rosch & Mervis 1975), a group of prototypical concrete objects may exist, but there may also be fuzzy areas where objects can either be concrete or abstract. Different psychological studies have different definitions of concreteness. For example, in the study of Sloutsky, Kaminski & Heckler (2005) regarding learning and transfer, it was found that the concreteness of symbols facilitates learning. Their concrete stimuli are defined as “perceptually rich images of concrete objects.” Other psychologists, such as deGroot (1965), suggest that concrete words are able to form an “image” in our thoughts, although having an “image” may not be a good criterion to measure concreteness. For instance, while it is possible to have an “image” of God, God is never concrete. Likewise, the definitions of concreteness will naturally vary, thus, indicating that concreteness is not easily measured or defined. The problem with concreteness in terms of metaphors, as suggested by Lakoff & Johnson (1980), is that source domains are usually more concrete. Humans map concrete source domains to abstract target domains unconsciously based on their life experiences, and it is a given that no two people in this world have had identical experiences defining their lives. For example, Lakoff & Johnson (1980) create the source domain of WAR, but not many people have had the concrete experience of war. For many people, the knowledge of war comes second-hand from images garnered in movies, from historical descriptions and, most regularly, from what they see or read in the news regarding wars in other, usually distant, countries. If few people have actually experienced war, the question of how individuals measure the concreteness of their images of war or if, indeed, it is even possible remains open. A similar problem is that of the source domains’ specificity. Sometimes there are just too many possibilities, resulting from a multitude of individual human judgments, which begs the question of whether or not an automatic ‘system,’ one unhampered by the biases from an individual’s personal life experiences, will help in the selection of source domains.

It is, therefore, the stated aim of this book to work towards refining the definitions and criteria necessary to determine source domains by using both computational and lexical methods. These two problems, outlined previously, will be dealt with in Chapters Seven through Ten. What follows first is a clarification of the terminology for source domains and of the terms used in this book.
2.3 Terminology

In order to avoid the generality of the scenario approach of metaphors caused by the unspecified definition of source domains, this book seeks to use certain guidelines regarding source domains. The scopes of source domains have been outlined by Ahrens (2002), shown in (2). In Ahrens (2002), several questions are asked in order to collect metaphorical expressions used by native speakers. These questions serve as cues that will lead the respondents to brainstorm the characteristics of a particular source domain. The sample answers for (2) are answers for the source domain of BUILDING (as in IDEA IS BUILDING).

(2) Sample answers for BUILDING:

1. What entities does the SD have?
   -- foundation, structure, base, model, layout, cement, brick, steel bar,
   -- sandstone, scaffolding, roof, wall, worker, window, door, plumbing, decoration

2. What qualities does the SD or the entity in the SD have?
   -- shaky, high, short, strong, weak, flimsy

3a. What does the SD do?
   -- to protect, to shield, to shelter

b. What can S/O do to or in the SD?
   -- to live in, to build, to construct, to tear down

(Ahrens 2002: 276-277; bolds added)

Based on this definition, we can analyze source domains in terms of their participants, actions, functions, etc. Previously, in models such as the Attributive Categorization Model by Glucksberg & Keysar (1990), the elements of source domains are not clearly stated. For example, in the conceptual metaphor of MY JOB IS A JAIL, the attributes that are compared are not specified. The same goes for MY SURGEON IS A BUTCHER, where additional interpretation of the mapped domains (i.e., JOB, JAIL, SURGEON and BUTCHER) is required. Comparatively, source domains identified using the previous questions are more clearly stated. These questions make clear that metaphorical expressions involve not only the entities but also other elements in the source domains. Su (2002) also makes similar predictions regarding metaphors, believing that they not only exist in the form of ‘nouns’ but that they also include ‘qualities’ and ‘functions.’

3 A similar (‘element-based’) definition can be found in the ‘frame semantics’ by Fillmore (1977, 1982) as well as Fillmore & Atkins (1992). However, since the purpose of this book is not to
For the purpose of terminology clarification, examples such as IDEA IS A BUILDING are called ‘conceptual metaphors,’ and under these conceptual metaphors, there are various ‘types of metaphorical expressions,’ such as ‘foundation,’ ‘structure’ and ‘base,’ as suggested in (2). The frequency of each of these types will form the ‘tokens of metaphorical expressions’ in the corpora. The studies of conceptual metaphors in this book exclude image metaphors and orientational metaphors (such as GOOD IS UP or BAD IS DOWN).

2.4 Summary of Chapter

This chapter discusses the issue of ‘what constitutes a domain’ and looks, in particular, at (a) how previous works have defined general and specific source domains; and (b) how people define the concreteness of source domains based on their own human experience. Concreteness is quite difficult to define, even within the realm of psychology, where different studies seem to have different interpretations of concreteness. This chapter also gives suggestions for how these problems may be solved using approaches proposed in this book. As computational approaches will be employed in this work, the next chapter will review examples of computational work, as well as linguistic research, that has until now dealt with metaphor identification. Most of these studies have either ignored the importance of source domains or have assumed the source domains from Lakoff’s Conceptual Metaphor Homepage to be correct. These previous studies will be discussed in terms of the top-down approach and the bottom-up approach. These approaches have appeared separately in different studies but have never been compared in terms of their performance of source domain determination. This book will be the first work to compare both of these approaches in determining source domains, an aspect of metaphor analysis that has often been taken for granted in the vast majority of previous works. This book will not only place emphasis on computational and lexical methods, it will also emphasize the importance of source domains as a starting point for the formation of conceptual metaphors.

compare the varied definitions for “concepts, domains, construals, and categories,” as found in Clausner & Croft (1999:3) and Croft (2003), the controversies of these definitions will not be discussed in depth. Another work that also discusses the definitions of ‘concepts’ can be found in Leezenberg (2001:251-304), where ‘concepts’ in terms of the Wiggensteinian and Vygotskyan approaches are discussed.
Chapter 3: Metaphor Identification versus Source Domain Determination

In Chapter Two, we have taken a look at how source domain determination is an important issue in metaphor research (and yet it is often ignored). In this chapter, we will review some studies in which the purpose is the identification of metaphors. Similar to the studies in Chapter Two, these studies do not focus on defining the source domains. Moreover, as will be revealed in this chapter, their source domains are usually ‘borrowed’ from the existing Conceptual Metaphor Theory (Lakoff & Johnson 1980, Lakoff 1993). These ‘borrowed’ source domains, as discussed in Chapter Two, are sometimes inconsistent in generality and concreteness, which is the reason why different studies show different or incomparable source domains.

In this chapter, we will examine several studies and their treatment of source domains in the process of metaphor identification. Most of the studies mentioned in this chapter incorporate computational techniques; there are, however, some corpus-based studies that aim to identify metaphors. Like many disciplines of research, studies in metaphor identification can be viewed in terms of both the top-down approach and the bottom-up approach.

As mentioned in Chapter One, the Conceptual Metaphor Theory (Lakoff & Johnson 1980, Lakoff 1993) identifies metaphors using a top-down approach, meaning that conceptual metaphors (and source domains) are pre-determined based on intuition. Most top-down studies discussed in this chapter use the source domains of the Conceptual Metaphor Theory, especially those studies in the computational fields that use readily stored source domains as cues to extract metaphors automatically from corpora. Furthermore, previous computational approaches to metaphor have aimed to identify metaphors automatically and, thus, have not focused on “how” the conceptual mappings between the target-source domain pairings are related. This is because the goal of these studies is to produce a computational system that can “identify” metaphor, and so it would appear that what constitutes a source domain does not overly concern the authors of these studies.¹

On the other end of the pole, bottom-up approaches to metaphor identification usually identify source domains based on instances of metaphor use. These studies usually postulate source domains based on observation of the metaphorical patterns. Studies

¹ Another online database that also uses pre-determined source domains is that of Metalude at the Department of English, Lingnan University (http://www.ln.edu.hk/lle/cwd/project01/web/home.html).
using the bottom-up approach utilize the preference restriction paradigm. The preference restriction paradigm stipulates that the preferred arguments of a particular word be estimated statistically.

Many metaphor analyses can be divided into either top-down or bottom-up approaches, but none have compared these two paradigms using computational methods. This book aims to compare the top-down and bottom-up approaches, not in terms of metaphor “identification,” but rather in terms of “source domain determination.” This chapter looks at metaphor “identification” (and not “source domain determination”) because no previous work has emphasized the importance of source domain in metaphor analyses. In this chapter, we see how source domains are treated during the process of metaphor identification. The importance of source domain is emphasized in this book for the reason that source domain information is mapped onto the target domains to bring out the metaphorical meanings. The following sections will explain how previous studies have used the top-down approach (§3.1) and the bottom-up approach (§3.2) in metaphor identification.

### 3.1 Top-down Approach to Metaphor Identification

Studies that employ the top-down approach are usually united by a common similarity; they require lexical knowledgebases in order to determine source domains. In fact, there have been only two previous works that have made use of lexical knowledgebases for this approach: Martin (1990) and Wang (2006). Both of these works use forms of knowledge hierarchy to identify metaphors.

Martin (1990) proposes a system called MIDAS (Metaphor Interpretation, Denotation, Acquisition System). MIDAS is a computer system that is able to interpret metaphor and represent it, in addition to being capable of dynamical learning as soon as a new metaphor is inputted in the system. The MIDAS system recognizes a metaphor through the KODIAK semantic network created by Wilensky (1986). For instance, if the metaphor ‘kill a process’ is entered into the MIDAS system, the template known as the ‘metaphor-map’ for ‘kill’ will be searched, and if ‘process’ does not match the literal route, the expression will be interpreted as a metaphor. Furthermore, the system also displays the source domain related to ‘kill’ in the metaphorical use. To illustrate this concept using another example, please refer to Figure 3.1, where ‘enter-lisp’ is categorized as a ‘CONTAINER’ metaphor (circled), and this source domain of CONTAINER is pre-stored in the system.
Chapter 3: Metaphor Identification versus Source Domain Determination

Figure 3.1: Finding Related Metaphor for ‘Enter-Lisp’ (Martin 1990: 69)

From Figure 3.1, it can be seen that the ‘CONTAINER’ metaphor interpreted by MIDAS is taken from the source domains found in Lakoff & Johnson (1980) and Lakoff (1993). Even though Martin claims that MIDAS can learn new metaphors, the source domains are always strikingly similar for the simple reason that the system will, on initial spec, determine whether or not an inputted expression is an extension of an existing metaphor. If the answer to this question is yes, source domains from the MIDAS database will be displayed. From this process, it is clear that the system does not aim to determine the source domain from the inputted data but, rather, to search for a suitable source domain within the existing database.

Given the example of ‘kill a process’ discussed previously, MIDAS will have to reject the literal route before arriving at the metaphorical meaning. This approach is not only uneconomical, but it also requires many templates for ‘metaphor-maps’ in order to be able to recognize that an expression is metaphorical. For example, if ‘kill’ has several literal meanings, all the templates for the literal meanings must first be rejected before a metaphor can be identified. However, if the source domain for ‘kill’ can first be determined (e.g., ‘kill’ belongs to the source domain of ANIMATE BEING), the identification of ‘kill’ as a metaphor can be made by way of identifying its arguments that are not animate objects (such as ‘process’ and ‘system’). This way of identifying metaphors may produce quicker results than one that necessitates the rejection of all possible literal meanings of ‘kill’ before beginning the process of metaphor identification.

In another study, Wang (2006) uses the Chinese Concept Dictionary (CCD) (Liu 2004) to create a Chinese Noun-Noun metaphor database. The Chinese Concept Dictionary is similar in hierarchical structure to the English WordNet (see Chapter Four for the introduction of WordNet) and was created by the School of Electronics Engineering and Computer Science in Peking University. The hierarchy of the Chinese Concept Dictionary can be seen in Figure 3.2. All metaphors identified in the dictionary are mapped to this hierarchy, and source domains, such as ANIMAL (動物), PERSON (人類), PLANT (植物), ARTIFACT (人工物), NATURAL OBJECT (自然物) and
NATURAL PHENOMENON (自然現象), are pre-determined based on the Chinese Concept Dictionary.


**Figure 3.2:** The Complete Hierarchy of Chinese Concept Dictionary Used by Wang (2006:46)

Wang (2006) then searches a parsed Chinese dictionary called “The Grammatical Knowledge-Base of Contemporary Chinese” or GKB (語法信息辭典) to look for lexical items that possibly fall under these source domains. Lexical items, such as 海洋 hǎiyáng ‘ocean,’ 海灣 hàiwān ‘bay,’ 花朵 huāduǒ ‘flower,’ 工程師 gōngchéngshī ‘architect’ and 道路 dào lù ‘road,’ are then selected as keywords to search in the dictionary. These lexical items have been found to possess metaphorical meanings with several target domains. Constructions that help to define metaphors are shown in (1) for the examples of 海洋 hǎiyáng ‘ocean.’

---

2 Chinese pinyin is generated using the Pintone software (Teng, Cheng & Lin 2006).
Chapter 3: Metaphor Identification versus Source Domain Determination

(1) a. 競爭 的 海洋
   jìngzhēng de hǎiyáng
   ‘The ocean of competition...’

   b. 市場 經濟 的 海洋
   shìchǎng jīngjì de hǎiyáng
   ‘The ocean of (the) market’s economy...’

All target domains (boxed) that can form metaphorical expressions with these lexical items (underlined) are identified from the dictionary. In (1), 海洋 hǎiyáng ‘ocean’ forms metaphors with different target domains, such as 競争 jìngzhēng ‘competition’ and 市場 經濟 shìchǎng jīngjì ‘market’s economy.’ These possible occurrences of target domains with 海洋 hǎiyáng ‘ocean’ are first manually analyzed and then later extracted through machine learning.

Wang’s (2006) methodology raises several pertinent issues regarding source domain determination, the first one being that the source domains are intuitively selected from the beginning. Even though this book agrees that source domains are part of a conceptual hierarchy, we maintain that these source domains should be selected based on certain criteria. For example, Wang categorizes lexical items such as 海洋 hǎiyáng ‘ocean’ under NATURAL OBJECT; however, this grouping of lexical items is not completed automatically but, rather, accomplished manually. This issue relates directly to one of the core aims of this book: to reduce the mitigating factor of human judgment in the process of determining source domains. The summaries of these two studies can be seen in Figure 3.3.

As stated previously, the two top-down studies on metaphor identification reviewed in Figure 3.3 either borrow the source domain terms from the Conceptual Metaphor Homepage or manually categorize source domains under a knowledge hierarchy, such as the Chinese Concept Dictionary by Wang (2006). These studies do not focus on source domain determination, as their purpose is still metaphor identification. With regards to Wang (2006), for instance, the study could have been improved in terms of source domain determination if the groupings of lexical items (such as 海洋 hǎiyáng ‘ocean’ and 海灣 hǎiwān ‘bay’) had been automated. As it stands, Wang’s study (2006) selects nodes at particular levels within the Chinese Concept Dictionary manually. This causes the initial problem discussed earlier in Chapter Two to arise: The issue of the specificity of a source domain and how it differs across a sea of individuals rears its head. This problem could be avoided if Wang (2006) were to not first specify the source domains but, instead, have them proven automatically through the observations made in the
corpora. This is the approach that this book will take in order to improve the automatic selection of source domains for metaphors.

![Top-down Approach to Metaphor Identification](image)

**Figure 3.3:** Top-down Approach to Metaphor Identification
In the following section, studies using bottom-up approaches are discussed in order to determine whether or not bottom-up approaches have treated source domain determination as an important issue.

3.2 Bottom-up Approach to Metaphor Identification

While almost all top-down approaches use a knowledge hierarchy, all bottom-up approaches conversely use the notion of collocation. In this section, six studies will be discussed in terms of their use of the bottom-up approach. Among these studies, four are linguistics-based and the other two are computational-based.

With regards to the linguistic analyses of metaphors, the importance of collocation can be seen in the work of White (2003), Deignan (2005), Chung (2005, 2008) and Stefanowitsch (2005, 2006). Stefanowitsch (2006) argues that metaphors can be extracted via the implementation of the Metaphorical Pattern Analysis (MPA) of the metaphors, where a target domain that lies within a string of words belongs to an identifiable source domain. For example, in the sentence ‘He shot down all of my arguments,’ ‘arguments’ can be replaced by words such as ‘planes’ and ‘missiles,’ suggesting the source domain of WAR (also pre-determined) to be a possible source domain. The MPA analysis is shown in (2).

(2) MPA:

He shot down all of my arguments
TD DISCUSSION: argument
SD WAR:
He shot down my planes/missiles/…
General mapping: DISCUSSION IS WAR
Specific relation: argument = plane/missile

(Stefanowitsch 2006:67)

Using this framework, only keywords such as ‘plane/missile’ and ‘arguments’ are substituted, but the environment (the words occurring around the keywords) remains the same. In other words, the collocation information of the environment and the keywords are used in tandem to predict the meaning of the metaphorical string. Intuitively, it seems logical that this same method should also work for Chinese. For instance, ‘經濟起飛 Jingji qifei ‘economy takes off’ can be substituted by 飛機起飛 feiji qifei ‘airplane takes off’ or 蝴蝶起飛 hudie qifei ‘butterfly takes off,’ indicating AIRPLANE and

---

3 Recently, more publications, such as Dobrovolskij & Piirainen (2005:47) and Fellbaum (2007), have discussed the relation between metaphors and collocations in depth.
BUTTERFLY are possible source domains for 經濟 jingji ‘economy.’ This test reveals collocation to be a possible method to determine source domains in Chinese and, moreover, shows that selectional restriction is one way to observe collocational patterns in language use.

Another corpus-based work that also proposes collocation is that of White (2003), who examines metaphors related to the keyword ‘growth.’ Although White (2003) does not use the term ‘collocation,’ his categorization of ‘growth’ is aligned according to semantic relations, such as ‘agent,’ ‘patient’ and ‘participant.’ Comparatively, Deignan’s (2005) association of grammar with metaphor is slightly different, i.e., the focus is placed at the morphological level where meanings carried by different parts-of-speech, word class shift and inflection become more important.

The categorization of collocates according to grammatical relation is also employed by Chung (2005, 2008), where metaphors related to MARKET are identified manually, and all metaphorical expressions are sorted according to syntactic categories, such as ‘subject,’ ‘object,’ ‘modifier,’ etc. Access to these syntactic patterns is now available online with a program called Sketch Engine (Kilgarriff & Tugwell 2001).4 Figure 3.4 shows the display of collocates according to grammatical relations in Sketch Engine when a search of the phrase 起飛 qifei ‘take off’ is undertaken in the corpus of Central News Agency (CNA) in Taiwan. The results display collocates of the searched word in terms of grammatical relations, such as ‘subject,’ ‘modifier,’ etc. From here, it is evident that the occurrence of the term 經濟 jingji ‘economy’ (as in 經濟起飛 jingji qifei ‘economy takes off’), coincides with the occurrence of literal collocates, such as 飛機 feiji ‘airplane’ and 班機 banjii ‘flight’ (see boxed area in Figure 3.4). In Figure 3.4, the ‘subject’ relation notes the appearance of 經濟 jingji ‘economy’ and of other literal collocates possessing the same syntactic position of 經濟 jingji ‘economy,’ illustrating the original meaning of 飛機 feiji ‘airplane’ when 起飛 qifei ‘take off’ is not used metaphorically. The possibility of finding literal collocates to replace 經濟 jingji ‘economy’ illustrates that Stefanowitsch’s (2006) Metaphorical Pattern Analysis can, in turn, be used in source domain determination (more about Sketch Engine can be found in Chapter Four).5 However, more operational steps need to be undertaken to

---


5 We place inverted commas between the names of the grammatical relations (‘subject,’ ‘object,’ etc.) to indicate that these are the terms provided by Sketch Engine. Noise can sometimes be found among these grammatical relations (see Chapter Ten for the weaknesses of Sketch Engine).
name the source domains, as this is not emphasized in the Metaphorical Pattern Analysis.

**Figure 3.4:** The Search for 起飛 qǐfēi ‘take off’ in Central News Agency (CNA) in the Chinese Sketch Engine

With an eye to a purpose similar to that of Sketch Engine, White (2003) and Chung (2005, 2008) also arrange collocates under certain syntactic patterns. Their approach is also comparable to that of Stefanowitsch (2006), in so far that they all try to group collocates according to certain grammatical constructions. The only difference is that White (2003) does not carry out a statistical measure of collocates, whereas Stefanowitsch (2006) and Chung (2005, 2008) do. Although we propose that Stefanowitsch’s (2006) Metaphorical Pattern Analysis can be used to find source domains, Stefanowitsch (2006) does not use his own model to determine his source domains. Instead, he uses Metaphorical Pattern Analysis for identifying constructions, stating that metaphors such as *X fill with anger*, *X be full of/filled with anger*, *X keep lid on/contain anger*, *X held-in/pent-up anger* and *X be unable to contain anger* fall under the source domain of ‘ANGER/BEING ANGRY, IS A SUBSTANCE IN A CONTAINER (UNDER PRESSURE)’ (Stefanowitsch 2006:76). However, it should be noted that the source domain of ‘A SUBSTANCE IN A CONTAINER (UNDER PRESSURE)’ is still determined manually using his method.
In terms of computational works, which make use of a bottom-up approach, Mason (2004) stands out as one example that also works within the paradigm of collocation. Mason (2004) was the first dissertation that used large corpora (i.e., the Web) and domain-specific documents to determine the selectional preferences of verbs (he does not call it collocation) that are used metaphorically. For instance, when Mason (2004) examines words, such as ‘pour,’ ‘flow’ and ‘freeze,’ he finds selectional preference for ‘liquid’ and ‘assets.’ He then, accordingly, decides that the conceptual metaphor is FINANCE IS LAB based on a polarity measure. However, the way Mason defines source domains raises a labeling issue: Specifically, the final labeling for the conceptual metaphor FINANCE IS LAB (‘assets’ for FINANCE and ‘liquid’ for LAB) is subjectively determined. Mason (2004) does not explain why ‘assets’ cannot be a source domain by itself but, instead, must be assigned to another source domain of FINANCE. Likewise, he also does not explain why he has grouped ‘liquid’ as a part of LAB. These are the questions asked in this book regarding source domain determination, which no studies have discussed to date, as none have taken a serious look at the issue of source domain.

Another computational work that also makes use of the selectional restriction paradigm is that of Wilks (1975). A newer version of this work can be seen in Fass & Wilks (1983:179). Fass & Wilks implement a system called ‘Preference Semantics,’ a model that suggests a “semantic formula” to be the representation of each word-sense. For instance, the word ‘drink’ in (3) shows the sense-frame for the word ‘drink.’

\[(3) \quad (((\text{FLOW STUFF}) \text{ OBJE}) \quad \text{(MOVE CAUSE))} \]

This formula suggests that ‘drink’ “is an action, preferably done by animate things (*ANI SUBJ) to liquids ((FLOW STUFF) OBJE). The SUBJ (subject) displays the preferred agents of actions while the OBJE (object) displays the preferred objects or patients” (Fass & Wilks 1983:179). Sense-frames such as (3) serve as templates to decide whether a query sentence violates the template. If the literal templates are violated, a metaphorical interpretation will then be derived. This system, which was built in the late 1970s, also operates under the traditional pragmatic interpretation of Searle (1979), where a falsehood decision has to be made before a metaphorical interpretation is employed. A summary of the studies using bottom-up approaches can be seen in Figure 3.5.

---

6 The Preference Semantics suggests a sense-frame that utilizes selectional restriction, which Fass & Wilks (1983:180-181) called “preference-as-restriction.” The system, known as Met*, uses 500 sense-frames to decide whether a sentence is literal or non-literal based on whether or not this sentence violates the sense-frame.
Figure 3.5: Bottom-up Approach to Metaphor Identification

The bottom-up approach to metaphor identification has the advantage of having empirical evidence, but it also has a large disadvantage: the problem of how to label source domains. When lexical items have been collected under different groups using a bottom-up approach, a final category name must be assigned manually to the source domains in the end. This is the clearly seen weakness in all studies in Figure 3.5, where automatic naming of source domains becomes a problem. If we look at the sense-frame
A Corpus-driven Approach to Source Domain Determination

of Fass & Wilks (1983), there is no indication of what the source domain is. In fact, most metaphor identification like this usually downplays the roles of source domains because the aim is always to identify metaphors, not the source domains. This problem, however, will be limited in this book because the source domains will be determined at the outset of the study. For example, when Mason (2004) found the use of ‘pour’ with ‘money’ as well as ‘pour’ with ‘liquid,’ he could have searched for all other collocates of ‘pour’ in a general corpus so that clusters of words relating to ‘liquid’ could have been found in order to identify what the source domain is for ‘pour.’ For example, if ‘pour’ takes an argument that it belongs to the source domain of LIQUID, automatic extraction of metaphors can be easily accomplished by ruling out collocates of ‘pour’ that are not LIQUID. However, this step is not carried out in Mason (2004). Instead, he assigns ‘liquid’ to LAB by intuition without first proving the domain of ‘liquid.’ This book will improve on these methods and show that by first determining source domains, metaphors can be identified quickly in corpora.

3.3 Summary of Chapter

In sum, the issue of source domain boundary has often been ignored in studies that have attempted to identify metaphors. Source domains vary in different studies based on the reasons mentioned in Chapter Two, i.e., how people define general and specific source domains as well as how concrete source domains can be defined based on individual experience (as experience also differs among individuals). For these reasons, there is great difficulty in creating source domains that achieve consensus among all metaphor analyses. Most studies discussed in this chapter focus on metaphor identification, but none have focused on the importance of the selection of source domains. If source domains can be decided first, then the automatic extraction of metaphors becomes possible.

As has been touched upon, metaphorical expressions from corpora are required for proper source domain determination. This, however, is not to suggest that just any random sample of corpora will be helpful. In this book, corpora data have been carefully designed so that comparisons between datasets may be implemented. The design of different datasets and their analysis can be seen in Chapters Five and Six. For the present stage, metaphorical expressions will be identified manually so as to collect enough data for the processes of source domains in Chapters Seven through Ten. Before the discussion of datasets in Chapter Five, Chapter Four will first introduce the use of different lexical knowledgebases, as an understanding of these knowledgebases is crucial to the discussion of the methodology in the remaining chapters as well as for the computational purpose of source domain determination in Chapters Seven through Ten.
Chapter 4: Lexical Knowledgebases and Corpus

As seen in Chapter Three, studies on metaphor identification using a top-down approach require information from lexical knowledgebases, while studies using a bottom-up approach depend on corpora data. For example, in the top-down approach, Martin (1990) uses the KODIAK semantic network (Wilensky 1986) while Wang (2006) uses the Chinese Concept Dictionary (CCD). On the other hand, with the bottom-up approach, Mason (2004) uses the Web as a corpus, while both Chung (2005, 2008) and Stefanowitsch (2006) use corpora for the observation of collocations. Wang (2006) also uses corpora data from People’s Daily and various literary works, in addition to the Chinese Concept Dictionary. Since corpora and lexical knowledgebases are required in both top-down and bottom-up approaches, this chapter will introduce some of the lexical knowledgebases used in this book. These lexical knowledgebases are WordNet, SUMO, SinicaBow, the Chinese-English Merged Word List, Sketch Engine and the Chinese Gigaword Corpus 1.0.

4.1 WordNet

WordNet (Miller et al. 1990, Fellbaum 1998) is an English electronic knowledgebase that provides the semantic relations between word senses (or meanings). This knowledgebase was developed by the Cognitive Science Laboratory at Princeton University. When a word is searched for in WordNet, all of the senses of this word are subsequently displayed. Figure 4.1 shows the search result for ‘transport’ from the WordNet 1.7.1 browser. The numbers in brackets for each sense (see dotted arrow) are the sense frequencies, which are computed from a semantic concordance created by Landes, Leacock & Tengi (1998) using WordNet 1.6. The sense frequencies are based on two corpora (the Brown corpus and Stephen Crane’s novella entitled The Red Badge of Courage), i.e., the displayed sense frequencies show the occurrences of the different WordNet senses in these two corpora.

---

1 Screenshots from these databases can be found in Appendix A4. Only those functions needed in this book will be shown.
2 http://wordnet.princeton.edu/.
3 There are version 1.5, 1.6, 1.7, 1.7.1, 2.0, 2.1 and 3.0. Version 1.7.1 will be used in this book throughout because this is the version used in SinicaBow for the mapping of WordNet and SUMO. Online access is available, but downloads of the WordNet browser are also available at http://wordnet.princeton.edu/obtain.
In Figure 4.1, there are two parts-of-speech found for ‘transport’—noun and verb (see colored arrows). Under each part-of-speech, the number of senses is listed. For example, ‘transport’ as noun has five senses as does ‘transport’ as verb. Each of these senses can be expanded to find their hypernyms. One example is given in (1).
Similar to finding coordinate terms, hyponyms, troponyms (manner of) and synonym sets (synsets) for each sense (see Fellbaum 1998). In this manner, by looking at the relations between lexical items, WordNet can itself form a taxonomy of word senses (Fellbaum 1998), whereby all senses in WordNet are arranged in hierarchical form. WordNet has been used in various ways; a dictionary, a thesaurus and also as a lexical tool for computational research. In addition to the hierarchy in WordNet, upper hierarchies, known as ontologies, also exist. The following section will introduce SUMO (Suggested Upper Merged Ontology), as it has been used extensively by many researchers working with semantic knowledgebases, including Niles & Pease (2001), Ševčenko (2003) and Scott, Lewis & Langendoen (2002). Among these researchers, Niles & Pease (2001) and Ševčenko (2003) have successfully merged SUMO with other knowledgebases, such as WordNet.

## 4.2 SUMO (Suggested Upper Merged Ontology)

SUMO is an upper level ‘taxonomy’ that structures the common knowledge system shared by most people (Niles & Pease 2001). Since it is an upper ontology, SUMO

4 All nouns are regarded as ‘entity’ in WordNet. Abstract concepts, such as ‘politics’ and ‘economy,’ are also regarded as ‘entity’ in WordNet. Therefore, ‘entity’ does not necessarily refer to concrete objects.

5 SUMO is available at http://www.ontologyportal.org.
differs from WordNet by being more general.\textsuperscript{6} For example, Figure 4.2 shows the upper hierarchy for the concept of ‘transport’ in SUMO on the left (figure taken from KSMSA project by Ševčenko 2003) and in WordNet on the right (the upper nodes of both are arranged in a similar order for ease of comparison).\textsuperscript{7}

From Figure 4.2, it is apparent that SUMO is more general than WordNet. SUMO does not differentiate terms by part-of-speech (such as between verb and noun), as its major purpose is to display conceptual knowledge rather than linguistic knowledge. For example, ‘transport’ as either a noun or a verb will be categorized under the node of ‘transportation,’ regardless of part-of-speech. On the contrary, WordNet is more linguistic-based, and therefore, the hierarchies for different parts-of-speech are separated. The hierarchy in Figure 4.2 shows that of a ‘noun.’

Due to the different levels of specificity between SUMO and WordNet, there have been attempts to combine the two to enlarge the knowledge system. The KSMSA project by Ševčenko (2003) is a system that combines WordNet with SUMO.\textsuperscript{8} By merging these two taxonomies, the general ontological nodes are made more specific through the

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure4_2.png}
\caption{Upper Hierarchy for ‘transport’ in SUMO (Left) and WordNet (Right)}
\end{figure}

\textsuperscript{6} There are different levels of ontology, such as Mid-level Ontology (MILO) and domain specific ontology. However, this book will only use SUMO because it has been mapped to WordNet by different projects.


\textsuperscript{8} The KSMSA website is available at http://virtual.cvut.cz/smsaWeb/browser/title.
WordNet nodes. For example, when a term such as ‘finance’ is found in WordNet but not in SUMO (because the knowledge in SUMO is more general), KSMSA merges these two knowledgebases to form representations such as those found in Figure 4.3 for ‘finance’ (circled).

In Figure 4.3, the boxes with darker shading are from SUMO and those with lighter shading are from WordNet. We can see that the specificity of WordNet is complemented by the general domain information of SUMO and vice versa.

Another similar merging of WordNet and SUMO can be seen in the projects called the Sigma Knowledge Engineering Environment by Pease (2003) and SinicaBow (Academia Sinica Bilingual Ontological Wordnet or 中央研究院中英雙語知識本體)
SinicaBow provides a Chinese-English-Chinese interface that enables one to search for Chinese words that can then be mapped to WordNet and SUMO. The following section discusses the SinicaBow knowledgebase in greater detail.

**4.3 SinicaBow (Academia Sinica Bilingual Ontological Wordnet)**

SinicaBow is the direct translation of the English WordNet to Chinese, providing Chinese-English-Chinese mappings of searched words as well as their related WordNet explanations and SUMO nodes.

When a word is searched in SinicaBow, all WordNet senses for this word will be displayed. For example, 政治 zhèngzhì ‘politics’ has three senses in English WordNet; these senses will be displayed when the Chinese word 政治 zhèngzhì ‘politics’ is searched in SinicaBow. Table 4.1 shows the three senses of 政治 zhèngzhì ‘politics’ with their tabulated WordNet explanations as well as their respective mapped SUMO nodes.

<table>
<thead>
<tr>
<th>Chinese Word</th>
<th>(POS) WordNet Explanations</th>
<th>Related SUMO Nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>政治 zhèngzhì ‘politics’</td>
<td>(Noun) The profession devoted to governing and to political affairs</td>
<td>OccupationalRole (職業角色)</td>
</tr>
<tr>
<td></td>
<td>(Noun) The study of government of states and other political units</td>
<td>FieldOfStudy (研究領域)</td>
</tr>
<tr>
<td></td>
<td>(Noun) Social relations involving authority or power</td>
<td>SocialInteraction (人際互動)</td>
</tr>
</tbody>
</table>

These senses are collected from the information provided in SinicaBow. See Figure 4.4 for the representation of senses in SinicaBow.

---

9 The Sigma Knowledge Engineering Environment is available at http://sigma.ontologyportal.org:4010/sigma/KBs.jsp, while SinicaBow is available at http://bow.sinica.edu.tw.
In Figure 4.4, we can see that when 政治 zhèngzhì ‘politics’ is searched in SinicaBow, each sense of WordNet related to this word is provided. The English explanation of WordNet (marked (a) in Figure 4.4) and its related SUMO nodes (marked (b) in Figure 4.4) are also given. SinicaBow is extremely useful because it not only provides the different senses of Chinese words but also provides their ontological structure, which can then be compared to similar resources in other languages. For example, if other languages also have translations of their languages merged to WordNet and SUMO, the conceptual similarities and differences between the different languages can be discovered. In this book, SinicaBow provides the mappings to WordNet and SUMO for the metaphorical expressions analyzed later in Chapter Six.

Since SinicaBow displays direct translations of English WordNet, various special expressions in Chinese that are not used in English (such as four-character idioms) are not likely to be found in SinicaBow. Despite the lexical gaps, all translations in SinicaBow were validated by a group of bilingual translators during the mapping process (cf. Huang, Chang & Lee 2004). For the purpose of this book, if a metaphorical expression is found to be a missing word, another database can be implemented to complement SinicaBow. This will help to narrow the lexical gaps in SinicaBow. The following section will introduce the Chinese-English Merged Word List.
4.4 Chinese-English Merged Word List

Academia Sinica’s in-house database of the Chinese-English Merged Word List contains translated Chinese-English words collected from several bilingual dictionaries. This database consists of 527,132 Chinese-English translations (151,305 English words and 283,572 Chinese words). Figure 4.5 shows examples of mappings taken from the Chinese-English Merged Word List.

Figure 4.5: The Chinese-English Merged Word List
Each of the English translations in Figure 4.5 (see arrow) is mapped automatically to WordNet and SUMO for the purpose of this book. This is done through mapping each English translation to senses in WordNet, which is already linked to SUMO.

The Chinese-English Merged Word List, which is a useful tool for extending the knowledgebase of SinicaBow, will be used in later sections if a metaphorical expression is not found in SinicaBow. More uses of this database will be discussed in Chapter Seven in the top-down approach. Thus far, the previous sections have introduced four knowledgebases (WordNet, SUMO, SinicaBow and the Chinese-English Merged Word List). The following section will introduce Sketch Engine, a platform that helps to analyze corpora data effectively.¹⁰

### 4.5 Sketch Engine

Sketch Engine (Kilgarriff & Tugwell 2001) is a system that provides the collocations of words according to grammatical relations. It has been used to analyze large-scale corpora data, such as British National Corpus (BNC) and the Chinese Gigaword corpus. The Chinese Sketch Engine was created by Kilgarriff et al. (2005), and it has the same functions as the English Sketch Engine, i.e., it also arranges collocates for query words in grammatical relations. For example, when a query word is searched in Sketch Engine, the system will return with collocates for this query word arranged in grammatical relations, such as ‘objects of the query word,’ ‘subjects of the query word,’ ‘modifiers of the query word,’ etc. In addition, Sketch Engine also provides concordance and thesaurus functions as well as Sketch-Difference, a function that provides the comparisons of Wordsketches (of collocates) for two compared words.

Figure 4.6 shows an example of the search result for 經濟 jingji ‘economy’ in the Chinese Sketch Engine.

---

¹⁰ Since the Chinese-English Merged Word List is not available online, its snapshots will not be provided again in the appendix.
In Figure 4.6, the query word and its frequency within the entire Gigaword corpus are shown. The frequency for pairs of collocates, such as 經濟 jingji ‘economy’ and 振興 zhènxìng ‘to give life to’ under the ‘object_of’ relation (arrow in Figure 4.6), is given. In this case, it is 4,046 (in the second column), indicating that 經濟 jingji ‘economy’ appears as the ‘object of’ the verb 振興 zhènxìng ‘to give life to’ 4,046 times in the whole Gigaword corpus.

In addition to frequency, Sketch Engine provides an additional score for the ranking of saliency of collocates. This is because Kilgarriff & Tugwell (2001) have suggested that frequency alone may not be a reliable score, because the frequency of collocates is relative to the number of both words in the whole corpus. Therefore, they suggest the use of a more reliable account in order to standardize all frequencies for the collocates based on the overall performance of collocates in a particular condition. In Figure 4.6, the saliency value (the third column for each relation) for 經濟 jingji ‘economy’ and
振興 zhènxìng ‘to give life to’ is 73.13. This saliency value is relative to the performance of all collocates within the same relation, i.e., these two collocates are relatively high in their saliency level in all the collocates found within a similar relation.\(^ {11}\)

The thesaurus function in Sketch Engine arranges words in descending similarity scores with the query word. For example, the thesaurus results for 經濟 jīngjì ‘economy’ are shown in Figure 4.7.

\[ I(w_1, R, W2) = \log \left( \frac{|| R^* || \times || W1, R, W2 ||}{|| W1, R, || \times || R, W2 ||} \right) \]

---

\(^ {11}\) The calculation of saliency is an improvement from the Mutual Information (MI) value proposed by Church & Hanks (1990) because Kilgarriff & Tugwell (2001:34) found that the MI value “over- emphasises low frequency items.” In the Sketch Engine, the saliency value is calculated as the “product of Mutual Information \( I(\text{Church & Hanks 1990}) \) and log frequency” (Kilgarriff & Tugwell 2001:34). Kilgarriff & Tugwell follow the calculation in Lin (1998) and give the new saliency formula:
In Figure 4.7, all words that are found to have similar constructions (cf. construction grammar in Goldberg 1995) with 經濟jingjì ‘economy’ are displayed in descending similarity scores (squared). This function is very powerful, as it also provides words with similar linguistic use in descending order. For example, almost all instances in Figure 4.7 can be used with expressions such as 改革 gǎigé ‘reform/reformation’ and 成長 chéngzhǎng ‘grow/growth.’ Shared expressions such as these contribute to the similarity scores in the thesaurus function of Sketch Engine, which therefore can be seen to be based on linguistic patterns rather than on conceptual similarities.

On the other hand, the Sketch-Difference function displays differences in terms of constructions between two words. The sketch-difference for 經濟jingjì ‘economy’ and 政治zhèngzhì ‘politics’ is given in Figure 4.8. In (A), the shared collocates of the two words are given in different shadings in the Chinese Sketch Engine. The light shading indicates words shared by both 經濟jingjì ‘economy’ and 政治zhèngzhì ‘politics,’ while darker shadings indicate more-or-less uses with either one of the words. The more-or-less frequency of both 經濟jingjì ‘economy’ and 政治zhèngzhì ‘politics’ can also be seen from columns two and three of each relation (see arrows in (A): on the left is the frequency for 經濟jingjì ‘economy,’ and on the right is the frequency for 政治zhèngzhì ‘politics.’
Figure 4.8: Sketch-Difference for 經濟 jìngjì ‘economy’ and 政治 zhèngzhì ‘politics’
In (B) and (C) of Figure 4.8, the ‘economy’-only (經濟 jìngjì) and ‘politics’-only (政治 zhèngzhì) patterns are shown. From here, we can see collocates that are used in one of these two words only. Through using Sketch-Difference, too, we can observe linguistic patterns that are used in one of the two searched words. For example, the Sketch-Difference of 經濟 jìngjì ‘economy’ and 政治 zhèngzhì ‘politics’ reveals that only 經濟 jìngjì ‘economy’ can be used with the expression of 不景氣 bùjǐngqì ‘in recession’ ((B) in Figure 4.8), and only 獻金 xiànjīn ‘offered money’ can be used with 政治 zhèngzhì ‘politics’ ((C) in Figure 4.8), but not vice versa. As no other corpora have this function yet to date, this function of Sketch Engine is both unique and powerful.

This book will use the Chinese Sketch Engine and its functions to look for linguistic patterns for the benefit of analysis. As mentioned, the Chinese Sketch Engine is based on the Chinese Gigaword corpus. The following section will introduce the distinctive features of this corpus.

4.6 Chinese Gigaword Corpus Version 1.0

At present, two versions of Chinese Gigaword corpus are available, namely version 1.0 and 2.0. These two versions differ in terms of data size. Version 1.0 contains newspaper data from about 1991 until 2002, while version 2.0 contains data from 1991 until 2004. This book uses version 1.0 (Graff & Chen 2003) because this version is the first available on the Chinese Sketch Engine, and it has been stably used for a longer period of time. Both versions of the Chinese Gigaword corpus version 1.0 were produced by Linguistic Data Consortium (LDC) in 2003.12

The Chinese Gigaword corpus 1.0 contains an archive of newswire texts from the Central News Agency of Taiwan (CNA) and the Xinhua News Agency of Beijing (XIN). The data for CNA was collected between January 1991 and December 2002 while the data for XIN was collected between December 1990 and September 2002.13 This Gigaword corpus produces a huge database encompassing approximately one billion (1,000,000,000) characters and 667,498,342 segmented words (CNA comprises 12 Linguistic Data Consortium (LDC) is an open consortium that collects and distributes linguistic resources.
13 The content of this corpus consists of a recompilation of previous smaller corpora, such as some of the XIN news texts that have been released as LDC previous old corpora. Examples of previous releases are Mandarin Chinese News Text (Wu 1995) and the TREC (Text Retrieval Conference) Mandarin corpus (Rogers 2000).
a total of 429,094,907 words while XIN contains 238,403,435 words). The Chinese Gigaword corpora (both versions 1.0 and 2.0) constitute the largest Mandarin corpora to date.

The Chinese Gigaword version 1.0 has the sub-corpora shown in Table 4.2.

Table 4.2: Sub-Corpora in the Chinese Sketch Engine

<table>
<thead>
<tr>
<th>Sub-Corpora</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese_all_simp</td>
<td>All CNA and XIN data in simplified version</td>
</tr>
<tr>
<td>Chinese_all_simp: beijing_only</td>
<td>All XIN data in simplified version</td>
</tr>
<tr>
<td>Chinese_all_trd</td>
<td>All CNA and XIN data in traditional version</td>
</tr>
<tr>
<td>Chinese_all_trd: taiwan only</td>
<td>All CNA data in traditional version</td>
</tr>
</tbody>
</table>

The sub-corpora of ‘Chinese_all_simp’ and ‘Chinese_all_trd’ are the same except that one is in simplified characters and the other in traditional characters. Then, there are also the corpora of XIN-only and CNA-only. A suitable corpus from those available may be selected based on the purpose of the research. Search results may also be viewed in the format of concordance, word sketch, thesaurus and Sketch-Difference. These functions are available for all sub-corpora shown in Table 4.2.15

Based on the facilitation of the Chinese Sketch Engine and the Chinese Gigaword corpus, the selection of corpora has been carefully planned in this book, based on the options available in the Chinese Gigaword corpus. A more in-depth discussion of corpora selection follows in Chapter Five. The following section will first elaborate on how these different resources can be used in this book.

4.7 Use of Lexical Resources in the Book

Chapter Three has already shown that most previous works on metaphor identification ignore the importance of determining source domains, discussing also how the determination of source domains are able to help solve the problems of metaphor identification encountered in previous works.

This book will also select different corpora data for the purpose of comparing source domains between different datasets. For example, there will be datasets from Taiwan (CNA) and China (XIN). Within these different datasets, we will collect datasets with Single Target Domains (such as 經濟 jīngjì ‘economy’ only) and datasets with

---

14 Word segmentation was carried at the Academia Sinica CKIP (Chinese Knowledge and Information Process) group for version 2.0 before the corpus was automatically tagged (cf. Ma & Huang 2006, Huang 2006).

15 These sub-corpora were available at the time when this work was written.
Coordinated Target Domains (such as 經濟和政治 jīngjì hàn zhèngzhì ‘economy and politics’). No studies previously have compared coordinated and non-coordinated target domains. As we hypothesize that these two constructions may differ in their use of metaphorical expressions, we will compare these two constructions through the observation of their source domains selected. Moreover, facilitated by the displays of grammatical relations in Sketch Engine, these two different constructions can be extracted quickly and in large quantities.

Target domains are fixed in this book in order to observe how these fixed target domains interact with different source domains. The use of different datasets in this book will not only provide information regarding regional differences in the use of source domains, it will also provide information regarding the linguistic differences between different target domains. For example, while only 經濟 jīngjì ‘economy’ can be used with the expression of 不景氣 bùjǐngqì ‘in recession,’ the coordinate relations of 經濟和政治 jīngjì hàn zhèngzhì ‘economy and politics’ is not able to use this expression, probably because one of the coordinated target domains does not co-occur with the use of 不景氣 bùjǐngqì ‘in recession.’ On the other hand, if both target domains co-occur with the same metaphorical expressions (such as 改革 gǎigé ‘reform/reformation’ and 成長 chéngzhǎng ‘grow/growth’), these metaphorical expressions are likely to occur in the coordinate relation of 經濟和政治 jīngjì hàn zhèngzhì ‘economy and politics.’ The selection of target domains will be discussed in greater detail in Chapter Five.

From the different datasets and different target domains, types of metaphorical expressions and tokens of metaphorical expressions will be identified. As mentioned in Chapter Two, types of metaphorical expressions are lexical items, such as 改革 gǎigé ‘reform/reformation’ and 成長 chéngzhǎng ‘grow/growth.’ The frequency of each of these types will form the total tokens of metaphorical expressions in the corpora. Source domain determination will first analyze the types of metaphorical expressions collected manually from different datasets. Only then, after these types of metaphorical expressions have been sorted according to different source domains, will tokens of metaphorical expressions be considered in each group. Figure 4.9 gives an overview of the steps used in this book to determine source domains using both top-down and bottom-up approaches.
Figure 4.9: Top-down and Bottom-up Approaches to Source Domain Determination Used in Present Study
In Figure 4.9, a top-down approach is labeled using the boxed capital ‘T’ followed by the step number (T1, T2, T3, etc.), while a bottom-up approach uses the boxed capital ‘B’ followed by the step number (B1, B2, B3, etc.).

The top-down approach aims to define source domains by identifying the most concrete concepts in SUMO. As discussed in Chapter Three, determining source domains becomes difficult if the measure of concreteness is based on individual human judgments, which naturally vary across individuals. However, if these concrete nodes can be determined automatically, the influence of human judgment can be reduced. After selecting the most concrete SUMO node for each type of metaphorical expression (such as that from the various SUMO nodes of 改革 gāigé ‘reform/reformation’ and 成長 chéngzhǎng ‘grow/growth’), the definition from this concrete SUMO node will be used for grouping the different types of metaphorical expressions into groups of possible source domains. The details of the steps involved in this procedure are given below.

As shown in Figure 4.9, before step one of the top-down approach (T1) can be carried out, types of metaphorical expressions must first be manually extracted. T1 will first map these types of metaphorical expressions to WordNet and SUMO, using SinicaBow and the Chinese-English Merged Word List. In this step, each type of metaphorical expression in Chinese will be mapped to the English senses in WordNet through the bilingual interface of SinicaBow. After this is done, each English sense will then be automatically mapped to SUMO. Using SinicaBow, each type of metaphorical expression will yield different WordNet senses, depending on how many senses a metaphorical expression has. Through these mappings, each type of metaphorical expression will have one or more sense as well as corresponding SUMO nodes.

After the mappings to SinicaBow and the Merged Word List, the next step will then be to measure the conceptual distance between each of the mapped SUMO nodes (within each type of metaphorical expression) and several prototypical concrete nodes (T2). These prototypical concrete nodes are those that have been rated by human subjects in terms of concreteness scores. Therefore, based on the measurement of conceptual distance, one final concrete SUMO node will be selected from amongst the different senses of each type of metaphorical expression.

At this point, meaningful keywords in the SUMO definitions for these concrete SUMO nodes will be used to determine source domain names (T3). For example, the SUMO node of ‘Growth’ has the following definition (bold added).

The Process of **biological development** in which an **Organism** or **part** of an **Organism** changes its **form** or its **size**.
Bolded keywords can be used for determining source domain names (after removal of function words, neutral words, etc.). In the top down approach, these source domain names are first determined through the selected concrete nodes. After this is done, the final step will be to count the number of times these source domain names (such as ‘biological development’ and ‘organism’) appear in the definitions of all the SUMO nodes (concrete and non-concrete) for each type of metaphorical expression \(T_4\). It therefore follows that the source domains for each type of metaphorical expression depend on the concrete keywords in their SUMO definitions.

While these are the steps for the top-down approach, the bottom-up approach is notably different in the sense that the source domain names are determined at the final stage, after the collocates of the types of metaphorical expressions have been clustered. The bottom-up approach will first search for collocates in different grammatical relations in Sketch Engine for all of these types of metaphorical expressions \(B_1\). As the top-down and bottom-up approaches deal with the same data differently, the final source domains selected by each approach can be compared.

After the collocates for each type of metaphorical expression have been collected, the computation of cut-off points will be carried out in order to include only significant collocates in the next step of analysis \(B_2\). Cut-off points for each grammatical relation will be calculated based on a certain threshold value of means of saliency. These significant collocates above the threshold level will then be mapped to WordNet to find their WordNet hypernyms.

The next step will then be to cluster collocates of each type of metaphorical expression based on their WordNet hypernyms \(B_3\). These collocates form clusters of information regarding the source domain of each type of metaphorical expression. For example, a type of metaphorical expression has fifty collocates above a threshold level. These fifty collocates will be clustered together if they share similar hypernyms. If among these fifty collocates there are twenty-five that share a similar concept, this shared concept will be considered to have a high conceptual density, since the other shared concepts were lower in their number of collocates. The concept with the highest conceptual density is considered to be the source domain of this type of metaphorical expression. Finally, source domain names are selected based on the bottom nodes of the WordNet hierarchies with the highest conceptual density \(B_4\).

Using the approaches in this book, source domains will be determined automatically with level of specificity defined according to the WordNet hierarchy. Since the source domain name is within the hierarchy of WordNet, we can decide whether to choose a more specific source domain (selecting members from the lower hierarchy, as employed in this book) or a more general source domain (selecting members from the upper hierarchy). The reason for selecting the bottom nodes is because the bottom
nodes carry information of all their hypernyms, and as a result, one can view all hypernym information of these bottom nodes. Both approaches will be discussed further in later chapters.

4.8 Summary of Chapter

This chapter introduces four lexical knowledgebases (WordNet, SUMO, SinicaBow and the Chinese-English Merged Word List), a corpus (the Chinese Gigaword corpus) and a platform for analyzing collocations (Sketch Engine). These resources are crucial for the determination of source domains in this book. This chapter also takes an overview of the steps needed to determine source domains in later chapters. These steps clearly state the use of these different resources and their usefulness in regards to source domain determination. The following two chapters will discuss the manual analysis of metaphors in corpora: Chapter Five will discuss the selection of datasets based on the regional differences of Taiwan (CNA) and China (XIN) as well as the benefits of such datasets. Chapter Six will discuss the identification and analysis of metaphors.
Corpora have increasingly been recognized for their importance in linguistic analyses, not only by corpus linguists (Kennedy 1998, McEnery & Wilson 2001, and Gries 2003) but also by linguists in general. Analyses of the data found in corpora can be conducted from a number of different perspectives: There are studies that use corpora for the purpose of discourse analysis (Su 2002), teaching (Cameron & Low 1999), conversational analysis (Semino, Heywood & Short 2004), English for specific purposes (Goatly 2002) as well as in Austronesian languages studies (Huang 2002). In addition, studies that use different corpora to compare languages have also been undertaken (see Charteris-Black & Ennis (2001) for comparisons of Spanish and English; Özçalışkan (2003) for comparisons of English and Turkish; and Chung (2005, 2008) for comparisons of Chinese, Malay and English).

Regarding the comparison of different languages, the sampling is usually small due to factors such as the limitations of small corpora. In general, it is not necessary to collect a large amount of data, as small sets of data can sometimes show the same linguistic patterns. Another reason for a small sampling is the lack of a priori purpose when selecting corpora data, as linguistic phenomena are only discovered after the exploration of corpora data. In this book, the problems caused by these factors are avoided because the following steps have been taken. First, the selection of large corpora has become more feasible because, in recent years, corpora have steadily increased in size. For example, the Chinese Gigaword corpus, which is the largest Chinese corpora to date, contains roughly one billion characters. Second, a large amount of data has been deliberately collected for this book in order to further validate the linguistic phenomena observed using smaller sets of data. Third, the selection of corpora data in this book is based on a common purpose, and furthermore, data is specifically selected so as to ensure that comparisons between sub-corpora are sufficiently meaningful. In addition, for the purpose of comparing languages used in different regions, this book also ensures that the two datasets from separate regions are parallel in terms of the period of time in which the data was collected and as well in the number of instances collected in each.

When corpora data are not parallel or differ greatly in size, great methodological difficulties will arise when comparing linguistic patterns between cultures (and between different regions). For example, Chung (2005, 2008) examines the use of ‘market’ in Chinese, English and Malay with 100 instances of ‘market’ in each language. The reason for this small sampling is that the study was constrained by the lack of corpora data in one of the languages being compared. In the event of such an obstacle, should
the deficient language be dropped from linguistic research because of a lack of corpora data? The answer for this is a resounding ‘no’: One solution for this dilemma that has been successfully adopted in the past is to reduce the sampling size for all of the languages in the study, even those with a huge corpora database. This tactic was employed by Chung (2005, 2008) when she found that Malay had no official corpus of modern language, and that manual collection of newspaper texts was required to establish a corpus.\(^1\) Due to this reason, only a small number of keywords could be found and this, in turn, greatly constrained the sampling size of the other languages in the study.

In other instances, some studies were unable to find synchronized corpora for different languages. For example, in the examination of metaphors by Özçalışkan (2003) in English and Turkish, Özçalışkan (2003) had to use a special methodology in order to obtain comparable sample corpora for these two languages. To create this corpus, Özçalışkan (2003) made use of ten novels and five widely known newspapers for each language and collected corpora data based on the following methodology:

Each novel was opened 10 times randomly, and at each opening, the first five instances of metaphorical motion events were recorded, resulting in about 50 such instances from each novel. Each newspaper was followed for 3 consecutive days…The total number of motion verbs was 827 for the newspapers in English and 894 for the newspapers in Turkish (Özçalışkan 2003:194).

Great difficulties, therefore, are encountered when attempting to find parallel corpora data in order to compare different languages, and due to this problem, different methodologies are employed in the selection of corpora data, as stated above. The cases of Özçalışkan (2003) and Chung (2005, 2008) are examples of compromise when creating corpora for different languages where a size discrepancy exists.

Despite the use of different methodologies, there has been some agreement among researchers using corpora for research purposes. For example, researchers using corpora are usually aware that it is best if the sampling of corpora “implies a design and a purpose” (Tognini-Bonelli 2001:52) and that the corpora data should avoid “the idiolect of a speaker” (Huang 1994:168). However, these agreed upon guidelines for selecting corpora data are sometimes hard to follow, particularly when the corpora data is small in size. Specifically, when the same keyword is repeated in the same document that has been written by a single person, problems such as idiolect may still arise. Therefore, in

\(^1\) There is a corpus of classic Malay available at http://www.anu.edu.au/asianstudies/mcp/ under the Malay Concordance Project, the Australian National University.
order to create corpora data that will be meaningful for linguistic analyses, great care must be taken regarding the sampling selected.

This book not only uses corpora data with a large sampling size but also ensures that the sub-corpora from which this data has been collected is parallel in terms of its size and the number of years taken to collect it, thereby eschewing the previously mentioned concerns. Using a large corpus with carefully designed datasets, this book will present every step of corpora analyses, each of which reflects a common aim—to determine source domains and compare how these source domains are used across datasets.

The following section will describe the selection of the corpus and its sampling methodology. Special attention will be given to the discussion regarding how corpora data from different regions can be searched with keywords of different constructions. All of these carefully planned steps will contribute to the ultimate goal of this book, namely, the automatic determination of source domains.

5.1 Selection of Corpus: Chinese Gigaword Corpus 1.0

In terms of gathering a large quantity of corpora data for Taiwan and China, the Chinese Gigaword corpus provides a good option for text-based data. The advantages of the Chinese Gigaword corpus are numerous. First, the Chinese Gigaword corpus uses one type of genre (newspapers) from both Taiwan (Central News Agency of Taiwan or CNA) and China (Xinhua News Agency of Beijing or XIN). Second, this corpus is parallel in terms of the period of data collection (newspapers from 1991 to 2002) for both Taiwan and China. In this way, the aforementioned problems regarding corpora that are not synchronized are successfully avoided. A third advantage of the Chinese Gigaword corpus (series) is that it is the largest corpus to date. Even though the analysis in this chapter will select a small portion of the corpus, the large amount of corpora data allows the Sketch Engine to derive linguistic patterns in both languages, which leads to its fourth advantage: The Chinese Gigaword corpus is compiled for the Chinese Sketch Engine and, therefore, can be compared in terms of thesaurus, Word Sketch, and Sketch-Difference in the Chinese Sketch Engine (see Chapter Four for the functions of Sketch Engine).

In this book, corpora data from the Chinese Gigaword corpus will be selected for the manual identification of metaphors (the manual identification of metaphors is a necessary a priori step: In order to commence with automatic source domain

---

2 This could also be a disadvantage for those who would like to have a balanced corpus with different genres. However, for text-based use of metaphors, the Chinese Gigaword corpus provides a good choice of texts.
determination, a sufficient database of metaphorical expressions must first manually be identified). The different types of metaphorical expressions (such as 成長 chéngzhǎng ‘grow/growth’ and 起飛 qǐfēi ‘take off’) from CNA and XIN, respectively, will be sorted according to the different source domains in Chapters Seven through Ten. In this chapter (and as well in Chapter Six), the procedures for corpora analysis will be presented so that it may be seen that the sub-corpora have been designed to maximize the effectiveness of the results, in terms of source domain determination for later chapters. In the following section we will discuss the keywords (i.e., the target domains used to extract corpora data). These target domains will be used to collect data from CNA and XIN, respectively.

5.2 Selection of Target Domains

As mentioned, corpora data will be selected carefully in this book. The target domains have also been selected with great attention to detail. Chung, Ahrens & Huang (2005), in a previous study, found that 經濟 jìngjì ‘economy’ yields a variety of source domains and, thus, is a robust example of a target domain. Their analysis was based on a smaller corpus; the analysis of the same target domain using the Gigaword corpus will likely yield results that are even more robust and reliable.

The term 經濟 jìngjì ‘economy’ has a total of 1,295,965 words (0.19 per cent) in the entire Gigaword corpus (both CNA and XIN). In order to select data with Coordinated Target Domains, 經濟 jìngjì ‘economy’ is first examined to see how it interacts with other target domains. The choice of target domains must be carefully made based on specific predictions of conceptual structure in order to achieve meaningful comparisons of source domains for these target domains in later chapters. To achieve this objective, target domains that are related to 經濟 jìngjì ‘economy’ but have different conceptual levels are probable target domains that will yield meaningful results for the analysis of source domains. Therefore, the search for related target domains focuses on target domains that are either at similar or different conceptual levels with 經濟 jìngjì ‘economy.’ All information regarding conceptual levels has been obtained by examining the ontological structure of different target domains. Coordinate uses of target domains can be seen in the construction of ontology, where ontologists look for syntactic cues, such as ‘is a kind of’ for class inclusion and ‘and/or’ for listing. Therefore, the related target domains with 經濟 jìngjì ‘economy’ should have constructions similar to listing, i.e., and/or/comma in Chinese. Therefore, instances such as ‘經濟 jìngjì ‘economy’ and X,’ and ‘經濟 jìngjì ‘economy’ or X,’ are sought out as well as the use of commas in Chinese (between 經濟 jìngjì ‘economy’ and other target domains).
Examples of these constructions can be found in the Gigaword corpus shown in (1) (taken from CNA data).

(1) a. **AND**

就 國家 預算 、 財政 與 經濟 成長

jiù guójiā yùsuàn cáizhèng yǔ jīngjí chéngzhǎng

with.regard.to country budget finance and economy growth

表現 進行 報告 (CNA)

biǎoxiàn jìnxìng bàogào

perform proceed report

‘To make a report with regard to the performance of national budget, finance and economic growth...’

b. **OR**

印度 的 經濟 或 外交 政策

yīndù de jīngjí huò wàijīào zhèngcè

India DE economy or foreign.affairs tactic

都 不 會 有 重大 改變 (CNA)

dōu bù huì yǒu zhòngdà gǎibiàn

all Neg. will have major change

‘The tactics for both the economy and foreign affairs in India will not have major changes.’

c. **COMMAS**

紐約 時報 今天 從 政治 、 經濟 、 教育

niǔyuè shíbào jīntiān cóng zhèngzhì jīngjí jiàoyù

New.York daily.news today from politics economy education

等 角度 探討
děng jiǎodù tànštào
e tc. angle probe.into

‘Today, the *New York Times* probes into the angles of politics, economy, education, etc.’

In order to extract constructions such as those in (1), the Chinese Sketch Engine provides a powerful tool for displaying all collocates that fall under the relation of ‘and/or.’ Under this coordinate relation, all words in the above constructions will be listed according to descending saliency values. A snapshot of the Chinese Sketch Engine results for 經濟 jīngjí ‘economy’ is shown in Figure 5.1 (taken from Figure 4.6 in Chapter Four).
A Corpus-driven Approach to Source Domain Determination

As illustrated in Figure 5.1, the sub-corpus chosen for the Wordsketches of 

<table>
<thead>
<tr>
<th>Query word</th>
<th>Collocates of query word</th>
<th>Frequency of each collocate and query word</th>
<th>Saliency of each collocate and query word</th>
</tr>
</thead>
<tbody>
<tr>
<td>經濟 jīngjì ‘economy’</td>
<td>經濟</td>
<td>54576 1.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>若興</td>
<td>1946 73.13</td>
<td>場利</td>
</tr>
<tr>
<td></td>
<td>私務</td>
<td>3910 70.21</td>
<td>創作</td>
</tr>
<tr>
<td></td>
<td>選舉</td>
<td>3923 69.66</td>
<td>消息</td>
</tr>
<tr>
<td></td>
<td>制約</td>
<td>1604 66.25</td>
<td>才能</td>
</tr>
<tr>
<td></td>
<td>拉動</td>
<td>1074 66.1</td>
<td>表達</td>
</tr>
<tr>
<td></td>
<td>帶動</td>
<td>3410 64.64</td>
<td>全球化</td>
</tr>
<tr>
<td></td>
<td>利</td>
<td>780 62.75</td>
<td>發展</td>
</tr>
<tr>
<td></td>
<td>銷售</td>
<td>1128 61.9</td>
<td>挖掘合作</td>
</tr>
<tr>
<td></td>
<td>刺激</td>
<td>2460 60.84</td>
<td>合作</td>
</tr>
<tr>
<td></td>
<td>運動</td>
<td>6256 59.94</td>
<td>領域</td>
</tr>
<tr>
<td></td>
<td>發展</td>
<td>22520 57.95</td>
<td>達到</td>
</tr>
</tbody>
</table>

**Figure 5.1:** Collocates for the Query Word 經濟 jīngjì ‘economy’ in the Chinese Sketch Engine

Among these words, four collocates have been selected as possible target domains to appear in coordinate relation with 經濟 jīngjì ‘economy.’
Figure 5.2: ‘And/Or’ Relation of 經濟 jǐngjì ‘economy’

The choice of these four target domains was made after observing the different ontological representation of these target domains in SUMO when compared to 經濟 jǐngjì ‘economy’ for source domain analysis. Table 5.1 shows the selected target domains from the list of words found under the ‘and/or’ relation in Figure 5.1 (the frequency contains both CNA and XIN data).

Table 5.1: Selected Target Domains in the ‘And/Or’ Relation with 經濟 jǐngjì ‘economy’ in Both CNA and XIN Datasets Provided by Chinese Sketch Engine

<table>
<thead>
<tr>
<th>Target Domains</th>
<th>Gloss</th>
<th>Frequency of Collocates with 經濟 jǐngjì ‘economy’</th>
<th>Saliency Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>政治 zhèngzhì</td>
<td>‘Politics’</td>
<td>23,934</td>
<td>75.55</td>
</tr>
<tr>
<td>財政 cāizhèng</td>
<td>‘Finance’</td>
<td>2,355</td>
<td>44.35</td>
</tr>
<tr>
<td>外交 wàijiāo</td>
<td>‘Foreign Affairs’</td>
<td>1,520</td>
<td>35.63</td>
</tr>
<tr>
<td>教育 jiàoyù</td>
<td>‘Education’</td>
<td>1,214</td>
<td>22.75</td>
</tr>
</tbody>
</table>

Ontological concepts were found for each of the selected target domains in Table 5.1 through SinicaBow, which provides the Chinese-English-Chinese mappings for both the
A Corpus-driven Approach to Source Domain Determination

searched words and their related WordNet explanations and respective SUMO (Suggested Upper Merged Ontology) nodes.

As mentioned in Chapter Four, when the target domains in Table 5.1 are searched for in SinicaBow, all the WordNet senses for these words are given, and each sense will be linked to their ontological concepts through SUMO. For example, when 政治 zhèngzhì ‘politics’ is searched for in Sinica Bow, its three senses are given, namely, (a) a social relation, (b) a field of study and (c) a profession. Since each searched word will yield more than one sense in SinicaBow (and each sense will have respective SUMO node), an a priori step is to manually select the correct sense from the senses presented so that only one suitable sense remains for the target domains. The selected sense for the target domains will be linked to SUMO through SinicaBow so as to observe their respective ontological levels.

Table 5.2 shows the tabulated information from SinicaBow for the five target domains (including 經濟 jīngjì ‘economy’). Each target domain (first column) has one or more than one sense (second column), and each sense has its respective link to its SUMO node (third column). As stated, only one sense will be selected for source domain analyses. For example, 政治 zhèngzhì ‘politics’: As ‘social relations’ is the only meaning that could possibly yield metaphors and serve as a target domain, it is therefore this sense that is selected. A secondary reason for the selection of this sense is because it has the highest sense frequency (cf. §4.1 of Chapter Four) as given in WordNet (shown in column four). As mentioned in Chapter Four, the sense frequencies presented by Landes, Leacock and Tengi (1998) are based on analyses of an older version of WordNet. Therefore, there are new senses that exist but are not found in the old versions (marked ‘nil’) in Table 5.2. The senses that are marked ‘nil’ will not be considered. Regarding all target domains, only senses with the highest levels of sense frequency will be selected. For those with a single sense (such as 財政 cāizhèng ‘finance’ and 外交 wàijiāo ‘foreign affairs’), their selected sense is easily determined (even though they have ‘nil’ sense frequency in WordNet).

Table 5.2 shows the tabulated information from SinicaBow for the five target domains (including 經濟 jīngjì ‘economy’). Each target domain (first column) has one or more than one sense (second column), and each sense has its respective link to its SUMO node (third column). As stated, only one sense will be selected for source domain analyses. For example, 政治 zhèngzhì ‘politics’: As ‘social relations’ is the only meaning that could possibly yield metaphors and serve as a target domain, it is therefore this sense that is selected. A secondary reason for the selection of this sense is because it has the highest sense frequency (cf. §4.1 of Chapter Four) as given in WordNet (shown in column four). As mentioned in Chapter Four, the sense frequencies presented by Landes, Leacock and Tengi (1998) are based on analyses of an older version of WordNet. Therefore, there are new senses that exist but are not found in the old versions (marked ‘nil’) in Table 5.2. The senses that are marked ‘nil’ will not be considered. Regarding all target domains, only senses with the highest levels of sense frequency will be selected. For those with a single sense (such as 財政 cāizhèng ‘finance’ and 外交 wàijiāo ‘foreign affairs’), their selected sense is easily determined (even though they have ‘nil’ sense frequency in WordNet).

---

3 These newly added senses are usually due to several reasons. One reason is the lack of such words in either WordNet or the two corpora used for sense tagging (Brown corpus and Stephen Crane’s novella entitled The Red Badge of Courage); the other reason could be because there are new uses of words that are later realized by the analyzers. However, newly added senses seldom affect the most frequently appearing sense, as the most frequent senses are usually stable in the corpora.
### Table 5.2: Selected Senses from SinicaBow for Metaphor Analysis

<table>
<thead>
<tr>
<th>Target Domains</th>
<th>(POS) Word Net Explanations</th>
<th>Related SUMO Nodes</th>
<th>Sense Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>政治 zhèngzhì  ‘politics’</td>
<td>(Noun) Social relations involving authority or power</td>
<td>Social Interaction (人際互動)</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>(Noun) The study of government of states and other political units</td>
<td>Field Of Study (研究領域)</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>(Noun) The profession devoted to governing and to political affairs</td>
<td>Occupational Role (職業角色)</td>
<td>3</td>
</tr>
<tr>
<td>財政 càizhèng  ‘finance’</td>
<td>(Noun) The management of money and credit and banking and investments</td>
<td>Financial Transaction (金融交易)</td>
<td>nil</td>
</tr>
<tr>
<td>外交 wàijiāo  ‘foreign affairs’</td>
<td>(Noun) Negotiation between nations</td>
<td>Meeting (會面)</td>
<td>nil</td>
</tr>
<tr>
<td>教育 jiàoyù  ‘education’</td>
<td>(Noun) The activities of educating or instructing or teaching; activities that impart knowledge or skill</td>
<td>Educational Process (教育歷程)</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>(Noun) Knowledge acquired by learning and instruction</td>
<td>Knows (知道)</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>(Verb) Give an education to</td>
<td>Educational Process (教育歷程)</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>(Noun) The gradual process of acquiring knowledge</td>
<td>Educational Process (教育歷程)</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>(Noun) The profession of teaching (especially at a school or college or university)</td>
<td>Educational Process (教育歷程)</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>(Verb) Create by training and teaching</td>
<td>Learning (學習)</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>(Verb) Educate in or as if in a school</td>
<td>Educational Process (教育歷程)</td>
<td>nil</td>
</tr>
<tr>
<td></td>
<td>(Verb) Train to be discriminative; as of taste or judgment</td>
<td>Educational Process (教育歷程)</td>
<td>nil</td>
</tr>
<tr>
<td></td>
<td>(Noun) Properties acquired during a person’s formative years</td>
<td>Past Fn (過去時間函式)</td>
<td>nil</td>
</tr>
<tr>
<td></td>
<td>(Noun) The result of good upbringing (especially knowledge of correct social behavior)</td>
<td>Learning (學習)</td>
<td>nil</td>
</tr>
<tr>
<td></td>
<td>(Noun) The properties acquired as a consequence of the way you were treated as a child</td>
<td>Past Fn (過去時間函式)</td>
<td>nil</td>
</tr>
<tr>
<td></td>
<td>(Noun) Raising someone to be an accepted member of the community</td>
<td>Social Interaction (人際互動)</td>
<td>nil</td>
</tr>
<tr>
<td>經濟 jīngjì  ‘economy’</td>
<td>(Noun) The system of production and distribution and consumption</td>
<td>Social Interaction (人際互動)</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>(Noun) The efficient use of resources</td>
<td>Subjective Assessment Attribute (主觀評價屬性)</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 5.2 also provides the relevant links to SUMO for each selected sense for each of the five target domains. These nodes are elaborated upon in (2), and the ontological hierarchy is concisely illustrated in Figure 5.3.

(2) SUMO Nodes for the Five Target Domains

a. 經濟 jīngjì ‘economy’ occurs at the level of ‘Social Interaction.’

b. 政治 zhèngzhì ‘politics’ occurs at the same level as 經濟 jīngjì ‘economy’ (‘Social Interaction’).

c. 外交 wàijiāo ‘foreign affairs’ appears at a node (‘Meeting’) below 經濟 jīngjì ‘economy’ and 政治 zhèngzhì ‘politics.’

d. 財政 càizhèng ‘finance’ appears at a further lower node (‘Financial Transaction’) of 經濟 jīngjì ‘economy’ and 政治 zhèngzhì ‘politics,’ but 財政 càizhèng ‘finance’ also has another branch of ‘dual object process,’ indicating 財政 càizhèng ‘finance’ may use metaphors differently from those under 經濟 jīngjì ‘economy’ and 政治 zhèngzhì ‘politics.’

e. 教育 jiàoyù ‘education’ appears at the node of ‘EducationalProcess,’ which shares the common upper nodes of ‘IntentionalProcess’ with all the other target domains (but it is not under ‘SocialInteraction’).

![Figure 5.3: Related SUMO Nodes for 經濟 jīngjì ‘economy,’ 政治 zhèngzhì ‘politics,’ 外交 wàijiāo ‘foreign affairs,’ 教育 jiàoyù ‘education’ and 財政 càizhèng ‘finance’](image)
Based on the different ontological representations pictured above, the following local hypothesis regarding the similarities and differences of these target domains in their choice of metaphorical expressions, as well as the choice of source domains, can be made.4

(3) The **higher** a target domain on the **ontological hierarchy** is, the more **general** this target domain is. More general target domains will produce **more metaphorical expressions** in metaphor analyses.

Based on the hypothesis in (3), we can predict that 經濟 jīngjì ‘economy’ and 政治 zhèngzhì ‘politics’ will reveal a higher number of metaphorical expressions. There will be differences between the target domains at lower nodes (外交 wàijiāo ‘foreign affairs,’ 財政 cáizhèng ‘finance’ and 教育 jiàoyù ‘education’) when compared to 經濟 jīngjì ‘economy’ and 政治 zhèngzhì ‘politics.’ This study, however, constrains comparisons between nodes that are obviously distant in the hierarchy. Therefore, the hypothesis in (3) will not apply to predicting how target domains at lower nodes vary. This constraint is made because, in the Coordinated Target Domains, the four target domains are only compared to 經濟 jīngjì ‘economy.’ Comparisons among themselves are not reliable, since the method of data collection focuses on target domains that are in ‘and/or’ relations with 經濟 jīngjì ‘economy’ only: Stated another way, only Single Target Domains allow comparisons between target domains to be made. Additionally, since most of the target domains under the node of 經濟 jīngjì ‘economy’ and 政治 zhèngzhì ‘politics’ branch off differently, comparisons between them become difficult. Furthermore, the behavior of a node that has two upper branches (such as 財政 cáizhèng ‘finance’) may be influenced by both of the upper nodes. For these reasons, all other target domains will be compared to 經濟 jīngjì ‘economy’ and 政治 zhèngzhì ‘politics’ only, since both these appear at a similar level.

Using the method described above, the linguistic performance of the different target domains may be safely predicted based on the lexical knowledgebase of SUMO. In addition, Sketch Engine also provides a powerful tool for the selection of suitable constructions between different target domains. All of the selected target domains will be compared in terms of their choice of source domains across regions (i.e., between Taiwan and China) and also in terms of their different constructions. The following section provides the number of corpora instances selected for each target domain.

---

4 Global hypotheses are made against the book as a whole. Local hypotheses are made specifically for the performance of the target domains.
5.3 Corpora Instances for Different Target Domains

The corpora data for the five target domains selected (経濟 jīngjī ‘economy,’ 政治 zhèngzhì ‘politics,’ 財政 cáizhèng ‘finance,’ 外交 wàijiāo ‘foreign affairs’ and 教育 jiàoyù ‘education’) is shown in Table 5.3.

Table 5.3: Five Target Domains x CNA/XIN x Coordinated/Single Target Domains

<table>
<thead>
<tr>
<th></th>
<th>經濟 jīngjī</th>
<th>政治 zhèngzhì</th>
<th>財政 cáizhèng</th>
<th>外交 wàijiāo</th>
<th>教育 jiàoyù</th>
</tr>
</thead>
<tbody>
<tr>
<td>經濟 jīngjī</td>
<td>ECONOMY</td>
<td>CNA/XIN</td>
<td>CNA/XIN</td>
<td>CNA/XIN</td>
<td>CNA/XIN</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Single (1)</td>
<td>Coordinated (2)</td>
<td>Single (3)</td>
<td>Single (4)</td>
</tr>
<tr>
<td>政治 zhèngzhì</td>
<td>POLITICS</td>
<td>CNA/XIN</td>
<td>Coordinated (6)</td>
<td>CNA/XIN</td>
<td>CNA/XIN</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Coordinated (6)</td>
<td>CNA/XIN</td>
<td>Coordinated (8)</td>
<td>CNA/XIN</td>
</tr>
<tr>
<td>財政 cáizhèng</td>
<td>FINANCE</td>
<td>CNA/XIN</td>
<td>Coordinated (7)</td>
<td>CNA/XIN</td>
<td>CNA/XIN</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Coordinated (7)</td>
<td>CNA/XIN</td>
<td>Coordinated (8)</td>
<td>CNA/XIN</td>
</tr>
<tr>
<td>外交 wàijiāo</td>
<td>FOREIGN AFFAIRS</td>
<td>CNA/XIN</td>
<td>Coordinated (8)</td>
<td>CNA/XIN</td>
<td>CNA/XIN</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Coordinated (8)</td>
<td>CNA/XIN</td>
<td>Coordinated (8)</td>
<td>CNA/XIN</td>
</tr>
<tr>
<td>教育 jiàoyù</td>
<td>EDUCATION</td>
<td>CNA/XIN</td>
<td>Coordinated (9)</td>
<td></td>
<td>Single (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Coordinated (9)</td>
<td></td>
<td></td>
<td>Single (5)</td>
</tr>
</tbody>
</table>

The shaded columns (1-5) represent the Single Target Domains. In these Single Target Domains, there are no instances with coordinate constructions as there are in example (1) previously. This is because ‘clean’ Single Target Domains are needed in order to contrast with the Coordinated Target Domains. The removal of all instances with coordinated relations in the Single Target Domains was carried out manually.5

Table 5.3 also shows the columns for the Coordinated Target Domains (6-9). Since the Coordinated Target Domains have been formed through the observation of the coordinate constructions of the other four target domains with 經濟 jīngjī ‘economy,’ 經濟 jīngjī ‘economy’ will, therefore, not have data for Coordinated Target Domains.

5 All corpora instances for the Single Target Domains are found using the concordance function of the Sketch Engine by searching for ‘keyword in context’ (KWIC) and in the concordance of Gigaword using relevant corpora (CNA-only, XIN-only or both). The default window size for the ‘save’ function is 40 words on the left and right of the keywords. The saved data is then analyzed and sorted using Microsoft Excel files.
Thus, the analyses of the Single Target Domains and Coordinated Target Domains will tell us: (a) whether all five target domains are similar or different in terms of source domain analyses; and (b) whether the same target domain will behave similarly or differently when appearing in Single Target Domains versus Coordinated Target Domains.6 The following excerpt details the number of instances found for Single Target Domains.

### 5.3.1 Single Target Domains

In terms of the number of instances for each target domain, it is physically impossible to manually analyze all instances of Single Target Domains for all five target domains, as a daunting total of 2,267,631 corpora instances in the whole Gigaword corpus exists. For this reason, only the first 4 per cent of the instances in each Single Target Domains have been included. This first 4 per cent of data comes from the year 1991 in the Chinese Gigaword corpus 1.0 (see §5.4 for the discussion of year difference). The fact that in the early 1990s the economy began to improve makes data from the early 1990s linguistically valuable.

Table 5.4 lists the number of instances in the first 4 per cent of each target domain in CNA and XIN, respectively. The number of instances for the first 4 per cent in CNA (54,563) and XIN (48,340) are not significantly different in a Kruskal-Wallis test: \( \chi^2(1)=0.535, p=0.465 \) (mean rank for CNA=6.20; mean rank for XIN=4.80).7 This indicates that CNA and XIN do not differ in terms of their number of instances.

---

6 As mentioned, 經濟 jìngjì ‘economy’ is the key target domain in this book (as it is compared to the other four target domains) because it is found to have robust source domains in previous analyses of Chung, Huang & Ahrens (2003) and Chung, Ahrens & Huang (2003, 2004ab, 2005).

7 A Kruskal-Wallis test is a one-way analysis of variance that is used when normal distribution and heterogeneity of variables need not to be assumed. This test examines the mean ranks of groups instead of means, i.e., the ranks of all raw data will be compared. For example, for the number of corpora instances in Table 5.4, the following mean rank is obtained. The Kruskal-Wallis test compares the mean rank of the two groups:

<table>
<thead>
<tr>
<th>CNA/XIN</th>
<th>Corpora Instances</th>
<th>Ranks</th>
<th>Mean Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10,549</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2,493</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4,174</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7,295</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>24,951</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>XIN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4,378</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2,213</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1,436</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6,329</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>26,888</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

Mean Rank=6.2 Mean Rank=4.8
Table 5.4: First 4 Per cent of CorporaInstances in the Chinese Gigaword Corpus for Single Target Domains in CNA and XIN

<table>
<thead>
<tr>
<th>Target Domains</th>
<th>CNA</th>
<th>XIN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4 Per cent / Total Instances</td>
<td>4 Per cent / Total Instances</td>
</tr>
<tr>
<td>政治 zhèngzhì ‘politics’</td>
<td>10,549 / 263,732</td>
<td>4,378 / 109,446</td>
</tr>
<tr>
<td>財政 càizhèng ‘finance’</td>
<td>2,493 / 62,328</td>
<td>2,213 / 55,326</td>
</tr>
<tr>
<td>外交 wàijiāo ‘foreign affairs’</td>
<td>4,174 / 104,341</td>
<td>1,436 / 35,891</td>
</tr>
<tr>
<td>教育 jiàoyù ‘education’</td>
<td>7,295 / 182,384</td>
<td>6,329 / 158,218</td>
</tr>
<tr>
<td>經濟 jīngjí ‘economy’</td>
<td>24,951 / 623,775</td>
<td>26,888 / 672,190</td>
</tr>
<tr>
<td>Total</td>
<td>54,563 / 1,236,560</td>
<td>48,340 / 1,031,071</td>
</tr>
</tbody>
</table>

As mentioned, ‘clean’ Single Target Domain data is necessary, as it ensures that their comparisons with the Coordinated Target Domain data will not be affected by any coordination constructions in the Single Target Domains. Therefore, all coordinate constructions (with the other four target domains or with any other words) have been removed manually from all data in the Single Target Domains. Table 5.5 displays the results after removal of all coordinated data in the first 4 per cent of corpora instances in CNA and XIN has taken place.

Table 5.5: Total Corpora Instances for Single Target Domains in CNA and XIN (after Removing All Coordinated Constructions)

<table>
<thead>
<tr>
<th>Target Domains</th>
<th>CNA ‘Clean’ 4 Per cent / Total Instance</th>
<th>XIN ‘Clean’ 4 Per cent / Total Instance</th>
</tr>
</thead>
<tbody>
<tr>
<td>政治 zhèngzhì ‘politics’</td>
<td>8,793 / 263,732</td>
<td>3,274 / 109,446</td>
</tr>
<tr>
<td>財政 càizhèng ‘finance’</td>
<td>2,025 / 62,328</td>
<td>1,978 / 55,326</td>
</tr>
<tr>
<td>外交 wàijiāo ‘foreign affairs’</td>
<td>3,687 / 104,341</td>
<td>1,274 / 35,891</td>
</tr>
<tr>
<td>教育 jiàoyù ‘education’</td>
<td>7,029 / 182,384</td>
<td>5,177 / 158,218</td>
</tr>
<tr>
<td>經濟 jīngjí ‘economy’</td>
<td>22,061 / 623,775</td>
<td>22,144 / 672,190</td>
</tr>
<tr>
<td>Total</td>
<td>43,595 / 1,236,560</td>
<td>33,847 / 1,031,071</td>
</tr>
</tbody>
</table>
The ‘clean’ Single Target Domain data in Table 5.5 also shows that there are no significant differences between the CNA and XIN datasets in a Kruskal-Wallis test: $\chi^2(1)=0.884$, $p=.347$ (mean rank for CNA=6.4; mean rank for XIN=4.6). This result again illustrates that the numbers of removed coordinated constructions (about 15 per cent) are similar in both (details can be seen in Appendices A5.1 and A5.2). From this table, it can also be seen that the total instances remaining for metaphor analyses number 77,442 (the sum of 43,595 and 33,847).

5.3.2 Coordinated Target Domains

The total number of corpora instances for the Coordinated Target Domains is shown in Table 5.6.

**Table 5.6:** Total Instances for Coordinate Target Domains with 經濟 jīngjì ‘economy’ in Both CNA and XIN

<table>
<thead>
<tr>
<th>Target Domains</th>
<th>CNA+XIN</th>
<th>CNA</th>
<th>%</th>
<th>XIN</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>政治 zhèngzhì ‘politics’</td>
<td>23,934</td>
<td>11,698</td>
<td>48.88%</td>
<td>12,236</td>
<td>51.12%</td>
</tr>
<tr>
<td>財政 càizhèng ‘finance’</td>
<td>2,355</td>
<td>1,541</td>
<td>65.44%</td>
<td>814</td>
<td>34.56%</td>
</tr>
<tr>
<td>外交 wàijiāo ‘foreign affairs’</td>
<td>1,520</td>
<td>1,097</td>
<td>72.17%</td>
<td>423</td>
<td>27.83%</td>
</tr>
<tr>
<td>教育 jiàoyù ‘education’</td>
<td>1,214</td>
<td>649</td>
<td>53.46%</td>
<td>565</td>
<td>46.54%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>29,023</strong></td>
<td><strong>14,985</strong></td>
<td><strong>51.63%</strong></td>
<td><strong>14,038</strong></td>
<td><strong>48.37%</strong></td>
</tr>
</tbody>
</table>

Since a large amount of data is required for the creation of Wordsketches, the data for Coordinated Target Domains has been taken from the whole Chinese Gigaword corpus 1.0. The second column in Table 5.6 shows the total instances of coordination in the combined CNA+XIN. For example, 政治 zhèngzhì ‘politics’ has 23,934 instances that appear in coordination with 經濟 jīngjì ‘economy.’ This number comprises both CNA+XIN data in the whole Gigaword corpus. The breakdowns of the CNA and XIN
are given in the third to sixth columns. The comparisons between CNA and XIN in Table 5.6 can be clearly seen in Figure 5.4.8

In addition, Table 5.6 shows that only 政治 zhèngzhì ‘politics’ has coordinated data numbering more than ten thousand instances in both CNA and XIN, respectively. The other target domains have much fewer instances, with less than two thousand for each within the whole Gigaword corpus.9

![Figure 5.4: The Percentages of Corpora Instances for Coordinate Target Domains with 經濟 jīngjì ‘economy’ in Both CNA and XIN](image)

From Figure 5.4, we can see that the total 23,934 instances for 政治 zhèngzhì ‘politics’ constitutes 48.88 per cent of instances in CNA and 51.12 per cent in XIN. The number of instances for CNA and XIN, when tested in a Chi-square analysis, show a significant effect: $\chi^2(1)=12.09$, $p<.05$.10 This shows that the co-occurrence of 政治 zhèngzhì ‘politics’

---

8 Discussion of time span for both sub-corpora will be given in §5.4.

9 The coordinated data comprise corpora instances from the year 1991 to the year 2002. From the number of instances in Table 5.6, we notice that some Coordinated Target Domains show only about a thousand instances when in the whole corpus. This explains why sometimes it is impossible to create Wordsketches of a word based on data of a single year.

10 A Chi-square test is used because the number of corpora instances shows comparisons of nominal variables. Comparisons of raw corpora instances can be carried out because the instances in Table 5.6 show whether or not these instances coordinate with 經濟 jīngjì ‘economy’ in the whole Gigaword corpus 1.0. This means that given a similar amount of
Chapter 5: Selection of Corpus, Target Domains and Sub-Corpora for Target Domains

‘politics’ and 經濟 jīngjì ‘economy’ in a coordinate relation is more frequently found in newspapers in Taiwan than it is in newspapers in China. In fact, all comparisons between CNA and XIN show significant effects in Chi-square tests.\(^{11}\) The only difference occurs in 財政 cāizhèng ‘finance’ and 外交 wàijiāo ‘foreign affairs,’ where more coordinated instances with 經濟 jīngjì ‘economy’ were found in the China data (XIN) than in the Taiwan data (CNA) (the others are shown to have more instances in CNA). More similarities and differences between CNA and XIN can be found in Chapter Six.

Since all four target domains of 政治 zhèngzhì ‘politics,’ 外交 wàijiāo ‘foreign affairs,’ 教育 jiàoyù ‘education’ and 財政 cāizhèng ‘finance’ in 經濟 jīngjì ‘economy’ show coordinate relations with 經濟 jīngjì ‘economy,’ Figures 5.5 and 5.6 list the percentages of each of these four target domains within the total data of Coordinated Target Domains.

Figures 5.5 and 5.6 clearly reflect that only 財政 cāizhèng ‘finance’ and 政治 zhèngzhì ‘politics’ in XIN have more percentages than in CNA (the others show more percentages in CNA than in XIN). Particularly, the percentage for 財政 cāizhèng ‘finance’ is almost twice as much as when used with 經濟 jīngjì ‘economy’ in China than in Taiwan.

corpora data for CNA and XIN, respectively, the results in Table 5.6 show the probability of a particular target domain co-occurring with 經濟 jīngjì ‘economy.’

\(^{11}\) The Chi-square comparison between CNA and XIN for 財政 cāizhèng ‘finance’ is \(\chi^2(1)=224.43, p<.05\); the result for 外交 wàijiāo ‘foreign affairs’ is \(\chi^2(1)=298.87, p<.05\); and the result for 教育 jiàoyù ‘education’ is \(\chi^2(1)=5.82, p<.05\).
Figure 5.5: Percentages of Coordination of 政治 zhèngzhì ‘politics,’ 外交 wàijīào ‘foreign affairs,’ 教育 jiàoyù ‘education’ and 財政 cāizhèng ‘finance’ with 經濟 jīngjì ‘economy’ in Coordinated Target Domains in CNA

Figure 5.6: Percentages of Coordination of 政治 zhèngzhì ‘politics,’ 外交 wàijīào ‘foreign affairs,’ 教育 jiàoyù ‘education’ and 財政 cāizhèng ‘finance’ with 經濟 jīngjì ‘economy’ in Coordinated Target Domains in XIN
In addition, Figures 5.5 and 5.6 also call to attention the fact that 政治 zhèngzhì ‘politics’ constitutes more than 80 per cent of total coordinate instances with 經濟 jīngjī ‘economy’ in both CNA and XIN, respectively, indicating that 經濟 jīngjī ‘economy’ and 政治 zhèngzhì ‘politics’ appear together more frequently than with other target domains. Chi-square tests comparing 政治 zhèngzhì ‘politics’ to all other target domains show significant effects in both CNA and XIN. If viewed according to the respective ontological levels of both 經濟 jīngjī ‘economy’ and 政治 zhèngzhì ‘politics,’ which are equally general (as they occur at the same higher ontological levels than the others), it is not surprising that these two target domains appear together the most frequently.

This is also because these two target domains are at the same ontological level (‘SocialInteraction,’ see Figure 5.3), a level that is higher than other target domains, and their combination is high in frequency. Therefore, from the design of corpora like this, similarities and differences between Taiwan and China can be pinpointed by examining the frequency of occurrences of target domains.

5.4 Total CNA and XIN Data for the Different Target Domains

Following the selection of target domains, sub-corpora that are suitable for comparing these target domains must be selected. In order to forge a comparison between the language used in Taiwan (CNA) and in China (XIN), and to select corpora data for these target domains, instances of corpora for these two regions must balance. Based on the corpora instances of the Single Target Domains and the Coordinated Target Domains, the overall number of instances for CNA and XIN can be seen in Table 5.7.

<table>
<thead>
<tr>
<th>Sub-corpora</th>
<th>Instances of Corpora Data</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNA (Taiwan)</td>
<td>57,853</td>
<td>54.34</td>
</tr>
<tr>
<td>XIN (China)</td>
<td>48,612</td>
<td>45.66</td>
</tr>
<tr>
<td>CNA and XIN</td>
<td>106,465</td>
<td>100.00</td>
</tr>
</tbody>
</table>

The number of coordinates with 經濟 jīngjī ‘economy’ for 政治 zhèngzhì ‘politics’ is significantly different from the coordinates of 經濟 jīngjī ‘economy’ and 財政 cáizhèng ‘finance’ in CNA ($\chi^2(1)=7792.48$, $p<.05$) and XIN ($\chi^2(1)=9997.10$, $p<.05$); from 經濟 jīngjī ‘economy’ and 外交 wàijiāo ‘foreign affairs’ in CNA ($\chi^2(1)=8782.21$, $p<.05$) and XIN ($\chi^2(1)=11023.54$, $p<.05$); and from 經濟 jīngjī ‘economy’ and 教育 jiàoyù ‘education’ in CNA ($\chi^2(1)=9887.46$, $p<.05$) and XIN ($\chi^2(1)=10640.75$, $p<.05$).
In total, 106,465 instances of corpora data from both CNA (54 per cent) and XIN (46 per cent) have been collected. In terms of manual analysis, these numbers are huge. Previous studies have not been able to analyze such a large quantity of data manually. Furthermore, these instances are parallel in terms of text-type as well as in the period of time from which they have been selected. An extensive comparison of metaphorical expressions in Taiwan and China, particularly one using such a huge number of corpora data, has not been carried out in previous works. Furthermore, the comparison of source domain selection between the two regions being studied has also not been seen in previous works. The following examples in (4) show possible differences in terms of the metaphorical expressions found in CNA and XIN, respectively.

(4) a. 江澤民 「作為 總書記 也要 抓 經濟」 (CNA)
   Jiang Zemin to.become secretary.general also want grab economy
   ‘To be a Secretary General, Jiang Zemin (President of the Republic of China) also needs to grab (the prosperous) economy.’

   b. 搞 社會主義 現代化 建設 不 抓 經濟 (XIN)
   engaged.in socialism modernize construct Neg. grab economy
   ‘To be engaged in modernizing the construction of socialism, (one) must grab (the) economy.’

Even though the metaphorical expression of 抓 zhuǎ ‘grab’ can be found both in Taiwan (4a) and China (4b), its uses in Taiwan usually refer to the situation in China. Therefore, when this type of metaphorical expression is compared in terms of its frequency of use, the number of tokens found in Taiwan is likely to be fewer than in China. Similarly, the use of 硬 yìng ‘hard’ in (5) is found more in China than in Taiwan. The instances in Taiwan usually reflect the uses in China.

(5) 經濟 和 政治 要 兩手硬 (XIN)
   economy and politics want two-hand-hard
   ‘The economy and politics need to be firm.’

---

13 The analyses of these data took about two continuous months of close reading, with an average of 8 hours per day, by a single analyzer.
Therefore, through the careful design of corpora data for CNA and XIN, phenomena such as those found in (4) and (5) can be observed. 

As explained in Footnote 9, the instances of CNA and XIN in the Coordinated Target Domains come from the years 1991 to 2002, whereas the first 4 per cent of the Single Target Domains is taken from data in the year 1991 only. The summary of the instances according to years is shown in Table 5.8.

Table 5.8: Period of Time for the Corpora Instances in CNA and XIN

<table>
<thead>
<tr>
<th>Sub-corpora</th>
<th>Datasets</th>
<th>Period of Time</th>
<th>Instances of Corpora Data</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNA</td>
<td>Single Target Domains</td>
<td>1991</td>
<td>43,595</td>
<td>75.35</td>
</tr>
<tr>
<td></td>
<td>Coordinated Target Domains</td>
<td>1991~2002</td>
<td>14,258</td>
<td>24.65</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>1991~2002</strong></td>
<td><strong>57,853</strong></td>
<td><strong>100.00</strong></td>
</tr>
<tr>
<td>XIN</td>
<td>Single Target Domains</td>
<td>1991</td>
<td>33,847</td>
<td>69.63</td>
</tr>
<tr>
<td></td>
<td>Coordinated Target Domains</td>
<td>1991~2002</td>
<td>14,765</td>
<td>30.37</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>1991~2002</strong></td>
<td><strong>48,612</strong></td>
<td><strong>100.00</strong></td>
</tr>
</tbody>
</table>

In general, about 70 to 75 per cent of the data is from 1991 for both CNA and XIN respectively. The remaining 25 to 30 per cent comes from the combination of years from 1991 to 2002. This means that more than 70 per cent of the data comes from Single Target Domains, where target domains such as 經濟 jinjì ‘economy’ serve as keywords for the search of their uses within the sub-corpora of CNA and XIN. The remaining 30 per cent consists of constructions of the Coordinated Target Domains, i.e., the co-occurrences of 經濟 jinjì ‘economy’ with other target domains, such as 財政 cóizhèng ‘finance’ (which are ‘X and/or/comma Y’ constructions).14

As mentioned previously, most studies in metaphor analyses to date use corpora either to look for all instances of metaphors within a specific topic (such as that in Özçalıskan (2003) for the examination of emotional terms in English and Turkish) or to employ a single keyword for the selection of corpora instances in different languages (such as that in Chung (2005, 2008) with the keyword of ‘market’). No previous study has examined the similarities and differences between source domains when the searched keywords have different constructions. The use of ‘Single Target Domains’ and ‘Coordinated Target Domains’ in this book will be able to show whether or not target domains perform similarly in terms of linguistic behaviors when appearing in different constructions. For instance, this book will be able to explain why 經濟 jinjì

---

14 The number of instances of the 30 per cent of the Coordinated Target Domains collected between 1991 and 2002 differs significantly between years (see Appendix Table A5.3 for the number of instances). If a diachronic analysis of a target domain is needed, these differences will be taken into consideration.
‘economy’ shares some similar metaphorical expressions such as 成長 chéngzhǎng ‘grow/growth’ with other target domains such as 財政 cāizhèng ‘finance’ (see (6a) and (6b)) and why the coordinated use of 經濟 jīngjì ‘economy’ and 財政 cāizhèng ‘finance’ (see (6c) and (6d)) does not share certain metaphorical expressions such as 起飛 qǐfēi ‘take off.’ The peculiarity is clearly illustrated in example (6d), with the use of 起飛 qǐfēi ‘take off’ with the coordinated use of 經濟 jīngjì ‘economy’ and 財政 cāizhèng ‘finance’ but not in (6c), where 經濟 jīngjì ‘economy’ appears alone.

(6) a. 國內 經濟 將 可 穩健 成長 (CNA)
    guónlèi jīngjì jiāng kě wěnjiàn chéngzhǎng
    within.country economy will can stable grow
    ‘The economy in the country will be able to grow stably.’

b. 就 國家 預算、 財政 與 經濟 成長
    jiù guójiā yùsuàn cāizhèng yǔ jīngjì chéngzhǎng
    with.regard.to country budget finance and economy growth
    表現 進行 報告 (CNA)
    biǎoxiàn jīnxíng báogào
    perform proceed report
    ‘To make a report with regard to the performance of (the) national budget, finance and economic growth...’

c. 南非 一直 希望 經濟 起飛 (CNA)
    nánfěi yízhí xīwàng jīngjì qǐfēi
    South.Africa continuously hope economy take.off
    ‘South Africa has always been hoping to have rapid growth in the economy (to have the economy take off).’

d. * 南非 一直 希望 財政 與 經濟 起飛
    nánfěi yízhí xīwàng cāizhèng yǔ jīngjì qǐfēi
    South.Africa continuously hope finance and economy take.off
    ‘South Africa has always been hoping to have finance and the economy take off.’
From the examples in (6), it is clear that only certain types of metaphorical expressions work for both target domains in the coordinated constructions. The questions that need to be asked, therefore, are: What kind of cognitive mechanisms allows some metaphorical expressions to be accepted and some to be rejected? And how can one find more of these uses? The answers to these two questions can be obtained by analyzing corpora data using a specifically designed sampling. Chapter Six will provide the analyses of the corpora data.

5.5 Summary of Chapter

This chapter has examined some of the problems encountered when analyzing a small sampling of corpora data. The problems of a small corpus become more evident in studies comparing different languages. Therefore, to ensure a carefully designed corpus, this book carries out the following steps: First, a larger corpus (the Chinese Gigaword corpus) with a similar size of data collections is needed. Second, target domains of different ontological levels are selected in order to extract corpora data that is meaningful for the analysis of source domains. The ontological representations of these target domains are used to form generalizations regarding the performance of metaphors of these target domains. Third, sub-corpora of Taiwan newspapers (CNA) and China newspapers (XIN) from the Chinese Gigaword corpus are selected for each target domain. From each of these sub-corpora, two types of data are collected, namely those with searched keywords as Single Target Domains and those with searched keywords as Coordinated Target Domains. Careful selection of corpora data has been put forth in this chapter as a solid method for the building of hypothesis, both in metaphor analyses and in source domain determination. This chapter has also discussed the number of corpora instances collected for the conditions of Single Target Domains and Coordinated Target Domains.

For both Single Target Domains and Coordinated Target Domains, the total number of instances available for metaphor analysis in the next chapter is 106,465 (77,442 from Single Target Domains and 29,023 from Coordinated Target Domains). These instances are comprised of both CNA and XIN data found in the Chinese Gigaword corpus. In the next chapter, these 106,465 instances will be analyzed in terms of tokens of metaphorical expressions (the number of times metaphorical expressions such as 成長 chéngzhāng ‘grow/growth’ and 起飛 qǐfēi ‘take off’ appear). These tokens of metaphorical expressions will be summarized in terms of types of metaphorical expressions (such as 成長 chéngzhāng ‘grow/growth’ and 起飛 qǐfēi ‘take off’), and these accumulated tokens will then be added up. As a result, only these types of metaphorical expressions will need to be sorted for source domain determination.
The purpose of this book is to determine source domains automatically for different types of metaphorical expressions. However, before this can be done, various metaphor analyses are required in order to gather a sufficient number of metaphorical expressions for each target domain so that source domain determination for each type of metaphorical expression can be carried out. As the aim is to determine source domains, the analysis of the corpora instances will be carried out manually at the present stage. After source domains have been determined (as mentioned in Chapter Three), metaphor identification will become easier, as the selection of the source domains is one of the key issues in conceptual metaphors. Chapter Six will explain both the steps of analyses as well as the results yielded.
Chapter Six: Metaphorical Expression Identification Using Target Domains

Chapter Five presented a detailed analysis of the problems encountered by some studies when faced with the challenge of small corpus and corpora data of varying sizes. While problems of small corpora exist, in recent years, studies with corpora have tended to use larger and larger corpora.¹ When the data is voluminous, more selection parameters are needed in order to extract datasets that are meaningful for a particular study. As John Sinclair (1991:13) has emphasized, “the bulk of any lay discussion regarding corpora concerns the criteria for text selection.” Chapter Five has discussed extensively the importance of corpora selection by looking specifically at the number of instances selected for each target domain (under the conditions of Single Target Domains versus those found under Coordinated Target Domains in both CNA and XIN).

As mentioned, no previous studies have designed corpora for the purpose of observing automatic determination of source domains. In addition, prior to this one, no study has compared the source domains used in Taiwan (CNA) and China (XIN) using large scale corpora data, parallel in terms of size and collection time span. Finally, no studies have implemented keywords with different constructions (single and coordinate) to extract corpora data for the purpose of the comparison of source domains. This book will fill these research gaps by using the carefully designed corpora data described in Chapter Five.

In this chapter, corpora instances will be analyzed in terms of their metaphorical expressions. As automatic identification of metaphors is not possible if source domains are not first defined, and for the reasons discussed in Chapter Three, these metaphorical expressions will be collected manually at this stage. After the manual identification of metaphorical expressions has been carried out, these metaphorical expressions will then be grouped in terms of source domains, using both the top-down and bottom-up approaches, respectively, in Chapters Seven through Ten.

¹ However, it must be noted that, depending on their purpose, some studies with small corpora data that have met the requirements of their area of research can be found. One example of effective yet small corpora can be found in the data from speeches of American presidents. Studies such as Charteris-Black (2005), Cienki (2005), as well as Ahrens (2006) use small corpora built from speeches of American presidents, and these small corpora were found to have been extremely useful in making generalizations regarding the speeches of politicians from different parties.
6.1 Global Hypotheses versus Local Hypothesis of Linguistic Behaviors

Chapter One has posited two hypotheses regarding the results of the top-down and bottom-up approaches to source domain determination. These hypotheses state that (a) a lexical and computational method is able to reduce human subjectivity in determining source domains through the use of lexical resources in both top-down and bottom-up approaches; and (b) the top-down and bottom-up approaches will perform differently in terms of source domain determination, with the top-down approach returning general source domains and the bottom-up approach returning specific source domains.

These two hypotheses are global hypotheses posited for the overall book, in contrast to the hypothesis posited previously in Chapter Five, which is a local hypothesis that predicts the following linguistic behaviors of the corpora data:

(1) The higher a target domain on the ontological hierarchy is, the more general this target domain is. More general target domains will produce more metaphorical expressions in metaphor analyses (Chapter Five; pp.78; ex. (3)).

The local hypothesis in (1) predicts the number of metaphorical expressions we may find in the analysis of the different target domains. The following section will describe how metaphorical expressions can first be identified in corpora data so that they can then be used in source domain determination analyses.

6.2 Procedures for Metaphorical Expression Identification

Chapter Five has shown that there are 106,465 instances for both Single Target Domains and Coordinated Target Domains in CNA and XIN spread over five target domains (經濟 jīngjì ‘economy,’ 政治 zhèngzhì ‘politics,’ 外交 wàijiāo ‘foreign affairs,’ 教育 jiàoyù ‘education’ and 財政 cáizhèng ‘finance’). These target domains will serve as ‘keywords in context’ (KWIC) and will then be searched in the concordance of the Chinese Gigaword corpus 1.0 by using the concordance function of the Chinese Sketch Engine (see Chapter Four for the functions of Sketch Engine). From the search results, their metaphorical expressions will be identified manually by looking at the contexts of the target domains.

The identification of metaphorical expressions (such as 成長 chéngzhǎng ‘grow/growth’ and 起飛 qǐfēi ‘take off’) in the corpora data is based on manual extraction, employing the criterion of deciding whether the target domains are used with expressions that are potentially from a different domain. For example, the use of 策略
cēlūè ‘tactics’ in the Coordinated Target Domains of 經濟和政治 jīngjì hàn zhèngzhì ‘economy and politics’ in (2a) is identified as a type of metaphorical expression because, intuitively, 策略 cēlūè ‘tactics’ does not belong to the literal uses of either 經濟 jīngjì ‘economy’ or 政治 zhèngzhì ‘politics.’ This means that 策略 cēlūè ‘tactics’ possibly comes from a different domain that may not be part of the literal meanings of 經濟 jīngjì ‘economy’ or 政治 zhèngzhì ‘politics.’

(2) a. 第三國家應遵循何種經濟 dì sān guójiā yìng zūnxún hé zhōng jīngjì
rank three country should follow which kind economy
或政治策略 (CNA)
huò zhèngzhì cēlūè
or politics tactics
‘Third World Countries should follow which kind of tactics of economy or politics?’

b. 德國外交部 同時又
déguó wàijiāoštì yòu
Germany department.of.foreign.affairs at.the.same.time again
發動外交攻勢
fādòng wàijiāo gōngshì
start(engine) foreign.affairs attack
‘At the same time, the Department of Foreign Affairs in Germany has again started their attack.’

At the same time, an additional criterion may be employed in the identification of metaphorical expressions by observing whether or not a target domain can be put in the place of the source domain. This follows the criteria of Stefanowitsch’s Metaphorical Pattern Analysis (2006). For example, 政治策略 zhèngzhì cēlūè ‘political tactics’ (2a) and 外交攻勢 wàijiāo gōngshì ‘the attack of foreign affairs’ (2b) are metaphors because the target domains of 政治 zhèngzhì ‘politics’ and 外交 wàijiāo ‘foreign affairs’ can both be replaced by words such as 戰爭 zhànzhēng ‘war’ when used literally. The examples in (2) may thus be identified as types of metaphorical expressions. An example of a non-metaphorical expression is the use of 學者 xuézhě ‘scholar’ in (3), in which 政治 zhèngzhì ‘politics’ is used literally.

(3) 這位著名的政治學者在接受
zhè wèi zhūmíng de zhèngzhì xuézhě zài jiēshòu
this Class. well-known DE politics scholar at receive
Within this Hong Kong daily news interview, ‘During this period, this well-known scholar on politics is being interviewed by the Hong Kong Times.’

In this example, the use of 學者 xuēzhě ‘scholar’ does not need to be replaced by another literal meaning because it is itself a literal meaning. To encapsulate different terminology (mentioned previously in Chapter Two), when terms such as 策略 cèlùè ‘tactics’ or 攻勢 gōngshì ‘attack’ appear in the corpora several times, their accumulated frequency is called tokens of metaphorical expressions, and these terms themselves are called types of metaphorical expressions. The results collected for both types of metaphorical expressions and tokens of metaphorical expressions will be presented in the next section. Other target domains are treated in a similar fashion, and the results of these analyses are also presented in the following section.

### 6.3 Results of Metaphorical Expression Identification

Target domains are analyzed and presented under the conditions of ‘Single Target Domains’ and ‘Coordinated Target Domains’ for CNA and XIN only. The combination of Single+Coordinated Target Domains will not be presented because the combinations of different datasets will affect the overall results (CNA+XIN will only be presented if characteristics of ‘overlapped’ data are discussed). Furthermore, Coordinated Target Domains have different constructions than Single Target Domains, and combining them will influence the results of the data. All coordinated data is never compared on its own but only in conjunction with Single Target Domains, for which it serves as a base of comparison.

The results of metaphor identification analyses will be presented in the order of percentages of tokens of metaphorical expressions (§6.3.1); tokens-per-type ratios of metaphorical expressions (§6.3.2); and target domains comparisons (§6.3.3).

#### 6.3.1 Percentages of Overall Tokens of Metaphorical Expressions

In this section, the percentages of tokens of metaphorical expressions will be presented. In Figure 6.1, the percentages of tokens of metaphorical expressions in CNA and XIN can be viewed (for the full table, see Appendix Table A6.1).
Figure 6.1: Percentages of Tokens of Metaphorical Expressions in CNA and XIN in the Single Target Domains and Coordinated Target Domains

These tokens of metaphorical expressions show how many times types of metaphorical expressions such as 成長 chéngzhǎng ‘grow/growth’ and 起飛 qǐfēi ‘take off’ appear in the different datasets. Figure 6.1 shows the percentages of the tokens of metaphorical expressions selected for CNA and XIN found in the overall corpora instances.

From Figure 6.1, the percentages for the tokens of metaphorical expressions in the different datasets are readily apparent. For example, within Single Target Domains, the percentages of tokens of metaphorical expressions (over the total corpora instances for Single Target Domains) are 32.29 per cent for CNA and 31.91 per cent for XIN. These percentages concisely illustrate that the number of tokens of metaphorical expressions found in both CNA and XIN newspapers account for approximately 30 per cent of the total instances for the sample collected, indicating that approximately 30 per cent of the language in newspapers consists of metaphorical expressions.

Kruskal-Wallis tests comparing the different datasets in Figure 6.1 show that significance was found between CNA and XIN in the Single Target Domains: $\chi^2(1)=97.48$, $p<.05$ (mean rank for CNA=809.27; mean rank for XIN=1056.24).²

² Kruskal-Wallis tests were used because the number of tokens of metaphorical expressions (in column two) was not normally distributed. Another issue involved in the comparisons of the different datasets was the different corpora instances across datasets. Therefore, for the statistical
Significance was also found between CNA and XIN in the Coordinated Target Domains: \( \chi^2(1)=13.76, p<.05 \) (mean rank for CNA=356.05; mean rank for XIN=298.03). These results indicate the data provided by CNA to be significantly different from that which was provided by XIN.  

From Figure 6.1, we know that in both Single Target Domains and Coordinated Target Domains, a higher percentage for Taiwan data was found than for China data. Examples of metaphorical expressions that can be found in newspapers are shown in (4).

(4) a. 在 政府 財政 健全 的 情況 下 進行 (CNA)  
在 government finance healthy DE situation below carry.out  
‘To be carried out while the finance of the government is healthy...’

b. 伊拉克 通過 外交 途徑 向 西方 透露 (XIN)  
Iraq pass.through foreign.affair s path toward Western divulge  
‘Iraq informed (something) to the Western countries through the path of foreign affairs.’

The examples in (4) show how types of metaphorical expressions such as 健全 ‘healthy’ and 途徑 ‘path’ can be used with target domains such as 財政 ‘finance’ and 外交 ‘foreign affairs.’ For (4b), the total occurrences of 外交 ‘foreign affairs’ and 途徑 ‘path’ number about

<table>
<thead>
<tr>
<th>Types of Metaphorical Expressions</th>
<th>Tokens of Metaphorical Expressions</th>
<th>Percentages (Divided by Total Tokens of Metaphorical Expressions=43,595)</th>
</tr>
</thead>
<tbody>
<tr>
<td>建設 jiànsè ‘construct/construction’</td>
<td>1,377</td>
<td>0.031586191</td>
</tr>
<tr>
<td>成長 chéngzhǎng ‘grow/growth’</td>
<td>957</td>
<td>0.021952059</td>
</tr>
<tr>
<td>改革 gǎigé ‘reform/reformation’</td>
<td>910</td>
<td>0.020873953</td>
</tr>
</tbody>
</table>

Since the ranks start from smaller values to larger values (see Footnote 7 in Chapter Five), the higher the mean rank, the more data are found gathering the smaller values (thus, a skewed distribution). The mean rank does not give information regarding how frequent a dataset occurs when compared to the others. Rather, it gives information regarding the distributional patterns of the data.
13.30 per cent (twenty-three tokens of metaphorical expressions for 途徑 tújǐng ‘path’) in the Single Target Domain of 外交 wàijiāo ‘foreign affairs.’

For Coordinated Target Domains in Figure 6.1, the results are slightly different: The percentages of tokens of metaphorical expressions drop from 29.58 per cent in CNA to 23.09 per cent in XIN, displaying a discrepancy in the number of metaphorical expressions found in CNA and XIN, respectively. The four target domains of 政治 zhèngzhī ‘politics,’ 財政 cáizhèng ‘finance,’ 外交 wàijiāo ‘foreign affairs’ and 教育 jiàoyù ‘education’ are combined with 經濟 jīngjì ‘economy’ to form the coordinated data. The lower percentage of coordinated data within XIN illustrates that these four target domains have an overall lower possibility of coordinating with 經濟 jīngjì ‘economy’ in XIN than in CNA. For example, the use of 建設 jiànsè ‘construct/construction’ in ways such as those seen in (5) constitute about 5.57 per cent (235 tokens of metaphorical expressions) in the total coordinated instances in CNA; however, it only constitutes about 3.40 per cent (116 tokens of metaphorical expressions) in the total coordinated XIN data.

(5) a. 但 民主 政治 的 建設 上 卻 遠
    dàn mínzhǔ zhèngzhī de jiànsè shàng què yuǎn
    but democracy politics DE construction above yet far
    落後 於 經濟 建設 (CNA)
    luòhòu yú jīngjì jiànsè
    left.behind at economy construction
    ‘But the building of the democratic economy is far left behind when compared to the building of the economy.’

b. 要 大力 加強 政治 建設 (XIN)
    yào dàlì jiáqiáng zhèngzhī jiànsè
    want big.strength strengthen politics construction
    ‘(One) needs to powerfully strengthen the building of politics.’

Therefore, even with the same type of metaphorical expression, such as 建設 jiànsè ‘construct/construction’ in (5), its distributional patterns in CNA and XIN may still differ. From the results in Figure 6.1, tokens of metaphorical expressions are usually found more in CNA than in XIN.

In Figure 6.1, when comparing the Single Target Domains with the Coordinated Target Domains, the drop in the tokens of metaphorical expressions in the Coordinated Target Domains may indicate that the target domains of 政治 zhèngzhī ‘politics,’ 外交 wàijiāo ‘foreign affairs,’ 財政 cáizhèng ‘finance’ and 教育 jiàoyù ‘education’ may also perform differently when they appear alone and when they appear with 經濟.
“economy’ in coordinate relation. Examples can be seen in (6) (taken from (6) in Chapter Five).

(6) a. 國內 經濟 將 可 穩健 成長 (CNA)
guónèi jīngjì jiāng kě wěnjìan chéngzhǎng
within.country economy will can stable grow
‘The economy in the country will be able to grow stably.’

b. 就 國家 預算、財政 與 經濟 成長
jiù guójiā yùsuàn cáizhèng yǔ jīngjì chéngzhǎng
with.regard. to country budget finance and economy growth
表現 進行 報告 (CNA)
bìàoxiǎn jìnchéng bàoɡào
perform proceed report
‘To make a report with regard to the performance of (the) national budget, finance and economic growth...’

c. 南非 一直 希望 經濟 起飛 (CNA)
nánfēi yīzhí xīwàng jīngjì qǐfēi
South.Africa continuously hope economy take off
‘South Africa has always been hoping to have rapid growth in the economy (to have the economy take off).’

d. 南非 一直 希望 財政 與 經濟 起飛
nánfēi yīzhí xīwàng cáizhèng yǔ jīngjì qǐfēi
South.Africa continuously hope finance and economy take off
‘!South Africa has always been hoping to have finance and the economy take off.’

In (6a) and (6b), we can see that 成長 chéngzhǎng ‘grow/growth’ can be used both with the Single Target Domain of 經濟 jīngjì ‘economy’ as well as with the Coordinated Target Domains as in 財政與經濟 cáizhèng yǔ jīngjì ‘finance and economy.’ Comparatively, 起飛 qǐfēi ‘take off’ cannot be used with the Coordinated Target Domains of 財政與經濟 cáizhèng yǔ jīngjì ‘finance and economy’ in (6d), but it can be used with the Single Target Domain of 經濟 jīngjì ‘economy’ in (6c).

While this section finds that tokens of metaphorical expressions are especially low in the Coordinated Target Domains, and specifically in XIN (cf. Figure 6.1), the next section intends to determine whether this phenomenon can be better understood by examining tokens-per-type ratios for each condition.
6.3.2 Tokens-per-Type Ratios for Overall Metaphorical Expressions

Figure 6.2 will explore the contrast between types and tokens of metaphorical expressions in the corpora data selected.

In Figure 6.2, the datasets for CNA, XIN and CNA+XIN are shown in terms of tokens-per-type ratios. Tokens-per-type ratios indicate the average tokens of metaphorical expressions per type of metaphorical expression. For example, in Figure 6.2, CNA has a ratio of 12.50 for Single Target Domains, indicating that each type of metaphorical expression in CNA comprises about thirteen tokens of metaphorical expressions in the overall Single Target Domains.

![Figure 6.2: Tokens-per-Type Ratios of Metaphorical Expressions for CNA, XIN and CNA+XIN in the Single Target Domains and the Coordinated Target Domains](image)

In Figure 6.2, the results show that the tokens-per-type ratios for XIN in both Single Target Domains (dotted line) and Coordinated Target Domains (solid line) are consistently higher than those of CNA. A high ratio means that each type of metaphorical expression in XIN consists of more tokens of metaphorical expressions on average than each type of metaphorical expression found in CNA. For example, 改革 gāigé ‘reform/reformation’ and 成長 chéngzhǎng ‘grow/growth’ exemplified in (7) to follow are types of metaphorical expressions that are often repeated.
(7) a. 對 墨西哥 經濟 和 教育 改革
duì měixīgē jìngjì hé jiàoyù gǎiguì
facing Mexico economy and education reformation
貢獻 良多 （CNA）
gòngxiàn liánghuǒ
contribute good many
‘To have contributed to the revolution of the economy and education of Mexico…’

b. 國內 經濟 將 可 穩健 成長 （CNA）
guónrén jìngjì jiāng kě wènjì àn chéngzhǎng
within country economy will can stable grow
‘The economy in the country will be able to grow stably.’

Each of these types of metaphorical expressions occurs one or more times in the datasets, and the tokens-per-type ratios can be produced by obtaining the average tokens of metaphorical expressions per type of metaphorical expression. Therefore, when the same metaphors are more often repeated, a smaller range of metaphor types will be found.

From Figure 6.2, too, we can predict that types of metaphorical expressions (such as 改革 gǎiguì ‘reform/reformation’ and 成長 chéngzhǎng ‘grow/growth’) in XIN are more centralized than those used in CNA (i.e., they take more tokens of metaphorical expressions on average than those of CNA). From Figure 6.1, we know that in both Single Target Domains and Coordinated Target Domains the mean rank for CNA is significantly higher than that of XIN, indicating that most of the frequencies of the tokens of metaphorical expressions in CNA are higher than those of XIN.

In this section, we proceed to compare the other paradigm, i.e., the Single Target Domains versus the Coordinated Target Domains in CNA, XIN and CNA+XIN. For CNA, a Kruskal-Wallis test shows that a significant effect was found in the Single Target Domains (mean rank=671.01) versus the Coordinated Target Domains (mean rank=1084.67): $\chi^2(1)=275.06, p<.05$. Even though the tokens-per-type ratios in Figure 6.2 were given in the form of ‘means,’ $t$-tests cannot be performed because the data were not normally distributed. Therefore, the Kruskal-Wallis tests were used instead. All comparisons across datasets must first divide the number of each token of metaphorical expression by the total token number in the respective datasets (cf. Footnote 2). This is to avoid comparing datasets of different sizes. The kurtosis and skewness values for the Single Target Domains and Coordinated Target Domains are shown in the following table.
Single Target Domains (mean rank=407.31) versus the Coordinated Target Domains (mean rank=578.75): $\chi^2(1)=74.89$, $p<.05$. As for CNA+XIN, the Kruskal-Wallis test also shows a significant effect between the Single Target Domains (mean rank=869.14) and the Coordinated Target Domains (mean rank=1281.96): $\chi^2(1)=213.47$, $p<.05$. The results, therefore, indicate that all comparisons between the Single Target Domains and the Coordinated Target Domains are significant, with the Single Target Domains possessing higher tokens-per-type ratios than the Coordinated Target Domains. In Figure 6.3, the tokens-per-type ratios for the overlapped and non-overlapped metaphorical expressions are displayed. Overlapped metaphorical expressions are metaphorical expressions that appear in both CNA and XIN, while non-overlapped metaphorical expressions are those that are not repeated in CNA and XIN. Examples of overlapped metaphorical expressions for the Single Target Domains can be seen in (8a) and (8b), while (8c) and (8d) show examples that do not overlap.

(8) a. 對大陸的經濟建設起了很大作用 (CNA)
大的function
‘To have impacted greatly the construction of the economy of Mainland China...’

b. 獨大陸的經濟建設起了很
face Mainland.China DE economy construction up LE very
då de zuòyòng
big DE function

---

Kurtosis and skewness values for the Single Target Domains

<table>
<thead>
<tr>
<th>CNA/XIN</th>
<th>Mean</th>
<th>N</th>
<th>Std. Deviation</th>
<th>Kurtosis</th>
<th>Std. Error of Kurtosis</th>
<th>Skewness</th>
<th>Std. Error of Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNA</td>
<td>2.867314920E-04</td>
<td>1126</td>
<td>1.474909753E-03</td>
<td>259.194</td>
<td>.146</td>
<td>14.712</td>
<td>.073</td>
</tr>
<tr>
<td>XIN</td>
<td>4.712730576E-04</td>
<td>677</td>
<td>2.024052847E-03</td>
<td>233.100</td>
<td>.188</td>
<td>13.097</td>
<td>.094</td>
</tr>
<tr>
<td>Total</td>
<td>3.560241375E-04</td>
<td>1803</td>
<td>1.703803150E-03</td>
<td>265.444</td>
<td>.115</td>
<td>14.227</td>
<td>.058</td>
</tr>
</tbody>
</table>

Kurtosis and skewness values for the Coordinated Target Domains

<table>
<thead>
<tr>
<th>CNA/XIN</th>
<th>Mean</th>
<th>N</th>
<th>Std. Deviation</th>
<th>Kurtosis</th>
<th>Std. Error of Kurtosis</th>
<th>Skewness</th>
<th>Std. Error of Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNA</td>
<td>6.545013496E-04</td>
<td>452</td>
<td>2.689154840E-03</td>
<td>165.527</td>
<td>.229</td>
<td>11.367</td>
<td>.115</td>
</tr>
<tr>
<td>XIN</td>
<td>1.044724932E-03</td>
<td>221</td>
<td>4.076000308E-03</td>
<td>65.206</td>
<td>.326</td>
<td>7.608</td>
<td>.164</td>
</tr>
<tr>
<td>Total</td>
<td>7.826431204E-04</td>
<td>673</td>
<td>3.213407505E-03</td>
<td>109.978</td>
<td>.188</td>
<td>9.590</td>
<td>.094</td>
</tr>
</tbody>
</table>

5 For all of these results, the Single Target Domains consistently possess lower mean rank than the Coordinated Target Domains, indicating that most of the data of the Single Target Domains appears before the data of the Coordinated Target Domains (and, thus, is skewed).
b. 公安 工作 是 搞 好 經濟 建設 (XIN)
     gōngān gōngzuò shì gǎo hǎo jīngjì jiànsè
     public.security job be work.out good economy construction
     ‘The job of public security is to work out the construction of the economy.’

c. 要求 我國 給予 政治 庇護 (CNA)
     yāoqiú wǒguó gěiyǔ zhèngzhì bìhù
     request our.country provide politics protection
     ‘To request our country provide political protection...’

d. 更 爲 西藏 經濟 的 騰飛
     gèng wèi xīzàng jīngjì de téngfēi
     more for Tibet economy DE fly.swiftly
     尽 一 份 力 (XIN)
     jìn yī fèn lì
     to.the.utmost one Class. strength
     ‘To have contributed more (or to provide strength to the upmost) to the swift rising of the Tibetan economy...’

From (8a) and (8b), we can see that 建設 jiànsè ‘construct/construction’ is a type of metaphorical expression that appears in both CNA and XIN (and, thus, an example of overlapped data). Overlapped types of metaphorical expressions usually contain a high number of tokens of metaphorical expressions within each corpus. For example, 建設 jiànsè ‘construct/construction’ is almost equally numerous within both CNA (1,612) and XIN (1,487), suggesting that metaphorical expressions that are used in both Taiwan and China are usually high-frequent metaphorical expressions. This can also be seen from their higher tokens-per-type ratios in Figure 6.3.
On the contrary, types of metaphorical expressions that do not overlap (such as that in (8c) and (8d)) are usually low in frequency. This can also be seen in Figure 6.3, where their tokens-per-type ratios are far lower than those that overlap. When Kruskal-Wallis tests were conducted on the results in Figure 6.3, the results found significance between overlapped (mean rank=1167.57) and non-overlapped (mean rank=553.86) for the Single Target Domains: $\chi^2(1)=661.42, p<.05$; and between overlapped (mean rank=443.98) and non-overlapped (mean rank=205.40) within the Coordinated Target Domains: $\chi^2(1)=278.20, p<.05$. These results indicate that overlapped metaphorical expressions are significantly different from non-overlapped metaphorical expressions.

As for the overlapped data in the Single Target Domains (mean rank=234.77) versus the Coordinated Target Domains (mean rank=318.83), significance was also found: $\chi^2(1)=32.51, p<.05$. For non-overlapped data, too, significance was found between the Single Target Domains (mean rank=600.58) and the Coordinated Target Domains (mean rank=1054.85): $\chi^2(1)=368.97, p<.05$. Therefore, the results show that all comparisons in Figure 6.3 were significant, indicating that not only does overlapped and non-overlapped data differ, the Single Target Domains and the Coordinated Target Domains also differ. The analysis in this section allows us to make certain predictions regarding highly frequent metaphorical expressions, i.e., these highly frequent metaphorical expressions are more likely to be used cross-regionally in Taiwan and China when compared to...
lower frequency metaphorical expressions.  In the next section, comparisons between different target domains will be illustrated.

### 6.3.3 Comparisons between Target Domains

In this study, five target domains have been examined under all conditions of the Single Target Domains and the Coordinated Target Domains with data from both CNA and XIN. These five target domains are 經濟 jìngjì ‘economy,’ 政治 zhèngzhì ‘politics,’ 財政 cāizhèng ‘finance,’ 外交 wàijiāo ‘foreign affairs’ and 教育 jiàoyù ‘education.’ The local hypothesis proposed from ontological representations of these target domains predicts that 經濟 jìngjì ‘economy’ and 政治 zhèngzhì ‘politics,’ both of which occur at higher ontological level, will yield more types of metaphorical expressions than other target domains, for the reason being that the higher a target domain is on the ontology, the more general the target domain is. Consequently, the number of tokens of metaphorical expressions for these two target domains will be expected to be higher than the other target domains. A comparison of the aforementioned target domains will be presented to illustrate the discrepancy between Single Target Domains and Coordinated Target Domains. Figure 6.4 reveals the percentages of tokens of metaphorical expressions over total corpora instances for the respective conditions in all five target domains. 經濟 jìngjì ‘economy’ has no coordinated data, as all of the other four target domains are in coordination with it. The Single Target Domain of 經濟 jìngjì ‘economy’ will be used as a point of comparison for the other four target domains in the Single Target Domain group.

---

6. We also found that overlapped metaphorical expressions constitute 85 per cent of the overall CNA data (i.e., amongst the 18,293 tokens of metaphorical expressions in CNA, those that overlap constitute 15,567 tokens of metaphorical expressions). The percentage in XIN is higher, i.e., 95 per cent (13,500 overlapped from a total of 14,208). These results mean that those overlapped items in XIN constitute almost 95 per cent of all the tokens of metaphorical expressions, while those in CNA constitute a lower percentage. This further confirms our statement that overlapped metaphorical expressions are high-frequent metaphorical expressions.
Chapter 6: Metaphorical Expression Identification Using Target Domains

Figure 6.4: Percentages of Tokens of Metaphorical Expressions in Five Target Domains with Different Datasets

All of the Coordinated Target Domains can only be compared to their respective counterpart within the Single Target Domains. Comparisons among the different Coordinated Target Domains will not be carried out, as the differences in metaphorical expressions between two target domains may be caused by factors other than the choice of metaphorical expressions. For instance, the frequency of one target domain may also constrain the possibility of its coordination with other target domains. Therefore, all comparisons of target domains are made in the paradigm of the Single Target Domains versus the Coordinated Target Domains as well as the comparison of different target domains within the Single Target Domains (see Appendix A6.2 for further detail).

Figure 6.4 shows the percentages of tokens of metaphorical expressions; almost all of the CNA bars are higher than their XIN bar counterparts in every case but three datasets (arrows in Figure 6.4). These datasets are the Single Target Domain of 政治 zhèngzhì ‘politics’ and both conditions of 教育 jiào yù ‘education.’ In particular, 教育
jiào yù ‘education’ can be highlighted for having more overall XIN data than other target domains, which indicates that this target domain may differ from others. The following examples in (9) show the most often occurring types of metaphorical expressions for the Single Target Domain of 教育 jiào yù ‘education’ in CNA and XIN, respectively.

(9) a. 加強 渔業 資源 保育 教育 (CNA)
    jiāqiáng yúyè zīyuán bǎoyù jiào yù
    strengthen fishery resource welfare education
    ‘To strengthen the education for the welfare of fishery resources...’

b. 深入 進行 全民 國防 教育 (XIN)
    shēnrù jìn xíng quán mín guόfáng jiào yù
    deep carry.out all.people national.defense education
    ‘To carry out intensive education on national defense for all people...’

The use of 加強 jiāqiáng ‘strengthen’ such as in (9a) constitutes 14 per cent of the total tokens of metaphorical expressions for 教育 jiào yù ‘education’ in CNA (182 tokens of metaphorical expressions from 1,300). On the other hand, the example of 進行 jìn xíng ‘carry out’ such as in (9b) constitutes about 22.35 per cent (317 tokens) of the total 1,419 tokens of metaphorical expressions in XIN. From here, we can see that these top types of metaphorical expressions differ between CNA and XIN, further emphasizing how minor differences between Taiwan and China can be revealed using corpora data.

From all the various target domains in Figure 6.4, the highest percentages of tokens of metaphorical expressions appear in the Single Target Domain of 經濟 jīngjī ‘economy’ in CNA. The most frequently occurring type of metaphorical expression for 經濟 jīngjī ‘economy (i.e., 建設 jiànshè ‘construct/construction’) is the same in CNA and XIN. The percentages of 建設 jiànshè ‘construct/construction’ are almost the same in CNA and XIN: it constitutes 15.56 per cent (1,346) from the total 8,648 tokens of metaphorical expressions in CNA for the Single Target Domain of 經濟 jīngjī ‘economy’ and 16.87 per cent (1,308) from the total 7,752 tokens of metaphorical expressions in XIN. Examples of 建設 jiànshè ‘construct/construction’ can be seen previously in (5).

On the other hand, the lowest percentage of metaphorical expressions is found in the Coordinated Target Domains of 外交 wài jiāo ‘foreign affairs’ (dotted arrow in

---

7 For the present chapter, all the top types of metaphorical expressions for each target domain will not be listed due to space limitions. However, they will be discussed after the source domains have been determined.
In addition, both CNA and XIN also have similar most frequently occurring types of metaphorical expressions, e.g., 制裁 zhīcài ‘impose sanction.’ Examples of 制裁 zhīcài ‘impose sanction’ in the coordinated use of 外交 wàijiāo ‘foreign affairs’ are given in (10).

(10) a. 將 給 自己 帶來 西方 的 外交 與
jiàng gěi zìjǐ dài lái xīfāng de wàijiāo yǔ
will give self bring western DE foreign.affairs and
经济 zhīcài
jingjì zhīcài
economy impose.sanction
‘Will bring to oneself sanctions on foreign affairs and (the) economy...’

b. 包括 禁止 其 民航 飛行 國際 航線
bāokuò jīnzhǐ qī mínzhāng fēixíng guójì hángxiàn
include forbid it civil.aviation flight international air.route
的 一 系列 経済 和 外交 制裁 (XIN)
de yī xíliè jīngjì hàn wàijiāo zhīcài
DE one series economy and foreign.affairs impose.sanction
‘...(which) includes a series of sanctions on economy and foreign affairs, forbidding its (Libya’s) national airlines to fly international air routes.’

For Figure 6.4, a comparison between the percentages of metaphorical expressions only provides information in terms of how frequently metaphorical expressions are found within the dataset of different target domains. These percentages provide pertinent information, such as what percentage of newspaper content within different target domains is comprised of metaphorical expressions. It seems that almost all target domains have a greater percentage of metaphorical expressions in the Single Target Domains than in the Coordinated Target Domains, with the exception of 外交 wàijiāo ‘foreign affairs,’ which is found to have more metaphorical expressions when it is used in conjunction with 経済 jīngjì ‘economy’ than when it is by itself. Examples of 外交 wàijiāo ‘foreign affairs’ used as a Single Target Domain can be seen in (11).

(11) a. 日本 積極 的 參與 外交 努力 (CNA)
riběn jījí de cānyǔ wàijiāo nǔlì
Japan vigorous DI participate foreign.affairs strive
‘Japan is participating vigorously in the strive for foreign affairs.’
b. 探討 新的 外交 努力的 可能性 (XIN)

探討新外交努力的可能性（XIN）

To probe into new possibilities in the striving for foreign affairs...

The exception of 外交 wàijiāo ‘foreign affairs’ is rather unexpected, as from the previous section we can see that the data in the Coordinated Target Domains is usually smaller in quantity than in the Single Target Domains. The exception of 外交 wàijiāo ‘foreign affairs’ illustrates that even with a smaller number of corpora instances of Coordinated Target Domains for 外交 wàijiāo ‘foreign affairs’ (1,097 for CNA; 423 for XIN) than its Single Target Domains (3,687 for CNA; 1,274 for XIN), the percentages of metaphorical expressions in the Coordinated Target Domains are more numerous than that of the Single Target Domains. This demonstrates that when in coordinate relation with 經濟 jìngjì ‘economy,’ the behavior of the target domains may differ from their respective Single Target Domains.

6.3.3.1 Comparisons between Single Target Domains

Figure 6.5 shows a comparison of tokens-per-type ratios within the Single Target Domains.

Figure 6.5: Tokens-per-Type Ratios in the Single Target Domains
For the different target domains in Figure 6.5, comparisons through Kruskal-Wallis tests were carried out (by dividing each of the tokens of metaphorical expressions by the respective total number of corpora instances for each target domain in CNA and XIN).\(^8\) The results show that all target domains were significantly different in terms of CNA and XIN except 財政 cāizhèng ‘finance:’ \(\chi^2(1)=3.02, p=.082\) (mean rank for CNA=87.13; mean rank for XIN=100.59).\(^9\) The most often occurring type of metaphorical expression for 財政 cāizhèng ‘finance’ (which is the same for both CNA and XIN) is shown in (12).

(12) a. 要 想 改善 政府 的 財政 困難 (CNA)
    yào xiǎng gǎiánhàn zhèngfǔ de cāizhèng kùnnán
    want think make.better government DE finance difficult
    ‘...has (the) intention to improve the financial difficulties of the government.’

b. 美國 地方政府 財政 困難 將 影響 經濟 (XIN)
    měiguó dìfāngzhèngfǔ cāizhèng kùnnán jiàng yǐnxǐng jīngjù
    America local.government finance difficult will affect economy
    ‘The local government of America is having financial difficulties, which
    will affect the economy.’

困難 kùnnán ‘difficult/difficulty’ is the most frequently used type of metaphorical expression in CNA (157 tokens of metaphorical expressions, which constitute 25.20 per cent of the total 623 tokens of metaphorical expressions) and in XIN (124 tokens of

\(^8\) This is carried out because the data have different sizes, even though they are the similar first 4 per cent of the total corpora instances. Kruskal-Wallis tests are used because the data are not normally distributed. Skewness and kurtosis values are shown below.

<table>
<thead>
<tr>
<th>Single Target Domains</th>
<th>Mean</th>
<th>N</th>
<th>Std. Deviation</th>
<th>Skewness</th>
<th>Std. Error of Skewness</th>
<th>Kurtosis</th>
<th>Std. Error of Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>外交 ‘foreign affairs’</td>
<td>1.679423E-03</td>
<td>175</td>
<td>4.868486E-03</td>
<td>8.139</td>
<td>.184</td>
<td>80.453</td>
<td>.365</td>
</tr>
<tr>
<td>政治 ‘politics’</td>
<td>1.012589E-03</td>
<td>665</td>
<td>2.776832E-03</td>
<td>6.520</td>
<td>.095</td>
<td>52.513</td>
<td>.189</td>
</tr>
<tr>
<td>財政 ‘finance’</td>
<td>2.572925E-03</td>
<td>186</td>
<td>8.660336E-03</td>
<td>7.232</td>
<td>.178</td>
<td>54.293</td>
<td>.355</td>
</tr>
<tr>
<td>教育 ‘education’</td>
<td>1.334323E-03</td>
<td>344</td>
<td>4.563115E-03</td>
<td>8.870</td>
<td>.131</td>
<td>98.519</td>
<td>.262</td>
</tr>
<tr>
<td>經濟 ‘economy’</td>
<td>6.782157E-04</td>
<td>1094</td>
<td>3.401130E-03</td>
<td>12.805</td>
<td>.074</td>
<td>197.488</td>
<td>.148</td>
</tr>
</tbody>
</table>

\(^9\) \(\chi^2(1)=136.33, p<.05\) for 政治 zhèngzhì ‘politics’ (mean rank for CNA=287.26; mean rank for XIN=487.38); \(\chi^2(1)=32.64, p<.05\) for 外交 wàijiāo ‘foreign affairs’ (mean rank for CNA=77.66; mean rank for XIN=127.70); \(\chi^2(1)=14.59, p<.05\) for 教育 jiàoyù ‘education’ (mean rank for CNA=152.20; mean rank for XIN=192.56); and \(\chi^2(1)=13.63, p<.05\) for 經濟 jīngjù ‘economy’ (mean rank for CNA=576.29; mean rank for XIN=505.20).
metaphorical expressions, which constitute 36.69 per cent of the total 338 tokens of metaphorical expressions).

In order to forge a comparison between the target domains, similar Kruskal-Wallis tests were employed. However, since there will be multiple comparisons for the different target domains, the alpha level has been corrected to .01 based on the Bonferroni method (rather than .05). The overall comparisons within CNA show that there was significant effect between the target domains: \( \chi^2(4)=436.73, p<.01 \). Similarly, within XIN, there was also significant effect found between the target domains: \( \chi^2(4)=294.65, p<.01 \). The mean ranks for the different target domains are given in Table 6.1.

Table 6.1: Mean Ranks for the Single Target Domains in CNA and XIN

<table>
<thead>
<tr>
<th>Single Target Domains</th>
<th>CNA</th>
<th>XIN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean Ranks</td>
<td>Order of Ranks</td>
</tr>
<tr>
<td>外交 wàijī ‘foreign affairs’</td>
<td>1115.84</td>
<td>5</td>
</tr>
<tr>
<td>政治 zhèngzhī ‘politics’</td>
<td>820.98</td>
<td>2</td>
</tr>
<tr>
<td>財政 cáizhēng ‘finance’</td>
<td>1314.71</td>
<td>4</td>
</tr>
<tr>
<td>教育 jiàoyù ‘education’</td>
<td>1016.53</td>
<td>3</td>
</tr>
<tr>
<td>經濟 jīngjī ‘economy’</td>
<td>548.37</td>
<td>1</td>
</tr>
</tbody>
</table>

Notice that the ranking for 經濟 jīngjī ‘economy’ always comes first in both CNA and XIN; this indicates that the data of 經濟 jīngjī ‘economy’ contains more metaphorical expressions with lower frequencies (cf. Footnote 2 regarding mean ranks). For CNA, the ranking after 經濟 jīngjī ‘economy’ is 政治 zhèngzhī ‘politics’ while for XIN, the second ranking is 教育 jiàoyù ‘education’ (see number 2 in ‘Order of Ranks’ in Table 6.1). The question we are interested in is whether or not 經濟 jīngjī ‘economy’ and 政治 zhèngzhī ‘politics’ (in bold in Table 6.1) differ from the other target domains. The hypothesis in (1) previously says that “the higher a target domain on the ontological hierarchy is, the more general this target domain is. More general target domains will produce more metaphorical expressions in metaphor analyses.” Therefore, we then compared 經濟 jīngjī ‘economy’ and 政治 zhèngzhī ‘politics’ to other target domains, respectively, in Kruskal-Wallis tests. The results show that both of these two target domains differ from all other target domains in both CNA and XIN, with the exception of the XIN data, where 政治 zhèngzhī ‘politics’ and 財政 cáizhēng ‘finance’ do not differ significantly: \( \chi^2(1)=1.95, p=.162 \) (the mean rank for 財政 cáizhēng ‘finance’ is 128.60; the mean rank for 政治 zhèngzhī ‘politics’ is 115.81). These insignificant

---

10 Mean rank in Kruskal-Wallis is dependent on the data compared. Therefore, the mean rank in Table 6.1 will not be similar to the mean rank when only CNA or XIN data of two target
results indicate that there is no difference to be found between these two target domains in terms of the occurrences of metaphorical expressions in the total XIN instances analyzed. The following examples in (13) show metaphorical expressions used with 政治 zhèngzhì 'politics' and 財政 cáizhèng 'finance,' respectively.

(13) a. 嚴格 把 好 思想 政治 關

yánge bǎ hǎo sǐxiǎng zhèngzhì guān
strict guard good thoughts politics gate
‘To guard the gate of the politics of thoughts...’

b. 保證 未來 特區 政府 有 一 個
bǎozhèng wèilái tèqū zhèngfǔ yǒu yī ge
ensure future special.area government have one Class.
健全 的 財政 (XIN)
jiùnquán de cáizhèng
sturdy DE finance
‘To ensure that the government of the special administrative region will have sturdy finances...’

The insignificance between 政治 zhèngzhì 'politics' and 財政 cáizhèng 'finance' in XIN can also be predicted by looking at the overall mean rank in Table 6.1, i.e., the mean rank of these two target domains (shaded) is close to one another (615.34 and 640.58). The overall results are given in Appendix A6.3. From the results, 經濟 jīngjì 'economy' differs significantly from all of the other four target domains, including the target domain of 政治 zhèngzhì 'politics.' This is unexpected because the ontology (see Figure 5.3 in Chapter Five) shows that 經濟 jīngjì 'economy' and 政治 zhèngzhì 'politics' occur at a similarly high level. It is, therefore, possible that they differ from all others but not among themselves. The results, however, found that they also differ from one another. This means that for a coarse categorization of the ontology, predictions regarding information in different hierarchical levels may become more robust than those at the close (or similar) levels. In other words, when 經濟 jīngjì 'economy' and 政治 zhèngzhì 'politics' were compared with other levels that are distinctively separated from them, their prediction is made better. However, for the target domains grouped domains are compared. This means that every time a different sub-set of the data is compared, these data will be re-ranked in each comparison.

Since the local hypothesis predicts only the differences of 經濟 jīngjì 'economy' and 政治 zhèngzhì 'politics' to other target domains, the comparisons of other target domains are not carried out in this book. The similarities and differences between other target domains are proposed to be looked at for future work.
closer to one another, their distinction cannot be predicted precisely, as they are treated as similar items in the ontology (although, in fact, they may also differ). This reflects an important discovery of linguistic categorization, whereby ontology can be seen as a coarse-grained categorization of concepts, and these categories can be examined in fine-grained corpora analyses. The following section will examine the results of the Coordinated Target Domains.

### 6.3.3.2 Comparisons between Coordinated Target Domains

For the Coordinated Target Domains, the tokens-per-type ratios are shown in Figure 6.6.

![Figure 6.6: Tokens-per-Type Ratios in the Coordinated Target Domains](image)

Since Coordinated Target Domains depend upon the coordination of the four target domains in Figure 6.6 with "經濟 jīngjì ‘economy,’" their corpora instances will naturally differ from one another. As a result, comparisons between Coordinated Target Domains will be made in terms of CNA versus XIN only within each target domain. Comparisons across target domains will not be carried out. Comparisons, however, can be made between the Single Target Domains (Figure 6.5) and the Coordinated Target Domains (Figure 6.6) within similar target domains, as we would like to know whether the tokens-per-type ratios are similar for the use of Single Target Domains (as in (14a)) and the Coordinated Target Domains (as in (14b)).
(14) a. 法國 將 竭盡 一切 外交 手段 來
fāguó jiàng jiéjìn yìqié wàijiāo shǒuduǎn lái
France will devote to the full whole foreign affairs trick come
避免 戰爭 (XIN)
bǐmiǎn zhànzhēng
avoid war
‘France will devote every trick of foreign affairs to avoid the happening of war.’

b. 美國 主張 盡量 採用 外交
měiguó zhǔzhāng jìnliàng cǎiyòng wàijiāo
America of the opinion to the utmost use foreign affairs
經濟 等 手段 和平 解決 南斯拉夫 危機 (XIN)
jìngjì děng shǒuduǎn hépíng jiéjué nánslāfū wéijī economy etc. trick peace solve Yugoslavia crisis
‘America’s stance is that they will try to use tricks of foreign affairs, economy, etc. to solve the crisis of Yugoslavia in a peaceful manner.’

In (14a), the use of 外交 wàijiāo ‘foreign affairs’ appears alone, while in (14b), the use of 外交 wàijiāo ‘foreign affairs’ appears with the target domain of 經濟 jìngjī ‘economy.’ Therefore, when we compare the use of the Single Target Domain of 外交 wàijiāo ‘foreign affairs’ with its Coordinated Target Domain with 經濟 jìngjī ‘economy,’ contrasts between these two constructions can be displayed.

For the results in Figure 6.6, we can see that the XIN data consistently has a higher number of tokens-per-type ratios than the CNA data. This result is also found previously in Figure 6.2, and this phenomenon was attributed to the tendencies of the metaphorical expressions in XIN being repeated more often than the metaphorical expressions in CNA (cf. Footnote 6). For the Coordinated results in Figure 6.6, Kruskal-Wallis tests were carried out (because all datasets were not normally distributed). All CNA versus

<table>
<thead>
<tr>
<th>Coordinated Target Domains</th>
<th>Mean</th>
<th>N</th>
<th>Std. Deviation</th>
<th>Kurtosis</th>
<th>Std. Error of Kurtosis</th>
<th>Skewness</th>
<th>Std. Error of Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>外交 ‘foreign affairs’</td>
<td>3.75181907E-03</td>
<td>101</td>
<td>6.73011507E-03</td>
<td>17.518</td>
<td>.476</td>
<td>4.089</td>
<td>.240</td>
</tr>
<tr>
<td>政治 ‘politics’</td>
<td>9.77947505E-04</td>
<td>594</td>
<td>3.96601931E-03</td>
<td>96.752</td>
<td>.200</td>
<td>9.061</td>
<td>.100</td>
</tr>
<tr>
<td>财政 ‘finance’</td>
<td>2.29870785E-03</td>
<td>91</td>
<td>3.03180419E-03</td>
<td>16.765</td>
<td>.500</td>
<td>3.840</td>
<td>.253</td>
</tr>
<tr>
<td>教育 ‘education’</td>
<td>5.59398612E-03</td>
<td>52</td>
<td>8.87097709E-03</td>
<td>4.879</td>
<td>.650</td>
<td>2.470</td>
<td>.330</td>
</tr>
</tbody>
</table>

12 Kurtosis and skewness values are provided in the following table.
XIN comparisons were found to be significant in the Coordinated Target Domains, indicating that the Taiwan data is not similar to the China data.\(^{13}\)

When comparing each target domain in the Coordinated Target Domains with similar target domains in the Single Target Domains in Figure 6.5, similar significant effects were found for all four target domains of 政治 zhèngzhì ‘politics,’ 外交 wàijiāo ‘foreign affairs,’ 財政 cáizhèng ‘finance’ and 教育 jiàoyù ‘education,’ detailed results for which are provided in Appendix A6.4.\(^{14}\) This means that the performance of the Coordinated Target Domains is indeed different from their respective Single Target Domains.\(^{15}\)

In addition, from the results in Figure 6.6, we can see that the ratios for 政治 zhèngzhì ‘politics’ are especially high. As the local hypothesis in (1) predicts that target domains that are at a higher level will produce more tokens of metaphorical expressions, high-number metaphorical expressions in 政治 zhèngzhì ‘politics’ may show that 政治 zhèngzhì ‘politics’ is in many ways similar, in terms of metaphorical expressions, to 經濟 jīngjì ‘economy’ (since they appear at a similar level).

In addition to making comparisons between the target domains, we are also interested in finding out whether there is a relationship between the frequencies of the instances in corpora and their use of metaphorical expressions. A Pearson correlation test comparing the number of corpora instances for each Coordinated Target Domain and their number of tokens of metaphorical expressions indicates a high correlation between these two, \(r(6)=0.988, p<.05\), meaning that the more instances that are analyzed for a Coordinated Target Domain, the more tokens of metaphorical expressions are found. This was also found for the number of corpora instances and the types of metaphorical expressions: \(r(6)=0.889, p<.05\). A similar Pearson correlation test comparing the number of corpora instances for each Single Target Domain and their number of tokens of metaphorical expressions also revealed there to be significant effect between these two: \(r(8)=0.990, p<.05\). Significance was also found for the correlation between the number of corpora instances and the number of types of metaphorical expressions: \(r(8)=0.878, p<.05\).

---

\(^{13}\) For 外交 wàijiāo ‘foreign affairs,’ \(\chi^2(1)=24.11, p<.05\) (mean rank for CNA=43.87; mean rank for XIN=76.59); for 政治 zhèngzhì ‘politics,’ \(\chi^2(1)=11.03, p<.05\) (mean rank for CNA=313.61; mean rank for XIN=264.80); for 財政 cáizhèng ‘finance,’ \(\chi^2(1)=6.89, p<.05\) (mean rank for CNA=50.98; mean rank for XIN=36.35); and for 教育 jiàoyù ‘education,’ \(\chi^2(1)=19.35, p<.05\) (mean rank for CNA=18.97; mean rank for XIN=36.77).

\(^{14}\) 經濟 jīngjì ‘economy’ was not included because it does not have coordinated data.

\(^{15}\) As mentioned, the Coordinated Target Domains cannot be compared among themselves; they will be constricted to their respective Single Target Domains.
As postulated in Wang’s (1969) ‘evolution model’ (and later attested in Huang [黃居仁] 2004b), the more frequency a word has, the more likely this word will have polysemous meanings, where there will be new meanings developed from the original meanings (diffusion) while the original meanings still remain, perhaps in a lower frequency (residue). This may be due to the fact that the same frequency effect is also at play in the creation of types of metaphorical expressions as well as in the tokens of metaphorical expressions in the different target domains. These high correlations also show that the higher the number of corpora instances in the corpus, the higher the number of metaphorical expressions to be found. In other words, the more frequency a word has, the more likely this word will be used metaphorically. These observations regarding frequency and metaphors have not been discussed in previous studies.

In the results from both the Single Target Domains and the Coordinated Target Domains, it can be seen that 經濟 jīngjì ‘economy’ and 政治 zhèngzhì ‘politics’ display a higher numbers of metaphorical expressions. These results may also be interpreted using the ‘evolution model’ (Wang 1969), which states that the more frequency a construction has, the likelier this construction will be used metaphorically. In Figure 6.6, in particular, the data for 政治 zhèngzhì ‘politics’ stands out as the most frequently co-occurring target domain with 經濟 jīngjì ‘economy.’ The examples in (15) show the uses of 大國 dàguó ‘powerful country,’ which occur with 經濟 jīngjì ‘economy’ and 政治 zhèngzhì ‘politics’ only but not with other target domains (see (15b), (15c) and (15d)).

(15) a. 日本和德國想成為政治和經濟 日本和德國想成為政治和經濟
riběn hàn déguó xiǎng chéngwéi zhèngzhì hàn jīngjì
Japan and Germany want to become politics and economy
大國
dàguó
powerful.country
‘Japan and Germany want to become powerful countries in politics and economics.’

---

16 Even though the Coordinated Target Domains are not compared among themselves, the corpora instances of coordination in Figure 6.6 come from the entire Chinese Gigaword corpus, i.e., in the whole corpus, 政治 zhèngzhì ‘politics’ appears to share the most coordination constructions with 經濟 jīngjì ‘economy,’ when compared to the other target domains.

17 A question mark in front of the sentence indicates that this sentence is doubtful in terms of occurrences. An exclamation mark indicates an impossible use.
A Corpus-driven Approach to Source Domain Determination

b. 日本和德国想成为外交和经济
   日本和德国想成为外交和经济
   大国
dàguó
   ‘Japan and Germany want to become powerful countries in foreign affairs and economics.’

c. 日本和德国想成为财政和经济
   日本和德国想成为财政和经济
   大国
dàguó
   ‘Japan and Germany want to become powerful countries in finance and economics.’

d. 日本和德国想成为教育和经济
   日本和德国想成为教育和经济
   大国
dàguó
   ‘Japan and Germany want to become powerful countries in education and economics.’

The reason why 大国 ‘powerful country’ is only seen with 政治 ‘politics’ and 经济 ‘economy’ in our data is because these two target domains are often used to compare power between different countries. Other target domains, in contrast, are seldom used in this manner.

The examples shown in (15) illustrate that some metaphorical expressions are found with 政治 ‘politics’ and 经济 ‘economy’ but not with other target domains. This supports the local hypothesis regarding the effect caused by the general ontological concept of these two target domains in terms of their frequencies. The coordinated data for 政治 ‘politics’ and 经济 ‘economy’ maximizes the similarities between these two target domains, as it is only when they can occur with the same types of metaphorical expressions that their tokens will increase in the coordinate relation, which is also in agreement with their similar ontological level.
Based on this reasoning, more occurrences of metaphorical expressions are observable for the Coordinated data of 政治 zhèngzhì ‘politics’ than for other target domains.

However, as also shown in this section, the local hypothesis previously in (1) (which is created based on the ontology) is not able to predict the similarities and differences between CNA and XIN, as the ontology assumes a shared conceptual knowledge in all cultures. It is therefore only through linguistic analyses that the advantages and disadvantages of using ontology are made apparent.

To sum up the comparisons of target domains, the local hypothesis regarding the higher number metaphorical expressions is seen in the target domains of 政治 zhèngzhì ‘politics’ and 經濟 jīngjì ‘economy’ in both CNA and XIN, with one exception of 財政 cáizhěng ‘finance’ in XIN. This could mean that certain adjustments should be made in terms of conceptual knowledgebases such as that provided by SUMO. Some data may not exactly match the ontological representation of human knowledge. No previous works have evaluated the ontology of metaphors with the aid of linguistic data; these observations will also be meaningful for predicting the uses of ontology in linguistics.

By undertaking a comparison between the Single Target Domains and the Coordinated Target Domains within metaphor analysis, we were able to provide linguistic evidence of how corpus can reveal different language use. In addition, the results regarding tokens of metaphorical expressions clearly support the usefulness of ontology in predicting linguistic behaviors of the target domains. More comparisons can be made after the source domains for the types of metaphorical expressions have been determined using the top-down and bottom-up approaches.

6.4 Summary of Chapter

The aim of this book is to compare the top-down and bottom-up approaches in determining source domains. However, before this can be done, a large amount of types of metaphorical expressions (such as 成長 chéngzhǎng ‘grow/growth’ and 起飛 qǐfēi ‘take off’) have to be collected so that source domain analyses can be carried out.

Based on the analyses of the datasets from both the Single Target Domains and Coordinated Target Domains, it is apparent that a hypothesis formulated using an ontology would be able to predict certain linguistic behaviors of the target domains in terms of frequency (as more general concepts tend to accept a wider range of metaphors, resulting in a higher frequency of metaphorical expressions). There are, however, some mismatches, as ontology is unable to predict regional differences (CNA versus XIN) and constructional differences (Single Target Domains versus Coordinated Target
Domains), but these differences can be found through corpus-based observations of linguistic behaviors.

This chapter has also illustrated the ability of linguistic data to complement the disadvantages of an upper ontology, i.e., an upper ontology usually assumes a shared cognitive concept between communities. However, through empirical data, this chapter is able to prove that even though ontology can aid in displaying the ‘big picture,’ detailed linguistic data can help bring out the minor differences between the two communities. For example, while the prediction of the use of metaphorical expressions in 経済 jīngjī ‘economy’ and 政治 zhèngzhì ‘politics’ is based on their respective ontological representations, the subtle differences in percentages between the two communities will only materialize from an analysis of empirical linguistic data, which is shown in the source domains analysis in the subsequent chapters of this study.

In addition, SUMO assumes a similar ontological concept to exist between different target domains within different communities, as SUMO does not specify whether 政治 zhèngzhì ‘politics’ in Taiwan is slightly different from the same concept in China. Using linguistic data is an effective way to compare and contrast minor discrepancies between the two. Nevertheless, this does not deny the workability of the ‘big picture’ of ontology; instead, ontology provides a platform for comparison between an assumed upper-level shared knowledge and actual corpora data, where a bottom-up linguistic approach complements the top-down domain approach. This book will further compare by analyzing the use of the top-down versus bottom-up approach when determining source domains.

The final outcome of the metaphorical expression identification process is defined by the total number of types of metaphorical expressions, which includes 1,282 types of metaphorical expressions from CNA and 736 from XIN, the combination of which produces 1,591 types of metaphorical expressions, since there are 427 examples of overlapping expressions. These types of metaphorical expressions will be used in Chapters Seven through Ten for the top-down and bottom-up approaches, respectively.18 Comparisons across target domains as well as datasets will be conducted after the source domains have been determined.

---

18 The combination of CNA+XIN is for the ease of programming only and not for linguistic analyses, as combinations of some data (such as those between different regions) will lose important information regarding the combined data.
Chapter 7: Top-down Approach to Source Domain Determination

The first (global) hypothesis states that a lexical and computational method will be able to illuminate the differences between these two approaches. Several lexical resources (such as SinicaBow, WordNet and SUMO) have been used to determine source domains so that human intuition can be reduced. The second (global) hypothesis of this chapter compares the top-down and bottom-up approach with regards to source domain determination (cf. distinction of global versus local hypotheses in Footnote 4 in Chapter Five). This hypothesis predicts that a top-down approach will differ from a bottom-up approach in performance when determining source domains. In order to test these hypotheses, Chapters Five and Six have introduced corpora data along with their respective analyses used to extract metaphorical expressions. The overall types of metaphorical expressions (such as 成長 chéngzhǎng ‘grow/growth,’ 起飛 qǐfēi ‘take off’ and 瘫瘓 tānhuàn ‘paralytic’) will be used for source domain determination. After these source domains have been determined, the tokens of metaphorical expressions (frequency of 成長 chéngzhǎng ‘grow/growth,’ 起飛 qǐfēi ‘take off’ and 瘫瘓 tānhuàn ‘paralytic’) can be added up in order to calculate the actual frequency of each source domain.

7.1 A Top-down Approach to Source Domain Determination

As expanded upon in Chapter Four, a top-down approach to source domain determination uses the two knowledgebases of WordNet and SUMO to map metaphorical expressions to conceptual knowledge. Once these mappings have been achieved, steps will be taken to select the source domain names for all types of metaphorical expressions. Later sections in this chapter will explain how keywords in SUMO definitions may be employed in the naming of source domains. These steps are defined as top-down because (a) source domain names are determined at the outset, before the grouping of types of metaphorical expressions has been carried out; and (b) the determination of these source domain names is carried out at the upper conceptual level.

Figure 7.1 illustrates the top-down approach of this book (extracted from Figure 4.9 in Chapter Four).
This chapter will carry out all four steps mentioned in Figure 7.1. Lexical knowledgebases, such as SinicaBow and the Chinese-English Merged Word List, will be used (see Chapter Four for an in-depth explanation of these resources).

The types of metaphorical expressions collected from Single Target Domains and Coordinated Target Domains in CNA+XIN will be used in the determination of source domains. Figure 7.2 provides the overall number of types of metaphorical expressions present in CNA and XIN within Single+Coordinated Target Domains. This data has been combined for ease of computation. As used in Chapter Six, ‘overlapped’ data is defined as overlapped metaphorical expressions that occur when CNA and XIN are combined.

---

1 As mentioned in Chapter Six, combining Single+Coordinated Target Domains does not have linguistic significance because it collapses two kinds of data with different constructions. However, in this chapter, the combination is done for ease in programming.
Figure 7.2: Combined Types of Metaphorical Expressions in Single+Coordinated Target Domains for the Purpose of Computation

Figure 7.2 reveals that 1,282 types of metaphorical expressions have been taken from CNA and 736 from XIN, the combination of which has produced 1,591 types of metaphorical expressions. These types of metaphorical expressions have been collected from the target domains of 经济 jìngjì ‘economy,’ 政治 zhèngzhì ‘politics,’ 財政 cāizhèng ‘finance,’ 外交 wàijiāo ‘foreign affairs’ and 教育 jiàoyù ‘education.’ §7.2 will discuss the mappings of each type of metaphorical expression to WordNet and SUMO. §7.3 will then go over the concreteness measurement within SUMO, while §7.4 will elaborate upon how source domain names are decided. Finally, §7.5 will discuss how types of metaphorical expressions are grouped under these source domain names.

7.2 Mappings to WordNet and SUMO

All types of metaphorical expressions will be mapped to WordNet and SUMO through the interface provided by SinicaBow as well the bilingual dictionary by the Chinese-English Merged Word List. These mappings are a necessary step, as the conceptual information carried by the different types of metaphorical expressions must be determined by using them.

As demonstrated in Chapter Four, the mapping of a Chinese word to WordNet and SUMO will return a list of different senses, where each sense will have its respective WordNet definition in English, as well as its own mapping to a SUMO node (see Figure 4.4 in Chapter Four). Each type of metaphorical expression searched in SinicaBow and
the Chinese-English Merged Word List will, therefore, yield one or more WordNet sense and SUMO node.

### 7.2.1 Mappings through SinicaBow

For the mappings through SinicaBow, all types of metaphorical expressions are programmed so that their WordNet senses and SUMO nodes can be found automatically. SinicaBow contains a complete Chinese translation of the English WordNet. Programs are written to extract all search results based on the lists of types of metaphorical expressions found for CNA and XIN and for Single Target Domains and Coordinated Target Domains. Different senses of each type of metaphorical expression will be extracted through these mappings, and the results of these programs will then be tabulated accordingly. Table 7.1 shows examples of types of metaphorical expressions (成長 chéngzhǎng ‘grow/growth,’ 起飛 qǐfēi ‘take off’ and 癲癇 tānhuān ‘paralytic’) and their mappings to WordNet and SUMO.

In Table 7.1, types of metaphorical expressions are listed in the first column. Each type of metaphorical expression will have one or more WordNet sense, as seen in column three. The second column lists the number of WordNet IDs in the SinicaBow knowledgebase, and column four shows the English WordNet synsets (synonymous sets), from which the types of metaphorical expressions have been translated. The last column shows the corresponding SUMO node for each WordNet sense. A type of metaphorical expression like 成長 chéngzhǎng ‘grow/growth’ has six meanings. Other types of metaphorical expressions may have multiple meanings as well, depending on how many senses they have in SinicaBow. The senses may include metaphorical and literal senses as well as different parts of speech (‘V’ stands for ‘verb’ in column two, while ‘N’ stands for noun). These mappings provide crucial links to WordNet and SUMO, which in turn determine the conceptual knowledge represented in these two lexical knowledgebases.

---

2 Therefore, one can expect that certain Chinese-specific usage of terms may not be found in SinicaBow if these usages are not found in English.
### Table 7.1: Tabulated WordNet and SUMO Information for 成長 chéngzhǎng ‘grow/growth,’ 起飛 qǐfēi ‘take off’ and 癱瘓 tānhuàn ‘paralytic’ in SinicaBow³

<table>
<thead>
<tr>
<th>Types of Metaphorical Expressions</th>
<th>WordNet 1.7.1 IDs</th>
<th>WordNet Definitions</th>
<th>WordNet Synsets</th>
<th>SUMO Nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>成長 chéngzhǎng ‘grow/growth’</td>
<td>06123676N</td>
<td>a change resulting in an increase: “The increase is scheduled for next month.”</td>
<td>increase</td>
<td>Increasing</td>
</tr>
<tr>
<td></td>
<td>00124017V</td>
<td>become bigger or greater in amount: “The amount of work increased.”</td>
<td>increase</td>
<td>Increasing</td>
</tr>
<tr>
<td></td>
<td>11380033N</td>
<td>the process of an individual organism growing organically; a purely biological unfolding of events involved in an organism changing gradually from a simple to a more complex level: “He proposed an indicator of osseous development in children.”</td>
<td>growth growing maturation development ontogeny ontogenesis</td>
<td>Growth</td>
</tr>
<tr>
<td>起飛 qǐfēi ‘take off’</td>
<td>00197016V</td>
<td>grow old or older: “She aged gracefully.” “We age every day—what a depressing thought!”</td>
<td>age get_on mature maturate</td>
<td>Growth</td>
</tr>
<tr>
<td></td>
<td>00198841V</td>
<td>develop and reach maturity; undergo maturation: “He matured fast.” “The child grew fast.”</td>
<td>mature maturate grow</td>
<td>Growth</td>
</tr>
<tr>
<td></td>
<td>00435581V</td>
<td>grow emotionally or mature: “The child developed beautifully in her new kindergarten.” “When he spent a summer at camp, the boy grew noticeably and no longer showed some of his old adolescent behavior.”</td>
<td>develop grow</td>
<td>Intentional Psychologic al Process</td>
</tr>
<tr>
<td>癱瘓 tānhuàn ‘paralytic’</td>
<td>00234664N</td>
<td>a departure; especially of airplanes</td>
<td>takeoff</td>
<td>Motion</td>
</tr>
<tr>
<td></td>
<td>06138355N</td>
<td>the initial ascent of an airplane as it becomes airborne</td>
<td>takeoff</td>
<td>Motion</td>
</tr>
<tr>
<td></td>
<td>01584722V</td>
<td>depart from the ground, as of an aircraft or balloon: “The plane took off two hours late.”</td>
<td>take_off lift_off</td>
<td>Transportation</td>
</tr>
<tr>
<td></td>
<td>12088007N</td>
<td>a condition marked by uncontrollable tremor</td>
<td>palsy</td>
<td>Pathologic Process</td>
</tr>
<tr>
<td></td>
<td>12290943N</td>
<td>loss of the ability to move a body part</td>
<td>paralysis palsy</td>
<td>DiseaseOr Syndrome</td>
</tr>
</tbody>
</table>

³ There are also SUMO definitions, but they are not included in this table due to space limitations. Some WordNet synsets are written with an underscore, such as for ‘get_on’ and ‘take_off’ in Table 7.1. These synsets are taken directly from WordNet.
In order to evaluate the results of the mappings, as well as to improve upon them, Table 7.2 gives the results of the mappings through SinicaBow.

**Table 7.2:** Types of Metaphorical Expressions and Tokens of Metaphorical Expressions That Are Found in SinicaBow

<table>
<thead>
<tr>
<th></th>
<th>Found Types / Total Types</th>
<th>%</th>
<th>Found Tokens / Total Tokens</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNA</td>
<td>763 / 1,282</td>
<td>59.52</td>
<td>16,030 / 18,293</td>
<td>87.63</td>
</tr>
<tr>
<td>XIN</td>
<td>478 / 736</td>
<td>64.95</td>
<td>12,687 / 14,208</td>
<td>89.29</td>
</tr>
<tr>
<td>CNA+XIN</td>
<td>916 / 1,591</td>
<td>57.57</td>
<td>28,717 / 32,501</td>
<td>88.36</td>
</tr>
<tr>
<td>Overlapped</td>
<td>325 / 427</td>
<td>76.11</td>
<td>26,653 / 29,067</td>
<td>91.70</td>
</tr>
</tbody>
</table>

The results of the mappings are broken down into the four categories of overall CNA, XIN and CNA+XIN as well as overlapped types of metaphorical expressions. The percentages in Table 7.2 reveal successful returns of mappings from WordNet and SUMO in SinicaBow. Types of metaphorical expressions and tokens of metaphorical expressions have been presented separately in order to determine the overall frequency of the found types of metaphorical expressions. This table also shows the percentages of coverage of SinicaBow in terms of types of metaphorical expressions.4

Table 7.2 also reveals that most of the types of metaphorical expressions (‘Found_Types’ in the second and third columns) are found in more than 60 per cent in SinicaBow. Going down the percentage list of the third column, an increase from 59.52 per cent to 76.11 per cent can be seen, indicating that within XIN, there are more percentages of types of metaphorical expressions to be found in SinicaBow than in CNA. The highest percentage of ‘Found_Types’ comes from the overlapped types, which constitutes 76.11 per cent of the total overlapped types, indicating that the greatest number of overlapped types of metaphorical expressions are found in SinicaBow. These results can be easily explained because, as seen in Chapter Six, overlapped types of metaphorical expressions are most likely high frequency words. It is also possible that most lexical knowledgebases store more words with a higher occurrence rather than those with a lower frequency of occurrence.

With regard to XIN, in Chapter Six most of the metaphorical expressions in this knowledgebase were discovered to be more centralized than those found in CNA, as each type of metaphorical expression consists of more tokens of metaphorical expressions in XIN than in CNA (i.e., with higher tokens-per-type ratios; see analyses in Chapter Six). It is also possible that the higher percentage of ‘Found_Type’ for XIN is due to the fact that XIN utilizes more high frequency words (because they are repeated more times)

---

4 Thus, it is also an evaluation of the performance of SinicaBow when large queries of Chinese terms are carried out.
than CNA does, meaning that the oft repeated types of metaphorical expressions in XIN are high frequency Chinese words also stored in SinicaBow.5

When compared to their respective tokens of metaphorical expressions (the fourth and fifth columns), the successfully returned cases are shown to have higher percentages. While the types of metaphorical expressions found in SinicaBow constitute more than 60 per cent of the total types, their tokens of metaphorical expressions constitute more than 87 per cent (column four) of the total tokens. This means that the remaining types of metaphorical expressions, numbering less than 40 per cent, which are not found, are probably low frequency words because their corresponding tokens of metaphorical expressions that are not found only constitute 13 per cent of the total tokens. In other words, the types of metaphorical expressions in CNA and XIN that are not found have an average of only 13 per cent of the total tokens of metaphorical expressions.

This chapter also runs a comparison of SinicaBow with the Chinese-English Merged Word List (hereafter ‘the Merged Word List’). All types of metaphorical expressions, including those that have been found in SinicaBow, will be programmed in a similar way using the Merged Word List. This is because the number of found types in SinicaBow is low (only 60 per cent), as the collection of Chinese words in SinicaBow depends on the existence of English words in the English WordNet. However, the use of the Merged Word List may induce noise into the results because, as introduced in Chapter Four, this knowledgebase is a collection of Chinese-English bilingual dictionaries that has not been checked by human examiners, as was carried out for SinicaBow. Therefore, the senses in SinicaBow are controlled. The addition of the Merged Word List will provide a way to compare these two lexical knowledgebases. The following section will provide the steps used in the mapping process that occurs when using the Merged Word List.

### 7.2.2 Mappings through the Chinese-English Merged Word List

The Merged Word List is used when a type of metaphorical expression cannot be found in SinicaBow. The procedure for the mappings using the Merged Word List is the same as that which is used for SinicaBow. When a Chinese type of metaphorical expression is searched for in the Merged Word List automatically, all English translations for this expression will be extracted. These English translations will then be mapped to both WordNet and SUMO. In Table 7.3, the results for 成長 chéngzhǎng ‘grow/growth’ when searched in the Merged Word List are shown.

---

5 Breakdowns of Single and Coordinated in respective CNA, XIN and CNA+XIN have the same patterns as the results in the overall data. Therefore, their percentages of cases that are not found will not be listed in this book.
Table 7.3: Tabulated WordNet and SUMO Information from 成長 chéngzhāng ‘grow/growth’ in the Chinese-English Merged Word List

<table>
<thead>
<tr>
<th>Type of Metaphorical Expression</th>
<th>WordNet Definitions</th>
<th>SUMO Nodes</th>
<th>Upper SUMO Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>成長 chéngzhāng ‘grow/growth’</td>
<td>become larger, greater, or bigger; expand or gain</td>
<td>Increasing Process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>come to have or undergo a change of (physical features and attributes)</td>
<td>Growth     Process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>be gradually disclosed or unfolded; become manifest</td>
<td>Process     Process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>expand in the form of a series, in mathematics</td>
<td>Content Development Process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>make visible by means of chemical solutions; of photographic film</td>
<td>Content Development Process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>act of improving by expanding or enlarging or refining</td>
<td>Creation Process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>vegetation that has grown</td>
<td>Plant Object</td>
<td></td>
</tr>
<tr>
<td></td>
<td>grow, progress, unfold or evolve through a process of evolution, natural growth,</td>
<td>Increasing Process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>differentiation or a conducive environment</td>
<td>Growth      Process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>increase in size by natural process; of living matter, such as plants and animals</td>
<td>Growth      Process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(pathology) an abnormal proliferation of tissue (as in a tumor)</td>
<td>Anatomical Structure Object</td>
<td></td>
</tr>
<tr>
<td></td>
<td>cause to grow or develop</td>
<td>Growth      Process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>the process of an individual organism growing organically; a purely biological</td>
<td>Growth      Process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>unfolding of events involved in an organism changing gradually from a simple to a</td>
<td>Growth      Process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>more complex level</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>elaborate by the unfolding of a musical idea and by the working out of the rhythmic</td>
<td>Music       Process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>and harmonic changes in the theme</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>something grown or growing</td>
<td>Growth      Process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>cultivate by growing; often involves improvements by means of agricultural</td>
<td>Process     Process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>techniques</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>develop and reach maturity; undergo maturation</td>
<td>Growth      Process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>become attached by or as if by the process of growth</td>
<td>Growth      Process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>the act of moving forward toward a goal</td>
<td>Motion      Process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>become an adult</td>
<td>Growth      Process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>change the use of and make available or usable</td>
<td>Intentional Process</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In Table 7.3, there are altogether thirty-six senses found for 成長 ‘grow/growth.’ When compared to Table 7.2, which has only six senses found for the same type of metaphorical expression in SinicaBow, the number of senses found using the Merged Word List seem to be too numerous. This hits upon the weakness with the Merged Word List, but this weakness will be overcome in the following section when only one sense is selected from all senses based on their concreteness scores.

With the Merged Word List, the similar types of metaphorical expressions and tokens of metaphorical expressions that were searched for in SinicaBow are programmed using this knowledgebase. The returns of these searches will be discussed at two levels. First, we compare the number of found cases for the types of metaphorical expressions found in SinicaBow (to see whether the same types of metaphorical expressions can be found). The results show that about 96 to 99 per cent of all types and tokens of metaphorical expressions that are found in SinicaBow are also found in the Merged

|breed freely and abundantly| Sexual Reproduction| Process|
|become abundant; increase rapidly| Increasing| Process|
a movement forward| Motion| Process|
natural height of a person or animal in an upright position| Length Measure| Quantity|
grow and flourish| Increasing| Process|
acquire or build up traits or characteristics| Intentional Psychological Process| Process|
grow emotionally or mature| Intentional Psychological Process| Process|
develop in a positive way| Process| Process|
move one’s chess pieces into strategically more advantageous positions| Game| Process|
processing a photosensitive material in order to make an image visible| Surface Change| Process|
cause to grow and differentiate in ways conforming to its natural development| Growth| Process|
generate gradually| Intentional Process| Process|
*(definition not available for this sense)*| Making| Process|
a district that has been developed to serve some purpose| Land Area| Object|
superimpose a three-dimensional surface on a plane without stretching, in geometry| Process| Process|
move into a strategically more advantageous position, of a chess piece| Game| Process|
Word List. Second, we provide the found cases for additional types of metaphorical expressions that have not previously been found using SinicaBow. The overall information of the found cases and the added percentages found in the Merged Word List only (columns four and seven) are shown in Table 7.4.

<table>
<thead>
<tr>
<th>Found_Types / Total_Types</th>
<th>%</th>
<th>Added %</th>
<th>Found_Tokens / Total_Tokens</th>
<th>%</th>
<th>Added %</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNA</td>
<td>1,003 / 1,282</td>
<td>78.24</td>
<td>17,405 / 18,293</td>
<td>95.15</td>
<td>7.52</td>
</tr>
<tr>
<td>XIN</td>
<td>605 / 736</td>
<td>82.20</td>
<td>13,801 / 14,208</td>
<td>97.14</td>
<td>7.84</td>
</tr>
<tr>
<td>CNA+XIN</td>
<td>1,214 / 1,591</td>
<td>76.30</td>
<td>31,206 / 32,501</td>
<td>96.02</td>
<td>7.66</td>
</tr>
<tr>
<td>Overlapped</td>
<td>393 / 427</td>
<td>92.04</td>
<td>28,534 / 29,067</td>
<td>98.17</td>
<td>6.48</td>
</tr>
</tbody>
</table>

When compared to SinicaBow, it can be seen that the Merged Word List is approximately 18 per cent larger (column four) in terms of types of metaphorical expressions and about 8 per cent larger in terms of tokens of metaphorical expressions (column seven). The overall performance in Table 7.4 shows that more than 76 per cent (column three) of the total types of metaphorical expressions can be found in the Merged Word List. These types constitute more than 96 per cent of the total tokens of metaphorical expressions (column six). Therefore, the not found cases number less than 24 per cent of the total types of metaphorical expressions, with only less than 4 per cent of tokens of metaphorical expressions not found. As these remaining 24 per cent types of metaphorical expressions that are not found constitute only 4 per cent of the total remaining types of metaphorical expressions that are not found, this is a reconfirmation of the aforementioned statement that not-found expressions are low frequency words, as their tokens are not often numerous in number.

In SinicaBow, there are in total 4,671 senses for 916 types of metaphorical expressions found in SinicaBow (five senses per type of metaphorical expression on average). For a similar 916 types of metaphorical expressions, the Merged Word List returned 34,617 senses (thirty-eight senses per type of metaphorical expression on average), representing a huge difference, with the additional 18 per cent of types of metaphorical expressions that are only found in the Merged Word List being comprised of 5,423 senses replete with 298 types of metaphorical expressions (eighteen senses per type of metaphorical expression on average). The lower average in the additional types

---

6 The performance of the Merged Word List is given in Appendix A7.1. This shows that almost all the metaphorical expressions that are found in SinicaBow can also be found in the Merged Word List, indicating a similar performance with regard to the same group of words.
of metaphorical expressions may be due to the fact that these expressions (which were not collected by SinicaBow) are low-frequency expressions.

§7.2.3 will discuss why some types of metaphorical expressions are not found in either SinicaBow or the Chinese-English Merged Word List.

7.2.3 Reasons for Types of Metaphorical Expressions That Are Not Found

All types of metaphorical expressions that are not found cannot be analyzed in terms of source domains because they lack information necessary for their processing using top-down approach as elaborated upon in this chapter (less than 25 per cent).

Usually, types of metaphorical expressions cannot be found in either SinicaBow or the Merged Word List for two main reasons (other than incomplete information in the two knowledgebases, which is not possible here because both knowledgebases should be sufficiently large). The first reason is that some words have not been collected in either SinicaBow or the Merged Word List (which comes from bilingual dictionaries) because these words cannot be translated. An example of types of metaphorical expressions in Chinese that do not have direct translation is shown in (1).

(1) 當前 世界 政治 風雲 多 變 (XIN)

dāngqián shìjiè zhèngzhì fēngyún duō biàn
currently world politics wind-cloud much change

‘Currently, global politics is changing rapidly.’

In this example, 風雲 fēngyún ‘wind-cloud’ does not have a direct translation of ‘wind and cloud’ in English. Rather, this word is used to describe ‘some dramatic phenomena, rife with unexpected ups and downs.’ With no direct English translation available, this word is unlikely to be included in SinicaBow, as SinicaBow is a direct translation of the English WordNet. Likewise, it is also not easy to translate a word like this into English in the Merged Word List. Therefore, while words like this are not to be found in either knowledgebases, we must remember that the tokens of metaphorical expressions not found constitute only 4 per cent of the total tokens.

The second reason why some words are not found is that these words are idiomatic expressions, and not all idiomatic expressions are included in bilingual dictionaries. It is also not possible to find many idiomatic expressions in SinicaBow because of the prevalence and influence of English in WordNet. An example of an idiomatic expression not found in both knowledgebases is shown in (2).
Even though some idioms in Chinese can be matched to an English idiom roughly equivalent in meaning (in this case, ‘to add insult to injury’), their meanings will not match exactly. Furthermore, it is unusual to define an idiom with another idiom in lexicography. Therefore, types of metaphorical expressions such as these cannot be found in either knowledgebase.

For each of the types of metaphorical expressions that have been found, there will be one or more WordNet senses, depending on their mappings of meanings in WordNet. In order to find a source domain name for all of the senses for a particular type of metaphorical expression, this chapter will first measure the conceptual distance between the SUMO nodes of these senses with several prototypical concrete nodes that have been selected by human subjects, for the reason that the keywords in the definitions for these concrete nodes will be used as the source domain names for the different types of metaphorical expressions. This chapter will, therefore, concentrate on the selection of concrete nodes.

7.3 Concreteness of Source Domains

The Conceptual Metaphor Theory (Lakoff & Johnson 1980, Lakoff 1993) proposes that source domains are usually more concrete than target domains. According to this view, a source domain mapping in a metaphorical expression should have two meanings: an abstract meaning from the target domain and a concrete meaning from the source domain. A similar observation has been made by Knowles & Moon (2006:15) when they suggest that the core meaning or the “basic meaning” usually “refers to something concrete or physical, from which have developed further senses which are often metaphorical.” If this is the case, the different senses of a type of metaphorical expression may, in fact, be comprised of both concrete and abstract meanings. For example, a metaphorical expression such as 成長 chéngzhǎng ‘grow/growth’ has the following meanings (see Table 7.5). The last column marks whether each sense is literal or metaphorical as judged by intuition (literal senses are in bold italics). A question mark is placed for the last sense because of the uncertainty involved in determining whether or not this sense is literal or metaphorical.
The last column in Table 7.5 has been manually marked based on the definitions of either WordNet definitions (second column) or SUMO definitions (fifth column). From the senses displayed in Table 7.5, those marked ‘literal’ are possible concrete senses that can provide helpful clues when deciding the source domain best suited for 成長 chéngzhǎng ‘grow/growth.’ The last sense cannot be decisively determined, falling between a literal and a metaphorical meaning because this sense is related to ‘emotion’ or ‘reasoning’ in the brain, neither of which refer to actual concept of growing in size, as the literal meaning of ‘Growth’ does. The Conceptual Metaphor Theory does not provide clear criteria for how concreteness in source domains should be evaluated. If human judgment is used in determining source domains, it is therefore inevitable that indecisive cases, such as ‘IntentionalPsychologicalProcess’ (last sense), will always be found. Based on this problem, certain parameters are needed in order to select the literal, as opposed to the metaphorical, meaning. Worthy of note is the fact that all literal senses in Table 7.5 come from the same SUMO node of ‘Growth.’ This means that concreteness in SUMO nodes may be used as a reference when determining concrete concepts. Thus, the concreteness of SUMO nodes is also measured during this step because no previous studies have tried to verify Lakoff’s (1993) assertion that source domains are more concrete than target domains.

---

7 At this stage, we do not know what the source domain is. All we know at this point is which sense is literal and which is metaphorical (through manual analysis).
Table 7.5: Metaphorical and Literal Meanings of 成長 chéngzhǎng ‘grow/growth’

<table>
<thead>
<tr>
<th>Sense</th>
<th>WordNet Definitions</th>
<th>WordNet Synsets</th>
<th>SUMO Nodes</th>
<th>SUMO Definitions</th>
<th>Literal vs. Metaphorical Meanings</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>a change resulting in an increase: “The increase is scheduled for next month.”</td>
<td>increase</td>
<td>Increasing</td>
<td>Any QuantityChange where the PhysicalQuantity is increased.</td>
<td>Metaphorical</td>
</tr>
<tr>
<td>(b)</td>
<td>become bigger or greater in amount: “The amount of work increased.”</td>
<td>increase</td>
<td>Increasing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(c)</td>
<td>the process of an individual organism growing organically; a purely biological unfolding of events involved in an organism changing gradually from a simple to a more complex level: “He proposed an indicator of osseous development in children.”</td>
<td>growth, growing, maturation, development, ontogeny, ontogenesis</td>
<td>Growth</td>
<td>The Process of biological development in which an Organism or part of an Organism changes its form or its size.</td>
<td>Literal</td>
</tr>
<tr>
<td>(d)</td>
<td>grow old or older: “She aged gracefully.” “We age every day—what a depressing thought!”</td>
<td>age, get_on, mature, matureate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(e)</td>
<td>develop and reach maturity; undergo maturation: “He matured fast.” “The child grew fast.”</td>
<td>mature, matureate, grow</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(f)</td>
<td>grow emotionally or mature: “The child developed beautifully in her new kindergarten.” “When he spent a summer at camp, the boy grew noticeably and no longer showed some of his old adolescent behavior.”</td>
<td>develop, grow</td>
<td></td>
<td>Intentional Psychological Process</td>
<td></td>
</tr>
</tbody>
</table>

| Note: | The identification of ‘literal’ or ‘metaphorical’ meanings in Table 7.5 were based on the author’s intuition, and the next step will show how intuition like this can be computed using an ontology.
| Note: | Notice that the English sense of ‘grow old’ does not necessarily fit the use of 成長 chéngzhǎng ‘grow/growth’ in Chinese, where 成長 chéngzhǎng ‘grow/growth’ does not refer to the aging process in Chinese (it refers more to a child’s growing process). This is recognized as the English-Chinese translation problem of SinicaBow. This problem of language transfer is a limitation of SinicaBow, as it is not a full resource for Chinese. Nevertheless, the problems as such should not affect the results extensively if the majority of the senses are correct. |
For the previously stated reasons, and because each type of metaphorical expression such as 成長 chéngzhǎng ‘grow/growth’ yields one sense or more when mapped to WordNet and SUMO, the following sub-section will compare the concreteness of different senses for each type of metaphorical expression. A comparison of senses is achieved by measuring the distance between the SUMO node belonging to each sense and some prototypical concrete SUMO nodes selected by human subjects, illustrated in the next section.

### 7.3.1 Methodology of Selecting Concrete Senses through Measuring Concreteness of SUMO Nodes

As mentioned in the previous section, mappings to WordNet and SUMO will return several WordNet senses. The previous section also suggests that among these senses, there will be abstract senses (such as sense (a) of 成長 chéngzhǎng ‘grow/growth,’ which is “a change resulting in an increase”) as well as concrete senses (such as sense (d) of 成長 chéngzhǎng ‘grow/growth,’ which has the definition of “grow old or older”). In order to select concrete senses that may help to provide information regarding source domains, this section proposes a method of measuring the conceptual distance between each SUMO node in every sense and several prototypical nodes. This process is carried out so as to ensure that the concrete sense with the highest concreteness score will be selected for source domain analyses.

In order to select one concrete sense from a list of senses within each type of metaphorical expression, such as in Table 7.5, this book proposes the comparison of the corresponding SUMO node for each of these senses (fourth column in Table 7.5) with several prototypical concrete SUMO nodes. Through comparison, the hierarchical distance between each sense with the prototypical concrete nodes can be measured. The closer the distance, the more concrete a SUMO node is. This section will therefore provide relative scores of hierarchical distance (called conceptual distance) between each node to the prototypical concrete nodes. The higher the score of a SUMO node of each sense, the more similar this node is to the prototypical concrete nodes (called conceptual similarity) and, thus, the more concrete it is.

#### 7.3.1.1 Measuring Conceptual Distance

The calculation of the distance requires the modification of the original algorithm supplied by Wu & Palmer (1994:136), which gives the following explanation regarding conceptual similarity based on Figure 7.3.
In Figure 7.3, there are four nodes—Root, C1, C2 and C3. These nodes have distances marked with $N1$, $N2$ and $N3$, respectively (which may comprise several nodes in one measure of distance). The algorithm for the calculation of concept similarity (abbreviated as ‘ConSim’ in Wu & Palmer 1994:136) is given in (3).

\[
(3) \quad ConSim(C1,C2) = \frac{2 \cdot N3}{N1 + N2 + 2 \cdot N3}
\]

This formula is then further modified by Budanitsky & Hirst (2006:19), who added a step after (4a) in order to find the distance (instead of conceptual similarity) between C1 and C2, as shown in (4b).

\[
\begin{align*}
(4) \quad a. \quad ConSim(C1,C2) &= \frac{2 \cdot N3}{N1 + N2 + 2 \cdot N3} \\
\quad b. \quad Distance(C1,C2) &= 1 - ConSim \ (C1, C 2)
\end{align*}
\]

The formula used in this chapter will follow the one seen in (4). However, since both formulas in (3) and (4) first require the calculation of conceptual similarity before the calculation of conceptual distance can take place, the reverse information is required

---

10 Formula shown in (4) has been modified in this book to relate to example (3) more clearly.
11 This method of calculation was suggested by Laurent Prévot (personal communication). All computer programs in this chapter were written by Ming-Wei Hsu.
because we have our own initial definition of distance. Our definition of distance is a simple calculation of the number of nodes to the root of ‘Entity.’ For example, in Figure 7.4, the distance between C1 and C3 (D1) is three nodes (see arrows in Figure 7.4), while the distance between C2 and C3 (D2) is four.

![Figure 7.4: Measuring Conceptual Similarity between SUMO Nodes](image)

The distance between C1 and C2 is the sum of D1 and D2. This sum is further divided by the D3, which is the distance to the root of the hierarchy (i.e., 2 nodes in Figure 7.4). For D3, an ‘exponential’ value has been added because it transforms the number of D3 to a smaller value, and this will increase the sum of D1 and D2 in the calculation process. The formulas are given in (5). (5a) shows the calculation of ‘conceptual distance,’ which is based on our own definition of ‘distance.’ (5b) shows the calculation of ‘conceptual similarity,’ which is a transformation of the formula in (4b).12

---

12 ‘Exponential’ is added in order to avoid cases where D1 and D2 are short, indicating that C1 and C2 appear at nodes closer to the root. When this happens, calculation without ‘exponential’ will return a high conceptual similarity score that will create an artificially higher value, which is, in fact, incorrect.

(Exponential function \( x = b^n \); \( \log_b(x) = n \))

\[ \log_{10}(x) = \log_{10}(x) \]
A Corpus-driven Approach to Source Domain Determination

\[ \text{ConceptualDistance}(C_1, C_2) = \frac{D_1 + D_2}{\text{Exp}(D_3)} \]

\[ \text{ConceptualSimilarity}(C_1, C_2) = \frac{1}{1 + \text{Distance}} \]

The previous formula of Budanitsky & Hirst (2006:19) in (4) first determines conceptual similarity; only then is distance calculated. In our formula in (5), conceptual distance is first calculated in (5a), and only after this step has been completed will the conceptual similarity be calculated in (5b).¹³

The purpose of comparing conceptual similarity between different nodes, it must be stressed, is to select one concrete sense from among the senses for each type of metaphorical expression. However, in order to determine which sense among these various senses is concrete, several prototypical concrete nodes are required so that these nodes from the types of metaphorical expressions can be compared to them. We need to be able to identify several nodes that are concrete in the ontology so that comparisons of distance to these concrete nodes can be carried out in order to calculate conceptual similarity. In the following section, we also make sure that these prototypical concrete nodes are scattered throughout different places in the ‘tree,’ so that the distance between the nodes they are being compared to do not increase unnecessarily.

¹³ An example of the program is shown following (explanations are given after the # symbol). The boxed node (‘InternalChange’) is the node at C3, where the paths of D1 and D2 meet. For C1 (‘Growth’), it is the fourth node after ‘Process’ (thus, D1=4). C2 (‘Cooking’) is one of the prototypical concrete nodes selected through human ratings. It is the third node after the process (thus, D2=3).

\[ D_1: \text{Entity Physical Process } \textbf{InternalChange (C3)} \text{ BiologicalProcess PhysiologicProcess AutonomicProcess } \textbf{Growth (C1)} \]

\[ D_2: \text{Entity Physical Process } \textbf{InternalChange (C3)} \text{ Creation Making } \textbf{Cooking (C2)} \]

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1:</td>
<td>5</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>D2:</td>
<td>5</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Conceptual Distance: (5+3)/(7.389056)=1.0826823
Conceptual Similarity: 1/(1+1.0826823)=0.48015007
Average = (0.74155945 + 0.48015007)/2 = 0.61085474

If one of the nodes (e.g., ‘Cooking’) has multiple paths, the average for the different paths will first be found.

\[ D_1: \text{Entity Physical } \textbf{Process (C3)} \text{ InternalChange BiologicalProcess PhysiologicProcess AutonomicProcess } \textbf{Growth (C1)} \]

\[ D_2: \text{Entity Physical } \textbf{Process (C3)} \text{ IntentionalProcess Making } \textbf{Cooking (C2)} \]

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1:</td>
<td>5</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>D2:</td>
<td>5</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Conceptual Distance: (5+3)/(7.389056)=1.0826823
Conceptual Similarity: 1/(1+1.0826823)=0.48015007
Average = (0.74155945 + 0.48015007)/2 = 0.61085474
The selection of the prototypical concrete nodes is based on subjects’ ratings of the concreteness of the SUMO nodes, illustrated in the next section.

**7.3.1.2 Human Rating of Concreteness for SUMO Nodes**

The prototypical concrete nodes are comprised of several SUMO nodes with the highest concreteness ratings, as defined by human subjects. A rating experiment was undertaken in which subjects were asked to rate all SUMO nodes according to a scale of concreteness measuring 1 to 7, with 1 being least concrete and 7 being most concrete. The questionnaire is included in Appendix A7.2.

A total of 626 SUMO nodes were provided in a single questionnaire to subjects online. The rating task was completed by subjects at their leisure, and the questionnaires were returned through electronic mail. Instructions were given in a separate file, which also contained a form asking for a self-reported personal background. A total of ten subjects (two males and eight females) took part in this study, all of whom possessed a self-rated English and Chinese proficiency of 4 or above on a scale of 1 to 7, with 1 being least proficient and 7 being most proficient. These subjects have a mean age of 28.90, with participants ranging in age from 23 to 34. These subjects are above university level in terms of their formal education. Participants were asked to rate all 626 nodes in the questionnaire.

The 626 SUMO nodes that comprised the body of the questionnaire can be divided into several ontological levels. For example, under the root of ‘Entity’ in SUMO exist the upper levels of ‘Physical,’ and ‘Abstract’ (see Figure 7.5). In order to ensure that there are SUMO nodes from different upper levels, the selection of the prototypical concrete SUMO nodes from the rating experiment was based on upper levels, such as ‘Physical’ and ‘Abstract.’ From Figure 7.5, the upper level of ‘Physical’ can be seen to consist of ‘Object’ and ‘Process,’ with these two major nodes having four children nodes, respectively. On the other hand, ‘Abstract’ has only five children nodes under it.

---

14 The questionnaire includes all 626 SUMO nodes (excluding functions and relations). Due to the long questionnaire, the number of subjects is reduced to 10 in this study.

15 It is also because the measurement of distance will increase when measuring distance between two branches, such as between ‘Agent’ and ‘Collection’ (under ‘Object’ in Figure 7.5). A more detailed discussion of this issue can be seen later in this section.

16 If a node under ‘Abstract’ is rated concrete, this node will also be included in the measurement of similarity distance.
Due to the fact that there are more nodes in the ‘Physical’ branch, both nodes under ‘Object’ and ‘Process’ will be considered separately for the selection of most prototypical nodes underneath these parent nodes. ‘Abstract,’ on the other hand, does not need to be separated into smaller units because the nodes in ‘Abstract’ number fewer than those in ‘Physical.’

Another reason why ‘Object’ and ‘Process’ are considered separately is because most of the SUMO nodes found in the senses of the types of metaphorical expressions (in the mappings to SinicaBow and the Merged Word List) also fall under these two parent nodes. In contrast, the smaller ‘Abstract’ is considered to be a single unit, without further division into different levels. Ten prototypical concrete nodes were selected from ‘Abstract,’ ten other nodes were selected for ‘Object’ and an additional ten were selected for ‘Process.’ These selected nodes are, on average, high in terms of

---

17 From the mappings to SinicaBow and the Merged Word List, there are 943 (13.75 per cent) in ‘Object,’ 4,858 (70.85 per cent) in ‘Process’ and 1,056 (15.40 per cent) in ‘Abstract.’
their concreteness ratings.\textsuperscript{18} From all of the 626 SUMO nodes, 30 prototypical concrete nodes were selected based on their ratings by the 10 subjects.\textsuperscript{19}

The selected prototypical concrete nodes fall under three main parent nodes, shown in Table 7.6. We require these three different main parent nodes because our formula in (5) takes into consideration the distance between two compared nodes as well as their parent nodes (i.e., the root). If we select prototypical concrete nodes from only one parent node, when compared with nodes in other parent nodes, the distance will have been unnecessarily increased due to the shared father nodes that are far away (see Figure 7.4).

These three main parents nodes for the selected SUMO nodes in Table 7.6 have mean ratings that do not differ from one another: $F(2,297)=0.641$, $p=.527$ (mean for ‘Object’ is 4.96; mean for ‘Process’ is 5.15; and mean for ‘Abstract’ is 4.93). For the three major parent nodes of ‘Object,’ ‘Process’ and ‘Abstract’ (column one), their prototypical concrete nodes (column two) are varied, so these selected prototypical nodes do not come solely from a single node (column three).

\begin{table}[h]
\centering
\begin{tabular}{|l|c|}
\hline
\textbf{Prototypical Concrete SUMO Nodes} & \textbf{Mean Ratings} \\
\hline
RelationalAttribute & 2.00 \\
IntentionalProcess & 2.56 \\
PsychologicalProcess & 3.22 \\
Destruction & 4.00 \\
TransportationDevice & 5.10 \\
Fish & 6.90 \\
\hline
\end{tabular}
\caption{Prototypical Concrete SUMO Nodes and Mean Ratings}
\end{table}

From the results above, we can see that some SUMO nodes were considered less concrete (such as ‘RelationalAttribute’ with an average rating of 2.00) as their ratings were closer to 1 on a scale of 1 to 7. Comparatively, some SUMO nodes were rated as highly concrete (such as ‘Fish’ with an average rating of 6.90). For the prototypical concrete nodes, we need to select nodes that are rated highly concrete by subjects (those rated nearly 7, such as ‘Fish’).
### Table 7.6: Selected Prototypical Concrete Nodes and Their Mean Ratings

<table>
<thead>
<tr>
<th>Major Parent Nodes</th>
<th>Prototypical Concrete SUMO Nodes</th>
<th>Immediate Parent Nodes of Prototypical Concrete Nodes (Multi-paths are Separated by '/')</th>
<th>Mean Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Object’ (Mean=5.15)</td>
<td>Building SelfConnectedObject</td>
<td></td>
<td>5.80</td>
</tr>
<tr>
<td></td>
<td>BodyPart SelfConnectedObject</td>
<td></td>
<td>5.70</td>
</tr>
<tr>
<td></td>
<td>Canine SelfConnectedObject/Agent</td>
<td></td>
<td>5.40</td>
</tr>
<tr>
<td></td>
<td>EducationalOrganization Collection/Agent</td>
<td></td>
<td>5.00</td>
</tr>
<tr>
<td></td>
<td>Government Collection/Agent</td>
<td></td>
<td>4.70</td>
</tr>
<tr>
<td></td>
<td>Organization Collection/Agent</td>
<td></td>
<td>5.20</td>
</tr>
<tr>
<td></td>
<td>FamilyGroup Collection/Agent</td>
<td></td>
<td>4.90</td>
</tr>
<tr>
<td></td>
<td>LandArea Region</td>
<td></td>
<td>5.00</td>
</tr>
<tr>
<td></td>
<td>SaltWaterArea Region</td>
<td></td>
<td>4.90</td>
</tr>
<tr>
<td></td>
<td>GeographicArea Region</td>
<td></td>
<td>4.90</td>
</tr>
<tr>
<td>‘Process’ (Mean=4.98)</td>
<td>Surgery IntentionalProcess</td>
<td></td>
<td>5.40</td>
</tr>
<tr>
<td></td>
<td>Gesture IntentionalProcess/Motion</td>
<td></td>
<td>5.20</td>
</tr>
<tr>
<td></td>
<td>Speaking IntentionalProcess/Motion</td>
<td></td>
<td>4.90</td>
</tr>
<tr>
<td></td>
<td>Selling IntentionalProcess/DualObjectProcess</td>
<td></td>
<td>4.50</td>
</tr>
<tr>
<td></td>
<td>Cooking InternalChange/IntentionalProcess</td>
<td></td>
<td>5.00</td>
</tr>
<tr>
<td></td>
<td>Breathing InternalChange/InternalChange</td>
<td></td>
<td>5.10</td>
</tr>
<tr>
<td></td>
<td>RadiatingLight Motion</td>
<td></td>
<td>4.70</td>
</tr>
<tr>
<td></td>
<td>Shooting Motion</td>
<td></td>
<td>5.33</td>
</tr>
<tr>
<td></td>
<td>Covering Motion</td>
<td></td>
<td>4.90</td>
</tr>
<tr>
<td></td>
<td>Walking Motion</td>
<td></td>
<td>4.80</td>
</tr>
<tr>
<td>‘Abstract’ (Mean=4.96)</td>
<td>PrimaryColor InternalAttribute</td>
<td></td>
<td>4.50</td>
</tr>
<tr>
<td></td>
<td>ColorAttribute InternalAttribute</td>
<td></td>
<td>4.00</td>
</tr>
<tr>
<td></td>
<td>DiseaseOrSyndrome InternalAttribute</td>
<td></td>
<td>3.90</td>
</tr>
<tr>
<td></td>
<td>Plan InternalAttribute</td>
<td></td>
<td>5.00</td>
</tr>
<tr>
<td></td>
<td>ComputerProgram Procedure</td>
<td></td>
<td>4.90</td>
</tr>
<tr>
<td></td>
<td>CelsiusDegree Procedure</td>
<td></td>
<td>5.10</td>
</tr>
<tr>
<td></td>
<td>EuroCent PhysicalQuantity</td>
<td></td>
<td>5.60</td>
</tr>
<tr>
<td></td>
<td>EuroDollar PhysicalQuantity</td>
<td></td>
<td>5.60</td>
</tr>
<tr>
<td></td>
<td>Centimeter PhysicalQuantity</td>
<td></td>
<td>4.90</td>
</tr>
<tr>
<td></td>
<td>UnitedStatesCent PhysicalQuantity</td>
<td></td>
<td>6.10</td>
</tr>
</tbody>
</table>

For example, under ‘Object’ in Table 7.6, there are prototypical nodes that come from ‘SelfConnectedObject,’ ‘Agent,’ ‘Collection’ and ‘Region’ (‘Agent’ is in the multi-path after the symbol ‘/’). The reason for the variation among these nodes is to avoid an unnecessary increase in distance when a SUMO node is compared to another node in
Chapter 7: Top-down Approach to Source Domain Determination

another branch.\textsuperscript{20} The nodes selected in Table 7.6, therefore, also cover most of the nodes under ‘Object,’ ‘Process’ and ‘Abstract.’\textsuperscript{21}

In Table 7.6, ‘Building’ and ‘BodyPart’ have the immediate parent node of ‘SelfConnectedObject.’ ‘Educational Organization’ has two immediate parent nodes (‘Collection’ and ‘Agent’), which means that ‘Educational Organization’ has two paths (called ‘multi-paths’ in Table 7.6). For prototypical concrete nodes that have more than one path, an average score will be calculated based on these different paths (cf. Footnote 13).

Each corresponding SUMO node for each WordNet sense (as in previous Table 7.5) will be compared to their respective upper categories of ‘Object,’ ‘Process’ and ‘Attribute.’ For example, sense (d) in Table 7.5 for \textit{成長 chēngzhǎng} ‘grow/growth’ has the corresponding SUMO node of ‘Growth.’ This SUMO node falls under the parent node of ‘Process.’ Therefore, ‘Growth’ will be compared to all prototypical concrete nodes under the parent nodes of ‘Process,’ which are ‘Surgery,’ ‘Gesture,’ ‘Speaking,’ ‘Selling,’ etc. After ‘Growth’ has been compared to each of the parent nodes of ‘Process,’ a score will be calculated for each comparison, whereby each score will be obtained by comparing ‘Growth’ and ‘Surgery,’ ‘Growth’ and ‘Gesture’ and so on. ‘Growth’ will, therefore, have ten scores of conceptual similarity from ten prototypical concrete nodes. Among these ten scores, only the highest will be selected, and this will form the highest score for ‘Growth’ in terms of its concreteness. The higher this number is, the higher the concreteness is. As mentioned, if a prototypical concrete node has multi-paths, the average for the multi-paths will first be computed before finding out the highest score from among the averaged multi-paths.

\textsuperscript{20} Since the formulas in (5) calculate the distance between $C_1$, $C_2$ and $C_3$, if $C_2$ appears in a different branch than $C_1$, the distance will increase because their $C_3$ will be higher. This is also because comparing across ‘branches’ will bring down the concreteness scores, as these scores measure the similarity distance between each node and the prototypical nodes. The closer they are at a particular place of the hierarchy, the more likely they have similar concrete meanings. However, when calculating across ‘branches,’ the distance will increase and the concreteness scores will drop. Therefore, it is fairer to compare within each ‘branch’ in order to reduce the problems caused by technical flaws.

\textsuperscript{21} All nodes are covered except ‘Relation’ and ‘SetorClass’ (under ‘Abstract’) because ‘Relation’ comprises relational meanings that are not nodes and ‘SetorClass’ is not an upper category found among the mappings in the data obtained. Thus, including them will cause noise, as distance will be increased unnecessarily (although they were also rated by the subjects).
7.3.2 Results of Measuring Concreteness of SUMO Nodes

Based on the calculations completed in the previous section, each type of metaphorical expression found to possess one or more WordNet sense with corresponding SUMO node will have one final concrete SUMO node selected. The selected concrete sense is defined by the highest score of conceptual similarity when compared to the prototypical concrete nodes. This score is the score of conceptual similarity. All SUMO nodes for each type of metaphorical expression possess one score of conceptual similarity. We then select the SUMO node with the highest score of conceptual similarity in terms of the concreteness measure. This most concrete node will be used when analyzing the source domain of a particular type of metaphorical expression. Table 7.7 will show comparisons of conceptual similarity for the different senses of 成長 chéngzhǎng ‘grow/growth,’ 起飛 qǐfēi ‘take off’ and 麻痺 tānhuàn ‘paralytic.’

<table>
<thead>
<tr>
<th>Type of Metaphorical Expression</th>
<th>WordNet Definitions</th>
<th>SUMO Nodes</th>
<th>Upper SUMO Categories</th>
<th>Concreteness Scores</th>
<th>Selected by Highest Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>成長 chéngzhǎng ‘grow/growth’</td>
<td>a change resulting in an increase: “The increase is scheduled for next month.”</td>
<td>Increasing</td>
<td>Process</td>
<td>0.769988</td>
<td>V</td>
</tr>
<tr>
<td>起飛 qǐfēi ‘take off’</td>
<td>become bigger or greater in amount: “The amount of work increased.”</td>
<td>Increasing</td>
<td>Process</td>
<td>0.769988</td>
<td></td>
</tr>
<tr>
<td>麻痺 tānhuàn ‘paralytic’</td>
<td>the process of an individual organism growing organically; a purely biological unfolding of events involved in an organism</td>
<td>Growth</td>
<td>Process</td>
<td>0.984411</td>
<td></td>
</tr>
</tbody>
</table>

The experiment, however, has several limitations. First, the number of subjects is limited because there are many questions involved. Second, fatigue is a possible factor for these subjects, as they had to answer more than six hundred questions at once. A possible improvement on the experiment might be to break the questions into several lists and include more subjects in the experiments. The fatigue issue can be improved if the subjects are instructed to rest after a certain number of questions. The order of the questions in the questionnaire can also be reversed.

Although other ways of finding a final concrete score are possible, mean is used here because it is the most suitable method found.
**Chapter 7: Top-down Approach to Source Domain Determination**

<table>
<thead>
<tr>
<th>Example</th>
<th>Domain</th>
<th>Process</th>
<th>Similarity</th>
<th>POS</th>
</tr>
</thead>
<tbody>
<tr>
<td>changing gradually from a simple to a more complex level: “He proposed an indicator of osseous development in children.”</td>
<td>Growth</td>
<td>Process</td>
<td>0.984411</td>
<td>V</td>
</tr>
<tr>
<td>grow old or older: “She aged gracefully.” “We age every day—what a depressing thought!”</td>
<td>Growth</td>
<td>Process</td>
<td>0.984411</td>
<td>V</td>
</tr>
<tr>
<td>develop and reach maturity; undergo maturation: “He matured fast.” “The child grew fast.”</td>
<td>Growth</td>
<td>Process</td>
<td>0.984411</td>
<td>V</td>
</tr>
<tr>
<td>grow emotionally or mature: “The child developed beautifully in her new kindergarten.” “When he spent a summer at camp, the boy grew noticeably and no longer showed some of his old adolescent behavior.”</td>
<td>Intentional Psychological Process</td>
<td>Process</td>
<td>0.916105</td>
<td></td>
</tr>
<tr>
<td>起飛 qǐfēi ‘take off’</td>
<td>a departure; especially of airplanes</td>
<td>Motion</td>
<td>0.909443</td>
<td></td>
</tr>
<tr>
<td>起飛 qǐfēi ‘take off’</td>
<td>the initial ascent of an airplane as it becomes airborne</td>
<td>Motion</td>
<td>0.909443</td>
<td></td>
</tr>
<tr>
<td>起飛 qǐfēi ‘take off’</td>
<td>depart from the ground, as of an aircraft or balloon: “The plane took off two hours late.”</td>
<td>Transportation Process</td>
<td>0.931738</td>
<td>V</td>
</tr>
<tr>
<td>癲癇 tánhuàn ‘paralytic’</td>
<td>a condition marked by uncontrollable tremor</td>
<td>Pathologic Process Process</td>
<td>0.931738</td>
<td></td>
</tr>
<tr>
<td>癲癇 tánhuàn ‘paralytic’</td>
<td>loss of the ability to move a body part</td>
<td>Disease Or Syndrome Attribute</td>
<td>1.00000</td>
<td>V</td>
</tr>
</tbody>
</table>

As explained, there may be more than one WordNet sense with similar SUMO nodes (‘Growth,’ for example, has several senses of 成長 chéngzhǎng ‘grow/growth.’) These SUMO nodes may correspond to several WordNet senses, meaning that these few
WordNet senses under the same SUMO node should have a similar concept of concreteness, which does not contradict the overall results, as a final concrete SUMO node is what is needed in this step.

From Table 7.7, the selected concrete SUMO node for 成長 chéngzhǎng ‘grow/growth’ is ‘Growth.’ Three WordNet senses are mapped to this SUMO node (shaded). These nodes come under the category of ‘Process’ (fourth column). For 成長 chéngzhǎng ‘grow/growth’ searched in the Merged Word List, the results are shown in Table 7.8. However, as can be seen in Table 7.8, noise is found in these results. The final selected sense is ‘LandArea,’ which is a sense referring to the physical expansion of land. Since the results from the Merged Word List differ in their selection of concrete senses, we expect to find different results produced by these two knowledgebases. In Chapter Eight, we will compare these knowledgebases to determine which of them provides better results in terms of source domain determination.

Table 7.8: Selected SUMO Nodes for 成長 chéngzhǎng ‘grow/growth’ with Highest Concreteness Scores from the Chinese-English Merged Word List

<table>
<thead>
<tr>
<th>Type of Metaphorical Expression</th>
<th>WordNet Definitions</th>
<th>SUMO Nodes</th>
<th>Upper SUMO Categories</th>
<th>Concreteness Scores</th>
<th>Selected by Highest Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>成長 chéngzhǎng ‘grow/growth’</td>
<td>become larger, greater, or bigger; expand or gain</td>
<td>Increasing</td>
<td>Process</td>
<td>0.769988</td>
<td></td>
</tr>
<tr>
<td></td>
<td>come to have or undergo a change of (physical features and attributes)</td>
<td>Growth</td>
<td>Process</td>
<td>0.984411</td>
<td></td>
</tr>
<tr>
<td></td>
<td>be gradually disclosed or unfolded; become manifest</td>
<td>Process</td>
<td>Process</td>
<td>0.711235</td>
<td></td>
</tr>
<tr>
<td></td>
<td>expand in the form of a series, in mathematics</td>
<td>Content Development</td>
<td>Process</td>
<td>0.833925</td>
<td></td>
</tr>
<tr>
<td></td>
<td>make visible by means of chemical solutions; of photographic film</td>
<td>Content Development</td>
<td>Process</td>
<td>0.833925</td>
<td></td>
</tr>
<tr>
<td></td>
<td>act of improving by expanding or enlarging or refining</td>
<td>Creation</td>
<td>Process</td>
<td>0.800682</td>
<td></td>
</tr>
<tr>
<td></td>
<td>vegetation that has grown</td>
<td>Plant</td>
<td>Object</td>
<td>0.973756</td>
<td></td>
</tr>
<tr>
<td>Term</td>
<td>Process</td>
<td>Score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>---------------</td>
<td>------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>grow, progress, unfold, or evolve through a process of evolution,</td>
<td>Increasing</td>
<td>0.769988</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>natural growth, differentiation, or a conducive environment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>increase in size by natural process; of living matter, such as</td>
<td>Growth</td>
<td>0.984411</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>plants and animals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(pathology) an abnormal proliferation of tissue (as in a tumor)</td>
<td>Anatomical</td>
<td>0.997527</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cause to grow or develop</td>
<td>Growth</td>
<td>0.984411</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>the process of an individual organism growing organically; a purely</td>
<td>Growth</td>
<td>0.984411</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>biological unfolding of events involved in an organism changing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gradually from a simple to a more complex level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>elaborate by the unfolding of a musical idea and by the working</td>
<td>Music</td>
<td>0.947915</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>out of the rhythmic and harmonic changes in the theme</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>something grown or growing</td>
<td>Growth</td>
<td>0.984411</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cultivate by growing; often involves improvements by means of</td>
<td>Process</td>
<td>0.711235</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>agricultural techniques</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>develop and reach maturity; undergo maturation</td>
<td>Growth</td>
<td>0.984411</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expression</td>
<td>Domain</td>
<td>Process</td>
<td>Score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------------------------------------------------</td>
<td>-------------------------------</td>
<td>----------------------</td>
<td>-----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>become attached by or as if by the process of growth</td>
<td>Growth</td>
<td>Process</td>
<td>0.984411</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the act of moving forward toward a goal</td>
<td>Motion</td>
<td>Process</td>
<td>0.909443</td>
<td></td>
<td></td>
</tr>
<tr>
<td>become an adult</td>
<td>Growth</td>
<td>Process</td>
<td>0.984411</td>
<td></td>
<td></td>
</tr>
<tr>
<td>change the use of and make available or usable</td>
<td>Intentional Process</td>
<td>Process</td>
<td>0.870048</td>
<td></td>
<td></td>
</tr>
<tr>
<td>breed freely and abundantly</td>
<td>Sexual Reproduction</td>
<td>Process</td>
<td>0.980013</td>
<td></td>
<td></td>
</tr>
<tr>
<td>become abundant; increase rapidly</td>
<td>Increasing</td>
<td>Process</td>
<td>0.769988</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a movement forward</td>
<td>Motion</td>
<td>Process</td>
<td>0.909443</td>
<td></td>
<td></td>
</tr>
<tr>
<td>natural height of a person or animal in an upright position</td>
<td>Length Measure</td>
<td>Quantity</td>
<td>0.999883</td>
<td></td>
<td></td>
</tr>
<tr>
<td>grow and flourish</td>
<td>Increasing</td>
<td>Process</td>
<td>0.769988</td>
<td></td>
<td></td>
</tr>
<tr>
<td>acquire or build up traits or characteristics</td>
<td>Intentional Psychological Process</td>
<td>Process</td>
<td>0.916105</td>
<td></td>
<td></td>
</tr>
<tr>
<td>grow emotionally or mature</td>
<td>Intentional Psychological Process</td>
<td>Process</td>
<td>0.916105</td>
<td></td>
<td></td>
</tr>
<tr>
<td>develop in a positive way</td>
<td>Process</td>
<td>Process</td>
<td>0.711235</td>
<td></td>
<td></td>
</tr>
<tr>
<td>move one’s chess pieces into strategically more advantageous positions</td>
<td>Game</td>
<td>Process</td>
<td>0.800682</td>
<td></td>
<td></td>
</tr>
<tr>
<td>processing a photosensitive material in order to make an image visible</td>
<td>Surface Change</td>
<td>Process</td>
<td>0.800682</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cause to grow and differentiate in ways conforming to its natural development</td>
<td>Growth</td>
<td>Process</td>
<td>0.984411</td>
<td></td>
<td></td>
</tr>
<tr>
<td>generate gradually</td>
<td>Intentional Process</td>
<td>Process</td>
<td>0.870048</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>(definition not available for this sense)</em></td>
<td>Making</td>
<td>Process</td>
<td>0.833925</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
From the total 916 types of metaphorical expressions that are found in SinicaBow (also found in the Merged Word List), only 775 (84.61 per cent) are found with concrete senses (cf. Table 7.2). From the additional 298 types of metaphorical expressions found in the Merged Word List, only 279 (93.62 per cent) are found to possess concrete senses (cf. Table 7.4). Sometimes concrete senses are impossible to find because some types of metaphorical expressions contain senses that are ‘relation’ or ‘function,’ neither of which are nodes in the SUMO hierarchy. They only indicate relational or functional meanings between objects. For example, 距離 juli ‘distance’ is a type of metaphorical expression with three senses, where all of these senses are mapped to the relation of ‘distance’ (thus, a relational concept between two objects that cannot be measured in terms of conceptual distance between two concepts). As a result, 距離 juli ‘distance’ does not have a concrete sense selected, even though it is found in both SinicaBow and the Merged Word List. In the next step, we select source domain names based on the concrete definitions selected using the step previously stated.

### 7.4 Determining Source Domain Names

Source domain names are derived from keywords found within the definitions of concrete SUMO nodes that have been selected in the last step. In this section, we will discuss the specific process by which source domain names will be determined from these keywords.

A strict criterion is followed in the selection of source domain names, whereby all function words and other unnecessary words are first removed from the definitions automatically. For example, in the definition for ‘Growth’ in (6), function words (boxed), such as prepositions, conjunctions and determiners, are also removed. The remaining keywords (in bold) are then clustered in the next step.\(^{24}\)

\(^{24}\) An earlier version had removed all verbs as well as upper SUMO nodes, such as ‘Process,’
(6) a. ‘Growth’ (for 成長 chéngzhǎng ‘grow/growth’)

\[
\text{The Process of biological development in which an Organism or part of an Organism changes its form or its size.}
\]

b. ‘Motion’ (for 起飛 qǐfēi ‘take off’)

\[
\text{Process of movement.}
\]

After removal of the function words, only the bold keywords in (6) will be collected. Table 7.9 shows some of the selected source domain names. As mentioned, noise may be found, which in this case involves verbs (‘changes’) or adverbs (such as ‘temporarily’). These verbs and adverbs are then removed automatically. The examples shown in Table 7.9 are possible source domains extracted from several concrete definitions, as determined by this process.

**Table 7.9: Examples of Possible Source Domain Names for the Top-down Approach**

<table>
<thead>
<tr>
<th>product</th>
<th>organism</th>
<th>species</th>
<th>time</th>
<th>device</th>
<th>perception</th>
</tr>
</thead>
<tbody>
<tr>
<td>trait</td>
<td>cognitive agent</td>
<td>participant</td>
<td>content-bearing object</td>
<td>sound</td>
<td>constant quantities</td>
</tr>
<tr>
<td>life</td>
<td>movement</td>
<td>homo</td>
<td>destruction</td>
<td>government</td>
<td>animal</td>
</tr>
<tr>
<td>corpuscular</td>
<td>agent</td>
<td>genus</td>
<td>death</td>
<td>geopolitical</td>
<td>area</td>
</tr>
<tr>
<td>object</td>
<td></td>
<td></td>
<td></td>
<td>area</td>
<td>creation</td>
</tr>
<tr>
<td>language</td>
<td>state</td>
<td>body part</td>
<td>constant quantity</td>
<td>acquisition</td>
<td>action</td>
</tr>
<tr>
<td>area</td>
<td>patient</td>
<td>organ</td>
<td>thing</td>
<td>rule</td>
<td>destination</td>
</tr>
<tr>
<td>tissues</td>
<td>location</td>
<td>ground</td>
<td>world</td>
<td>reproduction</td>
<td>desire</td>
</tr>
<tr>
<td>interaction</td>
<td>physical quantity</td>
<td>communication</td>
<td>food</td>
<td>regulation</td>
<td>contact</td>
</tr>
<tr>
<td>stationary</td>
<td>artifact</td>
<td>body</td>
<td>direction</td>
<td>game</td>
<td>wave</td>
</tr>
<tr>
<td>artifact</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>self-connected</td>
<td>quantity change</td>
<td>intentional psychological process</td>
<td>boundaries</td>
<td>enforcement</td>
<td>future</td>
</tr>
</tbody>
</table>

‘Object’ and ‘Class,’ but this step was later found to have involved too much human judgment in the process. Therefore, only function words were removed in this current version. Some of the function words were found on the Web, and they were added to form a huge list (see http://www.einfoweb.com/grammar/pros/list.html; http://www.abcteach.com/abclists/prepositions.htm). The removal of the function words was programmed by Petr Šimon.

Most words are single words and the multiple-word expressions (such as ‘ConstantQuantities’) are usually written as one word in SUMO but they are separated for the ease of reading in Table 7.9.
Chapter 7: Top-down Approach to Source Domain Determination

The data in Table 7.9 shows examples of possible source domain names based on the concrete definitions collected from SinicaBow. Even though these source domains are content words within the concrete SUMO definitions, they are not all at the same level of concreteness. As shown in Table 7.9, some source domains seem to have a physical look (such as ANIMAL and BODY PART) while some do not (such as QUANTITY CHANGE and STATE). These source domains may also differ in terms of their generality as there are both general (SPECIES and THING) and specific (AMINAL and BODY) source domains produced by the same approach. However, in our later discussion in Chapter Twelve, we will show that these multiple source domains, which are collected through operational steps, are useful in building a hierarchical definition of source domain. Therefore, regarding the part of the second global hypothesis which states that “the top-down approach will return general source domains while the bottom-up approach will return specific source domains,” the results of the top-down approach do not seem to return source domains at the same level of generality and concreteness. This is easily explained since SUMO is arranged in a hierarchical form, and as such, it should contain both domains; those that are general and those that are specific. Therefore, it is not possible to find only general information on a taxonomic structure. Nevertheless, the source domains produced from the top-down approach may still be ‘comparatively’ more general than the source domains obtained through the bottom-up approach. We will look into this issue after the bottom-up approach has been discussed in Chapters Nine and Ten.

Regarding the first global hypothesis, we have observed that the use of computational and lexical methods can help us to determine source domain names based on strictly followed criteria. The process is replicable and, therefore, reduces certain levels of human subjectivity in the top-down approach. However, there are several decisions to be made in the process of the top-down approach. For example, the selection of the prototypical concrete SUMO nodes has to be done manually. In order to avoid using the author’s own judgment of prototypical concrete nodes, a group of human subjects was employed. The psycholinguistic experiment was made as objective as possible, and it was also designed to be replicable. Furthermore, the use of human evaluation is not uncommon in computational research, as all computational studies attempt to validate their results based on human judgment. Therefore, the top-down approach is able to reduce human subjectivity by proposing a principled way of selecting the source domain names. However, it can also be evaluated regarding how well it is able to produce source domains that are correct. The evaluation of the approaches will take place in Chapter Eleven.

26 The words in Table 7.9 will be capitalized when referred to as source domains.
As mentioned, certain source domains in Table 7.9 seem to be related (such as HOMO, ORGANISM, BODY PART, etc.). However, since the purpose of this book is to determine source domains, further issues regarding similarity between source domains will be reserved for future studies.27

In the next section, we will discuss how types of metaphorical expressions are grouped under the source domains, such as those shown in Table 7.9.

7.5 Grouping of Types of Metaphorical Expressions Based on Source Domain Names

After the source domain names have been determined, the next step involves examining which types of metaphorical expressions fall under the different source domains. In order to accomplish this, we group types of metaphorical expressions based on whether or not these source domain names appear in all the SUMO definitions (including concrete and non-concrete nodes) within each metaphorical expression. This step will be illustrated, following. All nodes are included because only through the thorough examination of all SUMO nodes are we able to determine which source domain names are repeated most frequently in all of the senses of a particular type of metaphorical expression.

First, a matrix table is created, and this matrix table tallies the number of times the different source domains appear in all of the SUMO definitions of a particular type of metaphorical expression.28 Next, we extract the results by sorting according to types of metaphorical expressions or source domains. For example, in Table 7.10, the results are sorted based on types of metaphorical expressions of 成長 chéngzhǎng ‘grow/growth,’ 起飛 qǐfēi ‘take off’ and 麻痺 tānhuàn ‘paralytic.’ As mentioned, the top-down approach uses both SinicaBow and the Merged Word List; Table 7.10 provides the results in the following two knowledgebases.

---

27 Although the similarity scores have been computed, they are not presented in this book. We reserve it for future research, as more investigation is needed.
28 We did not show the matrix tables here because they are too huge.
Table 7.10: Source Domains for 成長 chéngzhǎng ‘grow/growth,’ 起飛 qǐfēi ‘take off’ and 麻痺 tānhuàn ‘paralytic’

<table>
<thead>
<tr>
<th>Type of Metaphorical Expression</th>
<th>Source Domains</th>
<th>Frequency of Concrete Keywords in all SUMO Definitions</th>
<th>SinicaBow</th>
<th>Source Domains</th>
<th>Frequency of Concrete Keywords in all SUMO Definitions</th>
<th>The Chinese-English Merged Word List</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORGANISM</td>
<td>7</td>
<td>ORGANISM</td>
<td>22</td>
<td>PROCESS</td>
<td>3</td>
<td>PROCESS</td>
</tr>
<tr>
<td>DEVELOPMENT</td>
<td>3</td>
<td>CLASS</td>
<td>11</td>
<td>PROCESS</td>
<td>3</td>
<td>PROCESS</td>
</tr>
<tr>
<td>PHYSICAL QUANTITY</td>
<td>2</td>
<td>DEVELOPMENT</td>
<td>9</td>
<td>PROCESS</td>
<td>2</td>
<td>PROCESS</td>
</tr>
<tr>
<td>QUANTITY CHANGE</td>
<td>2</td>
<td>SUBCLASS</td>
<td>5</td>
<td>PROCESS</td>
<td>2</td>
<td>PROCESS</td>
</tr>
<tr>
<td>INTENTIONAL PROCESS</td>
<td>1</td>
<td>PURPOSE</td>
<td>4</td>
<td>PROCESS</td>
<td>1</td>
<td>PROCESS</td>
</tr>
<tr>
<td>MIND</td>
<td>1</td>
<td>THING</td>
<td>4</td>
<td>PROCESS</td>
<td>1</td>
<td>PROCESS</td>
</tr>
<tr>
<td>BRAIN</td>
<td>1</td>
<td>INTENTIONAL PROCESS</td>
<td>4</td>
<td>PROCESS</td>
<td>1</td>
<td>PROCESS</td>
</tr>
<tr>
<td>STAGES</td>
<td></td>
<td></td>
<td></td>
<td>PROCESS</td>
<td>1</td>
<td>PROCESS</td>
</tr>
<tr>
<td>PHYSICAL QUANTITY</td>
<td></td>
<td></td>
<td></td>
<td>PROCESS</td>
<td>1</td>
<td>PROCESS</td>
</tr>
<tr>
<td>QUANTITY CHANGE</td>
<td></td>
<td></td>
<td></td>
<td>PROCESS</td>
<td>1</td>
<td>PROCESS</td>
</tr>
<tr>
<td>成長 chéngzhǎng ‘grow/growth’</td>
<td></td>
<td></td>
<td></td>
<td>PROPERTIES</td>
<td>1</td>
<td>PROPERTIES</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>INDIVIDUAL</td>
<td>1</td>
<td>INDIVIDUAL</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CREATION</td>
<td>1</td>
<td>CREATION</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SURFACE</td>
<td>1</td>
<td>SURFACE</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LENGTH</td>
<td>1</td>
<td>LENGTH</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CONSTANT</td>
<td>1</td>
<td>CONSTANT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>QUANTITY</td>
<td>1</td>
<td>QUANTITY</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PRESENCE</td>
<td>1</td>
<td>PRESENCE</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>STRUCTURAL</td>
<td>1</td>
<td>STRUCTURAL</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ANATOMY</td>
<td>1</td>
<td>ANATOMY</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>OBJECT</td>
<td>1 each</td>
<td>OBJECT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ORGANIZATION</td>
<td>1 each</td>
<td>ORGANIZATION</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CELL</td>
<td>1 each</td>
<td>CELL</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SOUND</td>
<td>1 each</td>
<td>SOUND</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PROPERTIES</td>
<td>1 each</td>
<td>PROPERTIES</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>INDIVIDUAL</td>
<td>1 each</td>
<td>INDIVIDUAL</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CREATION</td>
<td>1 each</td>
<td>CREATION</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SURFACE</td>
<td>1 each</td>
<td>SURFACE</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LENGTH</td>
<td>1 each</td>
<td>LENGTH</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CONSTANT</td>
<td>1 each</td>
<td>CONSTANT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>QUANTITY</td>
<td>1 each</td>
<td>QUANTITY</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PRESENCE</td>
<td>1 each</td>
<td>PRESENCE</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>STRUCTURAL</td>
<td>1 each</td>
<td>STRUCTURAL</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ANATOMY</td>
<td>1 each</td>
<td>ANATOMY</td>
</tr>
<tr>
<td>起飛 qǐfēi ‘take off’</td>
<td>PROCESS</td>
<td>2</td>
<td>7</td>
<td>PROCESS</td>
<td>2</td>
<td>PROCESS</td>
</tr>
<tr>
<td>MOVEMENT</td>
<td>2</td>
<td>MOVEMENT</td>
<td>5</td>
<td>PROCESS</td>
<td>2</td>
<td>PROCESS</td>
</tr>
<tr>
<td>MOTION</td>
<td>1</td>
<td>MOTION</td>
<td>4</td>
<td>PROCESS</td>
<td>1</td>
<td>PROCESS</td>
</tr>
<tr>
<td>TRANSPORTATION DEVICE</td>
<td>1</td>
<td>CLASS</td>
<td>3</td>
<td>PROCESS</td>
<td>1</td>
<td>PROCESS</td>
</tr>
</tbody>
</table>
Therefore, once this step has been completed, the numbers of occurrences of these source domains can be counted. Note that in the top-down approach, all types of metaphorical expressions, whether from CNA or from XIN, are treated similarly, as their comparison of concreteness is based on the SUMO ontology, and SUMO is a shared upper ontology that does not differentiate between these two communities. Therefore, only a single result for all types of metaphorical expressions will be found in the top-down approach (with distinctions made in terms of SinicaBow and the Merged Word List only).

---

29 The bottom-up approach will separate both CNA and XIN because it is not based on a shared concept.
In Table 7.10, the source domains for 成長 chéngzhǎng ‘grow/growth,’ 起飛 qǐfēi ‘take off’ and 麻痺 tānhuàn ‘paralytic’ are shown. The huge discrepancy between the numbers of source domains found for each database is readily apparent. The top source domains, which are repeated in both SinicaBow and the Merged Word List, are given in bold. For sentences such as (7a), the conceptual metaphors for 經濟成長 jīngjì chéngzhǎng ‘economy grows’ are ECONOMY IS AN ORGANISM; and ECONOMY IS A PROCESS, etc. As for 經濟麻痺 jīngjì tānhuàn ‘the economy is paralytic,’ the conceptual metaphors are also ECONOMY IS A PROCESS; ECONOMY IS AN ORGANISM, etc. (7b).

(7) a. 唯有社會安定，經濟才能成長 (CNA)
only if society stable economy then can grow
‘Only when society is stable, (the) economy will grow.’

b. 以免蘇聯經濟麻痺 (CNA)
to avoid Soviet.Union economy paralytic
‘In order to avoid the economy of the Soviet Union from becoming paralytic…’

From the source domains of 成長 chéngzhǎng ‘grow/growth’ and 麻痺 tānhuàn ‘paralytic,’ we know that these two types of metaphorical expressions have the same source domains of ORGANISM and PROCESS and, therefore, must share certain similarities within their source domains. Sorting by source domains is also possible; however, in this case, each type of metaphorical metaphor will be repeated in several source domains at the same time. For example, ORGANISM appears in both 成長 chéngzhǎng ‘grow/growth’ and 麻痺 tānhuàn ‘paralytic.’ Table 7.11 shows some examples of types of metaphorical expressions that are sorted according to their source domains produced by SinicaBow.

In Table 7.11, the first column shows the source domains, followed by examples of types of metaphorical expressions and their English gloss. The last column shows the number of times these source domain names appear in the SUMO definitions of the types of metaphorical expressions.
In Table 7.11, we see that some types of metaphorical expressions (such as 壓抑 yàiyì ‘inhibit’ and 放鬆 fàngsōng ‘relaxed’) can be found in more than one source domain. The following example shows the use of 壓抑 yàiyì ‘inhibit’ with 教育 jiàoyù ‘education.’ The result shows that 壓抑 yàiyì ‘inhibit’ comes from conceptual metaphors such as EDUCATION IS A STATE OF MIND as well as EDUCATION IS BEHAVIOR (as STATE OF MIND and BEHAVIOR are two concrete content words found in the SUMO definitions of 壓抑 yàiyì ‘inhibit’).
Therefore, the particular target domain of EDUCATION, where 壓抑 yì ‘inhibit’ is found, is seen as a ‘state of mind’ as well as a ‘behavior.’ When one refers to EDUCATION IS A STATE OF MIND, one may want to emphasize the intellectual aspect of education. On the other hand, EDUCATION IS BEHAVIOR may refer to education as being a person behaving in a certain way; indeed, Ahrens (2002) has postulated that a target domain uses different source domains for different reasons. From the example in (8), it may, therefore, be proposed that the relation between the source domains selected and the reasons for the source-target domain pairings can be further examined in future research.

7.6 Summary of Chapter

A pre-determined source domain name is one of the characteristics of the top-down approach, whereby possible source domains are first determined before grouping types of metaphorical expressions under these source domains. This chapter has given an in-depth discussion of the possible four steps in the top-down approach. These steps involve (a) mapping all types of metaphorical expressions to WordNet and SUMO through SinicaBow and the Chinese-English Merged Word List; (b) measuring the conceptual distance between SUMO nodes and several prototypical concrete nodes that have been selected using a human rating system (in order to select one final SUMO node with the highest conceptual similarity); (c) collecting concrete keywords in the SUMO definitions of the concrete nodes; and (d) clustering all types of metaphorical expressions under these concrete keywords, which serve as source domain names.

The results of the top-down approach indicate that each type of metaphorical expression can indeed possess several source domains, and further, that some of these source domains are more general while others are more specific. These results are markedly different from previous studies (such as Charteris-Black & Ennis 2001, Kövecses 2002, Charteris-Black 2004, and Chung, Ahrens & Huang 2005), all of which appear to assign each metaphor to one single source domain. For example, ‘growth’ is categorized under the single source domain of PLANT in one study, while it is categorized under ORGANISM in another study. These studies seldom categorize a
metaphor under several source domains, even though both of the above source domains are possible.30

The findings of this book also add clarity to the debate over specific versus general source domains. While Kövecses (2002) and Charteris-Black (2004) think that ‘growth’ is part of the source domain of PLANT, Chung, Ahrens & Huang (2005) suggest that ‘growth’ belongs to the source domain of PERSON. In another study, Charteris-Black & Ennis (2001) suggest that ‘growth’ should be grouped under the source domain of ORGANISM. If viewed according to the findings in this chapter, their source domains are pointed at several different levels. If the more general level of ORGANISM is taken, all other types of metaphorical expressions will be included. However, if the more specific level of PLANT or PERSON is used, terms that do not belong to these source domains (such as ‘microorganism’) will be left out. It may, therefore, be concluded from the top-down approach that source domains are actually a relative concept, where the selection of a general or specific source domain is dependent on the metaphorical expressions categorized under this source domain. Therefore, we can predict that the general source domain of ORGANISM will comprise more types of metaphorical expressions, while the specific source domain of PERSON will only be a subset of ORGANISM. In the later chapters, we also propose that connectivity between source domains can be established in future research, which will give a clearer explanation as to which source domains fall under the subset of a larger set. In Chapter Eight, the results from the top-down approach will be looked at. We will start working on the bottom-up approach in Chapter Nine, and our findings will be discussed in Chapter Ten. In Chapter Eleven, an evaluation of the top-down and bottom-up approaches will be undertaken before the discussion of linguistic phenomena in Chapter Twelve and the conclusion in Chapter Thirteen.

30 This is mainly due to the limitation of manual analysis, as an analyzer tends to categorize according to the bigger categories that are often repeated throughout all the data. Most analyzers do not bother creating source domains that seldom occur in their data. Therefore, whether ‘growth’ is categorized under PLANT, ORGANISM or ANIMAL may depend on which of these source domains are more often repeated with the other metaphors.
Chapter 8: Results of the Top-down Approach

The top-down approach uses information from the ontology in order to label the source domain names. This approach is characterized as top-down because it utilizes conceptual knowledge at the upper level. Since the approach is top-down, similar types of metaphorical expressions in data from Taiwan (CNA) and China (XIN) should yield similar source domains, as the ontology assumes a shared knowledge in spite of minor differences across communities. Steps involved in the top-down approach are re-stated in Figure 8.1 (also shown in Figure 7.1 in Chapter Seven).

---

1 However, the discussion of the source domains will still be conducted in terms of CNA and XIN. It is only that the source domain names for similar types of metaphorical expressions in both CNA and XIN are the same while the source domain names can be different in the bottom-up approach.
The steps involved in the top-down approach are as follows: First, map all types of metaphorical expressions (such as 成長 chéngzhǎng ‘grow/growth,’ 起飛 qǐfēi ‘take off’ and 麻癱 tānhuān ‘paralytic’) to WordNet and SUMO (WordNet is needed because it provides the links to SUMO) by using both SinicaBow and the Chinese-English Merged Word List. These two knowledgebases make the mappings possible because they translate the English WordNet into Chinese. Second, select the most concrete SUMO node among all the corresponding SUMO nodes (of the senses) within each type of metaphorical expression so that source domain names can be decided upon. The selection of the concrete sense is based on the measurement of concreteness of each sense against several prototypical concrete nodes (which have been selected by human subjects). Third, select source domain names from the SUMO definitions of the selected concrete sense. In this approach, source domain names are first determined before sorting the types of metaphorical expressions according to these source domains. This is one difference between the top-down approach and the bottom-up approach. The final step is to group types of metaphorical expressions according to source domains by observing whether or not these source domain names appear in all the SUMO definitions of a particular metaphorical expression.

In this chapter, the results garnered using the top-down approach will be discussed. §8.1 offers a discussion on the overall performance of the top-down approach (i.e., how well this approach performs in finding a source domain). §8.2 will discuss the most frequently appearing source domains found in the top-down approach as well as the most frequently appearing types of metaphorical expressions under each source domain. §8.3 deliberates the strengths and weaknesses of the top-down approach, and finally, §8.4 shows the summary of the chapter.

8.1 Performance of the Top-down Approach

For the performance of the top-down approach, we compare the percentages of cases where a source domain can be determined. These percentages will be matched against the types of metaphorical expressions that are found in the top-down approach, i.e., found cases shown in Table 7.4 in Chapter Seven.² These found cases represent cases that are found through mappings in SinicaBow and the Chinese English Merged Word List. From previous analysis of the found cases (Table 7.4 in Chapter Seven), the Merged Word List has been shown to have found 18 per cent more types of metaphorical expressions and 8 per cent more tokens of metaphorical expressions than SinicaBow. The

---

² Rather than ‘how many percentages out of the total metaphorical expressions with source domains determined.’
number of the found cases will be used as the denominator to calculate the percentages of the metaphorical expressions with found source domains in each knowledgebase.\(^3\)

In Figure 8.2, the overall performance of types of metaphorical expressions that are found in SinicaBow and the Merged Word List, respectively, are shown.

From Figure 8.2, we can see that the Merged Word List seems to perform better than SinicaBow in CNA, XIN and CNA+XIN, indicating that among the cases that are found, the Merged Word List may accomplish this by being able to determine source domains for more found cases than SinicaBow. The percentages of the Merged Word List in Figure 8.2 are consistently about 5 per cent higher than SinicaBow.

As mentioned, the percentages displayed in this section are percentages against the number of found cases in each knowledgebase. None of the results in Figure 8.2 displays a 100 per cent performance, indicating that not all the found cases can be determined in terms of their source domains. In other words, even though a metaphorical expression is found in either SinicaBow or the Merged Word List, determination of their source domains may still prove to be unsuccessful. Unsuccessful cases may occur if all the SUMO links of these types of metaphorical expressions are concepts indicating relations.

---

\(^3\) As mentioned in earlier chapters, the cases that are not found are mainly due to technical problems, and words that are not found cannot be considered for later steps.
or functions. For example, 朝著 (著) cháo (zhe) ‘face (progressive)’ is one example where all of its SUMO nodes (two) indicate a function (‘FrontFn’) and a relation (‘orientation’). If this happens, the computation of conceptual similarity cannot be carried out, as neither of these concepts indicates a location in the hierarchy. Therefore, we can see in Figure 8.2 that not all found cases can be determined in terms of their source domains.

In Figure 8.3, the percentages of found source domains, displayed by using tokens of metaphorical expressions, are shown.

![Figure 8.3: Percentages of Tokens of Metaphorical Expressions (Found in the Overall SinicaBow and the Merged Word List, Respectively) with Source Domains Determined in the Top-down Approach](image)

Compared to Figure 8.2, the percentages of the tokens of metaphorical expressions with found source domains are only about 1 per cent higher in the Merged Word List. The following example in (1a) shows a type of metaphorical expression (孕育 yùnyù ‘breed’) whose source domain was unable to be determined in SinicaBow but was, rather, determined using the Merged Word List. Its source domains are given in (1b) below (only those occurring more than three times are shown).

(1) a. 單一 的 自然 牧業 經濟 沒 能
dānyì de zìrán mùyè jìngjì měi néng
   single DE natural livestock.farming economy Neg. can
Chapter 8: Results of the Top-down Approach

b. ECONOMY IS
A PROCESS (6)
SEXUAL REPRODUCTION (5)
AN ORGANISM (4)
BRAIN (3)
MIND (3)
AN INTENTIONAL PROCESS (3)

One reason why 孕育 yùnyù ‘breed’ is not found in SinicaBow may be due to the fact that SinicaBow is a translated version of the English WordNet. If this word is not collected as part of the lexicon in the English WordNet, it will also not appear in SinicaBow (this could also be another reason why the performance of the Merged Word List is better than that of SinicaBow; namely, the Merged Word List collects all translated meanings in both languages).

8.2 Source Domains in the Top-down Approach

In the previous section, we witnessed the performance of two knowledgebases. In this section, we examine the types of source domains determined for the top-down approach. We will discuss this topic by examining different datasets, i.e., Single Target Domains and Coordinated Target Domains. All comparisons made between SinicaBow and the Merged Word List will also be undertaken.

8.2.1 Source Domains for the Single Target Domains

For ease of comparison, the percentages given in this section are the percentages for a particular source domain appearing in the overall tokens of metaphorical expressions. For example, there are 35.84 per cent of the total 14,075 tokens of metaphorical expressions in the overall Single Target Domains in CNA containing the source domain of COGNITIVE AGENT in SinicaBow. Since it is also true that one type of metaphorical expression can appear in several source domains, the percentages of individual source domains in the overall types of source domains cannot be provided, as the total number of tokens of metaphorical expressions is always more than the original 14,075, and the same types of metaphorical expressions are repeated in the results. The results provided
in this section represent the results of the whole set of data when searched in both SinicaBow and the Merged Word List.⁴

Among the five target domains of 政治 zhèngzhì ‘politics,’ 外交 wàijiāo ‘foreign affairs,’ 財政 cáizhèng ‘finance,’ 教育 jiàoyù ‘education’ and 經濟 jīngjì ‘economy,’ the most frequently appearing source domains are consistently CLASS and PROCESS.⁵ General keywords (i.e., keywords that appear in almost every dataset) will not be reported in the discussion of this section. A criterion validating the exclusion of these general keywords is that we cannot see their differences by looking at their types of metaphorical expressions because their types of metaphorical expressions are practically identical if listed under these general keywords (since they contain almost every type of metaphorical expression). For example, CLASS and PROCESS appear in every type of metaphorical expression, and therefore, their differences cannot be determined in terms of their types of metaphorical expressions.⁶ Therefore, it is possible that these general keywords include almost everything. The results provided in Table 8.1 show the first meaningful source domain obtained for each dataset (other than CLASS, PROCESS, SUBCLASS, etc., which are general keywords that are excluded because they appear in almost every dataset).

The differences between SinicaBow and the Merged Word List for CNA and XIN, respectively, can be observed in Table 8.1, where the target domains are presented in the left-most column. As each target domain will be discussed in terms of CNA and XIN data, and within CNA and XIN, there will be two sets of results from the two knowledgebases, respectively.

---

⁴ In the earlier comparisons, we mentioned that SinicaBow and the Merged Word List would be compared on two levels: First is for metaphorical expressions found in SinicaBow only, and second is for all metaphorical expressions, regardless of whether they are found in SinicaBow. For the discussion in this section, all data is used.

⁵ See Appendix A8.1 for the list of general keywords that are frequently appearing in almost all datasets. These source domains are not included in the discussion of the top-down approach. However, they will still be tested in terms of correctness in the next section (so as to see whether or not these source domains still hold, even if they appear in many datasets).

⁶ This is because if a general keyword is taken as a source domain, its presence is likely to be found in every dataset. If this is the case, this source domain will not be meaningful, as it will only become a constant variable that does not tell the difference between the target domains. Examples of general keywords that are found are ABSTRACT, ACTIVITY, ATTRIBUTES, BASIS, CLASS, CONCEPT, DOMAIN, ELEMENT, ENTITY, FACT, etc. Others are listed in Appendix A8.1.
Chapter 8: Results of the Top-down Approach

Table 8.1: Most Frequently Appearing Source Domains for Each Dataset in the Single Target Domains in the Top-down Approach

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Source Domains</th>
<th>Percentages in the Total Tokens of Metaphorical Expressions</th>
<th>Source Domains</th>
<th>Percentages in the Total Tokens of Metaphorical Expressions</th>
<th>Source Domains</th>
<th>Percentages in the Total Tokens of Metaphorical Expressions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>CN</strong></td>
<td></td>
<td><strong>XIN</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>politics</td>
<td>COGNITIVE AGENT</td>
<td>1,038 / 2,931 (35.54%)</td>
<td>COGNITIVE AGENT</td>
<td>2,165 / 2,921 (74.12%)</td>
<td>COGNITIVE AGENT</td>
<td>208 / 1,117 (18.62%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PURPOSE</td>
<td>873 / 1,117 (78.19%)</td>
</tr>
<tr>
<td>foreign affair</td>
<td>PURPOSE</td>
<td>190 / 583 (32.59%)</td>
<td>COGNITIVE AGENT</td>
<td>481 / 583 (82.50%)</td>
<td>PURPOSE</td>
<td>107 / 173 (61.85%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>155 / 173 (89.60%)</td>
</tr>
<tr>
<td>finance</td>
<td>ORGANISM</td>
<td>176 / 623 (28.25%)</td>
<td>ORGANISM</td>
<td>512 / 623 (82.34%)</td>
<td>PURPOSE</td>
<td>61 / 338 (18.09%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PURPOSE</td>
<td>275 / 338 (81.36%)</td>
</tr>
<tr>
<td>education</td>
<td>COGNITIVE AGENT</td>
<td>618 / 1,300 (47.54%)</td>
<td>COGNITIVE AGENT</td>
<td>1,096 / 1,300 (84.31%)</td>
<td>COGNITIVE AGENT</td>
<td>1,259 / 1,419 (88.72%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>economy</td>
<td>COGNITIVE AGENT</td>
<td>3,073 / 8,648 (35.53%)</td>
<td>COGNITIVE AGENT</td>
<td>6,495 / 8,648 (75.10%)</td>
<td>COGNITIVE AGENT</td>
<td>2,818 / 7,752 (36.35%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6,088 / 7,752 (78.53%)</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td><strong>5,044 / 14,075 (35.84%)</strong></td>
<td><strong>10,687 / 14,075 (75.93%)</strong></td>
<td></td>
<td><strong>3,830 / 10,799 (35.47%)</strong></td>
<td><strong>8,640 / 10,799 (80.01%)</strong></td>
</tr>
</tbody>
</table>

In Table 8.1, we observe a huge contrast in the percentages of the most frequently appearing source domains between SinicaBow and the Merged Word List. If we compare between these two knowledgebases, we will find that the source domains in the Merged Word List are constantly higher in percentages (74 per cent and above) when compared to SinicaBow (18 per cent and above). We can explain this phenomenon by recalling the structure of SinicaBow and the Merged Word List. We know that the Merged Word List returns more senses than SinicaBow does. From a statistical point of view, when the sample is large, there will be more power in the results obtained, as is also the case for the Merged Word List. When more senses are returned, the results could become more reliable, as the salient concepts will be repeated more often than in a knowledgebase with smaller results. However, when the results from a smaller knowledgebase are the same as those returned by a larger knowledgebase, we can say that the results are most likely to be salient concepts that stand out in both knowledgebases. Based on this
reasoning, we can understand why the same source domains are found using the two knowledgebases, even though their percentages differ greatly.\(^7\)

In Table 8.1, the shaded columns display similar results between SinicaBow and the Merged Word List. Differences occur between COGNITIVE AGENT and PURPOSE (non-shaded) for 政治 zhèngzhì ‘politics’ in XIN and for 外交 wàijiāo ‘foreign affairs’ in both CNA and XIN. We can explain the position of COGNITIVE AGENT and PURPOSE when we look into the top ten source domains in SinicaBow and the Merged Word List, respectively.

In Table 8.2, the top ten source domains (the general keywords are ignored) are given, found in the overall Single Target Domains. These ten source domains are shaded in Table 8.2. The non-shaded source domains are general keywords. Most of the top ten source domains found in SinicaBow and the Merged Word List are similar. The darker (black) shading shows source domains that are not overlapped when the sets of results in Table 8.2 are cross-examined.

**Table 8.2:** The Top Ten Source Domains Most Frequently Used for the Sum of the Single Target Domains in the Top-down Approach

<table>
<thead>
<tr>
<th>Source Domains</th>
<th>SinicaBow</th>
<th>The Chinese-English Merged Word List</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tokens of Metaphorical Expressions under Each Source Domain</td>
<td>% (Overall Tokens of Metaphorical Expressions 24,878)</td>
</tr>
<tr>
<td>CLASS</td>
<td>14,750</td>
<td>59.29</td>
</tr>
<tr>
<td>PROCESS</td>
<td>12,773</td>
<td>51.34</td>
</tr>
<tr>
<td>COGNITIVE AGENT</td>
<td>8,874</td>
<td>35.67</td>
</tr>
<tr>
<td>PURPOSE</td>
<td>7,170</td>
<td>28.82</td>
</tr>
<tr>
<td>SUBCLASS</td>
<td>6,305</td>
<td>25.34</td>
</tr>
<tr>
<td>PART</td>
<td>5,647</td>
<td>22.70</td>
</tr>
<tr>
<td>ORGANISM</td>
<td>4,818</td>
<td>19.37</td>
</tr>
<tr>
<td>OBJECT</td>
<td>4,491</td>
<td>18.05</td>
</tr>
<tr>
<td>THING</td>
<td>4,471</td>
<td>17.97</td>
</tr>
<tr>
<td>STAGES</td>
<td>4,159</td>
<td>16.72</td>
</tr>
<tr>
<td>PHYSICAL QUANTITY</td>
<td>4,092</td>
<td>16.45</td>
</tr>
</tbody>
</table>

\(^7\) However, the caveat is that this may be noise in the results of the Merged Word List because there are too many senses (some unwanted) returned through the searches. More discussion will be given in §8.3.
Chapter 8: Results of the Top-down Approach

<table>
<thead>
<tr>
<th>INTENTIONAL PROCESS</th>
<th>STAGES</th>
<th>AGENT</th>
<th>MOVEMENT</th>
<th>QUANTITY CHANGE</th>
<th>PHYSICAL QUANTITY</th>
<th>PROPOSITION</th>
<th>ATTRIBUTES</th>
<th>ARTIFACT</th>
<th>PATIENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>3,826</td>
<td>15,400</td>
<td>3,737</td>
<td>14,947</td>
<td>3,641</td>
<td>14,259</td>
<td>3,580</td>
<td>13,329</td>
<td>13,179</td>
<td>13,150</td>
</tr>
<tr>
<td>15.38</td>
<td>61.90</td>
<td>15.02</td>
<td>60.08</td>
<td>14.64</td>
<td>57.32</td>
<td>14.39</td>
<td>53.58</td>
<td>52.97</td>
<td>52.86</td>
</tr>
</tbody>
</table>

** White shades are general keywords; grey shades are overlapped source domains; black shades are non-overlapped source domains.

The percentages displayed in Table 8.2 show how many percentages of the overall tokens of metaphorical expressions contain a particular source domain. For example, in the results of source domains obtained through SinicaBow, there are 35.67 per cent of all the tokens of metaphorical expressions that belong to the source domain of COGNITIVE AGENT. Comparatively, COGNITIVE AGENT in the Merged Word List shows a percentage of 77.69. This gap in percentages between the two knowledgebases can be explained using the previous reasoning, i.e., the sampling returned through the Merged Word List is huge, and therefore, it will only confirm the results garnered from SinicaBow. Furthermore, since the results of the Merged Word List contain more found cases than SinicaBow (see Figures 8.2 and 8.3), the source domains yielded through the Merged Word List should again be a confirmation of the results from SinicaBow.

For the source domain of COGNITIVE AGENT, the following examples are shown in (2).

---

8 Uses of the different source domains will be given in the next section.
9 Notice that the formulation of conceptual metaphors, such as ECONOMY IS A COGNITIVE AGENT, becomes awkward, as the expression of "經濟建設 " ‘the building of the economy’ means that the economy is a receiver rather than builder (unlike the example in (1) where the economy is the agent of the sentence). Therefore, there are further issues related to constructions and the formulation of conceptual metaphors from source domains, which we have to look into as future research (this further proves the importance of constructions in metaphor research). For the examples in (2b-c), we suggest we formulate the conceptual metaphor as: X IS A RECIPIENT OF ACTIONS FROM A COGNITIVE AGENT. This problem regarding the specific formulation of conceptual metaphor occurs when it comes to specific examples, such as (2b-c). When we refer to source domains, we will state the following: 建設 "jiàngshè ‘construct/construction’ belongs to the source domain of COGNITIVE AGENT. This matches the purpose of the book, i.e., to determine the source domains for metaphorical expressions found in corpora using lexical and computational methods.
(2) a. ECONOMY IS A RECIPIENT OF ACTIONS FROM A COGNITIVE AGENT

將向外賓展現我國經濟建設

 jiāng xiāng wàibīn zhǎnxiàn wǒguó jīngjì jiànshè

will toward foreign.guest display our.country economy construct

的實力 (CNA)

dē shílì

DE strength

‘(Someone) will demonstrate to the foreign guests the strength in building the economy in my country.’

b. EDUCATION IS A RECIPIENT OF ACTIONS FROM A COGNITIVE AGENT

耐心推動許多困難重重的

 nàixīn tuīdòng xīdùō kùnmánchóngchóng dē

patient push many full.of.difficulties DE

教育改革理念 (CNA)

 jiàoyù gǎigé lǐnàn
education reform idea

‘To promote many reformations of ideas in education, which will face with many difficulties,...’

c. FOREIGN AFFAIRS ARE COGNITIVE AGENTS

國際外交將聯合對日制裁 (CNA)

 guójì wàijiāo jiàng liánhé duì rì zhìcái

international foreign.affairs will unite face Japan impose.sanction

‘International foreign affairs will unite and impose sanctions against Japan.’

Among these examples, only (2c) seems to be clearly indicating 外交 wàijiāo ‘foreign affairs’ as cognitive agent (thus, FOREIGN AFFAIRS ARE COGNITIVE AGENTS). The other sentences show that the target domains are the recipient of the actions by the COGNITIVE AGENT (see Footnote 9): 經濟 jīngjì ‘economy’ (2a) is the patient of 建設 jiànshè ‘construct’ (as in ‘the building of the economy’); 教育 jiàoyù ‘education’ (2b) becomes the patient of 改革 gǎigé ‘reform.’

Based on the results in Table 8.2, too, we can see why COGNITIVE AGENT, PURPOSE and ORGANISM appear as the most frequently appearing source domains interchangeably; they are the top three source domains that appear often in the total Single Target Domains, as demonstrated in Table 8.3. Their ordering is also similar in
SinicaBow and the Merged Word List, indicating the consistency of these results. For
the other remaining source domains, their ordering is slightly different, although most
of them are repeated in both knowledgebases.

The fact that the source domains of COGNITIVE AGENT, PURPOSE and
ORGANISM are found most frequent in the total Single Target Domains can also be
related to the work of Charteris-Black & Ennis (2001) and Chung (2005, 2008). These
two works state that personification or metonymy, usually categorized under the source
domains of ORGANISM (Charteris-Black & Ennis 2001:256) and PERSON (Chung et
al. 2005:74), is frequently found in different languages (English and Spanish for
Charteris-Black & Ennis (2001); and English, Malay and Mandarin for Chung (2005,
2008)). Therefore, we can see some consistencies of results in this study when
compared to previous human analysis of source domains. However, what is emphasized
in this book is that the source domains produced using the top-down approach were
selected using computational and lexical methods. If we compare with those of
Charteris-Black & Ennis (2001) and Chung (2005, 2008), we will see that the level of
specificity of the source domains selected by Charteris-Black & Ennis (2001) may be
closer to the top-down approach (examples of other source domains they use are
PHYSICAL MOVEMENTS and NATURAL DISASTERS). Comparatively, the
source domains selected by Chung (2005, 2008) are more specific than those shown in
Table 8.2 (such as COMPETITION, FOREST, OCEAN and ANIMAL). Therefore, based
on this study, we are able to evaluate how general and specific one’s analysis of source
domains is by comparing the source domains selected. More discussion about source
domains can be seen in the bottom-up approach in Chapters Nine and Ten.

8.2.2 Most Frequently Appearing Types of Metaphorical Expressions in
Different Source Domains for the Single Target Domains

Each type of metaphorical expression, such as 建設 jiànsè ‘construct/construction’
and 改革 gǎigé ‘reform/reformation,’ can appear in several source domains. This
indicates that, when appearing in different source domains, different aspects of the
source domains may be emphasized (such as between ORGANISM and COGNITIVE
AGENT). Table 8.3 shows the top two types of metaphorical expressions that appear
most frequently among the source domains listed in Table 8.1.

10 Charteris-Black & Ennis (2001) also display more specific conceptual metaphors underneath
these more general source domains. However, the major categories are the general conceptual
metaphors.

11 Intuitively, we may guess that COGNITIVE AGENT emphasizes the cognitive aspect of the
doer, while ORGANISM needs not emphasize this aspect. However, the specific reasons why
The percentages shown in Table 8.3 are percentages of the tokens of each type of metaphorical expression appearing among the total tokens of metaphorical expressions within the same source domains. For example, in the CNA results of 政治 'politics' produced by SinicaBow, 改革 gaïgé ‘reform/reformation’ constitutes 26.98 per cent (280 tokens of metaphorical expressions) of all the total tokens of metaphorical expressions under COGNITIVE AGENT (1,038 tokens of metaphorical expressions). The percentage of 改革 gaïgé ‘reform/reformation’ appears to be the highest among all other types of metaphorical expressions.

Table 8.3: The Top Two Types of Metaphorical Expressions in the Single Target Domains in the Top-down Approach

<table>
<thead>
<tr>
<th>Datasets</th>
<th>CNA</th>
<th>XIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>政治 zhângzhî ‘politics’</td>
<td>改革 gaïgé ‘reform/reformation’ (26.98%)</td>
<td>建設 jiànshè ‘construct/construction’ (46.42%)</td>
</tr>
<tr>
<td></td>
<td>改革 gaïgé ‘reform/reformation’ (12.93%)</td>
<td>改革 gaïgé ‘reform/reformation’ (7.59%)</td>
</tr>
<tr>
<td></td>
<td>保护 bìhù ‘protection’ (12.43%)</td>
<td>保护 bìhù ‘protect’ (5.96%)</td>
</tr>
<tr>
<td>外交 wàijiān ‘foreign affairs’</td>
<td>努力 nǔlì ‘diligent/diligence’ (43.68%)</td>
<td>努力 nǔlì ‘diligent/diligence’ (64.40%)</td>
</tr>
<tr>
<td></td>
<td>途徑 tújìng ‘path’ (28.95%)</td>
<td>途徑 tújìng ‘path’ (21.50%)</td>
</tr>
<tr>
<td></td>
<td>途徑 tújìng ‘path’ (11.44%)</td>
<td>途徑 tújìng ‘path’ (14.84%)</td>
</tr>
<tr>
<td>财政 câizhîng ‘finance’</td>
<td>因難 kùnmíng ‘difficulty’ (72.16%)</td>
<td>因難 kùnmíng ‘difficulty’ (45.09%)</td>
</tr>
<tr>
<td></td>
<td>因難 kùnmíng ‘difficulty’ (30.60%)</td>
<td>因難 kùnmíng ‘difficulty’ (31.40%)</td>
</tr>
<tr>
<td></td>
<td>負擔 fùdān ‘burden’ (24.76%)</td>
<td>負擔 fùdān ‘burden’ (25.40%)</td>
</tr>
<tr>
<td></td>
<td>負擔 fùdān ‘burden’ (11.36%)</td>
<td>負擔 fùdān ‘burden’ (9.09%)</td>
</tr>
</tbody>
</table>

Certain source domains are selected need to be further examined. For instance, Ahrens (2002) postulates that ‘Mapping Principles’ are the underlying reason why mappings occur between source and target domains. Our results show that source domains are probably compositional, meaning that one type of metaphorical expression can be mapped to closely related concepts with different levels of specificity (such as COGNITIVE AGENT and ORGANISM). Whether the compositional level of the source domains is related to the ‘Mapping Principles’ is an issue we would like to pursue in future work.
Using the information in Table 8.3, we can compare the results from CNA and XIN in SinicaBow versus the Merged Word List. The lighter shades show similar top two types of metaphorical expressions found between the two knowledgebases. For example, under COGNITIVE AGENT of 政治 zhèngzhì ‘politics’ in CNA, both top two types of metaphorical expressions are similar when searched through SinicaBow and the Merged Word List. Some other results (bottom row, for example) show similar top results for the type of metaphorical expression (cf. 建設 jiànsè ‘construct/construction’ in the total Single Target Domains in Table 8.3). The secondary results are different (cf. 改革 gǎigé ‘reform/reformation,’ 成長 chéngzhǎng ‘grow/growth,’ 進行 jìxìng ‘proceed’ and 情勢 qíngshì ‘situation’ in the Total Single Target Domains of Table 8.3).

The darker shades (row of 外交 wàixiāo ‘foreign affairs’) show similar top two types of metaphorical expressions throughout all cells, even though their source domains differ between SinicaBow and the Merged Word List in CNA. This shows that the same types of metaphorical expressions (努力 nǔlì ‘path’ and 途徑 tújìng ‘path’) appear in both the source domains of COGNITIVE AGENT and PURPOSE. This is a finding of this book that most types of metaphorical expressions can appear in more than one source domain. This could mean that source domains determined using the top-down approach are compositional, i.e., they may refer to the participants in the source domain.
domains (such as COGNITIVE AGENT), but they may also refer to a particular aspect of the source domains (such as PURPOSE of the participants). Huang, Ahrens, Chang, Chen, Liu & Tsai (2000) have tried to capture the different aspects of verbal semantics through a model called MARVS (Module-Attribute Representation of Verbal Semantics). This model emphasizes the argument structures of a lexical item, and in future work, we can also relate the source domains we found to this model.

In the following section, we will see more results of source domain determination. A more detailed examination of source domains will be provided in later discussion when all results have been procured.

### 8.2.3 Source Domains for the Coordinated Target Domains

For the Coordinated Target Domains, the source domains for each dataset are shown in Table 8.4. The shaded cells show the most frequently appearing source domains, making the similar source domains between SinicaBow and the Merged Word List apparent.

#### Table 8.4: Most Frequently Appearing Source Domains in the Coordinated Target Domains (with 經濟 jīngjì ‘economy’) in the Top-down Approach

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Source Domains (SinicaBow)</th>
<th>CNA Percentages in the Total Tokens of Metaphorical Expressions</th>
<th>Source Domains (The Merged Word List)</th>
<th>XIN Percentages in the Total Tokens of Metaphorical Expressions</th>
<th>Source Domains (SinicaBow)</th>
<th>Source Domains (The Merged Word List)</th>
</tr>
</thead>
<tbody>
<tr>
<td>政治 zhèngzhì ‘politics’</td>
<td>COGNITIVE AGENT</td>
<td>1.255 / 3.803 (33.09%)</td>
<td>AGENT</td>
<td>2.616 / 3.803 (68.79%)</td>
<td>COGNITIVE AGENT</td>
<td>AGENT</td>
</tr>
<tr>
<td>外交 wàixiáojī ‘foreign affairs’</td>
<td>COGNITIVE AGENT</td>
<td>97 / 216 (44.91%)</td>
<td>ORGANISM</td>
<td>177 / 216 (81.94%)</td>
<td>COGNITIVE AGENT</td>
<td>27 / 77 (35.06%)</td>
</tr>
<tr>
<td>財政 càizhèng ‘finance’</td>
<td>COGNITIVE AGENT</td>
<td>29 / 109 (26.61%)</td>
<td>ORGANISM</td>
<td>79 / 109 (72.48%)</td>
<td>COGNITIVE AGENT</td>
<td>30 / 116 (25.56%)</td>
</tr>
<tr>
<td>教育 jiàoyù ‘education’</td>
<td>COGNITIVE AGENT</td>
<td>63 / 90 (70.00%)</td>
<td>ORGANISM</td>
<td>80 / 90 (88.89%)</td>
<td>COGNITIVE AGENT</td>
<td>47 / 86 (54.65%)</td>
</tr>
<tr>
<td>Total Coordinated Target Domains</td>
<td>COGNITIVE AGENT</td>
<td>1.444 / 4.218 (34.25%)</td>
<td>COGNITIVE AGENT</td>
<td>2.491 / 3.130 (79.58%)</td>
<td>COGNITIVE AGENT</td>
<td>2.731 / 3.409 (80.11%)</td>
</tr>
</tbody>
</table>
Notice that the Coordinated Target Domains indicate coordination with 經濟 jìngjì ‘economy’; therefore, the results in Table 8.4 show the most frequently appearing source domains for the coordinated use of these four target domains with 經濟 jìngjì ‘economy.’ Examples are given in (3), illustrating the coordination between 經濟 jìngjì ‘economy’ and 財政 cāi zhèng ‘finance’ (see Footnote 9 of this chapter for the formation of conceptual metaphors).

(3) a. ECONOMY AND FINANCE ARE RECIPIENTS OF ACTIONS FROM COGNITIVE AGENTS

并 可能 使 重大 的 經濟 和 財政
bing kěnéngh ěr zhòngdà de jìngjì hé cāi zhèng
as well may cause important DE economy and finance
改革 步調 為 之 延遲 下來 (CNA)
guīgé bǔ diào wéi zhī yán chí xià lái
reform/reformation pace for this slow down come down
‘...and may cause the slowdown of the pace of the major economic and financial revolution’

b. ECONOMY AND FINANCE ARE ORGANISMS

巴 自治區 經濟 及 財政 面臨
bā zì zhì qū jìng jì jí cāi zhèng miàn lín
Paraguay municipality economy and finance face
重大 困難 (XIN)
zhòng dà kùn nán
important difficulty
‘The economy and finances of the municipal of the Republic of Paraguay are facing major difficulties.’

The example in (3a) shows the most frequently appearing source domain (COGNITIVE AGENTS) from SinicaBow, while the example in (3b) shows the most frequently appearing source domain in the Merged Word List (ORGANISMS), which is different from SinicaBow.

From Table 8.4, we can see that there are more dissimilar source domains yielded by the two knowledgebases, as compared to the results of the Single Target Domains in Table 8.1. An examination of the top ten source domains in the Coordinated Target Domains produces the following results, shown in Table 8.5:
Table 8.5: The Top Ten Source Domains Most Frequently Used for the Sum of the Coordinated Target Domains

<table>
<thead>
<tr>
<th>Source Domains</th>
<th>SinicaBow Tokens of Metaphorical Expressions under Each Source Domain</th>
<th>% (Overall Tokens of Metaphorical Expressions 7,627)</th>
<th>The Chinese-English Merged Word List Tokens of Metaphorical Expressions under Each Source Domain</th>
<th>% (Overall Tokens of Metaphorical Expressions 7,627)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLASS</td>
<td>5,141</td>
<td>67.41%</td>
<td>CLASS</td>
<td>7,231</td>
</tr>
<tr>
<td>PROCESS</td>
<td>3,408</td>
<td>44.68%</td>
<td>PROCESS</td>
<td>6,380</td>
</tr>
<tr>
<td>OBJECT</td>
<td>3,128</td>
<td>41.01%</td>
<td>SUBCLASS</td>
<td>5,793</td>
</tr>
<tr>
<td>SUBCLASS</td>
<td>2,514</td>
<td>32.96%</td>
<td>OBJECT</td>
<td>5,723</td>
</tr>
<tr>
<td>COGNITIVE AGENT</td>
<td><strong>2,275</strong></td>
<td><strong>29.83%</strong></td>
<td>COGNITIVE AGENT</td>
<td><strong>5,627</strong></td>
</tr>
<tr>
<td>PART</td>
<td>1,858</td>
<td>24.36%</td>
<td>AGENT</td>
<td>5,126</td>
</tr>
<tr>
<td>THING</td>
<td>1,785</td>
<td>23.40%</td>
<td>THING</td>
<td>5,054</td>
</tr>
<tr>
<td>STAGES</td>
<td>1,767</td>
<td>23.17%</td>
<td>PART</td>
<td>5,042</td>
</tr>
<tr>
<td>PURPOSE</td>
<td>1,738</td>
<td>22.79%</td>
<td>ORGANISM</td>
<td>5,003</td>
</tr>
<tr>
<td>ATTRIBUTES</td>
<td>1,505</td>
<td>19.73%</td>
<td>PURPOSE</td>
<td>4,862</td>
</tr>
<tr>
<td>AGENT</td>
<td><strong>1,282</strong></td>
<td><strong>16.81%</strong></td>
<td>ATTRIBUTES</td>
<td>4,636</td>
</tr>
<tr>
<td>LEGALITY</td>
<td>1,071</td>
<td>14.04%</td>
<td>MOVEMENT</td>
<td>4,484</td>
</tr>
<tr>
<td>ETIQUETE</td>
<td>1,071</td>
<td>14.04%</td>
<td>STAGES</td>
<td>4,206</td>
</tr>
<tr>
<td>MORALITY</td>
<td>1,071</td>
<td>14.04%</td>
<td>LOCATION</td>
<td>4,108</td>
</tr>
<tr>
<td>ARTIFACT</td>
<td>1,038</td>
<td>13.61%</td>
<td>PATIENT</td>
<td>3,703</td>
</tr>
<tr>
<td>QUALITIES</td>
<td><strong>1,015</strong></td>
<td><strong>13.31%</strong></td>
<td>ORGANIZATION</td>
<td>3,485</td>
</tr>
<tr>
<td>MODIFICATION</td>
<td>927</td>
<td>12.15%</td>
<td>INTENTIONAL PROCESS</td>
<td>3,399</td>
</tr>
</tbody>
</table>

** White shades are general keywords; grey shades are overlapped source domains; black shades are non-overlapped source domains.

Table 8.5 shows that COGNITIVE AGENT, AGENT and PURPOSE are shared source domains. However, since we mentioned previously in Chapter Six that the Coordinated Target Domains can only be compared to their respective Single Target Domains, Table 8.5 will only tell us that the source domains produced by the Coordinated Target Domains are not all the time similar to the Single Target Domains (except COGNITIVE AGENT and PURPOSE at the top of the list). In order to see the types of metaphorical expressions in each dataset, the following section will display the results.
8.2.4 Most Frequently Appearing Types of Metaphorical Expressions in Different Source Domains for the Coordinated Target Domains

In Table 8.6, the top two types of metaphorical expressions for each dataset are presented. Similarly, lightly shaded cells show similar results obtained through SinicaBow and the Merged Word List. Darker shades are examples of similar types of metaphorical expressions with different source domain names, proving that these types of metaphorical expressions appear frequently in both source domains.

When we compare the results of the Coordinated Target Domains in Table 8.6 with those of the Single Target Domains in Table 8.3, we see that most of the types of metaphorical expressions used by the Coordinated Target Domains are dissimilar to their respective Single Target Domains. For example, in the Single Target Domains in Table 8.3, 外交 wàijiāo ‘foreign affairs’ (second row) shows similar results for its top two types of metaphorical expressions across all cells, which are 努力 nǔlì ‘diligent/diligence’ and 途徑 tújìng ‘path,’ respectively. However, when we see its coordinate uses with 经济 jīngji ‘economy’ in Table 8.6, 外交 wàijiāo ‘foreign affairs’ appears most frequently with 制裁 zhìcái ‘impose.sanction,’ 手段 shǒuduàn ‘trick,’ 交流 jiāoliú ‘interflow’ and 建设 jiànsè ‘construct/construction.’ In fact, most of the types of metaphorical expressions in Table 8.6 are metaphorical expressions that are similar in each of the four Coordinated Target Domains with 经济 jīngji ‘economy.’
Table 8.6: The Top Two Types of Metaphorical Expressions in the Coordinated Target Domains (with 經濟 jìngjì ‘economy’) in the Top-down Approach

<table>
<thead>
<tr>
<th>Datasets</th>
<th>CNA</th>
<th>The Merged Word List</th>
<th>XIN</th>
<th>The Merged Word List</th>
</tr>
</thead>
<tbody>
<tr>
<td>政治 zhòngzhì ‘politics’</td>
<td>COGNITIVE AGENT 彰革 gāngé ‘reform/reformation’ (47.25%) 建設 jiànshè ‘construct/construction’ (17.12%)</td>
<td>AGENT 改革 gǎireg ‘reform/reformation’ (22.67%) 建設 jiànshè ‘construct/construction’ (8.21%)</td>
<td>COGNITIVE AGENT 影視 xiàngshì ‘terrain’ (23.38%) 秩序 zhìxiù ‘orderliness’ (21.08%)</td>
<td></td>
</tr>
<tr>
<td>外交 wàijiā ‘foreign affairs’</td>
<td>COGNITIVE AGENT 削減 xiāojì ‘impose sanction’ (34.02%) 政策 zhèngcè ‘trick’ (9.41%)</td>
<td>COGNITIVE AGENT 削減 xiāojì ‘impose sanction’ (52.94%) 建設 jiànshè ‘construct/construction’ (18.52%)</td>
<td>COGNITIVE AGENT 削減 xiāojì ‘impose sanction’ (24.29%) 建設 jiànshè ‘trick’ (22.86%)</td>
<td></td>
</tr>
<tr>
<td>財政 cái zhèng ‘finance’</td>
<td>COGNITIVE AGENT 改革 gǎireg ‘reform/reformation’ (55.17%)</td>
<td>ORGANISM 改革 gǎireg ‘reform/reformation’ (20.25%)</td>
<td>COGNITIVE AGENT 難難 kùn nán ‘difficulty’ (8.86%)</td>
<td>AGENT 困難 kùn nán ‘difficulty’ (24.27%) 影響 xiàngyǐng ‘terrain’ (17.48%)</td>
</tr>
<tr>
<td>教育 jiàoyù ‘education’</td>
<td>COGNITIVE AGENT 改革 gǎireg ‘reform/reformation’ (34.92%)</td>
<td>ORGANISM 改革 gǎireg ‘reform/reformation’ (27.50%)</td>
<td>COGNITIVE AGENT 改革 gǎireg ‘reform/reformation’ (38.30%)</td>
<td>MOVEMENT 改革 gǎireg ‘reform/reformation’ (23.68%) 建設 jiànshè ‘construct/construction’ (38.30%)</td>
</tr>
<tr>
<td>总协调总体目标</td>
<td>COGNITIVE AGENT 改革 gǎireg ‘reform/reformation’ (44.04%)</td>
<td>COGNITIVE AGENT 改革 gǎireg ‘reform/reformation’ (21.96%)</td>
<td>COGNITIVE AGENT 改革 gǎireg ‘reform/reformation’ (29.09%)</td>
<td>COGNITIVE AGENT 影響 xiàngyǐng ‘terrain’ (23.04%) 影響 xiàngyǐng ‘order’ (19.26%)</td>
</tr>
</tbody>
</table>
For example, in the Single Target Domains, 制裁 zhùcái ‘impose.sanction’ (not amongst the top two shown in Table 8.3) is found only with 外交 wàijiāo ‘foreign affairs’ and 經濟 jīngjì ‘economy,’ as exemplified in (4a) and (4b). When they are in coordination (4c), 制裁 zhùcái ‘impose.sanction’ is naturally used with both.

(4) a. 胡適 判斷 國際 外交 將 聯合 對 日
Hu-Shih judge international foreign.affairs will unite face Japan
制裁 (CNA)
zhùcái sanction
‘Hu-Shih ruled that the international foreign affairs will unite to impose sanctions on Japan.’

b. 美國 解除 對 南非 的 經濟 制裁 將
America remove face South.Africa DE economy sanction will
有助於 南非 的 經濟 復甦 (CNA)
yóuzhùyú nánfēi de jīngjì fùsū
will help South.Africa DE economy recovery
‘America will remove the sanctions imposed upon South Africa, and this will help the economy of South Africa to recover.’

c. 北韓 一旦 試射 飛彈 將 遭致
North.Korea once try shoot missile will cause
經濟 和 外交 制裁 (CNA)
jīngjì hán wàijiāo zhùcái
economy and foreign.affairs sanction
‘Once North Korea tries to shoot their missiles, sanctions will be imposed upon them in the economy and in foreign affairs.’

In the following section, we will elaborate on the top ten source domains for the overall data procured using the top-down approach.

8.2.5 Top Ten Source Domains for the Top-down Approach

In this section, we summarize the top ten source domains (shaded) from both SinicaBow and the Merged Word List. The darker shades are source domains that are
not repeated in either database. We can see that only several source domains are dissimilar among these two knowledgebases.

Table 8.7: The Top Ten Source Domains Most Frequently Used for the Sum of the Single+Coordinated Target Domains in the Top-down Approach

<table>
<thead>
<tr>
<th>Source Domains</th>
<th>SinicaBow Tokens of Metaphorical Expressions under Each Source Domain</th>
<th>% (Overall Tokens of Metaphorical Expressions 32,501)</th>
<th>The Chinese-English Merged Word List Tokens of Metaphorical Expressions under Each Source Domain</th>
<th>% (Overall Tokens of Metaphorical Expressions 32,501)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLASS</td>
<td>19,891</td>
<td>61.20</td>
<td>CLASS</td>
<td>29,681</td>
</tr>
<tr>
<td>PROCESS</td>
<td>16,181</td>
<td>49.79</td>
<td>PROCESS</td>
<td>26,356</td>
</tr>
<tr>
<td>COGNITIVE AGENT</td>
<td>11,149</td>
<td>34.30</td>
<td>SUBCLASS</td>
<td>25,361</td>
</tr>
<tr>
<td>PURPOSE</td>
<td>8,908</td>
<td>27.41</td>
<td>COGNITIVE AGENT</td>
<td>24,954</td>
</tr>
<tr>
<td>SUBCLASS</td>
<td>8,819</td>
<td>27.13</td>
<td>PURPOSE</td>
<td>23,156</td>
</tr>
<tr>
<td>OBJECT</td>
<td>7,619</td>
<td>23.44</td>
<td>OBJECT</td>
<td>22,766</td>
</tr>
<tr>
<td>PART</td>
<td>7,505</td>
<td>23.09</td>
<td>ORGANISM</td>
<td>22,666</td>
</tr>
<tr>
<td>THING</td>
<td>6,256</td>
<td>19.25</td>
<td>AGENT</td>
<td>21,850</td>
</tr>
<tr>
<td>STAGES</td>
<td>5,926</td>
<td>18.23</td>
<td>PART</td>
<td>21,714</td>
</tr>
<tr>
<td>ORGANISM</td>
<td>5,608</td>
<td>17.25</td>
<td>THING</td>
<td>21,261</td>
</tr>
<tr>
<td>AGENT</td>
<td>5,019</td>
<td>15.44</td>
<td>STAGES</td>
<td>19,606</td>
</tr>
<tr>
<td>ATTRIBUTES</td>
<td>4,683</td>
<td>14.41</td>
<td>MOVEMENT</td>
<td>19,431</td>
</tr>
<tr>
<td>INTENTIONAL PROCESS</td>
<td>4,487</td>
<td>13.81</td>
<td>INTENTIONAL PROCESS</td>
<td>19,188</td>
</tr>
<tr>
<td>PHYSICAL QUANTITY</td>
<td>4,408</td>
<td>13.56</td>
<td>ATTRIBUTES</td>
<td>17,965</td>
</tr>
<tr>
<td>PROPOSITION</td>
<td>4,168</td>
<td>12.82</td>
<td>PHYSICAL QUANTITY</td>
<td>16,900</td>
</tr>
<tr>
<td>ORGANIZATIONAL PROCESS</td>
<td>3,921</td>
<td>12.06</td>
<td>PATIENT</td>
<td>16,853</td>
</tr>
<tr>
<td>QUANTITY CHANGE</td>
<td>3,734</td>
<td>11.49</td>
<td>ARTIFACT</td>
<td>16,367</td>
</tr>
</tbody>
</table>

** White shades are general keywords; grey shades are overlapped source domains; black shades are non-overlapped source domains.

The top five source domains are similar, albeit in different order, in both SinicaBow and the Merged Word List. These source domains are COGNITIVE AGENT, PURPOSE,
Chapter 8: Results of the Top-down Approach

ORGANISM, STAGES and AGENT. We will return to these source domains after we obtain the results from the bottom-up approach.

In the following section, we outline the strengths and weaknesses of the top-down approach. The discussion of the strengths and weaknesses will help us to crystallize the answers to both hypotheses: i.e., that the lexical and computational method will help to reduce human subjectivity; and that each of the top-down and the bottom-up approaches has inherent advantages and disadvantages.

8.3 Strengths and Weaknesses of the Top-down Approach

The strengths and weaknesses of the top-down approach can be summarized, as seen in Table 8.8.

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economic (one conceptual knowledge structure for all)</td>
<td>Leaves out important regional differences</td>
</tr>
<tr>
<td>Measurement of distance/conceptual similarity is allowed on a conceptual hierarchy</td>
<td>The measurement is possibly ‘tree-dependent’</td>
</tr>
<tr>
<td>Replicability (integration of lexical knowledgebases into linguistic research)</td>
<td>Lack of resources for the mapping to SUMO (problem of language transfer for different lexical knowledgebases)</td>
</tr>
<tr>
<td>Evaluation of SinicaBow and the Merged Word List is allowed</td>
<td>Levels of noise are not controlled (especially for the Merged Word List)</td>
</tr>
</tbody>
</table>

Each point in Table 8.8 will be elaborated in the following sub-sections.

8.3.1 Being Economic versus Regional Differences

The use of the ontology is economical because it does not differentiate between Taiwan (CNA) and China (XIN) data. Therefore, the same types of metaphorical expressions (such as 成長 chéngzhǎng ‘grow/growth’) will be grouped under the same source domains for data from both Taiwan and China. The rationale behind the use of ontology is that there are basic concepts shared by most humans, regardless of the communities they live in. 成長 chéngzhǎng ‘grow/growth,’ therefore, forms the concept of ‘Growth,’ which should be present in all languages.

This assumption, however, leaves out the important features of different communities that are often emphasized in sociolinguistic and anthropological studies.
Therefore, one limitation when using ontology is that it will be challenged by studies that value societal differences. Furthermore, there is also the question of ‘coverage,’ i.e., even though SUMO has been put together by speakers of different languages, and it has been discussed in great detail, it still does not have sufficient depth. Even though there might be one representative of speakers in each language, the ‘representativeness’ of this speaker may still be challenged.

8.3.2 Measurement of Conceptual Similarity versus ‘Tree-independence’

This book proposes an innovative way of measuring conceptual similarity between different concepts. We compare each SUMO node to several prototypical concrete concepts that have been rated by human subjects in order to produce their similarity scores. Applying this measurement of conceptual similarity to metaphor research was carried out for the first time ever in literature. We do, however, want to outline its possible weakness so as to improve on our methodology.

The measurement of conceptual similarity is heavily based on conceptual distance, i.e., the immediate node shared by two nodes and the distance of this immediate node to the root. Since only the SUMO ontology is used, we cannot compare the results against a second ontology. The measurement of conceptual similarity, thus, may be ‘tree-dependent.’

8.3.3 Replicability versus Lack of Lexical Knowledgebases

Since the top-down approach is carried out using a principled methodology, it should be replicable. As described in earlier chapters, most metaphor studies use different criteria for determining source domains. Replicability, therefore, historically has been a problem in most previous studies, as no consensus has been reached as to how to standardize the way source domains are determined. However, in the top-down approach, we clearly lay out the steps needed and, therefore, the study can be replicated. In this way, the approach should work for languages other than Chinese, as long as these languages have been mapped to the English WordNet. The only problem is that these languages may face difficulties with ‘language transfer’ between databases, similar to those that occurred with the Chinese-English translation in this chapter. When language transfer is not needed (such as when applying the top-down approach to the English language or using the newly created Chinese WordNet (Huang [黃居仁] 2004a, 2005, 2006 and 2007)), which is still in progress, the top-down approach is expected to perform even better.
Nevertheless, one possible problem that others may have encountered is lack of resources. For example, SinicaBow and the Merged Word List are used so that search queries in Chinese terms can be linked to SUMO. Where other languages are concerned, one must first possess at least one of the following:

(5) a. A translated English WordNet to one’s target language (like SinicaBow).

OR

b. An electronic bilingual dictionary translating target language to English, so that the English translation can be linked to WordNet (then SUMO) through readily available lexical knowledgebases, such as the KSMSA project by Ševčenko (2003) (like the Merged Word List).12

If none of the resources listed in (5) are available, the replication of the top-down approach may become problematic. Therefore, for other languages, the availability of resources becomes a crucial component in making this approach a successful one.

8.3.4 Evaluation of Resources versus Noise

The top-down approach compares two knowledgebases (SinicaBow and the Merged Word List) in terms of their success in determining source domains. This is the first study of its kind, one that attempts to evaluate these two knowledgebases based on issues concerning conceptual metaphors. This study is innovative and an important addition to the field of the application of lexical knowledgebases in linguistic research.

However, the evaluation of the two knowledgebases cannot be treated on par. This is because one of the knowledgebases (the Merged Word List) introduces noise into the results that may influence the evaluation. The following example shows noise created by the Merged Word List for the type of metaphorical expression of 棋 qi ‘chess.’ 棋 qi ‘chess’ appears in the translation of the three English words ‘chess,’ ‘chess_game’ and ‘solitaire,’ shown in (6).

(6) 棋 qi → ‘chess’
→ ‘chess_game’
→ ‘solitaire’
The last meaning of ‘solitaire’ contains three senses: (a) a single gem; (b) card games; and (c) extinct flightless birds. Therefore, when ‘solitaire’ is mapped to WordNet, all these senses will be retrieved, and they are all found to be concrete senses. As a result, the source domains for 棋 qi ‘chess’ contain not only GAME and CONTEST but also CORPUSCULAR OBJECT and PRODUCT as well as VERTEBRATE. Even though these are usually low in frequency, rumblings that usually drop out because they do not follow the trend, they should be reduced where possible in order to strengthen the evaluation of the two knowledgebases. The Merged Word List has introduced a large amount of data into the analysis, indirectly increasing the possibility of finding recurring concrete keywords (i.e., the source domains). However, the noise created by this methodology may cause problems in the evaluation of the two knowledgebases.

8.4 Summary of Chapter

In this chapter, we have compared the overall performance of the top-down approach in terms of the percentages of source domains it can determine within different datasets. These percentages are high. In the top-down approach, more than 90 per cent of all the found types of metaphorical expressions have their source domains determined using either SinicaBow or the Merged Word List (see Figures 8.2 and 8.3).

For a comparison of the types of metaphorical expressions that have had their source domains determined, we have contrasted SinicaBow and the Merged Word List in the top-down approach (which is also, by proxy, an evaluation of these two knowledgebases). The conclusion asserts that these two knowledgebases show (slightly) different collections of types of metaphorical expressions. Some of the types of metaphorical expressions found in SinicaBow are not found in the Merged Word List (and vice versa). Examples of types of metaphorical expressions that are found in SinicaBow only include 成長率 chéngzhǎnglǜ ‘rate of growth,’ 強大 qiángdà ‘big and powerful,’ 土石流 tǔshíliú ‘mudflows and landslides,’ 扎根 zhāgēn ‘take root,’ 創紀錄 chuàngjìlù ‘break record,’ etc. This indicates that these two knowledgebases are not similar in terms of their collection of lexical items.

In this chapter, too, we have compared the most frequently appearing source domains and the top two types of metaphorical expressions found under each target domain. We have also examined the top ten source domains from the Single Target Domains (Table 8.2), Coordinated Target Domains (Table 8.5) and overall top-down approach (Single+Coordinated Target Domains; Table 8.7) and found that COGNITIVE AGENT and PURPOSE are frequently appearing source domains in the top-down approach, indicating that personification is present in almost all target domains. There are also other source domains that appear within different target domains. The results of
the Coordinated Target Domains are not necessarily similar to the Single Target
Domains, which also confirms our previous discussion seen earlier in Chapter Six
regarding Single Target Domains versus Coordinated Target Domains.

This chapter also discusses the strengths and weaknesses of the top-down approach,
and this discussion can be related to both hypotheses of the book. We have found that
top-down approaches use lexical and computational methods, which reduce human
subjectivity when determining source domains, and have also determined that the
top-down approach has inherent advantages and disadvantages. Chapter Nine provides the
steps requisite to carrying out the bottom-up approach, and Chapter Ten subsequently
discusses the results of the bottom-up approach.
Chapter 9: Bottom-up Approach to Source Domain Determination

In terms of taking a bottom-up approach, this chapter emphasizes the importance of collocation in metaphor analysis. Previous studies that have examined this aspect of metaphor analysis include Deignan (1999, 2005) and Stefanowitsch (2005, 2006). In particular, Stefanowitsch (2006) argues that the literal meanings of the metaphors can be identified when a metaphor's target domain terms are replaced by its source domain terms. For example, in the sentence ‘He *shot down* all of my arguments,’ ‘arguments’ can be replaced by words such as ‘planes’ and ‘missiles’ to form the literal meaning of the sentence.1

A similar method also proves workable with Chinese metaphors, as illustrated in (1), where examples of metaphorical uses of 經濟 *jīngjí ‘economy’* with three types of metaphorical expressions (成長 *chéngzhǎng ‘grow/growth,’ 起飛 *qǐfēi ‘take off’ and 麻痺 *mábì ‘paralytic’) can be seen.

(1) a. 中銀 集團 與 港澳 經濟 zhōngyín jítuán yǔ gāngào jīngjí Bank.of.China group and Hong.Kong.Macau economy 一起 成長 (XIN) yìqǐ chéngzhǎng together grow ‘The Bank of China Group will grow simultaneously with the economy of Hong Kong and Macau.’

b. 以免 蘇聯 經濟 麻痺 (CNA) yǐmiǎn sūlián jīngjí mábì to.avoid Soviet.Union economy paralytic ‘In order to avoid the economy of the Soviet Union becoming paralytic…’

---

1 Deignan (2005:219-220) has made an observation regarding ‘all guns blazing’ and ‘shoot down in flames.’ She thinks that these expressions are used more often as metaphorical meanings than literal meanings. Literal uses, according to her, usually appear in different order, such as ‘flames were fanning outwards.’ However, an in-depth explanation is probably needed for this phenomenon.
In order to ascertain whether the replacement of target domain can indeed yield the literal meanings of these sentences, the following tests are carried out in order to see which terms can successfully replace the target domain of 經濟 jingjì ‘economy’ in each sentence. For example, in (1a) the target domain of 經濟 jingjì ‘economy’ in 經濟成長 jingjì chéngzhǎng ‘economy grows’ can possibly be replaced by stating 小孩成長 xiǎohái chéngzhǎng ‘children grow’ as well as 小貓成長 xiǎomāo chéngzhǎng ‘kitties grow’ (by intuition). Both 小孩 xiǎohái ‘child’ and 小貓 xiǎomāo ‘kitty’ are literal uses with 成長 chéngzhǎng ‘grow.’ Their literal meanings have replaced the metaphorical meanings of 經濟 jingjì ‘economy.’ Similarly, for (1b), 經濟癱瘓 jingjì tānhuān ‘the economy is paralytic’ can be replaced by 老人癱瘓 lǎorén tānhuān ‘the old man is paralytic’ as well as 病人癱瘓 bìngrén tānhuān ‘the patient is paralytic.’ As for (1c), 經濟 jingjì ‘economy’ can be substituted by 飛機 fēijī ‘airplane’ as well as 蝴蝶 húdié ‘butterfly.’ These replaced terms are possible collocates that provide clues as to what the source domains could be for these metaphors.

However, it is necessary that the replaced terms in the examples above are not only literal collocates, but that they also appear in particular grammatical relations to the target domains—for example, all of the examples in (1) have 經濟 jingjì ‘economy’ as the subject of their metaphorical expression. Replaced terms that are grammatically functional, in addition to being literal collocates, can be seen in other examples like 損害 jīnghài ‘to hurt the economy’ where 經濟 jingjì ‘economy’ is the ‘object’ of 損害 shānghài ‘hurt.’ Therefore, the grammatical positions of target domains appear to have some bearing when this approach is employed, in order to see the possible literal meanings that can replace 經濟 jingjì ‘economy’.

In order to determine the patterns of collocates that are arranged in different grammatical relations with 經濟 jingjì ‘economy,’ the saliency lists from Sketch Engine become of great use, first, because all collocates are arranged in this way in Sketch Engine, and second, because the construction information has been provided in the grammatical relations portion of Sketch Engine. Therefore, in order to find literal terms that can replace 經濟 jingjì ‘economy’ in a metaphorical phrase, such as 經濟癱瘓 jingjì tānhuān ‘the economy is paralytic,’ literal collocates that appear in the same grammatical relation as 經濟 jingjì ‘economy’ can be sought out in the Chinese Sketch Engine. If 癱瘓 tānhuān ‘paralytic’ can also be used in other relations (such as 癱瘓 tānhuān ‘paralytic’ as ‘modifier’ in 癱瘓的經濟 tānhuān de jingjì ‘economy that is paralytic’),
more than one grammatical relation will be collected for each type of metaphorical expression. The following section will proceed with an overview of the steps needed to enact a bottom-up approach.

9.1 A Bottom-up Approach to Source Domain Determination

Unlike what has been accomplished using the top-down approach in Chapters Seven and Eight, this chapter will take a different tack to determine source domains for the same types of metaphorical expressions used in the top-down approach of previous chapters. Figure 9.1 illustrates the requisite steps for the bottom-up approach (extracted from Figure 4.9 in Chapter Four) to be employed in both this chapter and in Chapter Ten.

![Figure 9.1: Steps Involved in the Bottom-up Approach to Source Domain Determination](image)

2 Some of the grammatical relations in the Chinese Sketch Engine are based on English grammar, which has applied to Chinese incorrectly. However, most of the results and collocates are correct. More discussion regarding this can be found in later sections of this chapter and in Chapter Ten.
As mentioned, collocates collected from the Chinese Sketch Engine for all types of metaphorical expressions will be considered in terms of determining possible source domains. The first step, shown in (B1) of Figure 9.1, is to extract different grammatical relations in the Chinese Sketch Engine that contain the target domain terms (any of the five target domains of 經濟 jīngjì ‘economy,’ 財政 cáizhèng ‘finance,’ 外交 wàijīào ‘foreign affairs,’ 教育 jiàoyù ‘education’ and 政治 zhèngzhì ‘politics’) to be collected automatically. After this has been completed, step two (B2) will then divide the collocate lists into significant and non-significant ones in order to reduce the number of collocates that are of no significance and will only create noise in the analysis. Step three (B3) involves the clustering of the selected collocates according to their WordNet synsets. After these collocates have been clustered, their source domain names will be determined by searching for shared WordNet hypernyms within the clusters (B4). Each of these steps will be discussed in this chapter.

The naming of source domains in the top-down approach and the bottom-up approach differ markedly in that the bottom-up approach selects source domain names after the clustering of all the types of metaphorical expressions has been carried out. In contrast, the top-down approach, as it has been seen, first determines the source domain names before the groupings of the types of metaphorical expressions can be completed. In addition, these two approaches differ in their use of lexical knowledgebases. The top-down approach uses an upper ontology that is more general than the bottom-up approach, which uses the WordNet hierarchy and is more specific. Therefore, the directions of the two approaches are entirely different: one works from the pre-determined source domains while the other works from collocations of metaphorical expressions. More comparisons of the two approaches will take place in later discussion. In the following section, however, the steps for the bottom-up approach will first be discussed.

9.2 Extraction of Collocates from the Chinese Sketch Engine

As mentioned in Chapter Six, a combined total of 1,591 types of metaphorical expressions have been collected, 1,282 from CNA and 736 from XIN, respectively. These types of metaphorical expressions have been collected on the basis of the analyses of five target domains (經濟 jīngjì ‘economy,’ 政治 zhèngzhì ‘politics,’ 財政 cáizhèng ‘finance,’ 外交 wàijīào ‘foreign affairs’ and 教育 jiàoyù ‘education’). Each of these types of metaphorical expressions forms a construction with its respective target domain. For example, 傷害 經濟 shāngài jīngjì ‘to hurt the economy’ has the construction of [Metaphor Type (verb)+Target Domain (object)] while 經濟成長 jīngjì chéngzhǎng ‘economy grows’ has the construction of [Target Domain (subject)+Metaphor Type

3 All programs used in this chapter were prepared by Petr Šimon.
Chapter 9: Bottom-up Approach to Source Domain Determination

(verb)]. The same constructions, save for the fact that literal terms replace the target domains, may be found by searching for the collocates at the same ‘subject’ position with 經濟 jīngjì ‘economy.’ When collocates of 成長 chéngzhāng ‘grow/growth’ are searched for in the Chinese Sketch Engine, different grammatical relations are found (see Figure 9.2).

For example, Figure 9.2 shows the various grammatical relations of collocates for 成長 chéngzhāng ‘grow/growth.’ Among these relations, some grammatical relations are not needed for the source domain analysis. For instance, in Figure 9.2, only the relation of ‘subject’ is required, as the target domain of 經濟 jīngjì ‘economy’ forms constructions with only 成長 chéngzhāng ‘grow/growth’ in this relation. The other relations will not be extracted.

**Figure 9.2:** Collocates of 成長 chéngzhāng ‘grow/growth’ in the Chinese Sketch Engine

---

4 This figure shows the snapshot of only the first half of 成長 chéngzhāng ‘grow/growth.’
One way to filter out the unwanted relations is to search for whether 經濟 jǐng jì ‘economy’ appears in any of the collocates of any relations. Once 經濟 jǐng jì ‘economy’ is spotted, the relations containing 經濟 jǐng jì ‘economy’ will be extracted, and those that do not contain it will be filtered out. It is also necessary to search for all the grammatical relations of 成長 chéng zhǎng ‘grow/growth’ in order to discover what other collocates also appear when 經濟 jǐng jì ‘economy’ appears. For instance, the collocates for all the ‘subjects’ of 成長 chéng zhǎng ‘grow/growth’ displayed in Figure 9.2 also includes 經濟 jǐng jì ‘economy’ (arrow), indicating that this relation will be extracted and its collocates analyzed.

As mentioned in Chapter Six, the data for CNA and XIN comprise data from Single Target Domains and Coordinated Source Domains, respectively. The combination of both sets of data in CNA (Single+Coordinated) and XIN (Single+Coordinated) is employed solely for the economy of computation (but never CNA+XIN for the bottom-up approach). When the source domains have been decided in this chapter, a discussion will be pursued in terms of Single Target Domains and Coordinated Target Domains in CNA and XIN, respectively, in Chapter Ten. In the top-down approach, where conceptual similarity is compared, the combination of CNA+XIN (including Single Target Domains and Coordinated Target Domains) was simultaneously examined because a conceptual approach does not differentiate between Mandarin in Taiwan and China. In the bottom-up approach in this chapter, however, the combined data of CNA+XIN cannot be achieved because the extraction of collocates from the Chinese Sketch Engine is dependent on sub-corpora. Therefore, using the methodology mentioned previously, types of metaphorical expressions from CNA will be searched only in the sub-corpus of CNA, and types of metaphorical expressions from XIN will be searched only in the sub-corpus of XIN. This is because collocates extracted from CNA and XIN may differ depending on the sub-corpora, and this is one way whereby CNA and XIN can be compared first in terms of collocates and later in terms of source domain determination. Regarding the collection of collocates, only the ones that contain any of the above specified target domains will be collected. The next section will discuss some of the results produced from this method.

9.2.1 Types of Metaphorical Expressions Found in the Chinese Sketch Engine

During the process of collecting grammatical relations from the Chinese Sketch Engine, it was discovered that certain types of metaphorical expressions that cannot be found exist due to reasons that will be explained in §9.2.2. The percentages of types of
metaphorical expressions found with target domains among their various collocates are broken-down in terms of CNA and XIN in Table 9.1.

**Table 9.1:** Overall Types of Metaphorical Expressions and Tokens of Metaphorical Expressions That Are Found in the Chinese Sketch Engine

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Found_Types / Total_Types</th>
<th>%</th>
<th>Found_Tokens / Total_Tokens</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNA</td>
<td>1,228 / 1,282</td>
<td>95.79</td>
<td>18,158 / 18,293</td>
<td>99.26</td>
</tr>
<tr>
<td>XIN</td>
<td>703 / 736</td>
<td>95.52</td>
<td>14,022 / 14,208</td>
<td>98.69</td>
</tr>
</tbody>
</table>

Table 9.1 (shaded columns) shows that more than 96 per cent of the searched types of metaphorical expressions can be found using the methodology mentioned previously, i.e., by searching for the target domains within the grammatical relations of the types of metaphorical expressions. The corresponding tokens of metaphorical expressions that have been found constitute about 99 per cent of total metaphorical tokens in CNA and XIN, meaning that the remaining 4 per cent of types of metaphorical expressions constitute only 1 per cent of the total tokens. This discovery serves to confirm previous findings in Chapter Seven that the items not found are, in fact, low frequency items.

The observations regarding the collocating relations between the target domains and the metaphorical expressions are, in point of fact, based on corpora data from the Chinese Gigaword Corpus. The advantage of using the Chinese Sketch Engine here is that the results are then based on the whole Chinese Gigaword corpus. Therefore, through observations made using a sampling of the corpus, overall generalizations can be made by a careful analysis of the relationship between the target domains and the types of metaphorical expressions.

### 9.2.2 Reasons for Types of Metaphorical Expressions That Are Not Found

The reasons for the majority of types of metaphorical expressions not being found in the Chinese Sketch Engine can vary, whereas the not-found cases in SinicaBow and the Merged Word List are mainly due to a lack of words. However, there are three main

---

5 Similar with the top-down approach, only types of metaphorical expressions that are found will be further analyzed for source domain determination. It may be seen that more percentages of found cases are present in the bottom-up approach because all types of metaphorical expressions that have been taken from the Chinese Gigaword corpus and Sketch Engine should naturally contain these expressions. It would, therefore, be unfair to compare the top-down approach and the bottom-up approach if their found cases are not similar. This is the reason why Chapter Eight (and, later, Chapter Ten) should consider that the found source domains will be measured against the found cases, rather than the total instances of metaphorical expressions in CNA and XIN.
reasons for types of metaphorical expressions not being found in the Chinese Sketch Engine, namely: technical problems, construction problems and segmentation problems. Technical problems refer, in particular, to not having enough corpora instances for a particular word to form Wordsketches that are statistically meaningful. Construction problems, on the other hand, arise when constructions that cannot be broken down into smaller units with the same meanings occur. One example is given in (2).

(2) 經濟 和 政治 要 兩手硬 (XIN)
经济 和 政治 要 two-hand-hard
‘The economy and politics need to be firm.’

In (2), the grammatical relation for the whole phrase 兩手硬 liàngshòuyìng ‘to be firm’ cannot be found. If only 硬 yìng ‘hard’ is searched in the Chinese Sketch Engine, the target domains of 經濟 jìngjì ‘economy’ and 政治 zhèngzhì ‘politics’ are not found among their collocates because they are not in direct relation with 硬 yìng ‘hard.’ Therefore, this metaphor type cannot be found in the Chinese Sketch Engine.

The third reason for cases not being found consists of those types of metaphorical expressions that do not have the target domains in their grammatical relations. This is because when the types of metaphorical expressions are manually analyzed, their relation with the target domains is not always clear. An example of this is shown in (3).

(3) 把 經濟 從 衰退 的 泥漬 中 拔出來 (XIN)
把 經濟 從 衰退 的 泥漬 中 拔出来 BA economy from dysfunction DE mud middle pull.out.come
‘To pull the economy out from the decaying mud...’

In this example, the grammatical relation between 經濟 jìngjì ‘economy’ and 拔出来 báchūlái ‘to put out’ cannot be specified because 拔出来 báchūlái ‘to put out’ is split in two different places of the sentence. The Chinese Sketch Engine sometimes encounters such segmentation problems due to the complexity of the constructions (衰退 shuāituì ‘dysfunction’ is not counted here because it modifies 泥漬 nītàn ‘mud’ and not 經濟 jìngjì ‘economy’). Therefore, even when 拔出来 báchūlái ‘to put out’ is found in manual analyses, it cannot be found when searched in the Chinese Sketch Engine.

Cases that are not found constitute less than 5 per cent in the overall types of metaphorical expressions, and they should not affect the overall patterns in the Chinese Sketch Engine. The following will show examples of the grammatical relations collected for the types of metaphorical expressions.
9.2.3 Selected Grammatical Relations from the Chinese Sketch Engine

Some types of metaphorical expressions will have more grammatical relations than others if the target domains are found in several relations. As mentioned, if a metaphor such as 瘫痪 tānhuàn ‘paralytic’ can appear in different grammatical relations with the target domain of 經濟 jīngjí ‘economy,’ more than one grammatical relation will be selected. Table 9.2 details the grammatical relations in which the different target domains are found. The words in the left column list the types of metaphorical expressions collected manually in Chapter Six. The first row on top contains the five target domains, and under each target domain the grammatical relations in which the target domains are found are listed.6

Table 9.2: Grammatical Relations in Each Type of Metaphorical Expression That Contain the Five Target Domains7

<table>
<thead>
<tr>
<th>成長 chéngzhǎng ‘grow/growth’</th>
<th>經濟 jīngjí ‘economy’</th>
<th>政治 zhèngzhì ‘politics’</th>
<th>財政 cāizhèng ‘finance’</th>
<th>外交 wàijiāo ‘foreign affairs’</th>
<th>教育 jiàoyù ‘education’</th>
</tr>
</thead>
<tbody>
<tr>
<td>subject</td>
<td>pp_ 在</td>
<td>pp_</td>
<td>pp_</td>
<td>pp_</td>
<td>pp_</td>
</tr>
<tr>
<td>政治 zhèngzhì ‘politics’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>財政 cāizhèng ‘finance’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>外交 wàijiāo ‘foreign affairs’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>教育 jiàoyù ‘education’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 9.2 shows the relations containing any of the five target domains as one of their collocates in their different relations. In Sketch Engine, each grammatical relation has a saliency value listed in descending order, from the most salient to the least salient. The lists are long, with many collocates, including significant ones and non-significant ones.

6 Since the data file for the grammatical relations is too large, the full list will not be provided in the appendix.

7 起飛 qǐfēi ‘take off’ and 成長 chéngzhǎng ‘grow/growth’ can also serve as nouns, but their occurrences may not be enough to produce Wordsketches. Therefore, grammatical relations related to their uses with the target domains as ‘nouns’ are not found.
In order to include only significant collocates in each list, the saliency lists will be cut into significant and non-significant lists. The next step after this, therefore, is to compute the cut-off points for each grammatical relation.

**9.3 Computing Cut-off Points**

In this step, all saliency lists are programmed, and their cut-off points are computed automatically. The purpose of computing cut-off points for the saliency lists is to find significant collocates for each grammatical relation to determine the source domain. In general, finding cut-off points for lists of results (such as frequency, saliency or Mutual Information (MI) values) is important in linguistic research. This is because most lexical resources, including British National Corpus, Academia Sinica Corpus of Mandarin Chinese (Chen, Huang, Chang & Hsu 1996) and Sketch Engine only provide either word lists or lists of collocates in descending order. Most people will only look at the top few and ignore the rest. There is no clear criterion for the selection of what constitutes a top result. As a result, different people include a different number of top results in their analyses.

This book will show how cut-off points can be computed based on several methods.

**9.3.1 Methods Based on Distributional Data of Saliency**

When researching cut-off points, three methods based on the distributional data, such as that in Figure 9.3, are of particular use.

![Figure 9.3: Patterns of Distributional Data](image)

The nonlinear regression featured in Figure 9.3 follows the pattern of Zipf’s law (Zipf 1932), which says that the most frequent value is most likely to be twice as much as the
second most frequent value. The purpose of knowing this distributional pattern is to find the correct mathematical function for this kind of distribution. For the pattern in Figure 9.3, the mathematical function is shown in (4), where \( f(x) \) is the value for the y-axis on the curve; it can be found using the following formula:

\[
(4) \quad f(x) = b(x^a)
\]

This also means that any point on the curve will be \((x, f(x))\) where \(f(x)\) represents the value on the y-axis.\(^8\)

When saliency lists in Sketch Engine are plotted, they are discovered to have almost identical patterns to those found in Figure 9.3. For instance, the results for ‘subject’ for the verb 起飛 qīfēi ‘take off’ in the CNA in the Chinese Sketch Engine are shown in Figure 9.4.

---

\(^8\) The outcome of this section is the product of discussions I had with Professor Chung-Ping Cheng, Professor Kathleen Ahrens, Professor Chu-Ren Huang and Petr Šimon.
In this saliency list, the distribution for the Chinese collocates (column one) versus the
saliency values (column three) can be seen in Figure 9.5. The Chinese collocates are
replaced by Chinese word rank on the x-axis because the first Chinese collocate in Figure
9.4 also means that this word lies at the top ranking in the list (thus, rank 1, 2, 3,…, etc.).
The y-axis represents the saliency values.

![Figure 9.5: Distributional Pattern for ‘Subjects’ of 起飛 qǐfēi ‘take off’ in CNA in the
Chinese Sketch Engine](image)

This also means that any point on the curve will be \((x, f(x))\) where \(x\) is the rank of the
Chinese collocates and \(f(x)\) is the function to calculate the value on the y-axis. Using
this formula, any positions for the curve can be calculated and the cut-off points will
also be a particular point on the curve (for example, the point of \((w, z)\) in Figure 9.6, to
follow) that can be found. This point demarcates the point where the curve is divided
into two sections, separating the significant list of collocates from the non-significant
list of collocates, where any collocates above this point are significant.

Based on the distribution data, there are three ways that the cut-off points in \((w, z)\)
can be found, as given in (5).

(5) **Method One:** Calculate the shortest distance from the start point \((0, 0)\)
(line 1 in Figure 9.6);

**Method Two:** Find the slope with the value -1 between the x-axis and
the y-axis (line 2 in Figure 9.6);

**Method Three:** Find a baseline for the co-occurrence of a target domain
and a metaphor type that are not in a grammatical relation
(line 3 in Figure 9.6).
Methods One and Two are based on a mathematical model. Both of these methods aim to find a point on the curve where there is an obvious change between the values. These two methods can be seen in Figure 9.6 to follow, where the y-axis shows the saliency values and the x-axis shows the rank of Chinese words from highest to lowest.

Method One can be carried out by calculating the position of \((w, z)\), where \((w, z)\) is the shortest distance from \((0, 0)\). This is because when every line departs from the starting point of \((0, 0)\), there will be a line that is the shortest distance from the curve. The point where this line touches the curve is the point where the curve changes the most from the y-axis to the x-axis.\(^9\)

On the other hand, Method Two calculates the slope with a value of -1 between the x-axis and the y-axis, because this is the slope that is most slanted. When the slope is -1, there is a high possibility where the curve changes the most at a certain point \((w, z)\). This is because the higher the curve on the y-axis, the more vertical the slope will be. Moreover, the further the curve moves away from \((0, 0)\) on the x-axis, the more horizontal the slope will be. Therefore, the most slanted slope between the vertical and horizontal axis will be the possible cut-off point representing where the curve has changed the most.\(^{10}\) The following section explains the actual procedure undertaken to compute the cut-off points.

In Method Three, the occurrence of the target domain and the metaphor type (within a certain distance) is random, and only the points in Figure 9.6 that are above the baseline will be considered significant. However, during the programming of the third method, a technical problem was encountered, whereby insufficient information was available for a certain variable (i.e., the number of grammatical relations in the whole Chinese Sketch Engine), crucial to the success of this method.\(^{11}\) Therefore, in this chapter, we discuss the results of only Methods One and Two.

\[^{9}\] The formula for Method One is: 
\[
i = \left[ \left( -ab \right)^{2} \left( \frac{1}{2 - 2a} \right) \right]; \text{ where } a \text{ and } b
\]
are the variables in the function of the nonlinear regression \(y = bx^{a}; i\) is the cut-off points; and \(n\) is the total number of collocates in the relation. For details of these two formulas, see Appendix A9.1.

\[^{10}\] The formula for method Two is: 
\[
i = \left[ \left(-ab\right)^{\frac{1}{1-a}} \right]. \text{ See Appendix A10.1.}
\]

\[^{11}\] This problem has been mentioned to Adam Kilgarriff in personal communication.
Note, however, that Methods One and Two have a common problem, i.e., their results may be affected by how we normalize the values on both axes. In Figure 9.6, the scales of ‘Rank of Chinese Collocates’ and ‘Collocates’ are not yet comparable. ‘Rank of Chinese collocates’ on the x-axis is a constant variable with numbers from 1 to n. These numbers are only indicators of ranks; they do not indicate calculable values and, therefore, cannot be included in the calculation of cut-off points. One way to solve this problem is to transform the constant values of the x-axis to normalize the data. The following sub-section will elaborate on this transformation.

9.3.1.1 Transformation of Data

The normalization of results has both its advantages and its disadvantages. One of the advantages of normalizing data is that the scales will become more parallel, as both are in decimal. However, the disadvantage to normalization is that the data will be transformed to a certain extent by this process. For example, the transformed data for Figure 9.6 will probably look like the dotted line in Figure 9.7 (labeled as ‘normalized data’).
The normalized data will have a different distance from (0, 0) (Method One) and different measures of slope (Method Two) as well as different baseline results (Method Three). The following will show some results using Methods One and Two based on transformed data.

**9.3.1.2 Results of Methods One and Two**

Normalization involves different calculations. The calculation process used in Methods One and Two of this book is one of the ways of transforming data. The normalization measures are given in (6).

\[
(6) \begin{align*}
\text{Y-axis:} & \quad \frac{\text{Collocate}_{\text{Rank1}, \ldots, \text{Rank}n}}{\text{Rank}_n} \\
\text{X-axis:} & \quad \frac{\text{Saliency}_{1, n}}{\text{Sum}(\text{Saliency}_1, \text{Saliency}_2, \ldots, \text{Saliency}_n)}
\end{align*}
\]

For the y-axis (saliency values), each collocate from rank 1 to \( n \) will be divided by rank, from highest to lowest. For example, if a Chinese word has 200 collocates in a particular relation, the normalization will divide collocates ranked 1 to 200 with 200 (thus, \( \frac{1}{200}, \frac{2}{200}, \ldots, \frac{200}{200} \)). Therefore, the output of the y-axis is a list of numbers ranging from 0 to 1. As for the x-axis, each saliency value will be divided by the sum of all 200 saliency values. The output of the x-axis is also displayed on a scale ranging from 0 to 1 (which is also the percentage of the saliency values).
Based on the normalization formula in (6), each saliency list will be calculated based on the formula of cut-off points given in Footnotes 9 and 10, previously. A trial of these two methods found the results detailed below.

For all the grammatical relations, their cut-off points have been found to lie between words number one and six, indicating that in the majority of grammatical relations, the most salient collocates usually appear within the first six words only. Examples of results for collocates in CNA (collocates of CNA and XIN are separated because different regions may use different collocates) are shown in Table 9.3. In this table, the first column shows the types of metaphorical expressions, followed by their relations in the second column. The third column shows the total collocates that the relations possess. ‘Pseudo-R-Square’ in column four shows the percentages of the curve that fit the non-linear regression (or, in colloquial term, ‘curve fitting’).

### Table 9.3: Calculation of Cut-off Points Using Methods One and Two (CNA)

<table>
<thead>
<tr>
<th>Types of Metaphorical Expressions</th>
<th>Relations</th>
<th>Total Collocates</th>
<th>Pseudo-R-Square</th>
<th>Method One (Short Distance from (0,0))</th>
<th>Method Two (Slope=-1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>成長 chéngzhǎng ‘grow/growth’</td>
<td>Subject</td>
<td>1490</td>
<td>0.906935</td>
<td>5.472613</td>
<td>4.211427</td>
</tr>
<tr>
<td></td>
<td>pp_、</td>
<td>288</td>
<td>0.901897</td>
<td>2.928936</td>
<td>2.189770</td>
</tr>
<tr>
<td></td>
<td>pp_比</td>
<td>14</td>
<td>0.97708</td>
<td>1.938021</td>
<td>1.869558</td>
</tr>
<tr>
<td></td>
<td>pp_在</td>
<td>34</td>
<td>0.829636</td>
<td>1.440731</td>
<td>1.030891</td>
</tr>
<tr>
<td></td>
<td>pp_受</td>
<td>4</td>
<td>0.931119</td>
<td>1.40818</td>
<td>1.418494</td>
</tr>
<tr>
<td></td>
<td>pp_於</td>
<td>25</td>
<td>0.936921</td>
<td>1.507649</td>
<td>1.153737</td>
</tr>
<tr>
<td></td>
<td>pp_較</td>
<td>13</td>
<td>0.723689</td>
<td>0.915610</td>
<td>0.573336</td>
</tr>
<tr>
<td></td>
<td>pp_對</td>
<td>32</td>
<td>0.961804</td>
<td>1.607662</td>
<td>1.228279</td>
</tr>
<tr>
<td>起飛 qǐfēi ‘take off’</td>
<td>Subject</td>
<td>268</td>
<td>0.933048</td>
<td>3.630461</td>
<td>2.926560</td>
</tr>
<tr>
<td></td>
<td>pp_、</td>
<td>31</td>
<td>0.939796</td>
<td>1.428015</td>
<td>1.032451</td>
</tr>
<tr>
<td>麻痺 tānhuàn ‘paralytic’</td>
<td>Subject</td>
<td>276</td>
<td>0.935357</td>
<td>4.384251</td>
<td>3.748123</td>
</tr>
<tr>
<td></td>
<td>Modifies</td>
<td>221</td>
<td>0.967868</td>
<td>3.787687</td>
<td>3.173571</td>
</tr>
<tr>
<td></td>
<td>pp_、</td>
<td>44</td>
<td>0.812178</td>
<td>1.357190</td>
<td>0.913848</td>
</tr>
</tbody>
</table>

12 The program will first find the variables $a$ and $b$ in the function of the nonlinear regression $y = b(x^a)$. These two variables provide information regarding the curve in the nonlinear regression. Both of these values are needed to find any points along the curve. Then, these two values will be included in the calculation of the cut-off points (formulas in Footnotes 9 and 10).

13 In later analyses, the prepositions ‘pp’ are removed because prepositions do not provide informative data regarding the source domains. For example, any words can be used with the prepositions listed in Table 9.3. Therefore, these are confusing factors that should be removed.
For example, the first relation (subject) of 成長 chéngzhǎng ‘grow/growth’ shows a ‘curve fitting’ of 91 per cent. The columns called ‘Method One’ and ‘Method Two’ are the columns for the cut-off points. The ‘subject’ of 成長 chéngzhǎng ‘grow/growth’ is shown to have cut off points at word number five in Method One and word number four in Method Two. Similar results can be seen in the examples of 起飛 qǐfēi ‘take off’ and 瘫瘓 tànhuàn ‘paralytic’ in Table 9.3.

An immediate observation of the cut-off points is that they cut off at the top few collocates of each list. In fact, different normalization methods may give different cut-off points. This makes the results of cut-off points questionable. For example, the cut-off point for the ‘subject’ relation of 起飛 qǐfēi ‘take off’ in Method One (short distance) is shown in Table 9.4 with a dotted line.

An immediate observation of the cut-off points is that they cut off at the top few collocates of each list. In fact, different normalization methods may give different cut-off points. This makes the results of cut-off points questionable. For example, the cut-off point for the ‘subject’ relation of 起飛 qǐfēi ‘take off’ in Method One (short distance) is shown in Table 9.4 with a dotted line.

### Table 9.4: Method One (Short Distance): ‘Subject’ 起飛 qǐfēi ‘take off’ (CNA)

<table>
<thead>
<tr>
<th>Chinese Collocates</th>
<th>English Gloss</th>
<th>Frequency</th>
<th>Saliency</th>
</tr>
</thead>
<tbody>
<tr>
<td>飛機 fēijī</td>
<td>airplane</td>
<td>538</td>
<td>60.19</td>
</tr>
<tr>
<td>班機 bānjī</td>
<td>airliner</td>
<td>248</td>
<td>48.40</td>
</tr>
<tr>
<td>路 pàodào</td>
<td>runway</td>
<td>71</td>
<td>39.86</td>
</tr>
<tr>
<td>經濟 jīngjī</td>
<td>economy</td>
<td>591</td>
<td>37.79</td>
</tr>
<tr>
<td>夢想 mèngxiǎng</td>
<td>dream</td>
<td>35</td>
<td>36.09</td>
</tr>
<tr>
<td>客機 kèjī</td>
<td>passenger plane</td>
<td>67</td>
<td>33.92</td>
</tr>
<tr>
<td>航空母艦 hángkōngmǔjiàn</td>
<td>aircraft carrier</td>
<td>33</td>
<td>32.32</td>
</tr>
<tr>
<td>滑行道 huáxíngdào</td>
<td>taxiway</td>
<td>8</td>
<td>30.80</td>
</tr>
<tr>
<td>專機 zhuānjī</td>
<td>special plane</td>
<td>28</td>
<td>27.07</td>
</tr>
<tr>
<td>小時 xiǎoshí</td>
<td>hour</td>
<td>35</td>
<td>24.60</td>
</tr>
<tr>
<td>航機 hàngjī</td>
<td>flight</td>
<td>14</td>
<td>24.43</td>
</tr>
<tr>
<td>包機 bāojī</td>
<td>charter plane</td>
<td>15</td>
<td>21.96</td>
</tr>
<tr>
<td>航班 hàngbān</td>
<td>flight</td>
<td>14</td>
<td>21.50</td>
</tr>
<tr>
<td>軍機 jūnjī</td>
<td>military plane</td>
<td>16</td>
<td>21.41</td>
</tr>
<tr>
<td>戰機 zhànjī</td>
<td>fighter plane</td>
<td>26</td>
<td>21.19</td>
</tr>
<tr>
<td>直昇機 zhíshēngjī</td>
<td>helicopter</td>
<td>18</td>
<td>20.41</td>
</tr>
<tr>
<td>班次 bāncì</td>
<td>flight order</td>
<td>13</td>
<td>19.93</td>
</tr>
</tbody>
</table>

Intuitively, the cut-off point at word number four would seem to be unreasonable because many other collocates below the cut-off points exist and they are also used significantly in conjunction with 起飛 qǐfēi ‘take off.’ For example, those shaded collocates below the cut-off point are all related to the first collocate of 飛機 fēijī ‘airplane.’ It is therefore not reasonable to exclude the shaded collocates below the cut-off point in Table 9.4.
The results produced through the use of normalized data seem to affect the overall results. However, it is also not possible to use the raw data from Sketch Engine for the calculation of cut-off points due to the limitation of the constant variable of ‘rank of Chinese collocates,’ as mentioned previously. For this reason, cut-off points based on distributional data cannot be used. Researching cut-off points in this book so far has concentrated on distributional data and transforming the results from Sketch Engine to make it distributional. However, this direction is perhaps not viable for determining cut-off points. In order to validate the idea of computing cut-off points, the following section will suggest an alternative method that does not require transformation.

9.3.2 Methods Based on ‘Mean of Means’

An alternative method, as described following, is suggested because (a) it does not involve transformation of data; and (b) it provides a threshold value for the selection of cut-off points (in contrast to the previously viewed Methods One and Two, in which the cut-off points were calculated directly using formulas based on the patterns of distributional data). This new method is called ‘mean of means.’ Steps of this method will be discussed next.14

First, a series of means will be found based on the top collocates. For the following mean, collocates are added for the calculation of means. For example, if there are 100 collocates in a relation, the first mean is the mean of collocates one and two; the second mean is the mean of collocates one, two and three (add a new collocate every time); the third mean is the mean of collocates one, two, three and four and so on (until the end of the list). Therefore, for 100 collocates, there will be 99 means that are based on an increasing number of top collocates. When this task has been completed, the mean for the overall means will be used as a threshold value for the cut off point.

\[
\text{Threshold} = \frac{\text{Mean}_{(\text{Saliency}_1, \text{Saliency}_2)} + \text{Mean}_{(\text{Saliency}_1, \text{Saliency}_2, \text{Saliency}_3)} + \ldots + \text{Mean}_{(\text{Saliency}_{(n-1)}, \text{Saliency}_{(n)}, \text{Saliency}_{(n+1)})}}{n-1}
\]

This threshold value is more reliable because it is the mean of a series of means. When calculating the means descending from the top collocates, one actually finds the mean for possible cut-off points. When all of these means are laid out, the cut-off point is above the threshold value.

14 This method is proposed by Professor Shu-Chuan Tseng (personal communication).
Figure 9.8: Computing ‘Means’ for the Collocates of ‘Subjects’ of 起飛 qifēi ‘take off’ (CNA)

Figure 9.8 reveals that a series of means is produced by increasing the number of collocates each time in the calculation. Therefore, when the saliency values decrease, the means will also become smaller. The longer the list, the smaller the ‘mean of means’ will become, and thus, the cut-off point will likewise decrease. This is unlike the previous two methods, Method One (short distance) and Method Two (slope), where cut-off points appear at collocates number one to six, regardless of the length of the collocates. This new method is computed independent of collocate lists. An example of a computation of ‘mean of means’ is shown in Table 9.5 (continuous dots in the table indicate hidden collocates).
Table 9.5: Mean of Means: ‘Subject’ 起飛 qǐfēi ‘take off’ (CNA)\textsuperscript{15}

<table>
<thead>
<tr>
<th>Collocate Number</th>
<th>Chinese Collocates</th>
<th>English Gloss</th>
<th>Frequency</th>
<th>Saliency</th>
<th>Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>飛機 fēijī</td>
<td>airplane</td>
<td>538</td>
<td>60.19</td>
<td>---</td>
</tr>
<tr>
<td>2</td>
<td>班機 bānjī</td>
<td>airliner</td>
<td>248</td>
<td>48.40</td>
<td>54.30</td>
</tr>
<tr>
<td>3</td>
<td>跑道 pǎodào</td>
<td>runway</td>
<td>71</td>
<td>39.86</td>
<td>49.48</td>
</tr>
<tr>
<td>4</td>
<td>經濟 jīngjì</td>
<td>economy</td>
<td>591</td>
<td>37.79</td>
<td>46.56</td>
</tr>
<tr>
<td>5</td>
<td>夢想 mèngxiǎng</td>
<td>dream</td>
<td>35</td>
<td>36.09</td>
<td>44.47</td>
</tr>
<tr>
<td>6</td>
<td>客機 kèjī</td>
<td>passenger plane</td>
<td>67</td>
<td>33.92</td>
<td>42.71</td>
</tr>
<tr>
<td>7</td>
<td>航空母艦 hāngkōngmǔjùiàn</td>
<td>aircraft carrier</td>
<td>33</td>
<td>32.32</td>
<td>41.22</td>
</tr>
<tr>
<td>8</td>
<td>滑行道 huáxíngdào</td>
<td>taxiway</td>
<td>8</td>
<td>30.80</td>
<td>39.92</td>
</tr>
<tr>
<td>9</td>
<td>專機 zhuānjī</td>
<td>special plane</td>
<td>28</td>
<td>27.07</td>
<td>38.49</td>
</tr>
<tr>
<td>10</td>
<td>小時 xiǎoshí</td>
<td>hour</td>
<td>35</td>
<td>24.60</td>
<td>37.10</td>
</tr>
<tr>
<td>11</td>
<td>航機 hángjī</td>
<td>flight</td>
<td>14</td>
<td>24.43</td>
<td>35.95</td>
</tr>
<tr>
<td>12</td>
<td>包機 bāojī</td>
<td>charter plane</td>
<td>15</td>
<td>21.96</td>
<td>34.79</td>
</tr>
<tr>
<td>13</td>
<td>航班 hángbān</td>
<td>flight</td>
<td>14</td>
<td>21.50</td>
<td>33.76</td>
</tr>
<tr>
<td>14</td>
<td>單機 jūnjī</td>
<td>military plane</td>
<td>16</td>
<td>21.41</td>
<td>32.88</td>
</tr>
<tr>
<td>15</td>
<td>戰機 zhànjī</td>
<td>fighter plane</td>
<td>26</td>
<td>21.19</td>
<td>32.10</td>
</tr>
<tr>
<td>16</td>
<td>直昇機 zhíshēngjī</td>
<td>helicopter</td>
<td>18</td>
<td>20.41</td>
<td>31.37</td>
</tr>
<tr>
<td>17</td>
<td>班次 bāncì</td>
<td>flight order</td>
<td>13</td>
<td>19.93</td>
<td>27.85</td>
</tr>
<tr>
<td>87</td>
<td>駕駛員 jiāshìyuán</td>
<td>driver</td>
<td>3</td>
<td>7.94</td>
<td>15.28</td>
</tr>
<tr>
<td>88</td>
<td>特號 tèhào</td>
<td>special number</td>
<td>1</td>
<td>7.85</td>
<td>15.20</td>
</tr>
<tr>
<td>89</td>
<td>秋門 qiūmén</td>
<td>a state in Siberia</td>
<td>1</td>
<td>7.83</td>
<td>15.11</td>
</tr>
<tr>
<td>90</td>
<td>產業 chānyè</td>
<td>Industry</td>
<td>15</td>
<td>7.82</td>
<td>15.03</td>
</tr>
<tr>
<td>91</td>
<td>雙機 shuāngjī</td>
<td>dual machines</td>
<td>1</td>
<td>7.78</td>
<td>14.95</td>
</tr>
<tr>
<td>92</td>
<td>父親節 bàbājié</td>
<td>Father’s Day</td>
<td>1</td>
<td>7.66</td>
<td>14.87</td>
</tr>
<tr>
<td>184</td>
<td>力 nènglì</td>
<td>capability</td>
<td>1</td>
<td>0.04</td>
<td>7.45</td>
</tr>
<tr>
<td>268</td>
<td>目標 mùbiāo</td>
<td>goal</td>
<td>1</td>
<td>0.03</td>
<td>7.42</td>
</tr>
</tbody>
</table>

Mean of Means (Threshold) 15.03

\textsuperscript{15} A small number of words in Sketch Engine are wrongly tagged. For example, 秋門 qiūmén is a location where the airplane takes off but it is wrongly tagged. These errors are due to the problems of Sketch Engine, but they will be dropped automatically during clustering because they may not fall in any clusters within the list of collocates.
In Table 9.5, the cut-off point is marked by a dotted line across the table after collocate number 89. This method locates the cut-off collocate at number 89, roughly one third down, from a total 268 collocates. The threshold value is given as ‘mean of means’ at the bottom, which is the mean value for all the means in the last column.

In Table 9.6, the calculation of cut-off points for 成長 chéngzhǎng ‘grow/growth,’ 起飛 qǐfēi ‘take off’ and 癱瘓 tānhuàn ‘paralytic’ are shown. The last two columns (shaded) show the place after which the lists of collocates are cut off, as well as its percentages, as constituted by collocates above the cut-off points.

**Table 9.6: Calculation of Cut-off Points Using Mean of Means**

<table>
<thead>
<tr>
<th>Types of Metaphorical Expressions</th>
<th>Relations</th>
<th>Mean of Means</th>
<th>Total Collocates</th>
<th>Cut-off Points</th>
<th>Percentages</th>
</tr>
</thead>
<tbody>
<tr>
<td>成長 chéngzhǎng ‘grow/growth’</td>
<td>subject</td>
<td>11.88</td>
<td>1490</td>
<td>490</td>
<td>32.89%</td>
</tr>
<tr>
<td>起飛 qǐfēi ‘take off’</td>
<td>subject</td>
<td>15.03</td>
<td>268</td>
<td>89</td>
<td>33.21%</td>
</tr>
<tr>
<td>癱瘓 tānhuàn ‘paralytic’</td>
<td>modifies</td>
<td>13.80</td>
<td>276</td>
<td>82</td>
<td>29.71%</td>
</tr>
</tbody>
</table>

From Table 9.6, we can see that the cut-off points appear approximately one third down from the top of the list. These significant collocates will be used in the next step for the purpose of clustering.

### 9.4 Clustering

In the last section, we have discussed that in examples such as 經濟起飛 jīngjì qǐfēi ‘economy takes off,’ the target domain of 經濟 jīngjì ‘economy’ can possibly be replaced by stating 飛機起飛 fēijī qǐfēi ‘airplane takes off’ as well as 蝴蝶起飛 húdié qǐfēi ‘butterfly takes off’ (by intuition). The terms replacing 經濟 jīngjì ‘economy’ will appear in the same collocation list as 經濟 jīngjì ‘economy’ in the Chinese Sketch Engine. An encapsulation of how collocation lists from the Chinese Sketch Engine can be used is seen in Figure 9.9 (repeated from Figure 9.4).

---

16 This table shows the results after all preposition relations (‘pp’) have been removed. The drop of ‘pp’ will be discussed in further detail in §9.4.2 under example (9).
In this figure, the occurrences of 經濟 jīngjì ‘economy’ as well as other possible replacement terms, such as 飛機 fēijī ‘airplane’ and 班機 bānjī ‘airliner,’ are found in the same collocation list of 起飛 qīfēi ‘take off’ when searched for in CNA.

Table 9.9: Collocates of ‘Subjects’ of 起飛 qīfēi ‘take off’ in CNA in the Chinese Sketch Engine

By utilizing the results of computing cut-off points, we are better equipped to select collocates that are significant (above the cut-off points) that, in turn, can then be used to carry out the analyses in this chapter. This chapter will take collocates above the cut-off points and group them into different clusters of concepts. The correspondence concepts of collocates have been obtained via their mappings to WordNet in SinicaBow.
Chapter 9: Bottom-up Approach to Source Domain Determination

and the Merged Word List. After these collocates are clustered, the final step will then be the selection of source domain names for these clusters of concepts.

However, in order to proceed to these two steps, an *a priori* step is carried out to map all collocates to WordNet. This step is needed because only through the mappings can the WordNet concepts of each collocate be found. The steps for the mappings are the same as those discussed in Chapter Seven, and both SinicaBow and the Chinese-English Merged Word List are utilized for the purpose of mappings. The next section will first elaborate on the mappings.

**9.4.1 Mappings to WordNet through SinicaBow and the Chinese-English Merged Word List**

The total grammatical relations after computation of cut-off points number 103,795 collocates. Among these collocates, there are both proper nouns (for the ‘subject’ relations) and abbreviations. Examples of proper nouns and abbreviations are given in (8).

(8) 保態課 *bào tài kè* ‘class of ecology’
    綜合網 *zōng hé wǎng* ‘combined webpage’
    帕拉瓜 *pā lā guā* ‘para melon’
    彭添誠 *pěng tiān chéng* (name of a person)

Since many of the collocates under the ‘subject’ relation, such as those in (8), are proper nouns, we do not expect a high performance of the mappings to WordNet through SinicaBow and the Merged Word List. However, the successfully mapped collocates should be frequently appearing collocates. The statistics for the found collocates are shown in Table 9.7.

---

17 Reasons for using these two databases can be found in Chapter Seven.

18 From Chapter Seven, we know that the lexical items collected in these two databases are highly frequent words as they cover almost 78 per cent of the total types of metaphorical expressions and more than 95 per cent of the tokens of metaphorical expressions (see Table 7.4 in Chapter Seven). However, since the collocates are numerous (about 200 for each grammatical relation; each type of metaphorical expression has several relations; and there are different collocate lists for CNA and XIN), such a high performance of the two databases for the found collocates is not expected.

19 If we average by the total types of metaphorical expressions in CNA+XIN (i.e., 1,591; see Figure 7.2 in Chapter Seven), each type of metaphorical expression will consist of about 65 collocates only (in all grammatical relations), although combining CNA and XIN is not carried out in this bottom-up approach.
Table 9.7: Collocates That Are Successfully Mapped to WordNet through SinicaBow and the Chinese-English Merged Word List

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Total Types of Collocates</th>
<th>SinicaBow Found</th>
<th>SinicaBow %</th>
<th>Merged Word List Found</th>
<th>Merged Word List %</th>
<th>Overall Found</th>
<th>Overall %</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNA</td>
<td>92,465</td>
<td>10,986</td>
<td>11.88</td>
<td>9,629</td>
<td>11.82</td>
<td>20,615</td>
<td>22.29</td>
</tr>
<tr>
<td>XIN</td>
<td>68,237</td>
<td>8,536</td>
<td>12.51</td>
<td>7,928</td>
<td>13.28</td>
<td>16,464</td>
<td>24.13</td>
</tr>
</tbody>
</table>

From Table 9.7, the numbers of collocates that are found in SinicaBow and the Merged Word List lie between 8,000 and 11,000. This should show a high performance of mappings, as about 10,000 collocates can be found in both knowledgebases. However, as mentioned, since problems like the ones encountered with proper nouns and abbreviations exist, a large amount of collocates, constituting more than 78 per cent of both CNA and XIN, cannot be found. However, as seen in Chapter Seven, the found collocates are usually frequently appearing collocates, while most of the ones not found are collocates that do not contribute to the determination of source domains. The removal of non-meaningful collocates from the mappings is, therefore, beneficial to the analyses. The following section will proceed to explain the first step of the bottom-up approach, namely the clustering of these collocates above the cut-off points (i.e., mean of means) that are found in SinicaBow and the Merged Word List.

9.4.2 Clustering of Found Collocates above ‘Mean of Means’

After each collocate has been mapped to WordNet, each type of metaphorical expression, such as 成長 chēngzhǎng ‘grow/growth,’ 起飛 qǐfēi ‘take off’ and 僵瘓 tānhuàn ‘paralytic,’ will yield one or more WordNet sense. The results of these mappings are similar to the mappings in Chapter Seven. However, while mappings in Chapter Seven are directed towards all metaphorical expressions, the mappings in this chapter are focused on collocates of the metaphorical expressions. Furthermore, the top-down approach in Chapter Seven also differs from this chapter because the top-down approach uses knowledge information from the upper ontology SUMO. In this chapter, source domains will be determined through the observation of the conceptual density of these collocates.

The lists of collocates found in SinicaBow and the Merged Word List are clustered according to their concepts in the WordNet hierarchies. Conceptual density is then measured by the number of types of collocates (such as 飛機 fēijī ‘airplane,’ 班機 bānjī ‘airliner’ and 跑道 pǎodào ‘runway’) that fall under a particular branch in WordNet. The more types of collocates that fall under a WordNet branch, the higher the conceptual density is for this WordNet branch. Tokens of collocates will not be considered in terms of conceptual density because some highly frequent metaphorical expressions would...
monopolize all branches with high conceptual density. Furthermore, the collection of
types of collocates provides information regarding which concepts most often appear
among collocates. This information is not provided in Sketch Engine, which only
provides the frequency and saliency information of the collocates. For each type of
metaphorical expression, all collocates above the cut-off points will be mapped to
WordNet. Then, the collection of these collocates under each WordNet branch will be
carried out. WordNet has nine main branches of ‘noun,’ given in (9).

(9) a. Act, human_action, human_activity
b. Possession
c. Entity, physical_thing
d. Psychological_feature
e. State
f. Event
g. Group, grouping
h. Phenomenon
i. Abstraction

In Chapter Six, only the noun forms of the five target domains of 經濟 jīngjì ‘economy,’
政治 zhèngzhì ‘politics,’ 外交 wàijiāo ‘foreign affairs,’ 教育 jiàoyù ‘education’ and
財政 cáizhèng ‘finance’ are selected for metaphor analyses. Therefore, collocates that
appear in the same grammatical constructions with these five target domains should also
be nouns, and only the noun hierarchy of WordNet will be considered for conceptual
density.20 This is because when a word acts as both a noun and a verb (such as 投資
tóuzī ‘invest/investment’), senses with both nouns and verbs will be extracted from
WordNet. However, since only the noun meanings (which may act as subject, object,
noun_modifier, etc.) may replace the five target domains, which are also nouns, the
other parts-of-speech other than nouns (such as verbs and adjectives) will be filtered out
when measuring conceptual density.21

In the WordNet noun hierarchy, there may still be several senses remaining for a
similar type of collocate. The following Table 9.8 shows the senses for 飛機 fēijī
‘airplane,’ 班機 bānjī ‘airliner’ and 投資 tóuzī ‘invest/investment’ and their selected
senses, which are under the noun hierarchy of WordNet (those ‘Yes’s in column three

---

20 Chapter Seven compares the conceptual distance between the different SUMO nodes. Since
SUMO appears at a higher level, linguistic information, such as parts-of-speech, is not the
major concern of SUMO.

21 Adjectives are also filtered out because all target domains remain as nouns even when they
modify other nouns. By filtering out adjectives, we can also filter out unnecessary meaning,
such as ‘economic’ for 經濟 jīngjì ‘economy’. 
of Table 9.8). All senses that are verbs (those ‘No’s in column three) are removed in later analysis.

This table shows 飞机 fei ji ‘airplane’ to have two senses that are nouns. Both 班机 ban ji ‘airliner’ and 投资 tou zi ‘invest/investment’ have three senses that are nouns, respectively. These senses will be grouped separately in the clustering process. As a result, the clustering of collocates in each grammatical relation may depend on the number of senses each collocate has in addition to whether or not these senses fall under a similar top level. However, these different numbers of senses for each collocate should not be a conflicting factor for the clustering process because clustering using different senses will only make the analyses more precise. Furthermore, different senses may fall under different top levels of the WordNet hierarchy if they have different meanings. These levels are shown in the last column in Table 9.8.

### Table 9.8: Tabulated WordNet Information from SinicaBow

<table>
<thead>
<tr>
<th>Types of Collocates</th>
<th>WordNet 1.7.1 IDs</th>
<th>Noun</th>
<th>WordNet Definitions</th>
<th>Synsets</th>
<th>Top Level in the Noun Hierarchy</th>
</tr>
</thead>
<tbody>
<tr>
<td>飞机 fei ji ‘airplane’</td>
<td>02337628N</td>
<td>Yes</td>
<td>a vehicle that can fly</td>
<td>aircraft</td>
<td>entity, physical_thing</td>
</tr>
<tr>
<td></td>
<td>02341562N</td>
<td>Yes</td>
<td>an aircraft that has a fixed wing and is powered by propellers or jets</td>
<td>airplane, aeroplane plane</td>
<td>entity, physical_thing</td>
</tr>
<tr>
<td>班机 ban ji ‘airliner’</td>
<td>02551844N</td>
<td>Yes</td>
<td>a liner with cabins for passengers</td>
<td>cabin_liner</td>
<td>entity, physical_thing</td>
</tr>
<tr>
<td></td>
<td>02340817N</td>
<td>Yes</td>
<td>a commercial airplane that carries passengers</td>
<td>airliner</td>
<td>entity, physical_thing</td>
</tr>
<tr>
<td></td>
<td>00230281N</td>
<td>Yes</td>
<td>a scheduled trip by plane between designated airports</td>
<td>flight</td>
<td>act, human_action, human_activity</td>
</tr>
<tr>
<td>投资 tou zi ‘invest/investment’</td>
<td>00836949N</td>
<td>Yes</td>
<td>the act of investing; laying out money or capital in an enterprise with the expectation of profit</td>
<td>investing, investment</td>
<td>act, human_action, human_activity</td>
</tr>
<tr>
<td></td>
<td>01748713V</td>
<td>No</td>
<td>convert (short-term floating debt) into long-term debt that bears fixed interest and is represented by bonds</td>
<td>fund</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>00836658N</td>
<td>Yes</td>
<td>the act of financing</td>
<td>financing, funding</td>
<td>act, human_action, human_activity</td>
</tr>
<tr>
<td></td>
<td>00720048N</td>
<td>Yes</td>
<td>the act of capitalizing on an opportunity</td>
<td>capitalization, capitalisation</td>
<td>act, human_action, human_activity</td>
</tr>
<tr>
<td></td>
<td>01788571V</td>
<td>No</td>
<td>make an investment</td>
<td>invest, put, commit, place</td>
<td>---</td>
</tr>
</tbody>
</table>
For example, both senses of 飛機 fēijī ‘airplane’ are in the top level of ‘entity, physical_thing.’ On the other hand, 班機 bānjī ‘airliner’ has two senses that fall under ‘entity, physical_thing’ and one sense under ‘act, human_action, human_activity.’ 投資 tóuzī ‘invest/investment’ has three senses under ‘act, human_action, human_activity.’ Therefore, it is possible that some senses will fall in either a similar or different top level in the WordNet noun hierarchy, depending on the meaning of the noun. These different senses will be clustered according to the WordNet noun hierarchy.22

As mentioned at the beginning at this section, the clustering of collocates depends on the number of collocates in each grammatical relation. A grammatical relation with a few collocates might not show any patterns in the clustering process because these few collocates may be scattered throughout the different top levels. As a result, the types of metaphorical expressions that have successfully clustered may be less than their number that was found to be in the Chinese Sketch Engine in Table 9.1, previously seen in this chapter (i.e., 1,228 for CNA and 703 for XIN). Table 9.9 shows the number of types of metaphorical expressions (such as 成長 chéngzhǎng ‘grow/growth,’ 起飛 qǐfēi ‘take off’ and 癱瘓 tānhuàn ‘paralytic’ ) that are successfully clustered in CNA and XIN, respectively (regardless of how many grammatical relations they have). The combination of CNA+XIN is not possible in the bottom-up approach because CNA and XIN have different lists of collocates gathered from different sub-corpora, respectively (i.e., the sub-corpora of CNA will be used to create the Wordsketches if CNA data is used, and the same goes for XIN as well).

### Table 9.9: Percentages of Successfully Clustered Types of Metaphorical Expressions (Based on Their Found Collocates) in CNA and XIN

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Successfully Clustered Types / Found Types</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNA</td>
<td>1,049 / 1,228</td>
<td>85.42</td>
</tr>
<tr>
<td>XIN</td>
<td>541 / 703</td>
<td>76.96</td>
</tr>
</tbody>
</table>

From Table 9.9, we know that more than 77 per cent of the types of metaphorical expressions contain collocates that are found in SinicaBow and the Merged Word List. The numbers shown in Table 9.9 reflect those that have been successfully clustered based on the top levels of WordNet noun hierarchy, meaning that less than 23 per cent of the types of metaphorical expressions contain grammatical relations with either of the small number of collocates or do not belong to the noun hierarchy of WordNet. Only the successfully clustered types of metaphorical expressions will be analyzed in

---

22 If a collocate has a sense that is in contradiction to the meaning of the target domains, this sense will be an outlier, which does not form main clusters with meaning of senses that match the meaning when used with the target domains.
terms of source domains. The following section will describe the steps used in analyzing
the clustered collocates.

9.4.2.1 Application Used for Clustering of WordNet Nodes

Figure 9.10 shows how the application was used to view the allocation of collocates
to WordNet nodes. On the left is a list of types of metaphorical expressions and on
the right is the number of types of collocates and tokens of collocates, as well as other
relevant information for each level in the WordNet hierarchy of ‘noun.’

Figure 9.10: Allocation of Collocates to WordNet Nodes

Figure 9.10 shows the highlighted row on the right to be the branch with the highest
conceptual density in the WordNet hierarchy of ‘noun.’ Number ‘35’ in the ‘Types’
column illustrates that there are thirty-five types of collocates accumulated under this
branch of ‘entity, physical_thing’ (the other columns are additional information regarding
each branch, which will not be discussed in detail here).

Figure 9.11 shows how the top levels from Figure 9.10 can be expanded. Figure
9.11 displays the expansion until the second level, while Figure 9.12 shows expansion
until the end of the top level. In each level, if there are types of collocates falling under
it, collocates will be displayed in the second column (see ‘Cn Synset’ in Figure 9.12).

In order to view the clustering results of the WordNet noun hierarchy, a visualization application
was created by Petr Šimon for the purpose of this book.
As mentioned, the accumulated number of types of collocates will be shown in the column of ‘Types.’ For example, the accumulated number of types of collocates at the top level of ‘entity, physical_thing’ is thirty-five.

<table>
<thead>
<tr>
<th>En Synset</th>
<th>Cn Synset</th>
<th>Dens</th>
<th>Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>abstraction</td>
<td></td>
<td>511</td>
<td>7</td>
</tr>
<tr>
<td>act, human_action, human_activity</td>
<td></td>
<td>502</td>
<td>2</td>
</tr>
<tr>
<td>entity, physical_thing</td>
<td></td>
<td>37030</td>
<td>155</td>
</tr>
<tr>
<td>object, physical_obj</td>
<td></td>
<td>27508</td>
<td>26</td>
</tr>
<tr>
<td>location</td>
<td></td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>part, piece</td>
<td></td>
<td>65</td>
<td>5</td>
</tr>
<tr>
<td>substance matter</td>
<td></td>
<td>18</td>
<td>2</td>
</tr>
<tr>
<td>causal_agent, cause</td>
<td></td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>group, grouping</td>
<td></td>
<td>591</td>
<td>1</td>
</tr>
<tr>
<td>psychological_feature</td>
<td></td>
<td>660</td>
<td>8</td>
</tr>
<tr>
<td>state</td>
<td></td>
<td>1136</td>
<td>2</td>
</tr>
</tbody>
</table>

**Figure 9.11:** Expanded WordNet Hierarchy until the Second Level (a zoom-in of the window)

**Figure 9.12:** Expanded WordNet Hierarchy until the Bottom Level (a zoom-in of the window)
Once the top level with the highest conceptual density is found, its hyponyms with the highest accumulated values can also be found. For ease of reading, only collocates at the bottom of each top node will be displayed. See the methodology for selecting collocates at the bottom nodes, described following.

From Figure 9.11, it has been determined that the level under ‘entity, physical_thing’ is ‘object, physical_object.’ Under this level, the highest number of accumulated types of collocates is twenty-six (comparing ‘26’ with ‘1’, ‘5’, ‘2’ and ‘1’ in the ‘Types’ column). Therefore, this node will be selected. Subsequently, under ‘object, physical_object,’ comes ‘artifact, artefact’ (third level in Figure 9.12), which has the highest accumulated number of types of collocates (twenty-two) and will, therefore, be selected. This process continues on in this fashion until the bottom level. At the last two levels, ‘airplane, aeroplane, plane’ (tenth level), ‘airliner’ and ‘fighter, fighter_aircraft, attack_aircraft’ (both at eleventh level) can be found. In the final analysis, ‘airliner’ has ultimately been selected for the reason that it has more accumulated number of types of collocates (two) than ‘fighter, fighter_aircraft, attack_aircraft’ (one). In the event that there are two nodes with the same scores at the end, both nodes will be selected for further analyses. However, in this case, only ‘airliner’ is selected because its score is higher than those seen for ‘fighter, fighter_aircraft, attack_aircraft.’

Table 9.10 shows the two top levels with the highest and second-highest values of conceptual density, respectively. This means that ‘airliner’ (see arrow) is the bottom node of the hierarchy with the highest conceptual density. The second-highest is ‘fashion,’ whose hierarchy is also given in Table 9.10. The rank of conceptual density is given in the first column, while the second column shows collocates that appear at the bottom node of the hierarchy in column four. Column three shows the accumulated number of types of collocates, which is the reference for determining conceptual density.
Table 9.10: Examples of Two Hierarchies with the Highest Conceptual Density

<table>
<thead>
<tr>
<th>Rank of Conceptual Density</th>
<th>Bottom Node (Frequency/Saliency)</th>
<th>Accumulated ‘Types of Collocates’ in the Hierarchy</th>
<th>Hierarchy (with Accumulated ‘Types of Collocates’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>客機  kējī ‘passenger plane’ (67/48.4)</td>
<td>客機  kējī ‘passenger plane’ (67/48.4)</td>
<td>客機  kējī ‘passenger plane’ (67/48.4)</td>
</tr>
<tr>
<td></td>
<td>班機 bānī ‘airliner’ (248/33.92)</td>
<td>35</td>
<td>Entity, physical_thing (35)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>→ object, physical_object (26)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>→ artifact, artefact (22)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>→ instrumentality, instrumentation (14)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>→ conveyance, transport (10)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>→ vehicle (10)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>→ craft (9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>→ aircraft (6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>→ heavier-than-air_craft (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>→ airplane, aeroplane, plane (4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>→ airliner (2)</td>
</tr>
<tr>
<td></td>
<td>時尚  shìshāng ‘contemporary’ (4/11.78)</td>
<td>時尚  shìshāng ‘contemporary’ (4/11.78)</td>
<td>Psychological_feature (8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>→ cognition, knowledge, noesis (7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>→ process, cognitive_process, mental_process, operation, cognitive_operation (4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>→ basic_cognitive_process (4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>→ discrimination, discernment (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>→ taste, appreciation, discernment, perceptiveness (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>→ vogue, trend, style (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>→ fashion (1)</td>
</tr>
</tbody>
</table>

Table 9.10 clearly shows the branch that has the highest conceptual density to be represented by 客機  kējī ‘passenger plane’ and 班機 bānī ‘airliner’ under the branch of ‘entity, physical_thing.’ In the last column of Table 9.10, the hierarchy for ‘entity, physical_thing’ is spread out. The last node of ‘airliner’ is the place where 客機  kējī ‘passenger plane’ and 班機 bānī ‘airliner’ are found. Above this node, additional nodes can be seen, which may consist of other collocates such as 飛機  fēijī ‘airplane’ at the hypernyms, with the name ‘aircraft’ and ‘airplane, aeroplane, plane’ (these are the two senses of 飛機  fēijī ‘airplane’); however, they fall under the same top level of ‘entity, physical_thing.’

9.4.2.2 Selecting Top Levels with the Highest Values of Conceptual Density

Based on the previous criteria stated, regarding the selection of bottom nodes as a representation of a whole branch of the top levels, a program was created to automatically extract all relevant bottom nodes with a high conceptual density. However, selecting one single level with the highest conceptual density (shaded level in Figure 9.13) may not be adequate because some types of metaphorical expressions may have a high conceptual density for metaphorical expressions, which falls under the top level of ‘abstraction.’
Figure 9.13: Allocation of Collocates to WordNet Nodes

Table 9.11 shows the list of collocates for 复甦 fūsū ‘recover/recovery.’ We can see that most collocates are related to ‘non-physical’ recovery. The meaning of bodily recovery can only be seen when 脑 nǎo ‘brain’ appears (shaded below).

Table 9.11: Collocates of ‘Subject’ of 复甦 fūsū ‘recover/recovery’ in CNA in the Chinese Sketch Engine

<table>
<thead>
<tr>
<th>Collocates</th>
<th>English Gloss</th>
<th>Frequencies</th>
<th>Saliencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>景气 jǐngqì</td>
<td>prosperity</td>
<td>4039</td>
<td>85.75</td>
</tr>
<tr>
<td>經濟 jīngjī</td>
<td>economy</td>
<td>6290</td>
<td>53.78</td>
</tr>
<tr>
<td>軍國主義 jūnguó zhǔyì</td>
<td>militarism</td>
<td>43</td>
<td>30.25</td>
</tr>
<tr>
<td>房市 fāngshì</td>
<td>market of real property</td>
<td>33</td>
<td>27.68</td>
</tr>
<tr>
<td>觸底 chūdǐ</td>
<td>bottom out</td>
<td>7</td>
<td>18.59</td>
</tr>
<tr>
<td>產業 chǎn yè</td>
<td>property</td>
<td>204</td>
<td>18.58</td>
</tr>
<tr>
<td>萬物 wàn wù</td>
<td>the whole creation</td>
<td>9</td>
<td>18.52</td>
</tr>
<tr>
<td>旅遊業 lǚyóuyè</td>
<td>tourism industry</td>
<td>34</td>
<td>18.28</td>
</tr>
<tr>
<td>建築業 jīngjiùyè</td>
<td>construction work</td>
<td>10</td>
<td>17.67</td>
</tr>
<tr>
<td>觀光業 guāngguāngyè</td>
<td>tourism industry</td>
<td>14</td>
<td>17.66</td>
</tr>
<tr>
<td>基本面 jīběnmian</td>
<td>basis</td>
<td>17</td>
<td>15.95</td>
</tr>
<tr>
<td>才能 cái néng</td>
<td>capability</td>
<td>69</td>
<td>14.28</td>
</tr>
<tr>
<td>經濟體 jīngtǐ</td>
<td>economic system</td>
<td>14</td>
<td>14.01</td>
</tr>
<tr>
<td>房地產 fāngdìcūn</td>
<td>real property</td>
<td>30</td>
<td>13.03</td>
</tr>
</tbody>
</table>
In cases like this, the concept with the highest conceptual density is likely to be metaphorical.

In order to avoid selecting the top level that returns source domains that are non-concrete (such as the example in 復甦 fusi ‘recover/recovery’), we select more than one level for source domain analysis. Since there are nine levels in total (see (9)), we select the first three levels (30 per cent) as a start. This decision is arrived at because from our previous results of cut-off points (see Table 9.6), we have realized that the salient results usually appear within the first 30 per cent of the listing. The following will further elaborate upon our observations regarding the top three levels.

Figure 9.14 shows the conceptual density for the ‘subject’ relation of 成長 chéngzhǎng ‘grow/growth.’ The results indicate that the level with the highest conceptual density (column entitled ‘Types’) is ‘abstraction,’ with an accumulated 91 types of collocates. This reflects what has been discussed earlier, specifically that when the majority of collocates of a type of metaphorical expression are defined as abstract, their conceptual densities will tend to be abstract as well.

<table>
<thead>
<tr>
<th>En Synset</th>
<th>Cn Synset</th>
<th>Dens</th>
<th>Types</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>abstraction</td>
<td>abstraction</td>
<td>426699</td>
<td>01</td>
<td>4690</td>
</tr>
<tr>
<td>act,human_action</td>
<td>entity,physical_thing</td>
<td>84600</td>
<td>30</td>
<td>2320</td>
</tr>
<tr>
<td>event</td>
<td>92967</td>
<td>57</td>
<td>1631</td>
<td></td>
</tr>
<tr>
<td>group,grouping</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>phenomenon</td>
<td>284</td>
<td>4</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>possession</td>
<td>31355</td>
<td>23</td>
<td>1385</td>
<td></td>
</tr>
<tr>
<td>psychological_feature</td>
<td>6630</td>
<td>17</td>
<td>390</td>
<td></td>
</tr>
<tr>
<td>state</td>
<td>278628</td>
<td>11</td>
<td>25340</td>
<td></td>
</tr>
</tbody>
</table>

Figure 9.14: The First Level with the Highest Conceptual Density for the ‘Subject’ Relation of 成長 chéngzhǎng ‘grow/growth’

---

24 Figure 9.13 shows having less than nine levels because not all levels are involved in that particular use of 起飛 qifēi ‘take off.’

25 One can analyze more levels as a comparison. However, the inclusion of more levels will introduce noise into the results.
The level of ‘abstraction’ is not helpful in determining what the source domains could be, as source domains are usually more concrete. Therefore, in order to avoid selecting one single top level with the highest conceptual density (which may include abstract concepts), this section experiments with the top three levels possessing the highest conceptual density for each grammatical relation in each type of metaphorical expression. An example of selection from three levels of the WordNet noun hierarchy can be seen in Figure 9.15.

<table>
<thead>
<tr>
<th>En Synset</th>
<th>Cn Synset</th>
<th>Dens</th>
<th>Types</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>abstraction</td>
<td>長 chéngzhāng</td>
<td>425699</td>
<td>91</td>
<td>4689</td>
</tr>
<tr>
<td>act, human_action</td>
<td>成長 chéngzhāng</td>
<td>84600</td>
<td>39</td>
<td>2820</td>
</tr>
<tr>
<td>entity, physical_thing</td>
<td>成長 chéngzhāng</td>
<td>92967</td>
<td>57</td>
<td>1631</td>
</tr>
<tr>
<td>event</td>
<td>行为 xíngwéi</td>
<td>15</td>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td>group, grouping</td>
<td>群體 qún tǐ</td>
<td>253400</td>
<td>10</td>
<td>25340</td>
</tr>
<tr>
<td>phenomenon</td>
<td>現象 xiànxiàng</td>
<td>294</td>
<td>4</td>
<td>71</td>
</tr>
<tr>
<td>possession</td>
<td>擁有 yōngyǒu</td>
<td>31355</td>
<td>23</td>
<td>1365</td>
</tr>
<tr>
<td>psychological_feature</td>
<td>心理 shēn lǐ</td>
<td>6630</td>
<td>17</td>
<td>390</td>
</tr>
<tr>
<td>state</td>
<td>狀態 zhuàng tāi</td>
<td>273828</td>
<td>11</td>
<td>25348</td>
</tr>
</tbody>
</table>

**Figure 9.15:** The First Three Levels with the Highest Conceptual Density for the ‘Subject’ Relation of 成長 chéngzhāng ‘grow/growth’

Analyses of all the top levels show that about 20 per cent of the results are led by ‘abstraction.’

**Table A:** Percentages of the Different WordNet Top Levels (for Noun) with the Highest Conceptual Density in CNA and XIN

<table>
<thead>
<tr>
<th>CNA</th>
<th>XIN</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WordNet Top Levels for Noun Hierarchy</strong></td>
<td><strong>Freq.</strong></td>
</tr>
<tr>
<td>Entity, physical_thing</td>
<td>2,599</td>
</tr>
<tr>
<td>Abstraction</td>
<td>1,333</td>
</tr>
<tr>
<td>Act, human_action, human_activity</td>
<td>454</td>
</tr>
<tr>
<td>Psychological feature</td>
<td>338</td>
</tr>
<tr>
<td>Group, grouping</td>
<td>202</td>
</tr>
<tr>
<td>State</td>
<td>158</td>
</tr>
<tr>
<td>Phenomenon</td>
<td>21</td>
</tr>
<tr>
<td>Event</td>
<td>17</td>
</tr>
<tr>
<td>Possession</td>
<td>16</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>5,138</strong></td>
</tr>
</tbody>
</table>

Selecting two levels may also be problematic, as half of the source domains could have been concrete while the other half are abstract. Therefore, the decision has been made to select the top three.
The highest conceptual density value for the ‘subject’ relation of 成長 chēngzhǎng ‘grow/growth,’ i.e., ‘abstraction,’ is selected, along with the second- and third-highest values in conceptual density (i.e., ‘entity, physical_thing’ and ‘act, human_action, human_activity’). These three levels are circled in Figure 9.15. From this example of 成長 chēngzhǎng ‘grow/growth,’ it would seem safe to predict that most collocates replacing 經濟 jīngjì ‘economy’ in 經濟成長 jīngjì chēngzhǎng ‘economy grows’ will also be metaphors. To clarify the discussion regarding the selection of top levels in WordNet noun hierarchy, each type of metaphorical expression will have three top levels with the highest conceptual density for each of the grammatical relations it possesses. For example, Table 9.6 previously has shown that 瘫瘓 tānhuàn ‘paralytic’ has two types of grammatical relations that contain the five target domains, i.e., ‘subject’ and ‘modifies.’ For each of these relations, three top levels have been selected for each. Analyses of these levels will be the final step, as discussed in the following section.

9.5 Analyses of Source Domain Names

As stated in the previous section, the three top levels from the WordNet noun hierarchy with the highest conceptual density values will be extracted automatically with an especially designed program. The results will then be tabulated. The following Table 9.12 shows the results for 成長 chēngzhǎng ‘grow/growth.’ From Table 9.12, the same collocate (such as 貨幣 huòbì ‘coinage’) appears several times as the different senses of the words. 貨幣 huòbì ‘coinage’ appears twice under the top level of ‘abstraction’ (rank 1) with bottom nodes of ‘money’ as well as ‘coinage, mintage, specie, metal_money’ (also rank 1). These two bottom nodes have been selected because they have the same scores at the bottom of the hierarchy of ‘abstraction,’ and thus, both will be collected. The second top level (rank 2) with the second-highest conceptual density value is ‘entity, physical_thing,’ meaning that collocates that replace the target domains of 成長 chēngzhǎng ‘grow/growth’ are mostly clustered around the concept of ‘abstraction,’ followed by ‘entity, physical_thing.’
Table 9.12: Examples of Selected Three Top Levels with High Conceptual Density

<table>
<thead>
<tr>
<th>Types of Metaphorical Expressions</th>
<th>Grammatical Relation</th>
<th>Rank of Conceptual Density</th>
<th>Bottom Node (Frequency/Saliency)</th>
<th>Synsets at Bottom Nodes</th>
<th>Top Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>成長 ‘grow/growth’</td>
<td>Subject</td>
<td>1</td>
<td>貨幣 huò bì (89.9.79)</td>
<td>money</td>
<td>Abstraction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>貨幣 huò bì (89.9.79)</td>
<td>coinage, mintage, specie, metal_money</td>
<td>Abstraction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>存款 cún huán (58.9.17)</td>
<td>deposit, bank_deposit</td>
<td>Abstraction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>預算 yù suàn (187/12.88)</td>
<td>budget</td>
<td>Abstraction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>小孩 xiǎohái (21.9.48)</td>
<td>kiddy</td>
<td>Entity, physical_thing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>小孩 xiǎohái (21.9.48)</td>
<td>bairn</td>
<td>Entity, physical_thing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>產量 chǎn liàng (44.12.86)</td>
<td>fruitage</td>
<td>Act, human_action, human_activity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>產量 chǎn liàng (22/16.21)</td>
<td>capacity</td>
<td>Act, human_action, human_activity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>產量 chǎn liàng (44/12.86)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The second-largest cluster is present under the top level of ‘entity, physical_thing’ and the third top level (rank 3), with the third-highest conceptual density value, is ‘act, human_action, human_activity.’ Each type of metaphorical expression will have these three levels in each of their relations.

This most concrete concept will be selected from among ranks 1 to 3 in the example shown in Table 9.12. All bottom nodes of these three rankings are considered possible source domains for types of metaphorical expressions (arrow in Table 9.12), as these bottom nodes show the most specific concept of a branch that is conceptually salient.

In the case of 成長 chéngzhǎng ‘grow/growth’ in Table 9.12, the hierarchical structure for ‘entity, physical_thing,’ with the bottom node of 小孩 xiǎohái ‘kiddy/bairn,’ is shown in (10).
In (10), we see that 小孩 xiǎohái ‘kiddy/bairn’ has two senses in WordNet that fall under ‘entity, physical_thing,’ indicating that these two bottom nodes are the most specific concept that has upper concepts, such as ‘living_thing,’ ‘organism,’ etc. (from general to specific). In Chapter Two, we have posed questions regarding the specificity of source domains. However, as seen in the hierarchical structure of (10), it is possible for different studies to make different decisions regarding source domains (such as ORGANISM in Charteris-Black & Ennis (2001) and Stefanowistch (2005) as well as PERSON in Chung, Ahrens & Huang (2005)). However, unlike the English ‘growth,’ the senses of 成長 chéngzhǎng ‘grow/growth’ in WordNet do not encompass the meaning PLANT, as has been suggested by Kövecses (2002) and Charteris-Black (2004) for the source domain of ‘growth’ in English. The top-down approach, however, does not capture this difference, as the top-down approach has placed emphasis on source domains such as COGNITIVE AGENT and ORGANISM, which appear at a higher conceptual level than those in (10).

Based on the examples in Table 9.12, it can be seen that source domains selected that make use of collocates are for specific source domains. In Table 9.12, some of the source domains for 成長 chéngzhǎng ‘grow/growth’ are metaphorical (such as 貨幣 huòbì ‘coinage’). These metaphorical uses will be kept in the results because they show how both concrete (such as 小孩 xiǎohái ‘kiddy/bairn’) and metaphorical concepts can be found using the collocational approach. The precision of the results will be tested using human ratings of the correctness of answer in Chapter Eleven.

Another example of bottom nodes can be seen in Table 9.13 for 抑鬱 yìyù ‘depressed,’ where the source domain terms are the bottom nodes (with arrows).
In Table 9.13, we can see that 抑鬱 yiyù ‘depressed’ is used both with the psychological aspect as well as with the state of a person. Therefore, the results from the bottom-up approach also support our contention that one type of metaphorical expression may be mapped to more than one source domain, as there may be more than one bottom node that can be accepted as source domains.

### 9.6 Summary of Chapter

This chapter has discussed how collocation can be used in a bottom-up approach for source domain determination. As the first step of the bottom-up approach, automatic extraction of collocates is carried out based on possible positions of the target domains with regard to each type of metaphorical expression. Even though the same types of metaphorical expressions used in the top-down approach (Chapter Seven) were also used in the bottom-up approach in this chapter, different results may be expected from both approaches. A comparison of both approaches will be carried out after the source domains have been determined.
The second step implements several methods to determine cut-off points, which help set parameters in order to select significant data from lists of empirical results. This chapter emphasizes the importance of determining a significant list because most empirical studies do not know where to stop listing results from listings, such as a frequency list. In fact, most studies tend to list the top few results, and these top few are dependent on the choice of the researchers. If criterion-based methods to determine cut-off points for the frequency lists can be devised, the subjectivity will be reduced in terms of choosing which top few words will be included. Furthermore, most lexical resources provide wordlists according to different criteria, such as frequency, mutual information values, collocation, saliency values, etc. To date, no one has suggested which of the top few listed should be looked at. We propose the cut-off points to be a great contribution to computational linguistics, to researchers in need of statistical methods of analyzing linguistic data and, finally, to researchers who need to conduct psycholinguistic experiments related to word meaning.

The final two steps are: clustering all collocates above the cut of points (mean of means); and finding WordNet hypernyms to act as source domain names. Before these two steps can be carried out, collocates must be mapped to SinicaBow and the Chinese-English Merged Word List, so that their corresponding WordNet concepts can be extracted. In the third step, the clustering of collocates is a necessary step because it aids in the process of grouping collocates together with similar concepts. From here, we can then select the WordNet top levels with the highest conceptual density. Three top levels of WordNet hierarchy have been selected, and from these three top levels, the bottom nodes are used as representations of source domain terms in the fourth step. These source domains are specific source domains in which we can always see their upper hierarchy by examining the hypernyms of the bottom nodes. In Chapter Ten, the results of the bottom-up approach will be discussed. Chapter Eleven will provide the evaluation of the correctness of source domains in both the top-down and the bottom-up approaches; Chapter Twelve will discuss the linguistic phenomena related to the source domains found in this book; and conclusions will be made in Chapter Thirteen.
Chapter 10: Results of the Bottom-up Approach

Unlike the top-down approach, the bottom-up approach does not first assume a shared conceptualization. Rather, the bottom-up approach takes into consideration differences observed at the lower levels, such as those between collocated words, constructional information and co-occurring frequencies. Since this is so, data from Taiwan (CNA) and China (XIN) will be treated separately because both communities may have different collocated words, constructional information and co-occurring frequencies. The steps involved in the bottom-up approach are shown in Figure 10.1.

Specifically, the steps involved in the bottom-up approach are as follows: First, we extract collocates of each type of metaphorical expression in CNA and XIN separately; second, we must compute cut-off points for the lists of collocates so that significant top collocates can be used in the next step; and finally, we cluster collocates above the cut-off points through their WordNet concepts. In this manner, all collocates can be
allocated to different levels of the WordNet hierarchy. After these steps have been carried out, the WordNet top levels with the top three highest conceptual density (i.e., with most collocates under the same branch) will be selected. Last, source domain names are determined through the observation of the bottom nodes of these selected top WordNet levels.

In this chapter, the results of the bottom-up approach will be discussed. §10.1 will first discuss the performance of the bottom-up approach in the number of found source domains (i.e., how well this approach performs in finding a source domain). §10.2 will discuss the most frequently appearing source domains found in the bottom-up approach as well as the most frequently appearing types of metaphorical expressions under each source domain. §10.3 will discuss the strengths and weaknesses of the bottom-up approach. §10.4 shows the summary of the chapter.

10.1 Performance of the Bottom-up Approach

Similar to Chapter Eight, the overall performance of the bottom-up approach is represented by the percentages of cases where a source domain can be determined. These percentages are obtained by dividing the types of metaphorical expressions with found source domains with the total found cases of the bottom-up approach in the Chinese Sketch Engine.¹

Table 10.1 shows the percentages of metaphorical expressions with source domains determined using the bottom-up approach. As mentioned, the data for CNA and XIN may yield separate results, as collocates used in Taiwan and China may differ. Therefore, CNA and XIN will not be combined in a table, as a type of metaphorical expression with a found source domain in CNA does not necessarily mean that it will also have a found source domain in XIN. Since the results are different, the combined CNA+XIN, as well as their overlapped metaphorical expressions, will not be shown in Table 10.1.

<table>
<thead>
<tr>
<th>FoundSD_Types / Found_Types</th>
<th>%</th>
<th>FoundSD_Tokens / Found_Tokens</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNA</td>
<td>1,045 / 1,228</td>
<td>85.10</td>
<td>17,684 / 18,158</td>
</tr>
<tr>
<td>XIN</td>
<td>574 / 703</td>
<td>81.65</td>
<td>13,722 / 14,022</td>
</tr>
</tbody>
</table>

In Table 10.1, about 82 to 85 per cent of the found types of metaphorical expressions have source domains that have been determined. These found types of metaphorical

¹ Unlike the top-down approach in Chapter Eight where two databases are compared, the performance of the bottom-up approach simply shows the differences between CNA and XIN.
expressions constitute about 97 to 98 per cent of the found tokens of metaphorical expressions. This means that only a small number of the metaphorical expressions found in the Chinese Sketch Engine are unable to have their source domains determined. When compared to the percentages of the top-down approach in Chapter Eight (Figures 8.2 and 8.3), we can see that the percentages of source domains determined in the top-down approach also ranges between 92 and 99 per cent for both types and tokens of metaphorical expressions. In fact, the percentages of the types of metaphorical expressions in the top-down approach are slightly higher than those of the bottom-up approach. However, since we are measuring against the found cases (not the total metaphorical expressions), the differences of the two approaches may not show how good these lexical knowledgebases are at collecting metaphorical expressions. Rather, these percentages show how well the lexical knowledgebases in the different approaches return a source domain, using the possible steps proposed herein.

Chapter Nine states that most of the time the reason we cannot determine a source domain is because these metaphorical expressions have too few collocates for clustering. However, as we can see from Table 10.1, these cases are few and far between with most of the metaphorical expressions found in the Chinese Sketch Engine possessing source domains that have been found.

In the following section, we will discuss the types of source domains selected in different datasets.

10.2 Source Domains in the Bottom-up Approach

In this section, the source domains selected for each dataset (Single Target Domains, Coordinated Target Domains as well as between CNA and XIN) will be shown. The five target domains examined previously are 政治 zhèngzhì ‘politics,’ 外交 wàijiāo ‘foreign affairs,’ 財政 cāizhèng ‘finance,’ 教育 jiàoyù ‘education’ and 經濟 jīngjì ‘economy.’ The discussion in this section will consider only the results yielded from the computational approaches, and the evaluation of the correctness of source domains will be carried out in Chapter Eleven.

10.2.1 Source Domains for the Single Target Domains

The source domains of the bottom-up approach come from the bottom nodes of the top three levels of WordNet noun hierarchies that have the highest conceptual density. Since the collocates for types of metaphorical expressions in CNA and XIN may differ, the results found in the bottom-up approach will also differ greatly between CNA and XIN. For example, collocates for 建設 jiànsè ‘construct/construction’ under the relation
of ‘subject’ show differences between CNA and XIN (see Table 10.2), and this will affect the selection of the source domains, as their most condensed concept may differ.

Table 10.2: Collocates for 建設 jiànshè ‘construct/construction’ in CNA and XIN

<table>
<thead>
<tr>
<th>CNA</th>
<th>XIN</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>subject 838</strong></td>
<td><strong>subject 3562</strong></td>
</tr>
<tr>
<td>交通 83</td>
<td>文明 258</td>
</tr>
<tr>
<td>硬體 13</td>
<td>經濟 1227</td>
</tr>
<tr>
<td>市政 21</td>
<td>重點 234</td>
</tr>
<tr>
<td>寶石 7</td>
<td>巨資 29</td>
</tr>
<tr>
<td>重點 40</td>
<td>法制 43</td>
</tr>
<tr>
<td>文明 23</td>
<td>工程 138</td>
</tr>
<tr>
<td>經濟 145</td>
<td>水利 36</td>
</tr>
<tr>
<td>文化 56</td>
<td>隊伍 39</td>
</tr>
<tr>
<td>縣政 7</td>
<td>軍隊 42</td>
</tr>
<tr>
<td>基礎 32</td>
<td>設施 60</td>
</tr>
<tr>
<td>人民幣 7</td>
<td>航運 24</td>
</tr>
<tr>
<td>軟體 8</td>
<td>力量 45</td>
</tr>
<tr>
<td>地方 25</td>
<td>道德 22</td>
</tr>
<tr>
<td>水利 8</td>
<td>電力 30</td>
</tr>
<tr>
<td>馬 5</td>
<td>作風 17</td>
</tr>
<tr>
<td>工程 21</td>
<td>規劃 40</td>
</tr>
<tr>
<td>交通 5</td>
<td>力氣 9</td>
</tr>
<tr>
<td>鐵路 8</td>
<td>鐵路 38</td>
</tr>
<tr>
<td>投資 20</td>
<td>行風 5</td>
</tr>
</tbody>
</table>

For each dataset in the bottom-up approach, the top source domains can be seen in Table 10.3. For the source domains in the bottom-up approach, both Chinese and English translations are provided. However, primary focus should be given to the Chinese source domains because the source domains from the bottom-up approach are made up of the bottom nodes that have been extracted based on the Chinese collocates.²

² For the top-down approach, previously, only English is given because the terms were taken from the SUMO definitions, which are available in English.
Table 10.3: Most Frequently Appearing Source Domains in the Single Target Domains in the Bottom-up Approach

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Source Domains</th>
<th>CNA</th>
<th>XIN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Percentages in the Total Tokens of Metaphorical Expressions</td>
<td>Percentages in the Total Tokens of Metaphorical Expressions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Source Domains</td>
<td></td>
</tr>
<tr>
<td>政治 zhèngzhì ‘politics’</td>
<td>CAPACITY/能力</td>
<td>522 / 2,921 (17.87%)</td>
<td>JOB/工作</td>
</tr>
<tr>
<td>外交 wàijiāo ‘foreign affairs’</td>
<td>PROFIT/利益</td>
<td>125 / 583 (21.44%)</td>
<td>INQUIRY/研究</td>
</tr>
<tr>
<td>財政 cáizhèng ‘finance’</td>
<td>CAPACITY/能力</td>
<td>181 / 623 (29.05%)</td>
<td>STATE/國家</td>
</tr>
<tr>
<td>教育 jiàoyù ‘education’</td>
<td>TESTING/測驗</td>
<td>278 / 1,300 (21.38%)</td>
<td>ASSIGNMENT/任務</td>
</tr>
<tr>
<td>經濟 jīngjí ‘economy’</td>
<td>TROOP/軍隊</td>
<td>2,283 / 8,648 (26.40%)</td>
<td>TROOP/軍隊</td>
</tr>
<tr>
<td>Total Single Target Domains</td>
<td>TROOP/軍隊</td>
<td>2,922 / 14,075 (20.76%)</td>
<td>TROOP/軍隊</td>
</tr>
</tbody>
</table>

In Table 10.3, we can see that most of the target domains display differences between CNA and XIN, with the exception of 經濟 jīngjí ‘economy’ and the total Single Target Domains (shaded). Percentages for the target domains in XIN are consistently higher than those for CNA. This phenomenon, as we have discussed previously, is due to the centralized use of metaphorical expressions in China, i.e., the same types of metaphorical expressions are likely to be repeated more often in Chinese. As a result, the percentage

---

3 Each Chinese source domain can sometimes be mapped to several English translations (such as ASSIGNMENT and JOB for 工作). The translations given in this table are usually the first ones in a string of names (such as in ‘state, nation, country, land, commonwealth, res_publica, body_politic’). When there are several different strings of names provided, the most often occurring ones will be selected manually. Unless the names have been changed, they will be stated (as in Footnote 4, following).

4 The original translation is ACCOUNT for 利益 lìyì ‘profit,’ but it has been changed in this work to PROFIT because ACCOUNT carries ambiguous meaning.
of a particular source domain will likewise increase (when the number of types of metaphorical expressions under this source domain increases).\footnote{Even though the collocates come from different relations of the same type of metaphorical expression (such as ‘subject,’ ‘object,’ etc., which have their own results of source domains), the final top source domains shown in Table 10.3 are the most often occurring ones, regardless of differences between relations.}

Another impression we receive from this data regarding these source domains is that the names for the English source domains may not match those of the Chinese. This is because, in this approach, Chinese collocates are represented by the bottom nodes and their English translations come from their mappings in WordNet. As mentioned, the major reference to the source domains should be based on the Chinese terms, i.e., 力 instead of CAPACITY. These source domains are unlike those suggested by Lakoff & Johnson (1980) and Lakoff (1993), which are pre-determined (such as BUILDING, FOOD, etc.) insofar as these source domains are determined using computational methods. They are also different from those in the top-down approach insofar as they are specific. Examples of types of metaphorical expressions that fall under CAPACITY/能力 are shown in (1):\footnote{For the bottom-up approach, we do not encounter problems with the formulation of conceptual metaphors. This is probably because grammatical relations have already been identified at the beginning, and that the source domains are collocates at the similar grammatical position with the target domains. As a result, the formulation of conceptual metaphors, such as POLITICS IS CAPACITY/能力, becomes more natural than the top-down approach, such as POLITICS IS A RECIPIENT OF ACTIONS FROM A COGNITIVE AGENT (see Footnote 9 of Chapter Eight), where grammatical relations were not first identified.}

(1) a. ECONOMY IS CAPACITY/能力

加速 社會主義 經濟 建設 (XIN)

jiākuài shèhuìzhǔyì jīngjì jiànsè

speed.up socialism economy construct

‘To speed up the construction of a socialist economy...’

b. FINANCE IS CAPACITY/能力

不僅 加重 了 財政 負擔 (XIN)

bùjǐn jiāzhòng le cāizhèng fùdān

not.only add.heavy LE finance burden

‘(It has) not only increased the financial burden...’

In (1a), politics is seen as something that possesses capacity, especially when it can be built upon. In (1b), finance can increase in capacity in forms of monetary burdens, such as loans.
In the following Table 10.4 we show the top ten source domains (other than the general keywords) that appear most frequently in CNA and XIN in the Single Target Domains. Similar to the treatment of source domains in Chapter Eight, the light shades are for the source domains that repeat in both CNA and XIN. The only difference is that the top-down approach is comparing two lexical knowledgebases using the same sets of types of metaphorical expressions. In the bottom-up approach, CNA and XIN are distinguished. Darker shaded areas show the source domains that differ in both sets of results. Non-shaded areas show the general keywords, which are ignored.

From Table 10.4, we can see that many of the source domains are non-repetitive in CNA and XIN, as the bottom-up approach focuses on minor differences; this is also the reason why it can bring out the similarities and differences between the two communities.

**Table 10.4:** The Top Ten Source Domains Most Frequently Used for the Sum of the Single Target Domains in the Bottom-up Approach

<table>
<thead>
<tr>
<th>Source Domains</th>
<th>Tokens of Metaphorical Expressions under Each Source Domain</th>
<th>% (Overall Tokens of Metaphorical Expressions 14,075)</th>
<th>Source Domains</th>
<th>Tokens of Metaphorical Expressions under Each Source Domain</th>
<th>% (Overall Tokens of Metaphorical Expressions 10,799)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BEHALF/方面</td>
<td>3,522</td>
<td>25.02</td>
<td>BEHALF/方面</td>
<td>3,191</td>
<td>29.55</td>
</tr>
<tr>
<td>TROOP/軍隊</td>
<td>2,922</td>
<td>20.76</td>
<td>TROOP/軍隊</td>
<td>2,012</td>
<td>18.63</td>
</tr>
<tr>
<td>BASIS/基礎</td>
<td>2,754</td>
<td>19.57</td>
<td>CAPACITY/能力</td>
<td>1,939</td>
<td>17.96</td>
</tr>
<tr>
<td>MATTER/問題</td>
<td>2,700</td>
<td>19.18</td>
<td>THEORY/理論</td>
<td>1,884</td>
<td>17.45</td>
</tr>
<tr>
<td>BRIEF/綱要</td>
<td>2,311</td>
<td>16.42</td>
<td>POINT/特點</td>
<td>1,851</td>
<td>17.14</td>
</tr>
<tr>
<td>CAPACITY/能力</td>
<td>1,955</td>
<td>13.89</td>
<td>MIGHT/力量</td>
<td>1,764</td>
<td>16.33</td>
</tr>
<tr>
<td>VIRTUE/道德</td>
<td>1,707</td>
<td>12.13</td>
<td>ASPECT/局面</td>
<td>1,659</td>
<td>15.36</td>
</tr>
<tr>
<td>COINAGE/貨幣</td>
<td>1,693</td>
<td>12.03</td>
<td>RESIDENCE/住宅</td>
<td>1,487</td>
<td>13.77</td>
</tr>
<tr>
<td>PECULIARITY/特性</td>
<td>1,650</td>
<td>11.72</td>
<td>CAR WHEEL/車輪</td>
<td>1,474</td>
<td>13.65</td>
</tr>
<tr>
<td>FRUITAGE/產量</td>
<td>1,608</td>
<td>11.42</td>
<td>WATER FAUCET/龍頭</td>
<td>1,454</td>
<td>13.46</td>
</tr>
<tr>
<td>CONSTRAINT/限制</td>
<td>1,557</td>
<td>11.06</td>
<td>JOB/工作</td>
<td>1,438</td>
<td>13.32</td>
</tr>
<tr>
<td>STATE/國家</td>
<td>1,539</td>
<td>10.93</td>
<td>ATTRACTION/吸引力</td>
<td>1,404</td>
<td>13.00</td>
</tr>
<tr>
<td>WORLD WIDE WEB/網絡</td>
<td>1,519</td>
<td>10.79</td>
<td>HIGHROAD/公路</td>
<td>1,402</td>
<td>12.98</td>
</tr>
<tr>
<td>AIM/目標</td>
<td>1,481</td>
<td>10.52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAPACITY/生產量</td>
<td>1,471</td>
<td>10.45</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**
- White shades are general keywords; grey shades are overlapped source domains; black shades are non-overlapped source domains.
- ** See Appendix A10.1 for the list of general keywords for the bottom-up approach.
In Table 10.4, notice how there are only two source domains that are repeated in CNA and in XIN, namely TROOP/軍隊 and CAPACITY/能力. Examples of CAPACITY/能力 have been given in (1); the following (2) provides the examples for TROOP/軍隊.

(2) a. POLITICS IS A TROOP/軍隊
台灣非凡的經濟成就與政治改革，
Taiwan extraordinary DE economy achievement and politics reformation
為美國帶來極大的機會 (CNA)
för America bring-come very big DE chance
‘The extraordinary achievement of the economy of Taiwan and the revolution of politics will open a great chance with (investment from) America.’

b. ECONOMY IS A TROOP/軍隊
並認為我方經濟實力雄厚 (CNA)
also believe our side economy strength solid
‘(Someone) also believes that the strength of our economy is solid.’

In (2a), politics is seen as a troop, which involves a revolution (either being reformed or causing reformation). In (2b), the economy is seen as a troop in terms of its strength. TROOP/軍隊 is also the most frequently appearing source domain in the target domain of 經濟 jìngjì ‘economy’ as well as in the total Single Target Domains (see Table 10.3).

In the following section, we examine the top two types of metaphorical expressions that fall under each source domain. As mentioned, the same types of metaphorical expressions can appear in several source domains. This fact will be brought more clearly into focus in the discussion of the next section.

10.2.2 Most Frequently Appearing Types of Metaphorical Expressions in Different Source Domains for the Single Target Domains

Table 10.5 shows the top two most frequently appearing types of metaphorical expressions: Only the total Single Target Domains (shaded bottom row) are shown to have similar (two) types of metaphorical expressions in CNA and XIN.
Table 10.5: The Top Two Types of Metaphorical Expressions in the Single Target Domains in the Bottom-up Approach

<table>
<thead>
<tr>
<th>Datasets</th>
<th>CNA</th>
<th>XIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>政治 zhèngzhì ‘politics’</td>
<td>CAPACITY/能力 改革 gǎigé ‘reform/reformation’ (53.64%) 危機 wēijī ‘crisis’ (15.90%)</td>
<td>JOB/工作 穩定 wéndìng ‘stable’ (28.62%)局面 júmiàn ‘state.of.affairs’ (23.57%)</td>
</tr>
<tr>
<td>外交 wàijīao ‘foreign affairs’</td>
<td>PROFIT/利益 努力 nǔlì ‘hard.work’(66.40%) 戰略 zhànliè ‘ruses.of.war’(7.20%)</td>
<td>INQUIRY/研究 努力 nǔlì ‘hard.work’(90.79%) 手段 shǒudúduǎn ‘trick’ (7.90%)</td>
</tr>
<tr>
<td>財政 cáizhèng ‘finance’</td>
<td>CAPACITY/能力 負擔 fùdān ‘burden’ (70.17%) 危機 wēijī ‘crisis’ (18.23%)</td>
<td>STATE/國家 困難 kùnnán ‘difficulty’ (79.49%) 負擔 fùdān ‘burden’ (16.03%)</td>
</tr>
<tr>
<td>教育 jiàoyù ‘education’</td>
<td>TESTING/測驗 接受 jiēshòu ‘accept’ (41.01%) 進行 jīnxíng ‘carry.out’ (26.26%)</td>
<td>ASSIGNMENT/任務 開展 kāi zhǎn ‘spread’ (49.59%) 抓 zhuā ‘grab’ (15.70%)</td>
</tr>
<tr>
<td>經濟 jīngjì ‘economy’</td>
<td>TROOP/軍隊 建設 jiànshè ‘construct/construction’ (58.96%) 改革 gǎigé ‘reform/reformation’ (25.67%)</td>
<td>TROOP/軍隊 建設 jiànshè ‘construct/construction’ (50.72%) 穩定 wéndìng ‘stable’ (11.13%)</td>
</tr>
</tbody>
</table>

In Table 10.5, we can also see the difference between the CNA and XIN data. For example, 政治 zhèngzhì ‘politics’ is seen as capacity in Taiwan, but it is seen as a job in China. In China, the stability (穩定 wéndìng ‘stable’ and 局面 jùmiàn ‘state.of.affairs’) of politics is emphasized; this is different from Taiwan, which emphasizes

---

9 As mentioned, interpretation in this chapter is based on the source domains determined using the lexical and computational approaches in this book. Correctness of these source domains will be measured later when a portion of the data will be extracted for testing.
changes in politics (改革 găigé ‘reform/reformation’ and 危机 wēijī ‘crisis’). Therefore, comparisons based on regional differences such as these can be carried out using the bottom-up approach. In 教育 jiàoyù ‘education,’ for instance, Taiwan places emphasis on carrying out testing on educational knowledge (接受 jiēshòu ‘accept’ and 进行 jìnxing ‘carry out’), in sharp comparison to China, which focuses on an individual’s mission to expanding and grabbing at educational knowledge (開展 kāizhǎn ‘spread’ and 抓 zhuā ‘grab’). We can also see that 外交 wàijiāo ‘foreign affairs’ in Taiwan focuses on gain and profit; however, 外交 wàijiāo ‘foreign affairs’ in China focuses on studying and learning about the tricks of foreign affairs. 财政 cáizhèng ‘finance,’ on the other hand, is more similar in terms of CNA and XIN, i.e., both see it as a kind of difficulty, burden or crisis. CNA and XIN are also similar in terms of their uses of types of metaphorical expressions for 经济 jīngjì ‘economy.’

When we compared these results to the results produced by the top-down approach in Chapter Eight, we found that the results of the bottom-up approach to be better at demonstrating the differences between CNA and XIN. In the top-down approach, most of the types of metaphorical expressions are similar, either between SinicaBow and the Merged Word List or between CNA and XIN. This reflects the earlier assumption we made regarding the conceptual-knowledge-based approach—this approach assumes a shared conceptualization between different communities. Therefore, it is not surprising if the top-down approach produces source domains and top types of metaphorical expressions that are similar across CNA and XIN. As for the bottom-up approach, we assume that the prototypical approach will yield results that highlight the prototypical uses of language in different regions. Therefore, the results in Table 10.5 are also expected because these top types of metaphorical expressions should be the most prototypical uses of the different Single Target Domains in CNA and XIN. Therefore, through comparisons such as these, this book concisely illustrates how different presuppositions regarding linguistic phenomena can lead to different results and, hence, to different analyses. Only through comparisons such as these are we able to see how different approaches will lead to different results. This is also the main reason why there have been great controversies regarding what should constitute a source domain. For instance, the source domains in the top-down approach are different from the bottom-up approach. However, this does not mean to imply that any of the approaches are incorrect. In fact, our findings show that no type of metaphorical expression belongs to one single source domain. Most types of metaphorical expressions can be grouped according to general source domains (such as COGNITIVE AGENT) as well as specific source domains (such as BAIRN). We therefore propose that in further studies the hierarchical links between source domains should be established clearly.
In the following section, we first discuss the results of the Coordinated Target Domains.

### 10.2.3 Source Domains for the Coordinated Target Domains

As for the results in the Coordinated Target Domains, displayed in Table 10.6, only 教育 jiào yù ‘education’ (shaded) is shown to possess similar source domains for CNA and XIN (i.e., TROOP/軍隊). Most of the other target domains show different source domains for CNA and XIN.

**Table 10.6: Most Frequently Appearing Source Domains in the Coordinated Target Domains in the Bottom-up Approach (with 經濟 jìng jì ‘economy’)**

<table>
<thead>
<tr>
<th>Datasets</th>
<th>CNA Source Domains</th>
<th>Percentages in the Total Tokens of Metaphorical Expressions</th>
<th>XIN Source Domains</th>
<th>Percentages in the Total Tokens of Metaphorical Expressions</th>
</tr>
</thead>
<tbody>
<tr>
<td>政治 zhèngzhí ‘politics’</td>
<td>TROOP/軍隊</td>
<td>999 / 3,803 (26.27%)</td>
<td>ASSIGNMENT/任務</td>
<td>1,166 / 3,130 (37.25%)</td>
</tr>
<tr>
<td>外交 wàijiāo ‘foreign affairs’</td>
<td>PUISSANCE/勢力</td>
<td>42 / 216 (19.44%)</td>
<td>MENACE/威脅</td>
<td>33 / 77 (42.86%)</td>
</tr>
<tr>
<td>財政 cái zhèng ‘finance’</td>
<td>CAPACITY/能力</td>
<td>29 / 109 (26.61%)</td>
<td>ASSIGNMENT/任務</td>
<td>35 / 116 (30.17%)</td>
</tr>
<tr>
<td>教育 jiào yù ‘education’</td>
<td>TROOP/軍隊</td>
<td>39 / 90 (43.33%)</td>
<td>TROOP/軍隊</td>
<td>39 / 86 (45.35%)</td>
</tr>
<tr>
<td>Total Coordinated Target Domains</td>
<td>TROOP/軍隊</td>
<td>1,092 / 4,218 (25.89%)</td>
<td>ASSIGNMENT/任務</td>
<td>1,227 / 3,409 (35.99%)</td>
</tr>
</tbody>
</table>

Among the source domains in Table 10.6, only two prove to be similar to the Single Target Domains (in bold). As mentioned, the Coordinated Target Domains are expected to be different from the Single Target Domains because the coordinate use is usually constrained by two target domains. For example, as a Single Target Domain, 教育 jiào yù ‘education’ is seen as a kind of assignment (ASSIGNMENT/任務) in China (see (3a)). However, not all types of metaphorical expressions for ASSIGNMENT/任務 (such as 開展 kāi zhǎn ‘spread’ and 抓 zhuā ‘grab;’ see Table 10.5) can be used with
A Corpus-driven Approach to Source Domain Determination

經濟 jìngjì ‘economy.’ Therefore, types of metaphorical expressions that cannot appear for both 教育 jiàoyù ‘education’ and 經濟 jìngjì ‘economy’ are naturally absent in the coordinate uses in both of these target domains (3b).

(3) a. EDUCATION IS AN ASSIGNMENT/任務
快樂 is 保障 教育 順利 開展 and 取得
zhè shì bǎozhèng jiàoyù shùnì kāi zhǎn hàn qūdé
this be ensure education smooth spread and obtain
成效 的 關鍵 (XIN)
chéngxiào de guānjià
effect DE key
‘This is the key to ensure that education spreads smoothly, and that it will obtain the maximum effect.’

b. ?這 is 保障 教育 and 經濟 順利 開展
zhè shì bǎozhèng jiàoyù hàn jìngjì shùnì kāi zhǎn
this be ensure education and economy smooth spread
和 取得 成效 的 關鍵
hàn qūdé chéngxiào de guānjià
and obtain effect DE key
‘?This is the key to ensure that education and economy spread smoothly, and that they will obtain the maximum effect.’

See how in (3a) ‘education spreads’ is a kind of assignment, but in (3b), a similar expression cannot be used with both education and economy. Phenomena such as these occur, and they help explain why there are huge differences between most of the source domains in the Single Target Domains and the Coordinated Target Domains.

In Table 10.7, we show the top ten source domains (other than general keywords) that appear with the Coordinated Target Domains.
Table 10.7: The Top Ten Source Domains Most Frequently Used for the Sum of the Coordinated Target Domains in the Bottom-up Approach

<table>
<thead>
<tr>
<th>Source Domains</th>
<th>Tokens of Metaphorical Expressions under Each Source Domain</th>
<th>% (Overall Tokens of Metaphorical Expressions 4,218)</th>
<th>Source Domains</th>
<th>Tokens of Metaphorical Expressions under Each Source Domain</th>
<th>% (Overall Tokens of Metaphorical Expressions 3,409)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TROOP/軍隊</td>
<td>1092</td>
<td>25.89</td>
<td>ASSIGNMENT/任務</td>
<td>1227</td>
<td>35.99</td>
</tr>
<tr>
<td>BASIS/基礎</td>
<td>1040</td>
<td>24.66</td>
<td>CAPACITY/能力</td>
<td>1023</td>
<td>30.01</td>
</tr>
<tr>
<td>CAPACITY/能力</td>
<td>1004</td>
<td>23.80</td>
<td>JOB/工作</td>
<td>1009</td>
<td>29.60</td>
</tr>
<tr>
<td>BRIEF/綱要</td>
<td>873</td>
<td>20.70</td>
<td>POINT/特點</td>
<td>919</td>
<td>26.96</td>
</tr>
<tr>
<td>ASSIGNMENT/任務</td>
<td>775</td>
<td>18.37</td>
<td>PERFORMANCE/表現</td>
<td>894</td>
<td>26.22</td>
</tr>
<tr>
<td>MATTER/問題</td>
<td>752</td>
<td>17.83</td>
<td>FORWARDING/發展</td>
<td>885</td>
<td>25.96</td>
</tr>
<tr>
<td>IMMIGRANT/移民</td>
<td>668</td>
<td>15.84</td>
<td>BASIS/基礎</td>
<td>791</td>
<td>23.20</td>
</tr>
<tr>
<td>DISTRICT ATTORNEY/檢察官</td>
<td>666</td>
<td>15.79</td>
<td>AFGHANISTAN/阿富汗#</td>
<td>778</td>
<td>22.82</td>
</tr>
<tr>
<td>DEEPNESS/深度</td>
<td>636</td>
<td>15.08</td>
<td>SARAJEVO/薩拉熱窩#</td>
<td>726</td>
<td>21.30</td>
</tr>
<tr>
<td>CONCLUSION/結論</td>
<td>636</td>
<td>15.08</td>
<td>MARKET/市場</td>
<td>667</td>
<td>19.57</td>
</tr>
<tr>
<td>VIRTUE/道德</td>
<td>530</td>
<td>12.57</td>
<td>DIVERSITY/變化</td>
<td>667</td>
<td>19.57</td>
</tr>
<tr>
<td>PROFIT/利益</td>
<td>482</td>
<td>11.43</td>
<td>CAMBODIA/柬埔寨#</td>
<td>654</td>
<td>19.18</td>
</tr>
<tr>
<td>MARKET/市場</td>
<td>474</td>
<td>11.24</td>
<td>MENACE/威脅</td>
<td>636</td>
<td>18.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DIAGNOSTIC TEST/分析</td>
<td>626</td>
<td>18.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DISTINCTION/差別</td>
<td>617</td>
<td>18.10</td>
</tr>
</tbody>
</table>

** White shades are general keywords; grey shades are overlapped source domains; black shades are non-overlapped source domains.

#Proper names are removed manually.10

10 The removed keywords are general keywords, proper nouns and if any of these source domains happen to be similar to the target domains, i.e., when the target domains become the bottom node of the most condensed concepts in the steps in Chapter Seven.
Table 10.7 clearly shows that the source domains in the overall Coordinated Target Domains are not similar (dark shades). CNA and XIN share three source domains (CAPACITY/能力, ASSIGNMENT/任務 and MARKET/市場). However, since the combined Coordinated Target Domains show a combination of different target domains, the results in Table 10.7 will be used as a reference.

In the following section, the top two types of metaphorical expressions in each source domain are shown.

10.2.4 Most Frequently Appearing Types of Metaphorical Expressions in Different Source Domains for the Coordinated Target Domains

In Table 10.8, the top two types of metaphorical expressions in the Coordinated Target Domains are shown.

Table 10.8: The Top Two Types of Metaphorical Expressions in the Coordinated Target Domains in the Bottom-up Approach (with 經濟 jìngjì ‘economy’)

<table>
<thead>
<tr>
<th>Datasets</th>
<th>CNA</th>
<th>XIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>政治 zhèngzhì ‘politics’</td>
<td>TROOP/軍隊 改革 gāigé ‘reform/reformation’ (59.36%) 建設 jiànsè ‘construct/construction’ (21.52%)</td>
<td>ASSIGNMENT/任務 形勢 xíngshì ‘terrain’ (49.91%) 秩序 zhìxù ‘order’ (45.03%)</td>
</tr>
<tr>
<td>外交 wàijīāo ‘foreign affairs’</td>
<td>PUISSANCE/勢力 孤立 gūlì ‘isolated’ (19.05%) 影響力 yǐnxuǎnglì ‘influencing power’ (16.67%)</td>
<td>MENACE/威脅 制裁 zhìcái ‘impose.sanction’ (51.52%) 手段 shǒuduàn ‘trick’ (48.49%)</td>
</tr>
<tr>
<td>財政 cáizhìhèng ‘finance’</td>
<td>CAPACITY/能力 改革 gāigé ‘reform/reformation’ (55.17%) 危機 wéijī ‘crisis’ (17.24%)</td>
<td>ASSIGNMENT/任務 形勢 xíngshì ‘terrain’ (51.43%) 危機 wéijī ‘crisis’ (17.07%)</td>
</tr>
<tr>
<td>教育 jiàoyù ‘education’</td>
<td>TROOP/軍隊 改革 gāigé ‘reform/reformation’ (56.41%) 建設 jiànsè ‘construct/construction’ (41.03%)</td>
<td>TROOP/軍隊 改革 gāigé ‘reform/reformation’ (43.90%) 建設 jiànsè ‘construct/construction’ (43.90%)</td>
</tr>
<tr>
<td>Total Coordinated Target Domains</td>
<td>TROOP/軍隊 改革 gāigé ‘reform/reformation’ (58.24%) 建設 jiànsè ‘construct/construction’ (21.52%)</td>
<td>ASSIGNMENT/任務 形勢 xíngshì ‘terrain’ (49.06%) 秩序 zhìxù ‘order’ (42.87%)</td>
</tr>
</tbody>
</table>
As seen in the Single Target Domains, 改革 gāigé ‘reform/reformation’ and 建設 jiànsè ‘construct/construction’ are the most frequently appearing types of metaphorical expressions for 經濟 jīngjì ‘economy.’ In Table 10.8, these two types of metaphorical expressions appear in several datasets (both in 政治 zhèngzhì ‘politics’ of CNA; only one in 財政 cáizhèng ‘finance’ of CNA; and both in 教育 jiàoyù ‘education’ of CNA and XIN). We can explain these results by assuming them to be the results of coordination, where 改革 gāigé ‘reform/reformation’ and 建設 jiànsè ‘construct/construction’ are uses that can be (strongly) agreed upon by these different target domains. In addition, 形勢 xīngshì ‘terrain’ is repeatedly found in XIN, while CNA contains mostly 建設 jiànsè ‘construct/construction,’ 改革 gāigé ‘reform/reformation’ and 危機 wēijī ‘crisis’ (except for 外交 wàijiāo ‘foreign affairs’). The examples of 外交 wàijiāo ‘foreign affairs’ are shown in (4).

(4) a. FOREIGN AFFAIRS AND ECONOMY ARE PUISSANCE/勢力

実施 shíshī  外交 wàijiāo  經濟 jīngjì  孤立 gūlì  與 yú
carry.out  foreign.affairs  economy  isolated  and
軍事 jūnshì  包圍 bāowéi  戰略 zhànliè
military  envelopment  tactic
‘To carry out the isolation of foreign affairs and the economy as well as military envelopment tactics...’

b. FOREIGN AFFAIRS AND ECONOMY ARE PUISSANCE/勢力

美國 měiguó  領導 lǐngdǎo  de  zhèxiē  méngguó  jiāng  yòngyǒu
America  lead  DE  these  allied.country  will  possess
相當 xiāngdàng  大  de  jīngjì  yú  wàijiāo  yǐngxiǎnglì
fairly  big  DE  economy  and  foreign.affairs  influencing.power
‘These America-led allied countries will possess a fair amount of influential power in the economy and foreign affairs.’

In (4a), we see that the isolation in foreign affairs is a kind of puissance, i.e., power and strength; the same goes for influencing power in foreign affairs in (4b). These types of metaphorical expressions are both related to the competition or power of control in some way. 政治 zhèngzhì ‘politics’ and 經濟 jīngjì ‘economy’ do not use these kinds of metaphorical expressions as often as the coordination of 外交 wàijiāo ‘foreign affairs’ and 經濟 jīngjì ‘economy.’

219
From Table 10.8, as well as from the previous results collected for the Single Target Domains, it is brought to our attention that similar types of metaphorical expressions can occur as the top types of metaphorical expressions in several source domains. This is possible if a type of metaphorical expression (such as 改革 ‘reform/reformation’) appears in several source domains. The multiple source domains of 改革 ‘reform/reformation’ are shown in (5).

(5) 改革 ‘reform/reformation’
    ASSIGNMENT/任務
    IMMIGRANT/移民
    TROOP/軍隊
    DISTRICT ATTORNEY/檢察官
    CONCLUSION/結論
    CAPACITY/能力
    DEEPNESS/深度

In (5), we see the source domains that 改革 ‘reform/reformation’ belongs to, which have been decided upon based on the most frequently occurring collocates of 改革 ‘reform/reformation.’ These source domains, as mentioned, are specific, as the approach is bottom-up. For example, DISTRICT ATTORNEY/檢察官 and IMMIGRANT/移民 may both be regarded as a type of COGNITIVE AGENT, but in the bottom-up approach, only the bottom nodes are obtained. One possible explanation for this is that, since the approach is bottom-up, the source domains obtained through this approach always indicate one part of the information within a larger domain. In the example of DISTRICT ATTORNEY/檢察官, we obtain a specific source domain that may belong to a larger source domain of COGNITIVE AGENT or ORGANISM (produced via the top-down approach). Therefore, we can say that the results of the bottom-up approach are indeed different from those collected using the top-down approach, in terms of source domains.

The following section displays the top ten source domains for the overall bottom-up approach.

10.2.5 Top Ten Source Domains for the Bottom-up Approach

In Table 10.9, the top ten source domains from the bottom-up approach are provided.

---

11 General keywords, such as those in Appendix A10.1, have been removed.
Table 10.9 brings to light the fact that these source domains are similar to those in the Single Target Domains, as the Single Target Domains constitute the majority of the overall corpora data: As can also be seen in previous discussion, the source domains of TROOP/軍隊, CAPACITY/能力 and ASSIGNMENT/任務 are commonly seen in many datasets. In this table, we see that they are the only three source domains repeated in the overall Single+Coordinated Target Domains.

Table 10.9: The Top Ten Source Domains Most Frequently Used for the Sum of the Single+Coordinated Target Domains in the Bottom-up Approach

<table>
<thead>
<tr>
<th>Source Domains</th>
<th>Tokens of Metaphorical Expressions under Each Source Domain (Overall Tokens of Metaphorical Expressions 18,293) %</th>
<th>Source Domains</th>
<th>Tokens of Metaphorical Expressions under Each Source Domain (Overall Tokens of Metaphorical Expressions 14,208) %</th>
</tr>
</thead>
<tbody>
<tr>
<td>BEHALF/方面</td>
<td>5,090</td>
<td>BEHALF/方面</td>
<td>5,161</td>
</tr>
<tr>
<td>TROOP/軍隊</td>
<td>4,014</td>
<td>BASIS/基礎</td>
<td>3,340</td>
</tr>
<tr>
<td>BASIS/基礎</td>
<td>3,794</td>
<td>CAPACITY/能力</td>
<td>2,962</td>
</tr>
<tr>
<td>MATTER/問題</td>
<td>3,452</td>
<td>POINT/特點</td>
<td>2,770</td>
</tr>
<tr>
<td>BRIEF/綱要</td>
<td>3,184</td>
<td>ASSIGNMENT/任務</td>
<td>2,471</td>
</tr>
<tr>
<td>CAPACITY/能力</td>
<td>2,959</td>
<td>JOB/工作</td>
<td>2,447</td>
</tr>
<tr>
<td>VIRTUE/道徳</td>
<td>2,237</td>
<td>TROOP/軍隊</td>
<td>2,215</td>
</tr>
<tr>
<td>ASSIGNMENT/任務</td>
<td>2,119</td>
<td>THEORY/理論</td>
<td>2,050</td>
</tr>
<tr>
<td>PERCULIARITY/特性</td>
<td>1,940</td>
<td>MIGHT/力氣</td>
<td>1,883</td>
</tr>
<tr>
<td>STATE/國家</td>
<td>1,883</td>
<td>STATE OF AFFAIRS/局面</td>
<td>1,852</td>
</tr>
<tr>
<td>CONSTRAIN/限制</td>
<td>1,851</td>
<td>WATER FAUCET/龍頭</td>
<td>1,609</td>
</tr>
<tr>
<td>WORLD WIDE WEB/網路</td>
<td>1,793</td>
<td>RESIDENCE/住宅</td>
<td>1,603</td>
</tr>
<tr>
<td>COINAGE/貨幣</td>
<td>1,788</td>
<td>CAR WHEEL/車輪/</td>
<td>1,601</td>
</tr>
<tr>
<td>IMMIGRANT/移民</td>
<td>1,774</td>
<td>AIM/目標</td>
<td>1,752</td>
</tr>
<tr>
<td>AIM/目標</td>
<td>1,752</td>
<td>** White shades are general keywords; grey shades are overlapped source domains; black shades are non-overlapped source domains.**</td>
<td></td>
</tr>
</tbody>
</table>
These source domains have been determined through the prototypical approach, whereby the most condensed concept of each type of metaphorical expression is analyzed. In the next chapter, we will evaluate a sample of the results produced using these two approaches. However, before this is done, we will first discuss the strengths and weaknesses of the bottom-up approach in the section.

10.3 Strengths and Weaknesses of the Bottom-up Approach

The strengths and weaknesses of the bottom-up approach can be summarized in Table 10.10.

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implementing collocation-based method for source domain determination</td>
<td>Too depending on collocation may produce unexpected results for source domains</td>
</tr>
<tr>
<td>Emphasis on regional differences</td>
<td>Not economic</td>
</tr>
<tr>
<td>Powerful in using large amount of data</td>
<td>Contains noise in the data</td>
</tr>
<tr>
<td>Measurement of cut-off points is applied</td>
<td>Needs an evaluation of the cut-off points since there are several methods involved</td>
</tr>
<tr>
<td>Replicability</td>
<td>Involves human intuition</td>
</tr>
</tbody>
</table>

Each point in Table 10.10 will be elaborated in the following sub-sections.

10.3.1 Collocation-based versus Collocation-dependent

The bottom-up approach is very powerful, in the sense that it implements collocation into determining source domains. While this idea was first put forward in the Metaphorical Pattern Analysis (MPA) by Stefanowitsch (2006), it has never been used for determining source domains in his own analysis. It has also not been integrated into lexical resources such as Sketch Engine. For example, this chapter proves that source domains can be found by using collocates from Sketch Engine, replacing the metaphorical meaning (such as in (6a)) with the literal meaning (6b).
Chapter 10: Results of the Bottom-up Approach

(6) a. 反映出目前经济的持续衰弱 (CNA)
   reflect out current economy DE continuous weak
   ‘To reflect the weakness of the current economy...’

b. 老年人被吵得神经衰弱 (CNA)
   oldster BEI quarrel DE nerve weak/weakness
   ‘The oldster was disturbed by the noises until he had a nervous breakdown (literal translation: nervous weakness).’

In this book, we have proven that the literal meaning of a type of metaphorical expression such as 衰弱 shuāiruò ‘weak/weakness’ in (6a) can be recovered through observing collocational use of 衰弱 shuāiruò ‘weak/weakness,’ which are concrete (see (6b)), undeniably showing the strength of using collocation. However, in some cases, like in (7), we are unsure whether source domains such as PROFIT/利益 (see Table 10.3) are informative enough to become a source domain.

(7) FOREIGN AFFAIRS ARE PROFIT/利益 (CNA)
   South.Korea government believe foreign.affairs hard.work final
   must will succeed
   ‘The government of the South Korea believes that the hard work in/of the foreign affairs will finally yield success.’

It is possible that the constructional meaning comes from the uses such as in (8).

(8) 為謀求國家與民眾最大的利益努力奮鬥
   for seek country and people most big DE profit hard.work strive
   ‘To strive for the best advantage for the country and its people...’

In cases like this, we can relate the matter of foreign affairs with gaining profits but are
still uncertain as to whether or not additional interpretation is required to understand the source domains. This is also the reason why we need to evaluate the results in Chapter Eleven.

In sum, the use of collocation is a powerful method in determining source domains. However, a small number of the results may carry the problem stated above, i.e., collocation becomes an impeding factor in the process, as relationships between collocated words are sometimes unclear.

10.3.2 Regional Differences versus Being Economical

Unlike the top-down approach, the bottom-up approach takes into consideration all differences between two sets of data. As a result, the bottom-up approach is better suited for examining regional differences, such as those that occur between the use of source domains in Taiwan and China. This book is innovative with respect to the fact that no previous studies have examined the source domains of a similar language in different regions; it is the first to do so.

Unfortunately, taking into account all differences between two communities may also hinder generalization of phenomena. For example, in CNA, 建設 jiānshè ‘construct/construction’ falls under the source domains of 住宅 zhùzhái ‘residence’ and 公路 gōnglù ‘highroad’ (under the ‘subject’ relation with the first level of conceptual density). In XIN, the first level of conceptual density for 建設 jiānshè ‘construct/construction’ only falls under 公路 gōnglù ‘highroad’ (but not 住宅 zhùzhái ‘residence’).

(9) a. ECONOMY IS A RESIDENCE/住宅
   也 希望 多多 了解 台灣 經濟 建設 的
   also hope much understand Taiwan economy construct DE
   經驗 (CNA)
   jīngyàn
   experience
   ‘...also hope to better understand the experience in building the economy of Taiwan’

b. POLITICS IS A HIGHROAD/公路
   總之，我們 加強 政治 建設 保證
   in general we strengthen politics construction ensure
Differences like this may not be of a great concern when using the top-down approach (as both 住宅 zhùzhái ‘residence’ and 公路 gōnglù ‘highroad’ share the same hypernym of ARTIFACT). In the bottom-up approach, however, when differences like these are found, they are unlikely to be ignored. Therefore, the bottom-up approach is better at demonstrating regional differences, but at the same time, it contains too much detail, which makes it uneconomical to use when a generalized common phenomenon for closely related concepts is desired.

### 10.3.3 Large Sampling versus Level of Noise

The bottom-up approach is advantageous because it uses a large sample of data. When the sample is large, the power of statistics increases, and the results of the bottom-up approach will also become reliable.

However, the use of large sampling will also introduce noise into the results. Noise in the bottom-up approach comes mainly from the Chinese Sketch Engine, where there are some collocated patterns that are incorrectly programmed by the system. For example, Table 10.11 shows collocates of 成長 chéngzhǎng ‘grow/growth’ under the ‘subject’ relation. We can see that among these ‘subjects,’ there are certain collocates that are not ‘subjects’ (shaded).

**Table 10.11:** Collocates of ‘Subjects’ of 成長 chéngzhǎng ‘grow/growth’ in CNA in the Chinese Sketch Engine

<table>
<thead>
<tr>
<th>Collocates</th>
<th>English Gloss</th>
<th>Frequencies</th>
<th>Saliencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>倍數 bēishù</td>
<td>multiple</td>
<td>476</td>
<td>65.17</td>
</tr>
<tr>
<td>經濟 jīngjì</td>
<td>economy</td>
<td>22431</td>
<td>64.63</td>
</tr>
<tr>
<td>位數 wèishù</td>
<td>figure number</td>
<td>371</td>
<td>59.87</td>
</tr>
<tr>
<td>營收 yīngshōu</td>
<td>gross interests</td>
<td>307</td>
<td>44.09</td>
</tr>
<tr>
<td>毛額 máo é</td>
<td>gross</td>
<td>318</td>
<td>43.83</td>
</tr>
<tr>
<td>供給額 gōngjǐé</td>
<td>amount of supply</td>
<td>238</td>
<td>43.18</td>
</tr>
<tr>
<td>報表 yíji  achievement</td>
<td>335</td>
<td>42.36</td>
<td></td>
</tr>
<tr>
<td>個位數 gèwèishù units</td>
<td>57</td>
<td>40.99</td>
<td></td>
</tr>
<tr>
<td>同月 tóngyuè</td>
<td>same month</td>
<td>149</td>
<td>40.07</td>
</tr>
</tbody>
</table>
Examples of these uses are given in (10).

(10) a. 我國 製造業 產值 未來 十年
wōguó zhìzuòyè chǎnzhì wèilái shí nián
our.country manufacture.industry production.value future ten year
將 呈 倍數 成長 (CNA)
jiàng chéng bèishù chéngzhǎng
will present multiple growth
‘The production values of the manufacture industry of our country will display multi-fold growth in the coming ten years.’

b. 四 月份 工業 生產 指數 也 較
sì yuèfèn gōngyè shēngchǎn zhǐshù yě jiào
four month industry production index also more
去年 同月 成長 百分之六點四 (CNA)
quǎnian tóngyuè chéngzhǎng bāifēnzhīliùdiǎnsì
last.year same.month grow 6.4%
‘The industrial production index of April also grew 6.4% when compared to the same month in the previous year.’

In (10a), 倍數 bèishù ‘multiple’ is not the subject of 成長 chéngzhǎng ‘grow/growth’, even though it appears in the ‘subject’ position; it acts as a modifier instead. As for (10b), 同月 tóngyuè ‘same month’ is a mistake in Table 10.11, as it is neither a subject nor a modifier: It is simply a noun indicating duration. Therefore, noise like this occurs in the Chinese Sketch Engine, especially for constructions such as 經濟建設 jīngjì jiànsè ‘the building of economy’ and 教育改革 jiàoyù gāigé ‘the revolution of the education.’ In these constructions, 經濟 jīngjì ‘economy’ and 教育 jiàoyù ‘education’ are the objects of the verbs (not the subjects). For noise like this, it is non-threatening under two conditions: (a) if it is small in number and does not form a pattern during the clustering of collocates (i.e., the ones that do not form a cluster fall out); and (b) if the collocates are categorized only into relations that are wrongly termed (‘subject’ instead of ‘object’) but all the collocates that are aligned possess the same behaviors (such as 經濟建設 jīngjì jiànsè ‘the building of the economy,’ 文明建設 wénmíng jiànsè ‘the building of civilization’ and 道路建設 dàolù jiànsè ‘the building of roads;’ see Figure 10.1). In this case, only the name for the relation is wrong; collocates and their behaviors are consistent. For the current book, noise at the technical level is not removed manually, as the majority fall into the previously stated conditions, which makes the noise less threatening than it appears to be at first glance.
10.3.4 Measurement of Cut-off Points versus Evaluation of the Methods

The bottom-up approach is also an innovative approach that suggests using a large amount of data with the incorporation of cut-off points. The measurement of cut-off points is important for empirical research, especially in discovering the salient results from linguistic listings. For example, the measurement of cut-off points can be applied to all corpora, providing a summary of the search results using any form of calculation (Mutual Information values, saliency values, raw frequencies, percentages, etc.). It is, therefore, powerful by virtue of being able to help users select the salient results subjectively.

However, since the measurements of cut-off points can be accomplished using several methods (three methods were demonstrated in Chapter Nine), there should be several ways to evaluate the results. For example, the cut-off points for the different methods are different. One possible way to evaluate the cut-off points is through human experiments, i.e., to run psycholinguistic experiments to compare the results of the different cut-off points to the choices made by a sample of human subjects. This step needs to be carried out so as to validate the results of the cut-off points.12

10.3.5 Replicability versus Human Interference

Four steps are involved in the bottom-up approach, and they should be replicable, provided the availability of the required resources exists. Unlike the top-down approach, the bottom-up approach does not involve human ratings but involves human decision-making in the process. First, human decision-making is needed when selecting one final method for cut-off points. Although the results are still replicable if the same method is used, the decision regarding which method to use still exists (see previous section for the discussion of this topic). Second, human decision-making is also needed when selecting the top WordNet levels (with the highest conceptual density) for source domain analysis. Although supported by valid reasoning, human interference as such may still be questioned, and we propose to overcome these two problems by (a) using decisions from more subjects; and (b) analyzing the results of the different top WordNet levels and making evaluation regarding these levels. These two possible solutions will be proposed for follow-up research.

12 For the current book, we will propose this experiment as a follow-up study. We propose to provide human subjects with different saliency lists from Sketch Engine. We will then ask the subjects to select the salient results from each list. We will then calculate the average position where these subjects divide each list and compare the results of the different methods to see how close they are to the different methods.
10.4 Summary of Chapter

In this chapter, the performance of the overall bottom-up approach is discussed. We present the source domains as well as the types of metaphorical expressions for the Single Target Domains and the Coordinated Target Domains. The source domains obtained using the bottom-up approach are found to be specific, with the regional differences between CNA and XIN highlighted.

Since different approaches have their strengths and weaknesses, some problems are to be anticipated. Kövecses (2006:192), in his review of a book regarding figurative language from cross-cultural and cross-linguistic perspectives, suggests two problems that may exist between the top-down and bottom-up approaches: First, “top-down researchers do not see the trees for the forests, and bottom-up researchers do not see the forest for the trees.” That means that the top-down approach sees the big picture of conceptual metaphors, while the bottom-up approach sees the details in parts of the big picture without knowing what the big picture could be. Second, “top-down researchers primarily look for and find regularities, whereas bottom-up researchers tend to find irregularities in their data.” Therefore, the problems that may occur during the bottom-up approach are also discussed in this chapter.

In Chapter Eleven, we select a sampling of results from both approaches, and these results will be rated by human subjects in terms of the correctness of source domains. Chapter Twelve will re-focus on the source domains discussed using a linguistic framework, while Chapter Thirteen will summarize the whole book.

---

13 Squared brackets indicate correction of Kövecses’ (2006:192) use of “researches” instead of “researchers.”
Chapters Eight and Ten examined the results for the top-down and bottom-up approaches, displaying the most frequently occurring source domains, as well as the types of metaphorical expressions under each source domain. In this chapter, we provide an evaluation of the correctness of the source domains determined based on a sampling of the results. In the following §11.1, we discuss the design of the experiment. In §11.2, we discuss the results of specific analysis of the top-down approach, and in §11.3, analysis of the bottom-up approach will be carried out. In §11.4, the limitations of the experiment will be discussed, and §11.5 will summarize the chapter.

11.1 Experiment Design for Measuring Correctness of Source Domains

In order to measure the precision of the source domains that are determined using the computational methods in earlier chapters, a human rating task was carried out. The following discusses the design of stimuli.

11.1.1 Selection of Stimuli

The purpose of this rating task is to measure how well the top-down and bottom-up approaches succeeded in determining source domains, or, stated in another way, how well the automatic determination of source domains yields correct results. The selection of experimental stimuli requires several considerations. First, it is not possible to test all the results because this will create far too many questions for the subjects to answer. Each question is comprised of pairs of expressions such as (arrows) in Table 11.1 for both the top-down and the bottom-up approaches.
Table 11.1: Examples of Stimuli for Human Rating Task

<table>
<thead>
<tr>
<th>Approaches</th>
<th>Datasets</th>
<th>Types of Metaphorical Expressions</th>
<th>Source Domains Determined (English)</th>
<th>Source Domains Determined (Chinese)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top-down (Conceptual Knowledge Approach)</td>
<td>CNA+XIN</td>
<td>成長 chéngzhāng ‘grow/growth’</td>
<td>INTENTIONAL PROCESS</td>
<td>意向性歷程 yìxiàngxìnglìchéng</td>
</tr>
<tr>
<td></td>
<td>SinicaBow</td>
<td>穩定 wěndìng ‘stable’</td>
<td>PROCESS</td>
<td>歷程 lìchéng</td>
</tr>
<tr>
<td></td>
<td>The Merged Word List</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bottom-up (Prototypical Approach)</td>
<td>CNA</td>
<td>成長 chéngzhāng ‘grow/growth’</td>
<td>BAIRN</td>
<td>小孩 xiǎohái</td>
</tr>
<tr>
<td></td>
<td>XIN</td>
<td>穩定 wěndìng ‘stable’</td>
<td>TEENS</td>
<td>青少年 qīngshàoniàn</td>
</tr>
<tr>
<td></td>
<td>CNA</td>
<td>穩定 wěndìng ‘stable’</td>
<td>STRUT</td>
<td>支柱 zhīzhù</td>
</tr>
<tr>
<td></td>
<td>XIN</td>
<td>穩定 wěndìng ‘stable’</td>
<td>STILL</td>
<td>基石 jīshí</td>
</tr>
</tbody>
</table>

The pairs that will be selected for the experiment are those that are shaded. As mentioned, the results for the top-down approach are similar for CNA and XIN (because the results come from a conceptual structure that does not distinguish between differences of communities). Many of the source domains obtained in the Merged Word List are similar to those found in SinicaBow, as seen in Table 11.1 for 成長 chéngzhāng ‘grow/growth’ and 穩定 wěndìng ‘stable.’ As for the bottom-up approach, when collocation is an important feature, CNA is different from XIN because collocates of the same types of metaphorical expressions may differ if they co-occur with different words when used in Taiwan and when used in China (see 成長 chéngzhāng ‘grow/growth’ and 穩定 wěndìng ‘stable’ in Table 11.1).

However, the SinicaBow and the Merged Word List may differ in the number of source domains obtained. In the experiment to follow, it is also possible to compare the results from the keywords in the definitions as well as the results from the SUMO nodes. As for the bottom-up approach, the design of the stimuli is based mainly on their sources, i.e., CNA, XIN or CNA and XIN.
The pairs of results (types of metaphorical expressions/source domains) are large in number in the overall results. The following shows the number of pairs for both the top-down and the bottom-up approaches.²

(1) Total: 64,400 pairs³
   a. Top-down
      SinicaBow: 10,026 pairs
      Word List: 30,683 pairs
   b. Bottom-up
      CNA: 15,341 pairs
      XIN: 8,350 pairs

In total, there are 64,400 pairs of results collected from the top-down and the bottom-up approaches, a volume of data far too large to be checked by human beings. This is the first consideration for the design of stimuli. A more economic way to deal with the results is by selecting a sample of the pairs for human ratings. Second, the fact that there are different sets of results—SinicaBow versus the Merged Word List in the top-down approach as well as CNA versus XIN in the bottom-up approach—the selection of stimuli has to take these into consideration. We have to balance between these datasets so as to produce meaningful comparisons between them. Third, since we decided to use Chinese-Chinese for the pairs of types of metaphorical expressions/source domains, we also had to consider whether all source domains in all results had a Chinese translation.⁴ Last, the types of metaphorical expressions selected (such as 成長 chéngzhǎng ‘grow/growth’ and 穩定 wēndìng ‘stable’ from Table 11.1) for the top-down and bottom-up approaches have to be similar so as to evaluate which of these approaches is better in determining source domains. The considerations mentioned previously are summarized in (2).

(2) a. A sampling of the results (pairs of types of metaphorical expressions and source domains) has to be selected.
   b. The pairs based on different datasets have to be balanced across the top-down and the bottom-up approaches.

² Most of the types of metaphorical expressions are repeated if they yield more than one source domain.
³ The pairs from the bottom-up approach are based on the top-three rankings of the WordNet top hierarchies with the highest conceptual density (see Chapter Nine).
⁴ Some source domains (about 30 per cent) in the top-down approach do not have Chinese translations because these source domain terms come from words in the SUMO definitions that are not readily translated (only the nodes of SUMO have been translated and checked).
c. All selected pairs must have Chinese translation, especially for the source domains of the top-down approach.
d. Similar types of metaphorical expressions are needed in order to compare the top-down approach with the bottom-up approach.

Based on these four criteria, all pairs that meet these criteria were singled out as stimuli for the experiment. First, source domains without Chinese translations in the top-down approach were not considered for testing. The remaining source domains were considered only if they fulfilled the criteria in stimuli selection of the bottom-up approach (so as to select similar types of metaphorical expressions in both the top-down and the bottom-up approaches).

As for the bottom-up approach, we have selected ten types of metaphorical expressions from those that are found in CNA-only; another ten types of metaphorical expressions in XIN-only; and ten types of metaphorical expressions that are found in CNA and XIN, shown in (3).

(3) a. Ten types of metaphorical expressions (CNA-only)
   b. Ten types of metaphorical expressions (XIN-only)
   c. Ten types of metaphorical expressions (CNA and XIN; 5 CNA-based; 5 XIN-based)

For the types of metaphorical expressions that overlapped in CNA and XIN, their selection was controlled so that five of the types of metaphorical expressions are the top five overlapped in CNA while the remaining five are the top five overlapped in XIN, illustrated in Table 11.2. In this table, all overlapped types of metaphorical expressions in CNA and XIN are shown. This means that the types of metaphorical expressions in CNA and XIN are exactly the same (if they repeat one another), but they are arranged differently, depending on how frequent they are found in the sub-corpora selected.

In Table 11.2, the top five CNA-based types of metaphorical expressions are in bold. These five types of metaphorical expressions were selected because they appear as the top five positions in CNA (thus, CNA-based). From Table 11.2, we can also see that the top five metaphorical expressions in CNA are scattered throughout different positions in the XIN list (which is the reason why we selected the top five from each list). In contrast, the other five metaphorical expressions in XIN are shaded. The remaining top five were selected from XIN in addition to those five that were already selected (thus, XIN-based).

---

5 We started the selection from the CNA-based top five metaphorical expressions (to start the selection from XIN will yield different results) because the subjects are all Taiwanese, and we wanted to include the most frequently appearing overlapped types of metaphorical expressions in CNA.
### Table 11.2: Types of Metaphorical Expressions That Overlap in CNA and XIN

<table>
<thead>
<tr>
<th>Types of Metaphorical Expressions</th>
<th>Tokens of Metaphorical Expressions</th>
<th>Types of Metaphorical Expressions</th>
<th>Tokens of Metaphorical Expressions</th>
</tr>
</thead>
<tbody>
<tr>
<td>建設 jiànshè ‘construct/construction’</td>
<td>1,612</td>
<td>建設 jiànshè ‘construct/construction’</td>
<td>1,487</td>
</tr>
<tr>
<td>改革 gǎigé ‘reform/reformation’</td>
<td>1,546</td>
<td>秩序 zhìxù ‘order’</td>
<td>843</td>
</tr>
<tr>
<td>成長 chéngzhǎng ‘grow/growth’</td>
<td>979</td>
<td>穩定 wěndìng ‘stable’</td>
<td>608</td>
</tr>
<tr>
<td>情勢 qíngshì ‘situation’</td>
<td>473</td>
<td>形勢 xíngshì ‘terrain’</td>
<td>602</td>
</tr>
<tr>
<td>制度 zhìdù ‘system’</td>
<td>439</td>
<td>情勢 qíngshì ‘situation’</td>
<td>456</td>
</tr>
<tr>
<td>制裁 zhìcái ‘impose.sanction’</td>
<td>376</td>
<td>制度 zhìdù ‘system’</td>
<td>428</td>
</tr>
<tr>
<td>危機 wēiji ‘crisis’</td>
<td>365</td>
<td>困難 kùnnán ‘difficult/difficulty’</td>
<td>350</td>
</tr>
<tr>
<td>穩定 wěndìng ‘stable’</td>
<td>328</td>
<td>增長 zèngzhǎng ‘grow/growth’</td>
<td>332</td>
</tr>
<tr>
<td>風險 fēngxiǎn ‘risk’</td>
<td>251</td>
<td>改革 gǎigé ‘reform/reformation’</td>
<td>322</td>
</tr>
<tr>
<td>困難 kùnnán ‘difficult/difficulty’</td>
<td>232</td>
<td>進行 jìnghé ‘carry.out’</td>
<td>317</td>
</tr>
<tr>
<td>力量 lìliàng ‘strength’</td>
<td>221</td>
<td>衰退 shuāitù ‘decay’</td>
<td>304</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>交流 jiāoliú ‘interflow’</td>
<td>168</td>
<td>結構 jiégòu ‘structure’</td>
<td>195</td>
</tr>
<tr>
<td>秩序 zhìxù ‘order’</td>
<td>157</td>
<td>開展 kāi zhǎn ‘open.and.spread’</td>
<td>181</td>
</tr>
<tr>
<td>壓力 yālì ‘pressure’</td>
<td>146</td>
<td>交流 jiāoliú ‘interflow’</td>
<td>173</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>形勢 xíngshì ‘terrain’</td>
<td>79</td>
<td>接受 jiēshòu ‘accept’</td>
<td>66</td>
</tr>
<tr>
<td>手段 shǒuduàn ‘trick’</td>
<td>76</td>
<td>舞台 wǔtái ‘stage’</td>
<td>65</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>增長 zèngzhǎng ‘grow/growth’</td>
<td>32</td>
<td>風險 fēngxiǎn ‘risk’</td>
<td>15</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>進入 jìnnrù ‘enter’</td>
<td>21</td>
<td>成長 chéngzhǎng ‘grow/growth’</td>
<td>8</td>
</tr>
<tr>
<td>潛力 qiánlì ‘potential’</td>
<td>20</td>
<td>孤立 gūlì ‘isolate/isolation’</td>
<td>8</td>
</tr>
</tbody>
</table>

---

Three dots in Table 11.2 indicate that part of the rows have been removed for the presentation of selected stimuli.
The final stimuli that fulfill all criteria in (2) are given in Table 11.3.

<table>
<thead>
<tr>
<th>Types of Metaphorical Expressions</th>
<th>Tokens of Metaphorical Expressions</th>
<th>Types of Metaphorical Expressions</th>
<th>Tokens of Metaphorical Expressions</th>
<th>Types of Metaphorical Expressions</th>
<th>Tokens of Metaphorical Expressions</th>
</tr>
</thead>
<tbody>
<tr>
<td>衰弱 shuāiruò ‘weak/weakness’</td>
<td>140</td>
<td>熱情 rèqìng ‘zeal’</td>
<td>13</td>
<td>建設 jiàoshè ‘construct/construction’</td>
<td>1,612</td>
</tr>
<tr>
<td>庇護 bìhù ‘protect/protection’</td>
<td>129</td>
<td>睚悟 jùwù ‘conscious of/consciousness’</td>
<td>13</td>
<td>改革 gǎigé ‘innovate/innovation’</td>
<td>1,546</td>
</tr>
<tr>
<td>革新 géixin ‘innovate/innovation’</td>
<td>60</td>
<td>占 zhàn ‘occupy’</td>
<td>9</td>
<td>成長 chéngzhǎng ‘grow/growth’</td>
<td>979</td>
</tr>
<tr>
<td>迫害 pòhài ‘persecute/persecution’</td>
<td>49</td>
<td>實行 shíxíng ‘implement/implementation’</td>
<td>9</td>
<td>情勢 qíngshì ‘situation’</td>
<td>473</td>
</tr>
<tr>
<td>停滯 tóngzhì ‘cease moving’</td>
<td>29</td>
<td>堅定 jiāndìng ‘firm’</td>
<td>8</td>
<td>制度 zhídù ‘system’</td>
<td>439</td>
</tr>
<tr>
<td>傷害 shānghài ‘wound’</td>
<td>28</td>
<td>海洋 hǎiyáng ‘ocean’</td>
<td>7</td>
<td>秩序 zhìxù ‘order’</td>
<td>843</td>
</tr>
<tr>
<td>生態 shēngtài ‘ecology’</td>
<td>25</td>
<td>走廊 zǒuláng ‘corridor’</td>
<td>7</td>
<td>穩定 wěndìng ‘stable’</td>
<td>608</td>
</tr>
<tr>
<td>報復 bàofù ‘revenge’</td>
<td>20</td>
<td>合唱 hécháng ‘choir’</td>
<td>6</td>
<td>形勢 xíngshì ‘terrain’</td>
<td>602</td>
</tr>
<tr>
<td>和諧 héxié ‘harmony’</td>
<td>19</td>
<td>彰固 gānggu ‘solid/solidify’</td>
<td>6</td>
<td>困難 kùnnán ‘difficult/difficulty’</td>
<td>350</td>
</tr>
<tr>
<td>智慧 zhìhuì ‘wisdom’</td>
<td>18</td>
<td>上升 shàngshēng ‘rise’</td>
<td>5</td>
<td>增長 zēngzhǎng ‘increase’</td>
<td>332</td>
</tr>
</tbody>
</table>

Based on the results of the computational methods, each type of metaphorical expression can be mapped to different source domains because the same type of metaphorical expression can belong to different source domains (such that 衰弱 shuāiruò ‘weakness’ may belong to source domains such as ORGANISM/生物體 as well as ANIMNAL/動物). Examples of pairings of one type of metaphorical expression with several source domains can be seen in (4).
(4) a. Top-down (from SinicaBow)$^7$

**Type of Metaphorical Expression**
衰弱 *shuāiruò* ‘weakness’

**Source Domains**
- 生物體 *shēngwùtǐ* ‘organism’
- 動物 *dòngwù* ‘animal’
- 組織 *zǔzhī* ‘tissues’
- 屬性 *shùxìng* ‘attribute’

b. Bottom-up (from CNA-only)

**Type of Metaphorical Expression**
衰弱 *shuāiruò* ‘weakness’

**Source Domains**
- 神經 *shénjīng* ‘nerve’
- 心臟 *xīnzàng* ‘heart’
- 體力 *tǐlì* ‘energy’
- 肌肉 *jīròu* ‘flesh’

In (4), we can see that a type of metaphorical expression, such as 衰弱 *shuāiruò* ‘weak/weakness,’ may contain one or more source domain in both approaches. In the top-down approach, the source domains yielded are the keywords in SUMO definitions. Some of these keywords are SUMO nodes (屬性 *shùxìng* ‘attribute,’ 生物體 *shēngwùtǐ* ‘organism’ and 動物 *dòngwù* ‘animal’) while others are not (組織 *zǔzhī* ‘tissues’). Therefore, the thirty types of metaphorical expressions in Table 11.3 in total yield 415 questions in the top-down questionnaire and 413 questions in the bottom-up questionnaire.

In the top-down approach, the total types of metaphorical expressions in Table 11.3 yield the following number of pairs in Table 11.4:

<table>
<thead>
<tr>
<th></th>
<th>CNA-only</th>
<th>XIN-only</th>
<th>CNA and XIN</th>
<th>Repeated in CNA and XIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>SinicaBow</td>
<td>63</td>
<td>39</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>Merged Word List</td>
<td>91</td>
<td>79</td>
<td>102</td>
<td>6</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>154</strong></td>
<td><strong>118</strong></td>
<td><strong>149</strong></td>
<td><strong>6</strong></td>
</tr>
<tr>
<td><strong>Grand Total</strong></td>
<td><strong>415 (removed repeated 6)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^7$ It is only in this chapter during the experiment that the source domains for the top-down approach are given translations (for the purpose of testing). In other chapters, only English will be given.

$^8$ For ease of comparison, the numbers of the top-down stimuli in this table are shown in the datasets of CNA-only, XIN-only and CNA and XIN, which are only distinguished in the bottom-up approach.
The ten types of metaphorical expressions that have been selected from CNA-only yield sixty-three pairs (as those exemplified in Table 11.1) in SinicaBow and ninety-one pairs in the Merged Word List (for the same types of metaphorical expressions). In XIN-only, the other ten types of metaphorical expressions yield thirty-nine pairs in SinicaBow and seventy-nine pairs in the Merged Word List. In total, there are 154 from CNA-only and 118 from XIN-only. As for those found in both CNA and XIN, there are a total of 149 pairs, but 6 of these pairs are repeated in both CNA and XIN, and therefore, they were removed. This leaves a grand total of 415 pairs for the top-down approach.

As for the bottom-up approach, the total number of stimuli is shown in Table 11.5.

<table>
<thead>
<tr>
<th>CNA-only</th>
<th>XIN-only</th>
<th>CNA and XIN (data from CNA)</th>
<th>CNA and XIN (data from XIN)</th>
<th>Repeated in CNA and XIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>110</td>
<td>87</td>
<td>118</td>
<td>122</td>
<td>24</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>413 (removed repeated 24)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The stimuli from the bottom-up approach are slightly different from the top-down approach because, for the types of metaphorical expressions found in CNA and XIN (shown previously in Table 11.3), the results from CNA and XIN many differ. Therefore, the pairs that are different were tested separately, while the repeated ones (twenty-four pairs) were asked once only so that they would not be repeated in the questionnaire. In total, the bottom-up approach yields 413 pairs for the rating task.

The pairs in the bottom-up approach also differ from the top-down approach in the sense that they are selected based on two additional parameters (not stated in (2)). The first parameter is the different results between CNA and XIN. The second parameter is based on the saliency values of the source domains selected (which are the bottom nodes of the WordNet noun hierarchies with the highest conceptual density (see Chapter Nine). For example, Table 11.6 shows the top three rankings of WordNet noun hierarchies (selected earlier in Chapter Nine) for 建設 jiànshè ‘construct/construction.’ The bottom nodes in column five are taken as source domain names for the bottom-up approach, and these nodes are the most specific terms under the top hierarchies of ‘entity, physical_thing,’ ‘abstraction’ and ‘group, grouping’ (column three). As can be seen in Table 11.6, the bottom nodes in column five have different saliency values (column six). The experiment design in this chapter also takes this into consideration.

For example, among all the bottom nodes in column five in Table 11.6, 軍隊 jùnduì ‘troop’ has the highest saliency value (29.24). Therefore, 軍隊 jùnduì ‘troop’ is considered to be the source domain with the highest saliency value among all the source domains (i.e., bottom nodes) found under the top three WordNet hierarchies selected.
The second highest will be 公路 gōnglù ‘high road, trunk road’ and 通道 tōngdào ‘gangway’ with a saliency value of 25.67 for each.

Table 11.6: Top Three WordNet Noun Hierarchy Selected and Their Bottom Nodes

<table>
<thead>
<tr>
<th>Type of Metaphorical Expressions</th>
<th>Relation</th>
<th>Top Three WordNet Noun Hierarchies</th>
<th>English Synset</th>
<th>Bottom Nodes_Frequency</th>
<th>Saliency</th>
<th>Ranking of Saliency</th>
</tr>
</thead>
<tbody>
<tr>
<td>建設 jiànshè ‘construct/construction’</td>
<td>n_modifer entity, physical_thing</td>
<td>highroad, trunk_road 公路_747</td>
<td>25.67</td>
<td>Rank 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>gangway 通道 21</td>
<td>25.67</td>
<td>Rank 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>corridor 通道 21</td>
<td>9.65</td>
<td>Rank 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>areaway 通道 21</td>
<td>9.65</td>
<td>Rank 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>thoroughfare 通道 747, 道路 501</td>
<td>9.65</td>
<td>Rank 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>thoroughfare 公路_747, 道路_501</td>
<td>9.65</td>
<td>Rank 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>abstractions virtue, virtuousness, moral_excellence 道德_100</td>
<td>14.61</td>
<td>Rank 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>group, grouping troop 軍隊_581</td>
<td>29.24</td>
<td>Rank 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>legion 軍隊_581</td>
<td>29.24</td>
<td>Rank 1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Thus, each type of metaphorical expression selected as experimental stimuli comprises source domains with different saliency values. The lowest saliency ranking that was included in this experiment appears at the sixth level (if such a level is found with a type of metaphorical expression). In Table 11.6, only four levels of saliency ranking are found for 建設 jiànshè ‘construct/construction’ (column seven). The reason why the saliency information is included is because, provided with this information, we will be able to compare the precision of the saliency values based on whether or not ‘Rank 1’ will be rated better than ‘Rank 2’ to ‘Rank 6’ in the human judgment category. Therefore, the number of stimuli shown in Table 11.5 also contains this additional parameter for the bottom-up approach.

11.1.2 Questionnaires

After all the stimuli were selected, they were arranged so that pairs with similar types of metaphorical expressions were not repeated on the same page, i.e., there was only one 建設 jiànshè ‘construct/construction’ on each page (see Table 11.6). There were two sets of questionnaires, one for top-down and one for bottom-up. The top-down questionnaire contains 415 questions, while the bottom-up questionnaire contains 413 questions. Both sets have different instructions, as seen in (5).9

9 Questionnaires are given in Appendix A11.1.
(5) a. Top-down: 閣卷共有 415 題，每個題目裡都有 A 和 B 兩個詞組。A
是詞義概念，B 是詞義的領域。請你在看過詞組之後，根據你的語感
和直覺，決定每題裡 A 的詞義概念是否屬於 B 的領域，然後圈選作
答。作答時，請在是或否兩個選項間選你覺得最適當的答案，每題
都必須作答。
Translation: This questionnaire contains 415 questions. Each question is
comprised of two lexical items, A and B. A shows the meaning concept
while B shows the domain of meaning. For all lexical items, please use
your intuition to decide whether the meaning concept of A belongs to
the domain of B. Please select between ‘Yes’ and ‘No,’ and please
answer all questions.

b. Bottom-up: 閣卷共有 413 題，每個題目裡都有 A 和 B 兩個詞組。請
你在看過詞組之後，根據你的語感和直覺，決定每題裡 A 和 B 是否
屬於相同的概念領域，然後圈選作答。作答時，請在是或否兩個選項
間圈選你覺得最適當的答案，每題都必須作答。
Translation: This questionnaire contains 413 questions. Each question is
comprised of two lexical items, A and B. For all lexical items, please use
your intuition to decide whether A and B belong to the same concept
domain. Please select between ‘Yes’ and ‘No,’ and please answer all
questions.

The pairs of stimuli take the form demonstrated in (6). The question of the task was
displayed on the header of each page in a document file. Only one of these questions in
the header of (6) was asked, depending on the set of questionnaires used.

(6) Top-down: 請問，A 的詞義概念是否屬於 B 的領域？
Bottom-up: 請問，A 和 B 是否屬於相同的概念領域？

| A       | B       |  |  |
|---------|---------|  |  |
| 1. 手錶  | 戒指    | 是 | 否 |
| 2. 頭髮  | 文具    | 是 | 否 |
| 3. 打球  | 運動    | 是 | 否 |
| 4. 潛水  | 打架    | 是 | 否 |

The instructions were different in (5) for the top-down approach (whether A belongs to
B) and the bottom-up approach (whether A and B belong to the same concept domain)
because we realized the results obtained by each of the two approaches would be different. The source domains determined through SUMO in the top-down approach are at a higher level, and they include the types of metaphorical expressions (thus, A belongs to B) shown in (7).\(^{10}\)

(7) Top-down: Do the types of metaphorical expressions belong to the source domains?

<table>
<thead>
<tr>
<th>Types of Metaphorical Expressions</th>
<th>Source Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>成長 chéngzhǎng ‘grow/growth’</td>
<td>ORGANISM/生物體</td>
</tr>
<tr>
<td>成長 chéngzhǎng ‘grow/growth’</td>
<td>PROCESS/歷程</td>
</tr>
<tr>
<td>改革 gǎigé ‘reform/reformation’</td>
<td>CREATION/產生</td>
</tr>
<tr>
<td>改革 gǎigé ‘reform/reformation’</td>
<td>ARTIFACT/人造物</td>
</tr>
<tr>
<td>建設 jiànsè ‘construct/construction’</td>
<td>COGNITIVE AGENT/具認知力施事者</td>
</tr>
<tr>
<td>建設 jiànsè ‘construct/construction’</td>
<td>ORGANIZATIONAL PROCESS/組織性歷程</td>
</tr>
</tbody>
</table>

However, for the bottom-up approach, the results are different because the source domains are collocates of the types of metaphorical expressions with the highest conceptual density. Therefore, it is more often the case that the source domains in the bottom-up approach appear together in texts rather than the source domains of the top-down approach, which usually belong to upper concepts. For example, in example (8), it is more natural to see FOREIGN AFFAIRS belonging to COGNITVE AGENT than FOREIGN AFFAIRS belonging to CAPACITY.

(8) Top-down: FOREIGN AFFAIRS AS COGNITIVE AGENT/具認知力施事 (SinicaBow)
Bottom-up: FOREIGN AFFAIRS ARE CAPACITY/能力 (CNA)

國際外交將聯合對日制裁
guójì wàijīáo jiàng liánhé duì rì zhìcái
‘International foreign affairs will unite and impose sanctions against Japan.’

---

\(^{10}\) Notice that the source domains for the top-down approach are English-based because the source (SUMO definitions) from which these source domains are determined is English-based.
In fact, CAPACITY/能力 gives more information to 制裁 zhicài ‘impose sanction’ than to 外交 wàijiāo ‘foreign affairs.’ Based on this reason, we ask the question whether A (制裁 zhicài ‘impose sanction’) and B (能力 nênglî ‘capacity’) fall under the same concept domain. If they belong to the same concept domain, CAPACITY/能力 will be considered a suitable source domain for 外交 wàijiāo ‘foreign affairs.’ More pairs of stimuli are given in (9).11

(9) Bottom-up: Do the types of metaphorical expressions fall under the same concept domains as the source domains?

<table>
<thead>
<tr>
<th>Types of Metaphorical Expressions</th>
<th>Source Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>成長 chéngzhǎng ‘grow/growth’</td>
<td>BAIRN/小孩</td>
</tr>
<tr>
<td>成長 chéngzhǎng ‘grow/growth’</td>
<td>FRUITAGE/產量</td>
</tr>
<tr>
<td>改革 gâigé ‘reform/reformation’</td>
<td>TROOP/軍隊</td>
</tr>
<tr>
<td>改革 gâigé ‘reform/reformation’</td>
<td>ASSIGNMENT/任務</td>
</tr>
<tr>
<td>建設 jiànsè ‘construct/construction’</td>
<td>RESIDENCE/住宅</td>
</tr>
<tr>
<td>建設 jiànsè ‘construct/construction’</td>
<td>ROAD/公路</td>
</tr>
</tbody>
</table>

Similarly, when subjects agree that 改革 gâigé ‘reform/reformation’ belongs to the same concept domain as TROOP/軍隊, we can also conclude that TROOP/軍隊 is a suitable source domain for the target domain of EDUCATION.

(10) EDUCATION IS A TROOP/軍隊

提高 學生 學習 興趣 成為 改革
tígāo xuēshēng xuéxí xìngqù chéngwéi gâigé
raise student learning interest become reform

山地 教育 的 當務之急
shāndì jiàoyù de dāngwùzhìjí
mountainous.region education DE current.urgency

‘To be able to raise the interest of learning has become the current urgency in the reformation of aboriginal education.’

All pairs of stimuli (A and B) were organized in one set of questionnaires, which were answered within one single session without a break, respectively for top-down and bottom-up questionnaires. In order to avoid fatigue, half of the questions were arranged in the correct order, while the other half were arranged in reversed order.

11 Notice that the source domains for the bottom-up results are Chinese-based because the source domains are the collocates of the types of metaphorical expressions. Subjects were informed of this fact.
11.1.3 Subjects and Procedures

There were seventeen subjects (three males and fourteen females). All of the subjects work as assistants at the Institute of Linguistics, Academia Sinica. All subjects have formal education above university level. Among these seventeen subjects, nine come from the Chinese WordNet (CWN) group and are highly sensitive to the linguistic senses of Chinese words. The remaining eight subjects have some linguistic training, but they do not analyze word senses as the CWN group does (as it is their jobs). As a result, the ratings of these subjects were analyzed as CWN and non-CWN groups. The mean age for the CWN group is 27.44 (ranging from 24 to 32) while the mean age for the non-CWN group is 32 (ranging from 23 to 42). Their mean age was not significantly different from one another in an independent t-test: \( t(9)=1.96, p=.084 \). On a scale of 1 to 7, with 1 being least proficient and 7 being most proficient, the mean proficiency in Mandarin for the CWN group is 6.6 and for the non-CWN group it is 6.3. As for the mean proficiency in Taiwanese, the CWN group has a mean of 4.6 while the non-CWN group has a mean of 5.4.

All subjects answered both the top-down and the bottom-up questionnaires in two separate sessions, and each session lasted between 30 to 90 minutes, depending on the subjects’ self-pace. All subjects were gathered in a seminar room, and they answered the questions together. The first session was run a few days before the second session so that subjects would not be affected by the different questions asked in the two questionnaires. These subjects were paid NT$150 per hour for answering the questionnaires, and most of the subjects finished all questions within half an hour per session. The questionnaires were counter-balanced in each session. In the first session, half of the subjects answered the top-down questionnaires and half answered the bottom-up questionnaires. The order of the questions in the questionnaires was also counter-balanced. Among the subjects who answered the top-down questionnaire, half of them received questionnaires with the correct order while the other half received the reversed order. This was similarly done for the bottom-up questionnaire. The subjects who answered questionnaires with the correct order in the first session were then given questionnaires with reversed order in the second session.

11.1.4 Results of the Correctness Ratings

For the analysis of the ratings, percentages of ‘Yes’ were first computed for each subject in different conditions.\(^{12}\) We used a Kruskal-Wallis test to analyze the results.

\(^{12}\) From both sets of questionnaires, there were thirty-two pairs removed from the top-down questionnaires because these pairs were repeated twice among the 415 questions. This
All subjects were included. The following Table 11.7 provides the means for the overall percentages of ‘Yes’ in the top-down and bottom-up approaches.

### Table 11.7: Means for Overall Percentages of ‘Yes’¹³

<table>
<thead>
<tr>
<th>Approaches</th>
<th>Mean Percentages</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top-down</td>
<td>.3120</td>
<td>.1668</td>
</tr>
<tr>
<td>Bottom-up</td>
<td>.3712</td>
<td>.2329</td>
</tr>
<tr>
<td>Total</td>
<td>.3416</td>
<td>.2017</td>
</tr>
</tbody>
</table>

From Table 11.7, we can see that the percentage of ‘Yes’ is about 31 per cent in the top-down approach and about 37 per cent in the bottom-up approach. This means that only about 30 per cent of the time, the subjects answered ‘Yes’ to the questions in the questionnaires. However, when comparing between the top-down and bottom-up approaches using a mixed design, their means of the percentages are marginal in producing a significant effect, $F(1,34)=4.52, p=.05$, indicating that 37 per cent is being marginally significant to 31 per cent.

The results for the questionnaires were then analyzed separately between the top-down and the bottom-up approaches as well as between the CWN and non-CWN groups, and the means of the percentages of ‘Yes’ are given in Table 11.8.

¹³ Kurtosis and skewness values assume a normal distribution.

### Table 11.8

<table>
<thead>
<tr>
<th>Approaches</th>
<th>CWN/ non-CWN</th>
<th>Mean</th>
<th>N</th>
<th>Std. Deviation</th>
<th>Kurtosis</th>
<th>Std. Error of Kurtosis</th>
<th>Skewness</th>
<th>Std. Error of Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top-down</td>
<td>CWN</td>
<td>0.408736</td>
<td>8</td>
<td>0.149954</td>
<td>0.315401237</td>
<td>1.48088</td>
<td>0.317252</td>
<td>0.752101</td>
</tr>
<tr>
<td></td>
<td>Non-CWN</td>
<td>0.231692</td>
<td>9</td>
<td>0.153436</td>
<td>2.218356902</td>
<td>1.399708</td>
<td>1.292834</td>
<td>0.717137</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>0.315007</td>
<td>17</td>
<td>0.172934</td>
<td>-0.464404322</td>
<td>1.063198</td>
<td>0.455287</td>
<td>0.549747</td>
</tr>
<tr>
<td>Bottom-up</td>
<td>CWN</td>
<td>0.542148</td>
<td>8</td>
<td>0.168319</td>
<td>-0.724073074</td>
<td>1.48088</td>
<td>-0.77234</td>
<td>0.752101</td>
</tr>
<tr>
<td></td>
<td>Non-CWN</td>
<td>0.231418</td>
<td>9</td>
<td>0.228237</td>
<td>0.247489089</td>
<td>1.399708</td>
<td>1.281947</td>
<td>0.717137</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>0.377644</td>
<td>17</td>
<td>0.252981</td>
<td>-1.744052775</td>
<td>1.063198</td>
<td>-0.0086</td>
<td>0.549747</td>
</tr>
<tr>
<td>Total</td>
<td>CWN</td>
<td>0.475442</td>
<td>16</td>
<td>0.168704</td>
<td>-1.172095199</td>
<td>1.090774</td>
<td>-0.08778</td>
<td>0.564308</td>
</tr>
<tr>
<td></td>
<td>Non-CWN</td>
<td>0.231555</td>
<td>18</td>
<td>0.188661</td>
<td>0.544944832</td>
<td>1.037795</td>
<td>1.220062</td>
<td>0.536278</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>0.346325</td>
<td>34</td>
<td>0.215732</td>
<td>-1.248224118</td>
<td>0.787898</td>
<td>0.264648</td>
<td>0.403053</td>
</tr>
</tbody>
</table>
Table 11.8: Means for Overall Percentages of ‘Yes’ According to Conditions

<table>
<thead>
<tr>
<th>Approaches</th>
<th>Groups</th>
<th>Mean Percentage</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top-down</td>
<td>CWN</td>
<td>.2317</td>
<td>.1534</td>
</tr>
<tr>
<td></td>
<td>Non-CWN</td>
<td>.4024</td>
<td>.1381</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>.3120</td>
<td>.1688</td>
</tr>
<tr>
<td>Bottom-up</td>
<td>CWN</td>
<td>.2398</td>
<td>.2205</td>
</tr>
<tr>
<td></td>
<td>Non-CWN</td>
<td>.5190</td>
<td>.1456</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>.3712</td>
<td>.2329</td>
</tr>
<tr>
<td>Total</td>
<td>CWN</td>
<td>.2357</td>
<td>.1843</td>
</tr>
<tr>
<td></td>
<td>Non-CWN</td>
<td>.4607</td>
<td>.1497</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>.3416</td>
<td>.2017</td>
</tr>
</tbody>
</table>

From Table 11.8, we can see the means of the percentages of ‘Yes’ answered by the CWN and non-CWN groups. The non-CWN groups (shaded) are consistently higher in their percentages than the CWN group. The CWN group returned very low percentages of ‘Yes’ (23 per cent for the top-down and 24 per cent for the bottom-up). Comparatively, the non-CWN group was shown to have higher percentages of ‘Yes’ (40 per cent for the top-down and 52 per cent for the bottom-up). However, the comparisons between the top-down and bottom-up approaches do not show significant effects in independent t-tests, either within CWN ($t(14)=-1.64$, $p=.12$) or non-CWN ($t(16)=-0.09$, $p=.929$).

When the means for the overall CWN and non-CWN are compared, significance was found, $F(1,34)=58.83$, $p<.05$, indicating that the means for the percentages of ‘Yes’ differ significantly between the CWN and non-CWN group. Since the results show that the number of ‘Yes’s answered by the non-CWN group were found to be significantly higher than that of the CWN group, we interviewed some of the subjects in the CWN group. A majority of the subjects responded that they were unable to use their intuition in their answers (as required in the instructions in (5)), as they were too familiar with listing senses of a lexical item in their head. As a result, they were over-critical because they were accustomed to analyzing whether all senses of the types of metaphorical expressions matched those of senses of the source domains. For example, for the pair of 建設 jiànsè ‘construct/construction’ and ROAD/公路, the subjects in the CWN group rejected this pair if they found that one of the senses of 建設 jiànsè ‘construct/construction’ (such as that in 建設文化 jiànsè wénhuà ‘to construct the culture’) does not match all of the senses of ROAD/公路. Due to their responses, which did not fit our expectation of the experiment, the CWN group will not be analyzed further for the remaining sections in this chapter. Only the responses from the non-CWN group will be analyzed.

In the next section, specific analyses of the top-down and the bottom-up results will be discussed. This is because, for both results, there are additional factors, such as
the similarities and differences between SinicaBow and the Merged Word List for the top-down approach, the similarities and differences between CNA versus XIN, and the effect of saliency values in identifying source domains for the bottom-up approach. The discussion in the next two sections is important because it tells us whether or not computational results are able to predict the ranking of the source domains determined.

11.2 Top-down: Comparing SinicaBow and the Merged Word List

For the top-down approach, there is one factor that has been mentioned in the stimuli section, and this factor involves the similarities and the differences between SinicaBow and the Merged Word List. The overall comparison of these two knowledgebases will be given, following. Even though we showed in Table 11.4, earlier, that the types of metaphorical expressions selected for the top-down approach can be divided into words that come from CNA and XIN, these two divisions are based mainly on the divisions in the bottom-up approach. For the analyses of the top-down approach, however, the distinction between CNA and XIN is unnecessary, as source domains for the same types of metaphorical expressions in CNA and XIN are the same if they are based on similar conceptualized structures.

11.2.1 SinicaBow versus the Merged Word List

All the stimuli from the top-down questionnaire can be divided into SinicaBow and the Merged Word List, with the results compared in Table 11.9. Since the sample now is based on non-CWN, and normal distribution is difficult to assume, the analysis uses the Kruskal-Wallis test. The results do not show significant effect between the percentages of ‘Yes’ for SinicaBow and the Merged Word List: \( \chi^2(1) = .335, p = .563 \).

<table>
<thead>
<tr>
<th>Lexical Knowledgebases</th>
<th>Mean Percentages</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SinicaBow</td>
<td>.4400</td>
<td>.1554</td>
</tr>
<tr>
<td>The Merged Word List</td>
<td>.4037</td>
<td>.1486</td>
</tr>
<tr>
<td>Total</td>
<td>.4219</td>
<td>.1481</td>
</tr>
</tbody>
</table>

From Table 11.9, we see that means for the percentages of ‘Yes’ for SinicaBow is 44 per cent, and for the Merged Word List it is 40 per cent. When we compared the performance of both knowledgebases in Chapter Eight, we found that the Merged Word List gave a better performance. However, in terms of correctness of source domains,
neither knowledgebase differs significantly, meaning that the source domains provided by both knowledgebases do not differ significantly from one another.\textsuperscript{14}

In the next section, we will compare an additional factor that involves the ranking of the computed source domains.

### 11.2.2 Ranking of Source Domains

We show in Table 11.10 two types of metaphorical expressions (成長 chēngzhǎng ‘grow/growth’ and 改革 gǎigé ‘reform/reformation’) as well as a few examples of their source domains obtained through SinicaBow in the top-down approach. In the second column, we see the number of senses yielded by each type of metaphorical expressions in SinicaBow. In the fifth column, we see the number of times these source domains are found among the SUMO nodes of the senses of 成長 chēngzhǎng ‘grow/growth’ and 改革 gǎigé ‘reform/reformation.’ In the last column, we discover the ranking of the source domains by dividing the frequency with the total senses.

<table>
<thead>
<tr>
<th>Type of Metaphorical Expressions</th>
<th>Total Senses</th>
<th>Source Domains (English)</th>
<th>Source Domains (Chinese)</th>
<th>Frequency of Source Domains in the Definitions of All Senses</th>
<th>Ranking of Source Domains (Frequency/Senses)</th>
</tr>
</thead>
<tbody>
<tr>
<td>成長 chēngzhǎng ‘grow/growth’</td>
<td>6 senses</td>
<td>BIOLOGICAL PROCESS</td>
<td>生物歷程</td>
<td>3</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>INTENTIONAL PROCESS</td>
<td>意向性歷程</td>
<td>1</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ORGANISM</td>
<td>生物體</td>
<td>7</td>
<td>1.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHYSICAL QUANTITY</td>
<td>物理量</td>
<td>2</td>
<td>0.33</td>
</tr>
<tr>
<td>改革 gǎigé ‘reform/reformation’</td>
<td>9 senses</td>
<td>AGENT</td>
<td>施事者</td>
<td>2</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ARTIFACT</td>
<td>人造物</td>
<td>4</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CHANGE OF POSSESSION</td>
<td>擁有權的轉移</td>
<td>1</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PROCESS</td>
<td>歷程</td>
<td>3</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td>COGNITIVE AGENT</td>
<td>具認知力施事者</td>
<td>2</td>
<td>0.22</td>
</tr>
</tbody>
</table>

\textsuperscript{14} This is explainable because, as mentioned, both knowledgebases are based on a similar resource (SUMO). It is unlikely that a large difference will be produced in terms of source domains.

\textsuperscript{15} Although the questionnaires presented the source domains in Chinese, both English and Chinese versions of the source domains will be given for the discussion in this chapter.
In this way, we can see that among the four source domains shown in Table 11.10 for 成長 chéngzhāng ‘grow/growth,’ ORGANISM has the highest score and is, therefore, a more important source domain among all the other source domains. As for 改革 gǎiguī ‘reform/reformation,’ five source domains are shown in Table 11.10, and only ARTIFACT has a higher score, compared to the other four. The numbers in the last column simply indicate how often a source domain was found amongst all the senses of a particular type of metaphorical expression. Therefore, based on these rankings, we can predict that the ranking of source domains may correlate highly with the ranking of the subjects. However, the number of senses for each type of metaphorical expression in SinicaBow and the Merged Word List differs greatly (because the number of senses in the Merged Word List was mainly due to the number of translations possessed by each type of metaphorical expression). Examples of source domains for 成長 chéngzhāng ‘grow/growth’ and 改革 gǎiguī ‘reform/reformation’ taken from the Merged Word List are given in Table 11.11.

Table 11.11: Examples of Source Domains Obtained through the Chinese-English Merged Word List in the Top-down Approach

<table>
<thead>
<tr>
<th>Type of Metaphorical Expressions</th>
<th>Total Senses</th>
<th>Source Domains (English)</th>
<th>Source Domains (Chinese)</th>
<th>Frequency of Source Domains in the Definitions of All Senses</th>
<th>Ranking of Source Domains (Frequency/Senses)</th>
</tr>
</thead>
<tbody>
<tr>
<td>成長 chéngzhāng ‘grow/growth’</td>
<td>44 senses</td>
<td>CREATION 產生</td>
<td>1</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GAME 遊戲</td>
<td>2</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CONTENT DEVELOPMENT 內容發展</td>
<td>2</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LANDAREA 陸地</td>
<td>1</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>改革 gǎiguī ‘reform/reformation’</td>
<td>49 senses</td>
<td>FUNCTION 函數</td>
<td>2</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>INTENTIONAL PROCESS 意向性歷程</td>
<td>4</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PROCESS 歷程</td>
<td>17</td>
<td>0.35</td>
<td></td>
</tr>
</tbody>
</table>

Table 11.11 shows the number of senses for 成長 chéngzhāng ‘grow/growth’ and 改革 gǎiguī ‘reform/reformation’ to be more than forty each. Their rankings are provided in the final column.

Since the number of senses differs greatly between SinicaBow and the Merged Word List, these two datasets will be analyzed separately. Comparisons will be made between the rankings of source domains versus the possibility of ‘Yes’ answered by the total subjects (i.e., for a particular question, if seven out of nine subjects answered ‘Yes,’ the possibility of ‘Yes’ for this question will be 0.78). When comparing the
answer of each question as given by the non-CWN group in the Pearson correlation tests, the results show that significance was found only with the Merged Word List: \( r(124)=0.27, p<.05 \). Significance was not reached in SinicaBow: \( r(121)=-0.11, p=.227 \). This means that only in the Merged Word List, the higher the rankings are, the higher the possibility subjects would provide a ‘Yes’ as the answer for each pair of stimuli. This result is expected, as SinicaBow is a lexical knowledgebase whose senses have been checked by human analyzers. Repetition of senses or senses with close meanings is possibly removed in SinicaBow, whereas in the Merged Word List, close senses remain. Therefore, the more senses are owned by a type of metaphorical expression, the higher the possibility that several senses will be mapped to a similar SUMO node. When this happens, the possibility of repetition of keywords (i.e., the source domain terms) in the definitions will increase. As a result, we found that only the Merged Word List shows significant correlation with human ratings. Based on the fact that only the Merged Word List has shown itself able to predict the rankings of human ratings in terms of correctness of source domains, we found the Merged Word List to be more effective in source domain determination than SinicaBow.

In the previous sub-section, we did not find significance between SinicaBow and the Merged Word List. However, in this section, we found that, in terms of prediction, the Merged Word List provides better answers when predicting whether or not the highly occurring source domain terms in the SUMO definitions will be rated ‘Yes’ by more subjects. In other words, source domains with higher rankings in the Merged Word List indeed have been rated by many subjects to be a good source domain. On the contrary, source domains with higher rankings in SinicaBow cannot predict a similar answer. For example, for 改革 gǎigé ‘reform/reformation’ in Table 11.10, the source domain of ARTIFACT/人造物 is shown to have a higher ranking (0.44) than PROCESS/歷程 (0.33). However, in the answers by the non-CWN group, only 50 per cent (four subjects) answered ‘Yes’ for ARTIFACT/人造物 and 100 per cent (eight subjects) answered ‘Yes’ for PROCESS/歷程. This is one example where SinicaBow was not able to predict the ratings of the subjects. The Merged Word List, on the contrary, predicted this response because the ranking for PROCESS/歷程 was high initially (see Table 11.11). This phenomenon was discovered because the Merged Word List yielded an abundant number of senses for a particular type of metaphorical expression, and this increased the possibility of salient concrete keywords becoming more salient (when they are often repeated). However, SinicaBow has been manually checked, and therefore, the number of senses is well controlled. As a result, salient concrete keywords may not stand out immediately among the other concrete keywords. This is one possible reason why the ratings by the subjects correlate significantly with the rankings of source domains in the Merged Word List but not with those of SinicaBow.
A Corpus-driven Approach to Source Domain Determination

In the following section, we further examine the results in the bottom-up questionnaire.

### 11.3 Bottom-up: Saliency Effect in Source Domain Determination

As discussed previously, the design of the bottom-up questionnaire included CNA and XIN results as well as the role of saliency values in determining source domains. Within the CNA and XIN distinction, there are types of metaphorical expressions that fall under CNA-only, XIN-only and their overlapped types of metaphorical expressions. The following examples in (11a) show the types of metaphorical expressions that are found in CNA-only. Examples of XIN-only are shown in (11b), while their overlapped results of source domains are shown in (11c) and (11d), where (11c) is based on collocates in CNA, and (11d) is based on collocates in XIN. We can see from (11c) and (11d) that the results from CNA and XIN differ, even with the same type of metaphorical expressions.\(^{16}\)

In (11a), we see 弱弱 shuāiruò ‘weak/weakness,’ a type of metaphorical expression, found in CNA-only. Under its relation of ‘subject’ (there may be more than one relation for each), the source domain selected is 神经 shénjīng ‘nerve’ with the saliency value of 43.99.\(^{17}\) The second source domain is 心脏 xīnzàng ‘heart,’ followed by 能力 tǐlì ‘energy.’ The information regarding ‘Rank 1,’ ‘Rank 2’ and ‘Rank 3’ (the stimuli continued till ‘Rank 6’) will be analyzed in this section with respect to CNA-only, XIN-only and overlapped (combining both CNA-based and XIN-based) as well as their comparisons with the CWN and non-CWN groups.

(11) a. CNA-only

衰弱 shuāiruò ‘weak/weakness’
‘weakness’
Relation: ‘Subject’

| Rank 1 | 神经 shénjīng ‘nerve’ (43.99) |
| Rank 2 | 心脏 xīnzàng ‘heart’ (36.87) |
| Rank 3 | 能力 tǐlì ‘energy’ (33.41) |

b. XIN-only

觉悟 juéwù ‘aware of’
Relation: ‘Subject’

| Rank 1 | 思想 sīxiǎng ‘thought’ (45.93) |
| Rank 2 | 主义 zhǔyì ‘doctrine’ (13.49) |
| Rank 3 | 群众 qúnzhòng ‘collection’ (10.90) |

---

\(^{16}\) Since the results for the overlapped types of metaphorical expressions differ between CNA and XIN, they are examined separately.

\(^{17}\) Note that saliency values cannot be compared across different relations.
Chapter 11: Evaluation of the Correctness of Source Domains

c. CNA and XIN Overlapped (based on CNA results)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Rank 1</th>
<th>Rank 2</th>
<th>Rank 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>成熟 chéngshóu ‘ripe’</td>
<td>女人 nǚrén ‘woman’ (23.12)</td>
<td>魅力 mèilì ‘charm’ (12.83)</td>
<td>態度 tài dù ‘attitude’ (10.88)</td>
</tr>
<tr>
<td>Relation: ‘Modifies’</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Relation: ‘Modifies’

<table>
<thead>
<tr>
<th>Domain</th>
<th>Rank 1</th>
<th>Rank 2</th>
<th>Rank 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>成熟 chéngshóu ‘ripe’</td>
<td>小麥 xiǎomài ‘wheat’ (26.12)</td>
<td>技術 jìshù ‘skills’ (21.11)</td>
<td>稻子 dàozǐ ‘rice’ (11.68)</td>
</tr>
<tr>
<td>Relation: ‘Modifies’</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

d. CNA and XIN Overlapped (based on XIN results)

In the examples in (11), the most salient source domains determined through selecting the bottom nodes of the WordNet hierarchies that have the highest conceptual density are displayed (Chapter Nine). Their results of saliency ‘Rank 1’ to ‘Rank 6’ can be seen in Table 11.12.

Table 11.12: Mean Percentages of ‘Yes’ (Divided by the Number of Non-CWN Subjects=8) for All Stimuli with Source Domains with Saliency ‘Rank 1’ to ‘Rank 6,’ Respectively

<table>
<thead>
<tr>
<th>Saliency Ranking of Source Domains (E.g., Source Domains of 衰弱 shuāiruò ‘weak/weakness’)</th>
<th>Mean Percentages of ‘Yes’</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rank 1</strong> (E.g., 神經 shéngjīng ‘nerve’ (43.99))</td>
<td>.6184</td>
<td>.2096</td>
</tr>
<tr>
<td><strong>Rank 2</strong> (E.g., 心臟 xīnzàng ‘heart’ (36.87))</td>
<td>.5141</td>
<td>.1860</td>
</tr>
<tr>
<td><strong>Rank 3</strong> (E.g., 體力 tǐlì ‘energy’ (33.41))</td>
<td>.5325</td>
<td>.1910</td>
</tr>
<tr>
<td><strong>Rank 4</strong> (E.g., 精神 jǐngshén ‘spirit’ (22.64))</td>
<td>.5391</td>
<td>.1953</td>
</tr>
<tr>
<td><strong>Rank 5</strong> (E.g., 肌肉 jīròu ‘muscle’ (18.45))</td>
<td>.5197</td>
<td>.1883</td>
</tr>
<tr>
<td><strong>Rank 6</strong> (E.g., 腎臟 shènzàng ‘kidney’ (12.63))</td>
<td>.5341</td>
<td>.2214</td>
</tr>
</tbody>
</table>

In Table 11.12, the source domains of 衰弱 shuāiruò ‘weak/weakness,’ as those used in (12), are arranged in descending order from the saliency values at ‘Rank 1’ to ‘Rank 6.’

(12) 以 協助 改善 衰弱 的 調整 (CNA)
yì xiézhì gǎishìng shuāiruò de jìngzhěng
for assist make.better weak DE economy
‘In order to assist in making the weak economy better...’
Therefore, with regards to examples such as (12), the source domains for 'weak/weakness' are potentially 'nerve' (Rank 1), 'heart' (Rank 2), 'energy' (Rank 3), 'spirit' (Rank 4), 'muscle' (Rank 5) and 'kidney' (Rank 6). Table 11.13 shows the comparisons between ‘Rank 1’ to ‘Rank 6’ for all data using Kruskal-Wallis tests.

<table>
<thead>
<tr>
<th></th>
<th>Rank 1</th>
<th>Rank 2</th>
<th>Rank 3</th>
<th>Rank 4</th>
<th>Rank 5</th>
<th>Rank 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\chi^2 (1)=4.59,\  p&lt;.05$</td>
</tr>
<tr>
<td>Rank 2</td>
<td></td>
<td></td>
<td>$\chi^2 (1)=3.36,\  p=.067$</td>
<td></td>
<td></td>
<td>$\chi^2 (1)=2.56,\  p&lt;.10$</td>
</tr>
<tr>
<td>Rank 3</td>
<td></td>
<td>$\chi^2 (1)=1.185,\  p=.667$</td>
<td></td>
<td></td>
<td></td>
<td>$\chi^2 (1)=.127,\  p=.722$</td>
</tr>
<tr>
<td>Rank 4</td>
<td></td>
<td></td>
<td>$\chi^2 (1)=.083,\  p=.773$</td>
<td></td>
<td></td>
<td>$\chi^2 (1)=.002,\  p=.968$</td>
</tr>
<tr>
<td>Rank 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\chi^2 (1)=2.56,\  p&lt;.10$</td>
</tr>
<tr>
<td>Rank 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\chi^2 (1)=.127,\  p=.722$</td>
</tr>
</tbody>
</table>

As shown in Table 11.13, significance was found for the comparisons of ‘Rank 1’ versus ‘Rank 2’ as well as ‘Rank 1’ versus ‘Rank 5’ (in bold). This indicates that ‘Rank 1’ is indeed special, as it differs from two other ranks (while no other ranks differ from each other). This shows that ‘Rank 1’ indeed carries certain information compelling subjects to be more likely to provide a ‘Yes’ to ‘Rank 1’ than to other ranks. However, since ‘Rank 1’ does not differ from all ranks, the predictability of saliency in source domain determination is possible but not absolute (i.e., saliency values at the top rank do not necessarily carry source domain information).

We explained earlier in Chapter Nine that while some cases like ‘prosperity recovers’ and ‘the market of the real estate recovers’ and the collocates of ‘prosperity’ and ‘the market of the real estate’ may be highly salient, they may not be good source domains for ‘recover/recovery.’ When this happens, the more suitable source domains (most likely more concrete source domains, such as PERSON, ANIMAL or ORGANISM) probably appear at lower saliency rankings, and ‘Rank 5’ may prove to be the place where some concrete source domains appear (and thus is significant). In (13), examples of source domains in both ‘Rank 1’ and ‘Rank 5’ for ‘revenge’ and their percentages of ‘Yes’ as rated by the non-CWN group, are shown.
(13) Percentages of ‘Yes’ for ‘Rank 1’ and ‘Rank 5’ for 報復 bàofù ‘revenge’

<table>
<thead>
<tr>
<th>Rank</th>
<th>词</th>
<th>形容词</th>
<th>百分之比</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>打擊 dàjī ‘strike’</td>
<td>77.78%</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>威脅 wēixiè ‘threaten’</td>
<td>100.00%</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>武力 wǔlì ‘armed force’</td>
<td>100.00%</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>愛人 àirén ‘lover’</td>
<td>33.33%</td>
<td></td>
</tr>
</tbody>
</table>

From this example, we can see that there are source domains that are rated by all the subjects to be ‘Yes’ at ‘Rank 5’ (in bold), and these source domains at ‘Rank 5’ were still considered to be good source domains by the subjects. The unexpected higher percentages of ‘Yes’ in ‘Rank 5,’ however, should not be taken as a negative sign because from here we can see that, in terms of source domains, the highest saliency values are not necessarily the most concrete source domains.

In general, the previous results show that the saliency provided from Sketch Engine is a possible reference when determining the correct source domains, but it is not an absolute reference. This is because not all highly salient collocates are suitable source domains. 景氣復甦 jǐngqì fùsū ‘prosperity recovers’ and 房市復甦 fángshì fùsū ‘the market of the real estate recovers’ are two good examples to support this statement.

### 11.3.1 CNA versus XIN

When comparisons were made between types of metaphorical expressions from CNA and XIN by the non-CWN group, the results show that CNA has consistently higher means for the percentages of ‘Yes’ than XIN, and the differences are significant tested with Kruskal-Wallis tests: $\chi^2(1) = 9.38, p < .05$.

<table>
<thead>
<tr>
<th>CNA/XIN</th>
<th>Mean Percentages</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNA</td>
<td>.5862</td>
<td>.1906</td>
</tr>
<tr>
<td>XIN</td>
<td>.4997</td>
<td>.1997</td>
</tr>
<tr>
<td>Total</td>
<td>.5430</td>
<td>.1995</td>
</tr>
</tbody>
</table>

From Table 11.14, we can see that the answers for the non-CWN group for CNA contain a higher mean of the percentage for ‘Yes’ (59 per cent from the total CNA questions) than for XIN (50 per cent from the total XIN questions). The higher percentages in CNA may indicate that subjects may be more familiar with the metaphorical expressions used in Taiwan, as all subjects are speakers of Taiwan Mandarin.

---

18 CNA and XIN here include both overlapped and non-overlapped data.
11.3.2 Overlapped versus Non-overlapped in CNA versus XIN

When we compared the types of metaphorical expressions that appear in CNA-only or XIN-only (called ‘non-overlapped’ in Table 11.15), and those in both CNA and XIN (called ‘overlapped’), the results in Table 11.15 were found. Examples of overlapped types of metaphorical expressions are usually highly frequent metaphorical expressions that appear in both CNA and XIN (such as 成長 chéngzhǎng ‘grow/growth,’ 建設 jiànshè ‘construct/construction’ and 穩定 wěndìng ‘stable’).

Table 11.15: Means of the Percentages of ‘Yes’ for the Non-CWN Group in Overlapped versus Non-Overlapped

<table>
<thead>
<tr>
<th></th>
<th>Mean Percentages</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overlapped</td>
<td>.5580</td>
<td>.1907</td>
</tr>
<tr>
<td>Non-overlapped</td>
<td>.5279</td>
<td>.2077</td>
</tr>
<tr>
<td>Total</td>
<td>.5430</td>
<td>.1995</td>
</tr>
</tbody>
</table>

The results show the means for the percentages of ‘Yes’ are 56 per cent for overlapped and 53 per cent for non-overlapped. The results, however, are not significant: $\chi^2(1)=.988$, $p=.320$. This means that, whether or not a type of metaphorical expression appears in CNA-only or XIN-only, their results do not differ from those that appear in both CNA and XIN. Therefore, when referring to the previous examples in (11), which are repeated, differences are not found between (14a) through (14d).

(14)  a. CNA-only

衰弱 shuāiruò ‘weak’/ ‘weakness’
Relation: ‘Subject’

思緒 sīxiù ‘thought’ (43.99)

b. XIN-only

覺悟 juéwù ‘aware of’
Relation: ‘Subject’

c. CNA and XIN Overlapped (based on CNA results)

成熟 chéngshóu ‘ripe’
Relation: ‘Modifies’
d. CNA and XIN Overlapped (based on XIN results)

成熟 chēngshóu ‘ripe’
Relation: ‘Modifies’

<table>
<thead>
<tr>
<th>Rank</th>
<th>小麥 xiǎomài ‘wheat’ (26.12)</th>
<th>技術 jìshù ‘skills’ (21.11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank 3</td>
<td>稻子 dàozi ‘rice’ (11.68)</td>
<td></td>
</tr>
</tbody>
</table>

In sum, for the bottom-up approach, saliency values were found to be useful in determining source domains, although it was determined that higher salient collocates do not necessarily carry more source domain information. Other than that, significance was also found between CNA and XIN, with CNA containing a greater number of ‘Yes’s than XIN. However, significance was not found between whether or not a type of metaphorical expression appears in CNA-only, XIN-only or in both CNA and XIN.

11.4 Strengths and Limitations of the Experiment

The purpose of this experiment is to evaluate the source domains that are produced by the two approaches previously outlined. In doing so, we are also evaluating these two approaches by measuring the answers from the subjects. The results display marginal significance between the two approaches, with the bottom-up approach being slightly better than the top-down approach. The overall percentages of correctness for combined non-CWN lie between 40 and 52 per cent (shaded percentages in Table 11.8). The question, therefore, is whether these results are good enough for source domain determination. In this section, we will discuss some of the strengths of the experiment as well as its limitations.

11.4.1 Strengths: Selecting Source Domains That Are Salient

This experiment does more than just provide an evaluation of source domains. From the results, we obtain the best source domains selected for the following types of metaphorical expressions. For example, in Table 11.16, the source domains with the highest ratings are given for each type of metaphorical expression in the top-down stimuli.19

---

19 Two more types of metaphorical expressions (海洋 hǎiyáng ‘ocean’ and 情勢 qíngshì ‘situation’) are not in Table 11.16 because these two types of metaphorical expressions are affected by the stimuli problem stated earlier. Their source domains were reduced to less than three, and therefore, the highest scores are not reliable.
### Table 11.16: Source Domains with the Highest Ratings for Different Types of Metaphorical Expressions in the Top-down Stimuli

<table>
<thead>
<tr>
<th>Types of Metaphorical Expressions</th>
<th>Source Domains</th>
<th>Number of ‘Yes’s/ Total Number of ‘Subjects’ (Non-CWN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>上升 <em>shàngshēng</em> ‘rise’</td>
<td>ORIENTATION/相對方位</td>
<td>8/8</td>
</tr>
<tr>
<td>生態 <em>shēngtài</em> ‘ecology’</td>
<td>LIVING/有生命跡象的/有生命的</td>
<td>8/8</td>
</tr>
<tr>
<td>成長 <em>chéngzhǎng</em> ‘grow/growth’</td>
<td>PLANT/植物</td>
<td>8/8</td>
</tr>
<tr>
<td>改革 <em>gǎigé</em> ‘reform/reformation’</td>
<td>PROCESS/歷程</td>
<td>8/8</td>
</tr>
<tr>
<td>和諧 <em>hèxié</em> ‘harmony’</td>
<td>SOCIAL INTERACTION/人際互動</td>
<td>8/8</td>
</tr>
<tr>
<td>建設 <em>jiànshè</em> ‘construct/construction’</td>
<td>CONTENT DEVELOPMENT/內容發展</td>
<td>8/8</td>
</tr>
<tr>
<td>迫害 <em>pòhài</em> ‘persecute/persecution’</td>
<td>DESTRUCTION/毀滅</td>
<td>8/8</td>
</tr>
<tr>
<td>革新 <em>gēxin</em> ‘innovate/innovation’</td>
<td>ORGANIZATIONAL PROCESS/組織性歷程</td>
<td>8/8</td>
</tr>
<tr>
<td>報復 <em>bàofù</em> ‘revenge’</td>
<td>INTENTIONAL PROCESS/意向性歷程 SOCIAL INTERACTION/人際互動</td>
<td>8/8</td>
</tr>
<tr>
<td>實行 <em>shíxíng</em> ‘implement/implementation’</td>
<td>REALIZATION/實現</td>
<td>8/8</td>
</tr>
<tr>
<td>增長 <em>zēngzhǎng</em> ‘increase’</td>
<td>PROCESS/歷程</td>
<td>8/8</td>
</tr>
<tr>
<td>熱情 <em>rèqíng</em> ‘zeal’</td>
<td>HUMAN/人類</td>
<td>8/8</td>
</tr>
<tr>
<td>庇護 <em>bìhù</em> ‘protect/protection’</td>
<td>AGENT/施事者</td>
<td>7/8</td>
</tr>
<tr>
<td>走廊 <em>zǒuláng</em> ‘corridor’</td>
<td>GEOGRAPHIC AREA/區域 ARTIFACT/人造物</td>
<td>7/8</td>
</tr>
<tr>
<td>疲弱 <em>shuāiruò</em> ‘weak/weakness’</td>
<td>ORGANISM/生物體</td>
<td>7/8</td>
</tr>
<tr>
<td>堅定 <em>jiàndìng</em> ‘firm’</td>
<td>TRAIT ATTRIBUTE/人格特質</td>
<td>7/8</td>
</tr>
<tr>
<td>傷害 <em>shānghài</em> ‘wound’</td>
<td>INTENTIONAL PROCESS/意向性歷程 SOCIAL INTERACTION/人際互動</td>
<td>7/8</td>
</tr>
<tr>
<td>形勢 <em>xíngshì</em> ‘terrain’</td>
<td>ORIENTATION/相對方位</td>
<td>6/8</td>
</tr>
<tr>
<td>制度 <em>zhìdù</em> ‘system’</td>
<td>COLLECTION/群集 ORGANIZATION/組織</td>
<td>6/8</td>
</tr>
</tbody>
</table>
In the third column, the number of ‘Yes’s by the total non-CWN group is shown. We can see that those that have obtained 8/8 are cases where these source domains are rated ‘Yes’ by all of the subjects.

Therefore, the source domain with the highest ratings for 上升 shàngshēng ‘rise’ is ORIENTATION/相對方位; for 生態 shèngtài ‘ecology’ it is LIVING/有生命跡象 的/有生命的, for 成長 chéngzhǎng ‘grow/growth’ it is PLANT/植物, etc. Examples for 上升 shàngshēng ‘rise’ and 生態 shèngtài ‘ecology’ are shown in (15).

(15) a. 今年 以來, 各種 類型 經濟 均 呈 jīnnián yǐlái gèzhǒng lèixínghéng jùn chéng this.year since all.kinds type economy without.except show 上升 shàngshēng 趨勢 qūshì trend ‘Since this year, all sectors of the economy show rising trends without fail.’

b. 大陸 的 政治 生態 已經 出現 了 dàlù de zhèngzhì shèngtài yǐjīng chūxiàn le Mainland.China DE politics ecology already appear LE 一些 變化 (CNA) yīxī biànhuà some change ‘The ecology of China politics has already shown some changes.’
While each type of metaphorical expression is mapped to several source domains, and all of them are evaluated by the subjects, it is not surprising that the subjects would select some that they thought to be suitable source domains. For instance, 上升 shàngshēng ‘rise’ is mapped to the following source domains in (16), and they were all tested. The number of times they gave ‘Yes’ as an answer is given in brackets.

(16) Source Domains of 上升 shàngshēng ‘rise’ Tested in the Experiment:

- ORIENTATION/相對方位 (8/8)
- MOTION/移動 (7/8)
- QUANTITY CHANGE/量變 (6/8)
- BODY MOTION/身體移動 (5/8)
- PHYSICAL QUANTITY/物理量 (4/8)
- DEVICE/裝置 (4/8)
- ORGANIZATION/組織 (4/8)
- COOKING/烹飪 (1/8)

Looking at these source domains, we can see that they refer to the different aspects of 上升 shàngshēng ‘rise,’ some of which contain fewer numbers of ‘Yes’ responses. The source domains with fewer numbers of ‘Yes’ are usually source domains that are less salient. For example, COOKING/烹飪 may refer to the rising of steam while cooking, but this use is less salient (only one ‘Yes’ out of eight subjects) when compared to ORIENTATION/相對方位 (eight ‘Yes’s out of eight subjects). Therefore, from this example, we can see why the results of the experiment average only about 40 per cent rated ‘Yes’ by the subjects in the top-down stimuli (by the non-CWN only; see Table 11.8). This low percentage may reflect the fact that the results of the top-down approach provide all possibilities of source domains for a particular type of metaphorical expression, while subjects only selected the source domains that they thought to be the most salient.

With the bottom-up stimuli, we discover the most salient source domains, as rated by all subjects and shown in Table 11.17.20 In Table 11.17, we also provide the number of ‘Yes’ responses, shown in the third column.21

20 停滞 tingzhì ‘cease moving’ is not in Table 11.17 because it has only two source domains tested. Therefore, selecting the highest-rated source domain becomes unreliable.

21 All English translation comes from the WordNet translation of the Chinese source domains (see Chapter Ten for the explanation on the English translation).
## Table 11.17: Source Domains with the Highest Ratings for Different Types of Metaphorical Expressions in the Bottom-up Stimuli

<table>
<thead>
<tr>
<th>Types of Metaphorical Expressions</th>
<th>Source Domains</th>
<th>Number of ‘Yes’s/Total Number of ‘Subjects’</th>
</tr>
</thead>
<tbody>
<tr>
<td>生態 shèngtài ‘ecology’</td>
<td>SEA SWALLOW/燕鷗 VEGETATION/植物</td>
<td>8/8</td>
</tr>
<tr>
<td>成長 chéngzhǎng ‘grow/growth’</td>
<td>FRUITAGE/產量 BOYHOOD/少年期 KIDDY/小孩</td>
<td>8/8</td>
</tr>
<tr>
<td>形勢 xíngshì ‘terrain’</td>
<td>BEARING/關係 MEDIUM/環境 DIAGNOSTIC TEST/分析</td>
<td>8/8</td>
</tr>
<tr>
<td>制度 zhìdà ‘system’</td>
<td>LEGISLATIVE ACT/法令 SPECIES/形式 HUMAN RIGHT/人權</td>
<td>8/8</td>
</tr>
<tr>
<td>和諧 héxié ‘harmony’</td>
<td>SPIRIT/氣氛 TRIM/狀態</td>
<td>8/8</td>
</tr>
<tr>
<td>建設 jiànsè ‘construct/construction’</td>
<td>BLUEPRINT/藍圖 RESIDENCE/住宅 HIGHROAD/公路</td>
<td>8/8</td>
</tr>
<tr>
<td>迫害 pòhài ‘persecute/persecution’</td>
<td>PLOY/手段 OUTRAGE/暴行 DEVILTRY/惡行</td>
<td>8/8</td>
</tr>
<tr>
<td>革新 géxin ‘innovate/innovation’</td>
<td>PHILOSOPHY/觀念 EDUCATION/教育</td>
<td>8/8</td>
</tr>
<tr>
<td>秩序 zhìxù ‘order’</td>
<td>FORUM/議會</td>
<td>8/8</td>
</tr>
<tr>
<td>報復 bàofù ‘revenge’</td>
<td>MENACE/威脅 ARSENAL/武力</td>
<td>8/8</td>
</tr>
<tr>
<td>智慧 zhìhuì ‘wisdom’</td>
<td>INITIATE/學者</td>
<td>8/8</td>
</tr>
<tr>
<td>傷害 shānghài ‘wound’</td>
<td>CATTINESS/惡意 LAND MINE/地雷</td>
<td>8/8</td>
</tr>
<tr>
<td>覺悟 juéwù ‘conscious/consciousness’</td>
<td>CONSCIENTIOUSNESS/良心</td>
<td>8/8</td>
</tr>
<tr>
<td>摩固 gōnggu ‘solid/solidify’</td>
<td>HOLD/根據地</td>
<td>7/8</td>
</tr>
<tr>
<td>困難 kùnnán ‘difficult/difficulty’</td>
<td>EXTENT/程度</td>
<td>7/8</td>
</tr>
</tbody>
</table>
These source domains are found using collocation. In general, we found that subjects agree on more or less similar source domains in the bottom-up approach, as seen in Table 11.17, which demonstrates that sometimes more than one source domain is rated ‘Yes’ by all non-CWN subjects for a particular type of metaphorical expression. We also found the target domains (in bold) in some of the more highly rated source domains. These target domains were deliberately kept in the questionnaire because they are part of the automated results from the bottom-up approach. The fact that they were selected by many subjects again proves that the metaphorical meanings of certain types of metaphorical expressions are as salient as the literal meanings, which we intend to find. For example, for 情勢 qìngshì ‘situation,’ we found both literal (GAME PLAN/策略)
Chapter 11: Evaluation of the Correctness of Source Domains

and metaphorical meanings (POLITICS/政治 and FORWARDING/發展). These results show that the salient meanings are not always literal. In a few cases, too, we found that collocation becomes an inhibition when selecting the source domains. For example, the source domain of PERCENTAGE/百分比 for 占 zhàn ‘occupy’ (last row) is one case where the source domain for 占 zhàn ‘occupy’ is difficult to determine, as its most condensed concepts are heavily collocated. This is one limitation of the results of the bottom-up approach (which have been discussed in Chapter Ten), and it is clearly seen from the rating task in this chapter. Despite this limitation, Table 11.17 shows that most of the highly rated source domains selected by the subjects contain literal meanings. These source domains in both Tables 11.16 and 11.17 are useful because they have been agreed upon by human subjects as the source domains for these types of metaphorical expressions.

In some cases in Table 11.17, too, we see that the highly rated source domains are still abstract. ORDER/程序 (for 實行 shìxing ‘implement/implementation’) and FORWARDING/發展 (for 情勢 qíngshì ‘situation’ and 穩定 wěndìng ‘stable’) are two examples where the metaphorical uses are highly rated by subjects in being in the same concept domains with the types of metaphorical expressions. These metaphorical uses are probably frozen forms of metaphors. For cases like this, the concrete source domains are hard to ‘recovery’ for several reasons, one being that, as Grady (1999) suggests, primary conceptual metaphors, such as MORE IS UP and ORGANIZATION IS PHYSICAL STRUCTURE, which occur at higher conceptual levels than the other metaphors, exist. These primary metaphors, like the frozen forms, may not have source domains that can be clearly determined. Furthermore, 實行 shìxing ‘implement/implementation,’ 情勢 qíngshì ‘situation’ and 穩定 wěndìng ‘stable’ have probably lost their concrete meanings, and this is another reason concrete source domains are not found for these types of metaphorical expressions. Also, since the bottom-up approach is based on human usage, it is possible that when a concrete meaning is not often used, it is unlikely to appear in the collocated patterns. This may be the case for ‘foundation’ in English, whose source domain is difficult to determine when it is no longer used in the concrete way.

In addition, since there have been no previous studies that state clearly the criteria to define a primary metaphor, the ability of ‘recovery’ the concrete source domains can possibly be used as a criterion to define what a primary metaphor could be. In this way, the work of source domain determination will provide hints as to how types of metaphors can possibly be distinguished (such that those with easily found concrete meanings could form a type, while those with unfound concrete meanings constitute another type). Therefore, determining source domains using concrete meanings could be one way to provide statistical data regarding the proportion of source domains that
can be determined and those that cannot be determined. The issue of distinguishing types of metaphors will be saved for future work. In the following section, we discuss the weaknesses of the experiment in this chapter.

### 11.4.2 Weaknesses: Sampling and Subjects

The results from the experiment are useful in providing an evaluation of the source domains. However, the problem with the experiment is that it only tested a small sampling of source domains (although the questions are numerous). Since a type of metaphorical expression is repeated several times in the questionnaires, it is possible for the subjects to compare among the source domains. As a result, they show preference for certain source domains. The experiment also faced problems with one group of subjects (i.e., the CWN group) who were highly critical regarding the questions asked. These weaknesses can be improved upon by including more subjects, and these subjects will be tested on stimuli with fewer source domains repeated.

In addition, we can also test the correctness of source domains using other methods. For example, we can use generation tasks to ask subjects to generate as many expressions as possible for each source domain determined. After that, we can check the correlation between their answers and our answers. Similarly, we can also ask subjects to generate as many source domains as possible for different types of metaphorical expressions. Generation tasks as such are meaningful because they are evaluating results produced from computational methods. However, since both methods require a large number of stimuli and with as many subjects as possible, we therefore propose that it should be follow-up research.

### 11.5 Summary of Chapter

This chapter has provided an evaluation of the top-down and the bottom-up approaches based on human ratings of whether or not the source domains are correct. A small sampling of results was selected, and they were rated by two groups of subjects. The overall results showed that while the bottom-up approach is (slightly) better in source domain determination, it differs only marginally from the top-down approach. Differences within top-down and bottom-up approaches were also examined. For the top-down approach, the ratings for results from SinicaBow and the Merged Word List do not show significant differences. However, the Merged Word List is shown to correlate better with the correctness of source domains than SinicaBow does. This means that, in terms of selecting salient concrete concepts, the Merged Word List is better than SinicaBow, as the Merged Word List returns more senses than SinicaBow, i.e., the more senses that are returned, the higher the possibility that the salient concrete concepts will stand out easily.
For the bottom-up approach, the prediction of saliency values toward correct source domains is seen only between the top ‘Rank 1’ versus ‘Rank 2’ and ‘Rank 1’ versus ‘Rank 5.’ These results are inconsistent because the top saliency values could not be clearly distinguished from the other remaining ranking. This also means that, for source domain determination, saliency values from Sketch Engine might be possible references for the prediction of source domains, but they are not absolute references. In terms of CNA versus XIN, the ratings for CNA are better as compared with XIN, and this is probably because all subjects are Chinese speakers from Taiwan. In terms of overlapped types of metaphorical expressions in CNA and XIN versus non-overlapped ones, there is no significant effect found, indicating that no difference was found to occur between the types of metaphorical expressions used in Taiwan-only, China-only, and both Taiwan and China (although significance was found between data in Taiwan versus data in China).

In addition to the correctness of source domains, the results show better ratings for the bottom-up approach, probably because subjects were not as familiar with the source domains from the top-down approach (such as COGNITIVE AGENT and ORGANISM) as compared to those from the bottom-up approach (such as CAPACITY/能力 and JOB/工作). This shows the advantages of collocation, which is usage-based, where subjects can immediately answer whether a type of metaphorical expression falls under the same concept domain as these source domains (such as CAPACITY/能力 and JOB/工作). The better performance of the bottom-up approach also indicates that subjects tended to respond positively to source domains used at the bottom level. Source domains from the conceptual level (such as COGNITIVE AGENT) are not as familiar to the subjects, bringing to light one disadvantage of the top-down approach. Furthermore, having different instructions issued is also one possible factor that affects the results of the two approaches. However, since the nature of the two approaches is different, we hope to improve on the experiment by providing one single set of instructions for both approaches. Finally, this chapter has discussed the strengths and the limitations of the experiments. In addition, top-rated source domains were also provided.
Chapter 12: A Hierarchical Definition of Source Domains

Charteris-Black (2004:17) suggests a hierarchical cognitive model of metaphor (i.e., Critical Metaphor Analysis), which distinguishes amongst ‘conceptual metaphors,’ ‘conceptual keys’ and ‘metaphors.’ ‘Conceptual metaphors’ refers to forms such as AMERICA IS A PERSON. A ‘conceptual key’ is defined as “a higher level metaphor that explains how several conceptual metaphors are related.” For example, the conceptual key for this conceptual metaphor is A NATION IS A PERSON, and there are lower-level metaphors such as the United States of America is a friend (Charteris-Black 2004:16). The definition of metaphor as such explains certain phenomena of metaphors, but it does not explain the relationship between source domains. For example, the source domain is the same for both the conceptual metaphor and conceptual key (i.e., PERSON). Furthermore, there have been inconsistencies in the naming of ‘conceptual keys.’ In one example, shown in Figure 12.1, the conceptual key is SPORT IS A STRUGGLE FOR SURVIVAL while its generic conceptual metaphor is SPORT IS WAR (which includes FOOTBALL IS WAR, RUGBY IS WAR, CRICKET IS WAR, etc.) and the conceptual key is different this time from the source domain (WAR).

![Diagram](image)

Figure 12.1: Examples of Conceptual Key, Conceptual Metaphors and Metaphors by Charteris-Black (2004:129)
From Figure 12.1, the Critical Metaphor Analysis seems to provide hierarchical information regarding the target domains (from SPORTS to FOOTBALL, RUGBY, CRICKET, etc.). In terms of source domains, they remain the same for the different hierarchical levels (i.e., WAR). When we examine the definition of source domain in Charteris-Black (2004:15), it is considered to be “the base domains of literal and figurative concepts that are embodied in words,” and “that basic source domains are used to reflect in language how we experience more abstract target domains.” This definition does not provide precise information regarding what constitutes a source domain, nor does it shed light on why PERSON or SPORT is a good example of a source domain, leading us back to the questions of specificity and concreteness that were raised earlier in Chapter Two.

In this book, we propose that source domain by itself should be granted a hierarchical definition, i.e., each type of metaphorical expression can be mapped to different levels of specificity of source domains. For example, 成長 chéngzhǎng ‘grow/growth’ can be mapped to ORGANISM (top-down) and KIDDY/小孩 (bottom-up) at the same time. Usually, these two source domains form conceptual information in hierarchical structure, as demonstrated in Figure 12.2.

![Figure 12.2: The Hierarchical Structure of Source Domains](image)

Therefore, this definition of source domain reflects the directionality of both the top-down and the bottom-up approaches. For metaphor interpretation, Shen (1999) suggests a hybrid model that combines the schema paradigm and the categorical paradigm. According to Shen (1999), the schema paradigm views domains as ‘schemata’ while the categorical paradigm views domains as being taxonomic. The hybrid model is proposed based on subjects’ response to metaphors, such as ‘stems are drinking straws,’ and the characteristics of source domains, i.e., whether the features possessed by ‘drinking straws’ veer more toward the schema paradigm or the categorical paradigm. Shen’s results are based on psycholinguistic experiments, while our results come from computation of a
large amount of corpora data. We both arrive at a similar conclusion, namely that source domains are found to be both schematic and hierarchical in form.

In the following section, we will examine several ‘source domains’ that have been discussed in previous work, such as personification, journey metaphors, war metaphors, building metaphors, etc.¹ We will discuss the results with regards to how these ‘source domains’ were obtained in this book. The data used for the discussion of this chapter comes from the overall CNA or XIN (combined Single+Coordinated Target Domains), which is based on press clippings.

### 12.1 Personification

Personification in the top-down approach is grouped under the source domains of ORGANISM, COGNITIVE AGENT, AGENT, PATIENT, SENTIENT AGENT, etc. Under these source domains, uses such as 財政負擔 cāizhèng fùdān ‘the burden of finance,’ 經濟的恢復 jīngjì de huīfù ‘the recovery of economy’ and 經濟健全 jīngjì jiànquán ‘the sturdy economy’ are found. Examples of these uses are shown in (1).²

(1) a. 公保、農保、年年虧損，形成 gōngbǎo, nóngbǎo, niánnián kūsǔn xíngchéng central.insurance agricultural.insurance every.year lose form
中央財政沈重負擔 zhōngyáng cāizhèng chénzhòng fùdān central.government finance heavy burden
‘The central and agricultural insurance loses every year create a heavy burden on the finances of the central government.’

b. 國民經濟的恢復和發展出現了 guómín jīngjì de huīfù hàn fāzhǎn chūxiàn le national economy DE recovery and development appear LE
良好德勢頭 liánhǎo de shìtóu (XIN)
good DE terrain
‘The recovery and development of the national economy appear to be on good ground.’

¹ We use lower case here because these terms are not the same as the source domains defined in Chapters Seven through Ten.
² Since each type of metaphorical expression in (1) can be mapped to several source domains. The source domains are not listed.
c. 政府 將 以 不 損及 財政 健全 為
zhèngfǔ jiàng yǐ bù sǔnjī cáizhèng jiànhuán wéi
government will according to Neg. damage finance sturdiness for
原則 (CNA)
yuánzé
‘As a rule, the government will not cause damage to the sturdiness of (the)
finance.’

In the uses above, the target domains have been personified. For the bottom-up approach,
personification is seen in the source domains related to living things (YOUTH/年輕人,
GAL/女孩, HUMANNESS/人, OLDEST/老人, KIDDY/小孩, etc.). ³ Metaphorical
expressions for these source domains are 生命力 shēngmìnlì ‘life force,’ 交往 jiāowǎng ‘relationship,’ 氣度 qìdù ‘tolerance,’ 僵瘓 tānhuàn ‘paralytic’ and 成長 chéngzhǎng ‘grow/growth,’ respectively.

(2) a. POLITICS AND ECONOMY ARE YOUTHS/年輕人
中國 在 政治 和 經濟 上 都 有 了
zhōngguó zài zhèngzhì hàn jīngjì shàng dòu yǒu le
China at politics and economy above also have LE
新的 復力 (CNA)
xīn de huòlì
new DE vitality
‘China has gained new vitality in its politics and economy.’

b. EDUCATION IS A GAL/女孩
致力 生動 活潑、 附有 生命 的 教育 (CNA)
zhìlì shēngdòng huòpō fùyǒu shēngmìng de jiàoyù
devoted to vivid lively enclose life DE education
‘...devoted to education that is vivid, lively, and full of life.’

c. POLITICS IS HUMANNESS/人
現在 是否 已 到 採取 大膽 的 政治 氣度 (CNA)
xiànzài shìfǒu yǐ dào cǎiqì dàdàn de zhèngzhì qìdù
now whether already adopt bold DE politics tolerance
‘...whether it is time to adopt a bold political tolerance.’

³ However, a quantitative report of the different types of metaphorical expressions will not be
given in this chapter because the purpose is not to compare between percentages of types of
metaphorical expressions. Instead, a descriptive account of the source domains is aimed for.
d. ECONOMY IS AN OLDSTER/老人
以免苏联经济瘫痪(CNA)
yǐmǐàn súlíán jīngjì tānhuàn
to avoid Soviet Union economy paralytic
‘In order to avoid the economy of the Soviet Union becoming paralytic...’

e. ECONOMY IS A KIDDY/小孩
唯有社会安定, 经济才能成长 (CNA)
wéiyǒu shèhuì āndìng jīngjì cái néng chéngzhǎng
only if society stable economy then can grow
‘Only when society is stable will the economy grow.’

Naturally, the personification of the target domains will include PLANT, shown in (3).

(3) a. 不应破坏目前的政治生态 (CNA; Top-down)
bù yìng pòhuài mùqián de zhèngzhì shēntài
Neg. should destroy present DE politics ecology
‘(One) should not destroy the present political ecology.’

b. 作为政治花瓶的「人大」
zuòwéi zhèngzhì huàpíng de réndà
becoming politics vase DE the. national. people. congress
無論在选举方式、代表
wúlùn zài xuǎnjù fǎngshì dàibiǎo
regardless of at election method representative
組成分子方面, 都必须
zǔchéng fènzǐ fāngmiàn dōu bīxū
compose of member aspect all must
具有形式上的代表性 (CNA; Bottom-up)
jùyǒu xíngshì shàng de dàibiǎoxìng
possess form above DE representativeness
‘For a political vase (decoration) like the National People Congress, it must possess representativeness in all forms, regardless of whether in the method of election or in the composition of its members.’

We can therefore see that in the hierarchical definition of source domains, personification is found in source domains that are both general and specific. Usually, source domains obtained from the top-down approach are more general than those acquired using the
bottom-up approach. In some cases (like PLANT), we can see that both approaches arrive at a similar level of specificity.

### 12.2 Journey Metaphors

The exact source domain of JOURNEY is not found in the top-down approach. In the top-down approach, the source domains for types of metaphorical expressions such as 軌道 guǐdào ‘track,’ 路 lù ‘road’ and 十字路囗 shìzǐlùkǒu ‘crossroads’ are TRANSPORTATION, TRANSPORTATION DEVICE and LOCATION as well as ARTIFACT. These source domains emphasize different aspects of the journey. TRANSPORTATION and TRANSPORTATION DEVICE refer to specific examples of journey, while LOCATION refers to a larger concept that includes all locations. Examples of types of metaphorical expressions under TRANSPORTATION DEVICE are shown in (4), while examples of LOCATION are shown in (5).

(4) 起飛 qǐfēi ‘take off’
    軌道 guǐdào ‘track’
    出發 chūfā ‘depart’
    轉向 zhuănxiàng ‘change direction’
    操縱 cāozòng ‘manipulate’
    前進 qiánjìn ‘go forward’
    航空 hàngkōng ‘aviation’

(5) 基礎 jīchǔ ‘basis’
    前景 qiánjǐng ‘prospect’
    困境 kùnjìng ‘predicament’
    舞台 wǔtái ‘stage’
    途徑 tújìng ‘path’

We can see that in (5), the non-shaded types of metaphorical expressions do not refer to the journey. This means that LOCATION is a general source domain that can refer to all kinds of locations.

As for the bottom-up approach, source domains that are related to journey are specific examples, such as HIGHROAD/公路, SAILBOAT/帆船 and PASSENGER/乘客. Examples of types of metaphorical expressions for these source domains are shown in (6).

---

4 The English translation for 乘客 chéngkè ‘passenger’ was originally FARE, but it has been changed to PASSENGER in this book.
(6) a. POLITICS AND ECONOMY ARE HIGHROADS/

必然会 影響 到 台灣 的 政治

bound.to will affect reach Taiwan DE politics

與 經濟 遠景 (CNA)

and economy distant.prospects

'(It is) bound to affect the distant prospects of Taiwan politics and economy.'

b. ECONOMY IS SAILBOAT/

經濟 的 再 出發 將 指日可待 (CNA)

economy DE again depart will to.be.just.round.the.corner

'The re-departure of the economy will be just round the corner.'

c. ECONOMY IS PASSENGER/

中共 現行 的 經濟 路線

Republic.of China in.operation DE economy route

是 屬於 保守派 而 非

be belong conservative but Neg.

改革派 的 (CNA)

reformative DE

'The current route of the economy that is in operation belongs to the conservative and not reformative.'

We once again see that source domains for the top-down approach are more general, while those of the bottom-up approach are usually specific examples of source domains. Similar types of metaphorical expressions can be found in both a general and a specific source domain (such as 出發 chūfā ‘depart’ in both TRANSPORTATION DEVICE and SAILBOAT, where the latter is an instance of the former).

12.3 War Metaphors

According to Koller (2004), the metaphors of WAR, SPORTS and GAMES are most often found in “media discourse on marketing and sales.” Since our data consists
of newspaper reports with varied target domains (not only restricted to marketing and sales), war metaphors are found frequently along with other source domains, which also regularly occur. Comparatively, SPORTS and GAMES are not as numerous because some lexical items under these source domains in Koller’s work (2004:67-69), such as ‘to jump’ and ‘to kick’ for SPORTS and ‘die’ and ‘draw’ for GAMES, have ambiguous meanings. In fact, many lexical items overlapped in these three source domains, which confirms what has been argued in Ritchie (2003), that CHESS and GAME cannot be distinguished clearly.

In our study, war metaphors are found in both the top-down and the bottom-up approaches. In the top-down approach, we found war metaphors under the source domain of CONTEST with the following types of metaphorical expressions:

(7) 手段 shǒuduǎn ‘trick’
    衝擊 chōngjī ‘assault’
    打擊 dǎjī ‘strike’
    策略 cèluò ‘tactics’
    封鎖 fēngsuǒ ‘block/blockage’
    成功 chénggōng ‘success’

Other than CONTEST, there is also INJURY with types of metaphorical expressions shown in (8).

(8) 衝擊 chōngjī ‘assault’
    傷害 shānghài ‘injure/injury’
    損害 sūnhài ‘hurt’
    承受 chéngshòu ‘bear’
    震撼 zhènhàn ‘vibrate/vibration’
    扭轉 niǔzhuǎn ‘twist’
    創傷 chuàngshāng ‘trauma’
    拉扯 lāchě ‘pull and drag’
    損傷 sūnshāng ‘lesion’

However, we do not find many other source domains that are related to war. Most of them (such as 大屠殺 dàtúshā ‘massacre’ and 侵略者 qīnlüèzhě ‘intruder’) have been grouped under more general source domains, such as ORGANISM, MAN and HOMO).

As for the bottom-up approach, TROOP/軍隊, WAR/戰爭, MILITARY PERSON/部隊, BATTLESHIP/戰艦, etc. are found. Examples of these source domains are given in (9).
Chapter 12: A Hierarchical Definition of Source Domains

(9) a. FOREIGN AFFAIRS AND ECONOMY ARE TROOP/軍隊
在「中國」不斷對台灣進行外交
zài zhōngguó bùdúàn duì táiwān jìnxing wàijiāo
at China continuous face Taiwan carry.out foreign.affairs
及經濟侵略下 (CNA)
jí jìngjì qīnlüè xià
and economy invade under
‘Under the continuous attack of China toward the foreign affairs and economy of Taiwan...’

b. POLITICS IS WAR/戰爭
阿富汗內戰，今天成了部落、種族與
āfūhàn nèizhàn jīntiān chéng le bālùò zhōngzú yǔ
Afghanistan civil.war today become LE tribe race and
政治上相互敵對的複雜混合 (CNA)
zhèngzhì shàng xiānghù dídù de fùzá húnhé
politics above mutual antagonism DE complicated mixture
‘Today, the civil war of Afghanistan has become a complicated mixture of mutual tribal, racial and political antagonism.’

c. EDUCATION IS MILITARY PERSON/部隊
革命傳統教育的基地 (XIN)
gémìng chuántōng jiàoyù de jìdī
revolt traditional education DE base
‘...a base to revolt against traditional education.’

d. BATTLESHIP/戰艦
台商的事業成功發展
tǎishāng de shìyè chénggōng fāzhǎn
Taiwanese.businessmen DE career success development
就是支援我國外交的主力 (CNA)
jiǔshì zhīyuán wǒguó wàijiāo de zhǔlì
be aid our.country foreign.affairs DE anchor.man
‘The success of the career development of the Taiwanese businessmen is the main force (anchor man) to aid the foreign affairs of our country.’

With the bottom-up approach, we can find more types of source domains because the source domains in the bottom-up approach are specific, and they can refer to all aspects of war (such as the different people in WAR, not just COGNITIVE AGENT, as defined
in the top-down approach). Comparatively, the top-down approach usually groups the war metaphors under CONTEST.

12.4 Building Metaphors

As for building metaphors, we find source domains such as BUILDING, PRODUCT and MAKING from the top-down approach. BUILDING contains one type of metaphorical expression only (i.e., 走廊 zǒuláng ‘corridor’). Examples of types of metaphorical expressions for PRODUCT are given in (10) while those for MAKING are given in (11). Both (10) and (11) contain other types of metaphorical expressions that do not necessarily relate to the building metaphors (shaded).

(10)

結 構 jiégòu ‘structure’
加強 jiàqiáng ‘strengthen’
舞台 wùtái ‘stage’
支柱 zhīzhù ‘pillar’
建立 jiànlì ‘build.up’
背景 bēijǐng ‘background’
網路 wǎnglù ‘World Wide Web’
籌碼 chóumǎ ‘chips’
架構 jiàgòu ‘carcass’

(11)

結 構 jiégòu ‘structure’
舞台 wùtái ‘stage’
支柱 zhīzhù ‘pillar’
背景 bēijǐng ‘background’
網路 wǎnglù ‘World Wide Web’
命脈 mìngmài ‘lifeblood’
籌碼 chóumǎ ‘chips’
架構 jiàgòu ‘carcass’

As the top-down approach provides source domains of a larger scope, we can find a wide variety of results for PRODUCT and MAKING under (10) and (11). The bottom-up approach, on the other hand, is more specific. The source domains that are related to the building metaphors are ARCHITECTURE/建築物, HOUSE/房屋, LODGE/小屋 and RESIDENCE/住宅.5 Examples of uses for these source domains can be seen in (12).

5 See Dobrovolskij & Piirainen (2005:187-203, Chapter Nine) for the discussion of the source domain of HOUSE from the point of view of several cultures. See also Grady (1997) for the discussion of THEORIES ARE BUILDINGS.
Chapter 12: A Hierarchical Definition of Source Domains

(12) a. ECONOMY AND FOREIGN AFFAIRS ARE ARCHITECTURE/建築物
ECONOMY AND FOREIGN AFFAIRS ARE HOUSE/房屋

對中共而言是經濟和外交
dù zhōnggōng ér yán shì jīngjì hàn wàijiāo
face Republic.of.China to speak be economy and foreign-affairs

對中共而言是經濟和外交
dù zhōnggōng ér yán shì jīngjì hàn wàijiāo
face Republic.of.China to speak be economy and foreign-affairs

對中共而言是經濟和外交
dù zhōnggōng ér yán shì jīngjì hàn wàijiāo
face Republic.of.China to speak be economy and foreign-affairs

對中共而言是經濟和外交
dù zhōnggōng ér yán shì jīngjì hàn wàijiāo
face Republic.of.China to speak be economy and foreign-affairs

For the Republic of China, Japan is the main support (pillar) of the economy and foreign affairs.

b. FOREIGN AFFAIRS AND ECONOMY ARE LODGE/小屋

經濟空間 (CNA)
jīngjì kōngjiān

經濟空間 (CNA)
jīngjì kōngjiān

經濟空間 (CNA)
jīngjì kōngjiān

經濟空間 (CNA)
jīngjì kōngjiān

经济空间 (CNA)
jīngjì kōngjiān

...to suppress us in terms of international participation in foreign affairs and the economy.

c. POLITICS IS RESIDENCE/住宅

使政治建設更進步 (CNA)
shǐ zhèngzhì jiànshè gèng jǐnbù

使政治建設更進步 (CNA)
shǐ zhèngzhì jiànshè gèng jǐnbù

使政治建設更進步 (CNA)
shǐ zhèngzhì jiànshè gèng jǐnbù

使政治建設更進步 (CNA)
shǐ zhèngzhì jiànshè gèng jǐnbù

使政治建設更進步 (CNA)
shǐ zhèngzhì jiànshè gèng jǐnbù

...to cause the construction of politics to be better.

Using the bottom-up approach, we obtain specific examples of buildings with different types of houses being found. When compared to PRODUCT or MAKING, the source domains from the bottom-up approach are more familiar in usage because they come from the collocated patterns in corpora. In most of the uses in (12), we found that building metaphors usually refer to the structure or space of building.

12.5 A Hierarchical Definition of Source Domains

All previous results support one main idea—source domains should be seen in a hierarchical form. As illustrated in Figure 12.3, the bottom of the figure shows the types of metaphorical expressions we use in our daily language. When we use a bottom-up approach, we obtain a specific source domain for these types of metaphorical expressions.
However, when we go higher up into the hierarchy, we will obtain a general source domain. The situation is vice versa when we move down the hierarchy, from the more general source domains to the more specific source domains. Therefore, the hierarchical definition of source domains postulates that no single source domain is absolute for a type of metaphorical expression. When a specific source domain is obtained, this only means that we are located at a particular place in a hierarchy. If we move up or down, we will encounter different source domains with different levels of specificity.

Figure 12.3: A Hierarchical Definition of Source Domains

Therefore, the idea behind this definition of source domain is the notion of inclusion and membership, which can be traced back to the Wittgenstein’s (1978) family resemblance, whereby the source domains with resembled features can be grouped together. This model of source domain is also one that follows the notion of ontology, where specific concepts are descendents of more general concepts.

The advantage of this model is that it provides links between source domains in the form of networks, which not only can be applied to the study of conceptual metaphors but can also be used to establish domain networks for the purpose of computational linguistics, psycholinguistics and the Semantic Web. Specifically, studies in automatic metaphor identification (stated in Chapter Three) make use of knowledge hierarchy (Martin 1990, Wang 2006) or selectional restriction (or collocation) (Fass & Wilks 1983, Mason 2004). These studies first ensure that an expression is not literal before they decide it is a metaphorical use. While such a method is possible, it requires one decision to lead to another. Therefore, by providing pairings of types of metaphorical expressions and source domains (such as 起飛 qǐfēi ‘take off’: AIRLINER/飛機 and 成長 chéngzhǎng ‘grow/growth’: KIDDY/小孩), our results can assist in the speedier
identification of metaphors. For example, all uses that match these pairings can be identified as metaphors automatically, without the need to first prove that these uses are not literal. The traditional approach to metaphor identification can be seen in Figure 12.4.

The traditional approach takes both intuitive examples and expressions from corpora. For example, 經濟 起飛 jìngjì qǐfēi ‘economy takes off,’ 蝴蝶 起飛 húdié qǐfēi ‘butterflies take off,’ 經濟 孤兒 jìngjì gūér ‘the orphan of economy’ and 小組 開會 xiǎozǔ kāihuì ‘sub-group’s meeting’ are expressions that can be found in corpora as well as intuitively. In these examples, 開會 kāihuì ‘meeting’ is different from 起飛 qǐfēi ‘take off’ and 孤兒 gūér ‘orphan’ because 開會 kāihuì ‘meeting’ can only be used literally. In Figure 12.4, we will see how these different expressions will be treated in the traditional approach.

6 使 古巴 成為 幾乎 陷於 未 工業化 狀態 的 shǐ gǔbā chéngwéi jīhū xiànyǔ wèi gōngyèhuà zhùàngtài de make Cuba become almost sink.into not.yet industrialized state DE 經濟 jìngjì 孤兒 gūér (CNA) economy orphan ‘...to cause Cuba to become an ‘orphan of the economy’ who has sunked into a situation that is not yet industrialized.’
A Corpus-driven Approach to Source Domain Determination

**A Single Source Domain**

起飛 qǐfēi ‘take off’: AIRPLANE
(e.g., 經濟-起飛 economy-take off)

- Conventional Metaphor
- Non-Conventional Metaphor

<table>
<thead>
<tr>
<th>Non-Metaphors</th>
</tr>
</thead>
<tbody>
<tr>
<td>蝴蝶-起飛 butterfly-take off</td>
</tr>
<tr>
<td>小組-開會 sub-group-meeting</td>
</tr>
</tbody>
</table>

Check Whether Source Domain of Keyword Is in Database (i.e. 孤兒)

- Yes
- No

Check for Literal Sense in Dictionary or Introspection

Examine Uses of Expressions with the Keywords (in bold)
(經濟-起飛 economy-take off; 蝴蝶-起飛 butterfly-take off; 經濟-孤兒 economy-orphan; 小組-開會 sub-group-meeting)

Intuitive Examples/Expressions from Corpora

- 經濟-起飛 jīngjì qǐfēi ‘economy takes off’
- 蝴蝶-起飛 húdié qǐfēi ‘butterflies take off’
- 經濟-孤兒 jīngjì gūér ‘the orphan of economy’
- 小組-開會 xiǎozǔ kāihuì ‘sub-group’s meeting’

Figure 12.4: Traditional Approach to Metaphor Identification
After the expressions are found, the next step is to examine the literalness of these expressions. For this purpose, the traditional approach usually examines the uses of the keywords (i.e., 起飛 qīfēi ‘take off,’ 孤兒 gūér ‘orphan’ and 開會 kāihuì ‘meeting’) with its other words (such as 經濟 jīngjì ‘economy’ and 蝴蝶 húdié ‘butterfly’ for 起飛 qīfēi ‘take off’; 經濟 jīngjì ‘economy’ for 孤兒 gūér ‘orphan’; and 小組 xiǎozǔ ‘sub-group’ for 開會 kāihuì ‘meeting’). These uses will be checked to see whether the literal sense is listed in the dictionary or if it is available after introspective analysis. If a literal sense is found, a non-metaphor is assumed. Therefore, we can see in Figure 12.4 that both 蝴蝶 起飛 húdié qīfēi ‘butterflies take off,’ and 小組 開會 xiǎozǔ kāihuì ‘sub-group’s meeting’ will be categorized as non-metaphors. If a literal sense is not found, the traditional approach will then check whether the source domains of the metaphor keywords (i.e., 起飛 qīfēi ‘take off’ and 孤兒 gūér ‘orphan’) are in their database. The database usually comprises examples collected from others’ studies, such as that from the Conceptual Metaphor Homepage (http://cogsci.berkeley.edu/lakoff/).

If the keyword is found in the database, it will be considered as a conventional use, and its source domain will be shown (such as 起飛 qīfēi ‘take off’; AIRPLANE for 經濟 jīngjì qīfēi ‘economy takes off’). The source domain displayed by the traditional approach is usually comprised of one single source domain, i.e., one source domain for an expression. In most cases, a metaphor-identification system usually collects source domain information from one resource only. For instance, in Martin’s (1990:69) example, a single source domain is shown for the expression of ‘Enter-Lisp,’ shown in Figure 12.5.

Applying conventional metaphor Enter-Lisp.

\[\text{Enter-Lisp} \ (\uparrow \text{Container-Metaphor})\]
\[
\begin{align*}
\text{(enter-lisp-res enter-res} & \ \rightarrow \ \text{lisp-invoke-result)} \\
\text{(lisp-enterer enterer} & \ \rightarrow \ \text{lisp-invoke)} \\
\text{(entered-lisp entered} & \ \rightarrow \ \text{lisp-invoked)} \\
\text{(enter-lisp-map Entering} & \ \rightarrow \ \text{Invoke-Lisp})
\end{align*}
\]

(Taken from Chapter Three; Figure 3.1; Page 19)

**Figure 12.5:** Finding Related Metaphor for ‘Enter-Lisp’ (Martin 1990:69)

In Figure 12.5, the source domain of CONTAINER (circled) is found for the use of ‘enter-lisp.’ It is unusual for more than one source domain of various abstractions to be listed.

On the other hand, if an expression is not found in the database, some systems may consider it as a non-conventional expression, i.e., when an expression is not stored in the database, it is considered as a rare occurrence of metaphor. Therefore, for rare
occurrences such as 經濟 孤兒 jīngjì gūér ‘the orphan of economy,’ it will be considered as a non-conventional use). It is unlikely that the source domains for non-conventional metaphors will be collected in the database because these non-conventional uses are less salient conceptually, and therefore, they are unlikely to be stored if the database is based on intuitive source domains.

The corpus-driven approach suggested in this book is different from the traditional approach. The research in this book involves a huge database comprising more than ten thousand instances of analyzed corpora. Considering the amount of data analyzed in this study, we consider this study to be a corpus-driven study, where a corpus is “seen as more than a repository of examples to back pre-existing theories or a probabilistic extension to an already well defined system” (Tognini-Bonelli 2001:84). A corpus-driven study is suggested to possess the following criteria:

- **Examples** are normally taken verbatim, in other words they are not adjusted in any way to fit the pre-defined categories of the analyst; **recurrent patterns and frequency distributions** are expected to form the basic evidence for linguistic categories; the absence of a pattern is considered potentially meaningful. (Tognini-Bonelli 2001:84; bolds added).

In view of these criteria, this book comprises all the characteristics mentioned in the previous excerpt. First, examples of metaphorical expressions from the Chinese Gigaword corpus are not adjusted to fit any “pre-defined categories” of source domains. Second, “recurrent patterns and frequency distributions” are emphasized in every corpora analysis in this book. Third, the absence of a pattern in a dataset, if consistent and recurrent, as compared to another dataset (e.g., source domains found in the Single Target Domains but not in the Coordinated Target Domains), will be discussed as “meaningful” linguistic phenomena.

Steps for the corpus-driven approach in metaphor identification can be found in Figure 12.6.

---

7 A dotted line is used for the non-conventional metaphor in Figure 12.4 because this depends on how representative the database is. If it comprises a large number of metaphor-source domain pairings, it will be powerful enough to claim that whatever is not found is non-conventional. If it does not collect enough data, the system will display only the found results. 經濟 孤兒 jīngjì gūér ‘the orphan of economy’ is an example that is found using the corpora analysis employed in Chapter Six. It has a frequency of four in the whole Gigaword corpus. For the traditional approach, it is unlikely that the single occurrence as such will be collected in the database.

8 Tognini-Bonelli (2001:65) distinguishes between two kinds of studies, namely a corpus-based study and a corpus-driven study. A corpus-based study “refers to a methodology that avails itself of the corpus mainly to expound, test or exemplify theories and descriptions that were formulated before large corpora became available to inform language study.”
Multiple Source Domains
起飛 qīfēi ‘take off’: AIRPLANE
起飛 qīfēi ‘take off’: ARTIFACT
(經濟 jīngjì qīfēi ‘economy takes off’)

Non-Conventional Metaphor

Check Frequency of Occurrences in Corpora

Check Whether Source Domain of Keyword Is Found in Database

Examine Keywords
(i.e., 起飛 qīfēi ‘take off’;
孤兒 gūér ‘orphan’;
開會 kāihuì ‘meeting’)

Intuitive Examples/Expressions from Corpora
經濟 jīngjì qīfēi ‘economy takes off’
蝴蝶 húdié qīfēi ‘butterflies take off’
經濟 jīngjì gūér ‘the orphan of economy’
小組 xiǎozǔ kāihuì ‘sub-group’s meeting’

Figure 12.6: Corpus-driven Approach to Metaphor Identification
First, the corpus-driven approach takes examples from corpora only. Considering the same examples used in the traditional approach, the corpus-driven approach examines the keywords (in bold) from the start. These keywords are 起飛 qǐfēi ‘take off,’ 孤兒 gūér ‘orphan’ and 開會 kāihuì ‘meeting.’ For these keywords, they will be checked in the database to see if their source domains are found in the database. If these keywords and their source domains are not found in the database, they will be considered as non-metaphors. This can be done if the database is built from analyzing a large amount of corpora data, as is done in this book.

On the other hand, if these expressions are found, they will be considered as metaphors, and they will then be checked in terms of their frequencies in the corpora. If the frequency of a metaphor exceeds X number (which is the threshold value), this metaphor will be considered as a conventional metaphor. On the contrary, if the frequency of a metaphor is lower than X number, it will be considered as a non-conventional metaphor. The delineation of X number can be carried out through computing cut-off points, as demonstrated in Chapter Nine, where a threshold value can be calculated based on linguistic lists. For both conventional and non-conventional metaphors, if they are found in the database, their source domains will be shown. For example, in Figure 12.6, the source domains for 起飛 qǐfēi ‘take off’ are AIRPLANE and ARTIFACT, while the source domains for 孤兒 gūér ‘orphan’ are DESCENDANT and COGNITIVE AGENT. As explained previously, the corpus-driven approach allows a hierarchical definition of source domains. Therefore, this approach will display source domains that are conceptually linked.

The corpus-driven approach differs from the traditional approach because, first, it does not need to prove that a literal sense exists as is done in the traditional approach. We can see that the traditional approach requires this additional step, which is usually a complicated process. The corpus-driven approach does not need this step because our database is huge, as it contains a large number of metaphor-source domain pairings that are collected from a large corpora, i.e., the source domains stored in the database of the corpus-driven approach are determined automatically. As for the traditional approach, the source domains are usually manually determined or they are taken from someone else’s database. Second, since the traditional approach usually collects examples from others’ metaphor research (which usually comes from one resource), the source domains in their database are subjective to the resource. Unlike the traditional approach, the corpus-driven approach determines source domains automatically. Source domains, therefore, are not the results of human decision-making. Furthermore, the corpus-driven approach will have a more systematic organization of source domains than the traditional approach because our source domains are conceptually linked. Thus, they can show how different domains overlap (see Figure 12.7 for more discussion), which also provides
a clearer definition of source domain than the traditional approach does. Third, the
corpus-driven approach is also different from the traditional approach because we are
able to show the frequencies of the metaphors in corpora. These frequencies will then
be used to determine whether a conventional metaphor is assumed.

Through this type of comparison, we can see how the source domains determined
in this work come in handy for automatic metaphor identification. In the corpus-driven
approach, the source domain information stored in the database can be used to
automatically identify metaphors quickly. This saves one step (i.e., checking for literal
sense), as used in the traditional approach. All expressions that are found in the
database are identified as metaphors; otherwise, they are grouped under ‘non-metaphor’
in the corpus-driven approach. This process is made possible because the model is
driven by a large corpus, and that the source domains stored in the database have to be
abundant in order to identify metaphors more easily.

Furthermore, since we are able to provide information regarding hierarchies of
source domains, this will help to link the identified metaphors to a conceptual hierarchy,
which is useful to studies of semantic networks. The results from this book should,
therefore, prove to be very useful, not only for explaining linguistic phenomena, but
also for studies in other fields.

This model also explains areas pertaining to whether or not boundaries of source
domains can be established clearly. This is a theoretical question that can be answered
based on the hierarchical definitions of source domains. For example, we may find that
similar metaphorical expressions may appear in several source domains. From the result
of this book, for example, we found that 途徑 tūjīng ‘path’ has the following source
domains:

(13) Source Domains for 途徑 tūjīng ‘path’ (SinicaBow)

<table>
<thead>
<tr>
<th>TIME</th>
<th>STATIONARY ARTIFACT</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEQUENCE</td>
<td>PURPOSE</td>
</tr>
<tr>
<td>LOCATION</td>
<td>FUTURE</td>
</tr>
<tr>
<td>ARTIFACT</td>
<td></td>
</tr>
</tbody>
</table>

These source domains emphasize different aspects of 途徑 tūjīng ‘path,’ and therefore,
different conceptual hierarchies may be found for similar types of metaphorical
expressions. By examining the overlaps between domains, we can provide a better
understanding of hierarchical links between concepts. This is one important implication
of this model, as the book has found that one type of metaphorical expression can be grouped under one or more source domain. Therefore, there are possible overlaps between domains that we can explain with the hierarchical model, i.e., we may find hierarchical links through their shared domain information, shown in Figure 12.7.

Figure 12.7: Hierarchical Links of Concept Domains

In Figure 12.7, we can see that types of metaphorical expressions such as 患 huàn ‘to suffer from,’ 健康 jiànkāng ‘health/healthy,’ 麻癱 tānhuàn ‘paralytic’ and 抑鬱 yìyù ‘depressed’ can be linked to one another because they share certain specific source domains. All of these expressions are found under the specific source domain of OLD MAN, as well as the general source domain of ORGANISM, while only 抑鬱 yìyù ‘depressed’ is found under a different specific source domain of TEMPER, though it is still linked to ORGANISM. As mentioned, each type of metaphorical expression
can be mapped to several source domains. Therefore, the links between these domains can be enlarged, and they are linked in a complicated network.

Therefore, we can see that our model allows for network-like interpretation of domains. Relations between these domains will be suggested for future work.

12.6 Summary of Chapter

This chapter presented several source domains that are commonly discussed in conceptual metaphor research. These source domains are related to personification, journey metaphors, war metaphors and building metaphors. We have provided examples from the top-down and bottom-up approaches, and based on the results we propose a hierarchical definition of source domains. The discussion in this chapter is important to the area of linguistic research, as it helps to solve discrepancies in issues related to the varied specificity of source domains, such as how different scholars have found different source domains for the same type of metaphorical expression. The hierarchical definition suggested in this chapter is powerful because it is based on results of corpora analysis that have then been tested by human subjects.

The corpus-driven approach proposes the possibility of employing more than one source domain. These source domains, which comprise both general and specific ones, can be linked in a hierarchical form. These source domains are obtained through a metaphor framework, and they possess the information between concept domains. Due to these advantages, this approach will make a powerful cross-disciplinary contribution because the results will be useful not only for automatic metaphor identification in the computational field, it will also be useful to the understanding of the cognitive mechanisms behind the use of these conceptual domains. A model as such can also generate work in the field of psycholinguistics and neurolinguistics because the model can be attested via human experiments. Finally, the model is able to provide theoretical explanation regarding how lexical items are related conceptually. More discussion on implications will be carried out in Chapter Thirteen. In Chapter Thirteen, too, a general conclusion will be provided for the overall issues discussed in this book.
Chapter 13: Conclusions

In this chapter, we will conclude the study undertaken for this book. In §13.1, we will first discuss the global hypotheses postulated at the beginning of this book. §13.2 will discuss the significance and implications of this study. §13.3 will discuss the future work resulting from this study.

13.1 Global Hypotheses

We postulated the following global hypotheses, and we have shown in previous chapters that these two hypotheses are discussed.

(1) a. Lexical and computational methods are able to reduce human subjectivity in determining source domains through both top-down and bottom-up approaches. Several lexical resources (such as SinicaBow, WordNet, SUMO and Sketch Engine) will be used to determine source domains so that human intuition can be reduced.

b. The top-down and the bottom-up approaches each have their inherent advantages and disadvantages. As a result, both the top-down and the bottom-up approaches will perform differently in the determination of source domains: The top-down approach will return general source domains, while the bottom-up approach will return specific source domains.

As for the first hypothesis in (1a), although computational methods have their strengths and weaknesses, these computational methods are replicable and, thus, help to reduce the problem of subjective decision-making when determining source domains. Some of the weaknesses of using computational methods are related to the technical limitations of lexical resources. We have argued in previous chapters that our use of computational methods is able to define source domain names using a lexical knowledgebase rather than simply naming source domains based on intuition alone. We have also shown that our results regarding source domain determination support the fact that one type of metaphorical expression may have one or more than one source domain. In the top-down approach, even though source domain names are selected from the keywords in the SUMO definitions that are concrete, some of these source domain names, as we have demonstrated in Chapters Seven and Eight, are more general, while others are more specific. One type of metaphorical expression may belong to both general and specific source domains (such as ORGANISM and MIND). Therefore, the results
in this book may explain why arguments have occurred, regarding whether source domains should be general or specific. In this book, we are able to show that there is no single correct answer for the source domain name. In fact, if we place the source domains on a conceptual hierarchy, we will always find a range of source domains, some more specific and others more general. However, not only is this book able to prove the different specificity of source domains, we have also discovered factors that have previously caused great debate regarding source domains, going so far as having lexical and computational approaches suggesting that source domains have different specificity, and human judgments of source domain names are a reflection of the nature of the source domains. In other words, this book is able to provide explanation of human cognition through the empirical research of source domain determination.

From the comparisons of both approaches mentioned in (1b), we know that the source domains determined by the two approaches are different. The source domains determined by the top-down approach come from content words in the concrete SUMO definitions, whereas those determined using the bottom-up approach come from the WordNet synsets with highest conceptual density (i.e., the WordNet hierarchy with the highest number of collocates gathering around it). Even though the source domains in the two approaches differ in their naming process, this will not become a contributing factor to the different performance of the two approaches, as expected in (1b). In fact, the hypothesis in (1b) is more interested in the question of whether there are advantages and disadvantages in the two approaches and, furthermore, how these two approaches differ in their performance.

In this book, we have proven that the bottom-up approach has performed (slightly) better than the top-down approach. In Chapters Eight through Ten, we discussed the similarities and differences as well as advantages and disadvantages of both approaches. Although the top-down approach does not necessarily return source domains at the same level of specificity (such as BAIRN and ORGANISM), their source domains are usually more general than those of the bottom-up approach. The source domains found in the bottom-up approach are specific instances of the source domains from the top-down approach most of the time, such that KIDDY/小孩 is an instance of ORGANISM. In Chapter Twelve, we explained the relationship between the general versus specific source domains and postulated a hierarchical definition of source domains.

13.2 Significance and Implications

This book has four main implications. The first contribution of this book is its important implications regarding imposing work with corpora design. The majority of
previous work in metaphor analysis has focused on finding examples of metaphors in corpora, with source domains often being identified manually. This book has observed the patterns of types of metaphorical expressions (such as 成長 chéngzhāng ‘grow/growth’ and 癱瘓 tānhuàn ‘paralytic’) and tokens of metaphorical expressions (the number of times 成長 chéngzhāng ‘grow/growth’ and 癱瘓 tānhuàn ‘paralytic’ appear). Additionally, this book has also analyzed sub-corpora data with various factors, including the analysis of Coordinated Target Domains versus Single Target Domains. The analysis in this book compares target domains that appear alone (such as 經濟 jīngjī ‘economy’ and 政治 zhèngzhì ‘politics’ respectively) and target domains that appear in coordination (such as 經濟和政治 jīngjī huá zhèngzhì ‘economy and politics’).

In addition, this book has carried out many comparisons of metaphorical expressions between Taiwan and China. Based on factors such as similar corpora size and similar genre (newspapers) for Single Target Domains and Coordinated Target Domains, comparisons between Taiwan and China have been carried out from various dimensions. The results show that the use of corpora can be made more effective by selecting corpora data that matches the purpose of the study. In this case, carefully designed corpora data have been selected based on factors such as cross-regional analyses, comparisons of constructions and comparisons between types of metaphorical expressions versus tokens of metaphorical expressions.

The second implication of this book is related to the first hypothesis in (1a), i.e., the use of lexical and computational methods is able to reduce human subjectivity in determining source domains; for example, in the following quotation by Semino, Heywood & Short (2004:1283), we can see clearly that source domains are often taken to be “unproblematic.”

As for vehicle identification, we have inserted HORSE as the default agent of galloping away, which on the face of it at least, seems fairly unproblematic.

From this short statement, we can see that source domains are often taken to be correct if no one challenges them. Source domains have not been determined through principled criteria in previous studies. Therefore, through the use of lexical and computational methods, principled criteria can be created and the subjectivity of human judgment can be reduced, i.e., when humans find it hard to determine which metaphor belongs to which source domain, lexical and computational methods can be used to overcome problems such as inconsistency in human judgments.

The third implication is related to the second research question and involves comparing top-down and bottom-up approaches in source domain determination. The studies undertaken prior to this one have not compared these two important approaches
in metaphor analyses. As described in Chapter Three, while many studies have tried to “identify metaphors” in corpora using either a top-down (-like) approach (such as Martin 1990) or a bottom-up (-like) approach (such as Mason 2004),¹ there has been no actual comparison of these two approaches. This book goes on to offer ways to improve upon previous studies by suggesting that we need to first “determine source domains” before we “identify metaphors.” Source domain identification in this book is carried out with carefully defined criteria so that these source domains may then be used to facilitate metaphor identification. In the process of determining source domains, this book compares the top-down and the bottom-up approaches in terms of correctness measures as well as in terms of their predictability of human judgment. The top-down approach determines source domains through the use of ontologies. The bottom-up approach, on the other hand, determines source domains by observing the degree of conceptual density required for the collocation of metaphorical expressions. A comparison of these two approaches in determining source domains such as this has been carried out for the first time in the literature, and therefore, for the first time, source domains have been determined with reduced human subjectivity through the use of computational methods.

On top of the comparisons forged between the top-down and the bottom-up approaches, this book asks an important question regarding the view of metaphors (also discussed in Kövecses 2006), i.e., whether metaphors should be viewed using a top-down or a bottom-up approach. In the top-down approach, all occurrences of conceptual metaphors are attributed to embodied conceptualization, whereby different communities are expected to possess the similar basic conceptual metaphors (cf. Kövecses 2006). Because this is the case, the use of shared ontology is one way of supporting this view. On the other hand, the bottom-up approach makes generalizations using the observations gathered from a small amount of data. The bottom-up approach does not assume a global conceptualization, but it concludes based on observation of conceptual metaphors used (cf. Kövecses 2006). Both of these approaches reflect the cognitive mechanisms behind the creation of conceptual metaphors. This book provides a comparison of both approaches using a shared conceptual structure versus actual collocated patterns found in language. Comparisons such as these within a single study represent a large-scale project, and this is why this study is important in evaluating two views of metaphors.

The fourth implication is related to concrete sense disambiguation. In the top-down approach, the automatic selection of concrete senses from among the many senses of the metaphorical expressions was performed. For example, this approach selects concrete senses by measuring the conceptual similarity of the corresponding SUMO node of

¹ Top-down (-like) and bottom-up (-like) because the authors did not identify these studies as being top-down or bottom-up.
each sense of a type of metaphorical expression (such as the different senses of 成長 chéngzhǎng ‘grow/growth’) against several prototypical concrete senses selected by human subjects. These concrete senses are needed to determine source domain terms because source domains are, by definition, more concrete than target domain terms. Therefore, this book proposes one way to identify concreteness, based on SUMO and WordNet. This research is important because we have aimed to specify clearly what constitutes the source domains of conceptual metaphors, and furthermore, contend that conceptual metaphors should not remain being conceptualized “at the level of thought” (Deignan 1999:180) but, to the contrary, should be more clearly defined.

Each of the issues mentioned above leads into a major direction in metaphor research. In the next section, follow-up research areas will be discussed, as these research areas are all important for studies on metaphors.

13.3 Future Work

The first research area relates to the categorization of specific and general target domains into upper conceptual domains. This is possible to achieve because source domains from the top-down approach are general (such as in (2a) for 改革 gǎigé ‘reform/reformation’), but those from the bottom-up approach are specific (such as in (2b)).

(2) a. Top-down Approach (SinicaBow)
   ARTIFACT
   AGENT
   PURPOSE
   MODIFICATION
   COGNITIVE AGENT
   CHANGE OF POSSESSION
   DESTINATION

b. Bottom-up Approach (CNA)
   ASSIGNMENT/任務
   IMMIGRANT/移民
   TROOP/軍隊
   DISTRICT ATTORNEY/檢察官
   CONCLUSION/結論
   CAPACITY/能力
   DEEPNESS/深度

2 Source domains that are in Appendix A8.1 and A10.1 have been removed.
Even though these source domains are different, they are related because they all appear as source domains for the same type of metaphorical expression. Therefore, the first future research area will aim to examine the relatedness of these source domains as well as to investigate how many upper conceptual domains can be created based on the source domains found. For example, the relation between COGNITIVE AGENT and DISTRICT ATTORNEY/檢察官 may not appear as related, if based on the results of the two approaches above, and thus, it may be possible to relate different types of metaphorical expressions (such as those that fall under the same COGNITIVE AGENT and DISTRICT ATTORNEY/檢察官) and create a hierarchical structure for these source domains.

Our second research area will expand on the notion of concreteness because concreteness itself forms a huge area of research. The top-down approach uses concreteness as an important notion in determining source domains. The concreteness measure has not been carried out for the bottom-up approach because the bottom-up approach builds from evidence at the bottom, and top-level information such as ‘entity’ or ‘abstraction’ is not used to decide whether or not the bottom information is concrete. Table 13.1 provides all the possible source domains for 成長 chéngzhāng ‘grow/growth’ in CNA obtained through the bottom-up approach.

Table 13.1: Possible Source Domains for 成長 chéngzhāng ‘grow/growth’ (CNA) Obtained through the Bottom-up Approach

<table>
<thead>
<tr>
<th>Type of Metaphorical Expression</th>
<th>Relation</th>
<th>Ranks in WordNet Hierarchy</th>
<th>Source Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>成長 chéngzhāng ‘grow/growth’</td>
<td>subject</td>
<td>1</td>
<td>COINAGE/貨幣</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>DEPOSIT/存款</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>BUDGET/預算</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>KIDDY/小孩</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>FRUITAGE/產量</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>CAPACITY/生產量</td>
</tr>
</tbody>
</table>

From this table, it can be seen that the source domains for 成長 chéngzhāng ‘grow/growth’ may vary in terms of concreteness. However, the present approach only measures the rankings of the source domains based on saliency values. It is, therefore, possible to investigate whether or not concreteness can still help in the bottom-up approach. Research in this direction will not only be an extension of this book, it will also be another evaluation of Lakoff’s (1993) statement regarding concreteness of source domains.

Our third research area pertains to the issue raised in Chapter Twelve, i.e., whether or not boundaries of source domains can be established clearly. In this book, we found
that similar metaphorical expressions can be categorized under several source domains. For example, the similar example of 途經 tūjīng ‘path’ given in Chapter Twelve has many source domains:

(3) Source Domains for 途經 tūjīng ‘path’ (SinicaBow)
  TIME
  STATIONARY ARTIFACT
  SEQUENCE
  PURPOSE
  LOCATION
  FUTURE
  ARTIFACT

(Chapter Twelve; pp. 281; ex. (13))

As mentioned, these source domains emphasize different aspects of 途經 tūjīng ‘path.’ For example, the ARTIFACT may emphasize the physical material (i.e., tar) of path; SEQUENCE may emphasize time; PURPOSE may emphasize the person using the path; and LOCATION may emphasize a specific point on the path. These different aspects may be related to Pustejovsky’s Qualia structure (1991), which suggests that words are compositional. From example (3), we can see that certain source domains emphasize related but different aspects of the metaphorical expressions. Therefore, we can pose the question of whether or not boundaries between source domains are clear-cut. If source domains are fuzzy (Rosch 1973), we may find more overlapping metaphorical expressions in different source domains than metaphorical expressions that do not overlap.

As a metaphorical expression may have more than one source domain, it is also possible to investigate the relatedness of concepts in source domains belonging to similar or related metaphors. In this way, we can come closer to answering the question regarding whether or not clear boundaries between source domains exist. For example, we can discover the relatedness of concepts through a conceptual network such as that shown in Figure 13.1.
Figure 13.1: Examples of Metaphorical Expressions for STATIONARY ARTIFACT and Some of Its Related Concepts

In Figure 13.1, we can see the source domain concept of STATIONARY ARTIFACT at the centre of the network. 途徑 tūjìng ‘path’ (circled) is one of the metaphorical expressions that fall under it. Other metaphorical expressions that also fall under STATIONARY ARTIFACT are 橋樑 qiáoliáng ‘bridge,’ 基地 jìdì ‘base’ and 道路 dàolù ‘highroad,’ all of which are human-made objects. In Figure 13.1, too, we can see source domain concepts (i.e., HOUSE and BUILDING) that are also related to
STATIONARY ARTIFACT (see arrows). These concepts are generated based on the source domains that we found in this book. We can see that these source domain concepts are produced through a metaphor framework. In Figure 13.2, we see another conceptual hierarchy for DEVICE.

In Figure 13.2, we can see the metaphorical expressions that are related to DEVICE. Two related concepts of DEVICE are TRANSPORTATION and TRANSPORTATION DEVICE (see arrows). For DEVICE, there are concrete objects, such as 火藥桶

3 Note that HOUSE and BUILDING can also be related to one another, but at this level, as presented in Figure 13.1, their links are not yet seen because this figure is only part of the hierarchy. The whole hierarchy is too large to be displayed in this book.

4 The relatedness of source domains is an ongoing research developed from this book, and Figure 13.1 shows the preliminary results. The conversion of the data from this book in this section was programmed by Ming-Wei Hsu, and the software used was TouchGraph (http://www.touchgraph.com/; available at http://sourceforge.net/projects/touchgraph/).
Object-driven Approach to Source Domain Determination

For TRANSPORTATION, there is only one metaphorical expression shown (軌道 guīdào ‘track’). For TRANSPORTATION DEVICE, we can see verbs such as 起飛 qǐfēi ‘take off’ and 出發 chūfā ‘depart,’ both of which are related to the actions of the transportation device. As mentioned previously, the source domains found in this book can refer to different aspects of a domain, including the participants, processes, events, etc. (cf. Footnote 9 of Chapter Eight). The metaphorical expressions for TRANSPORTATION DEVICE reflect these results.

The relatedness of these concepts is found through examining the source domains of the metaphorical expressions. From here, we can see that source domain information can be used to establish domain information. The corpus-driven approach has advantages in this respect because (a) the data is abundant; and (b) the source domains are automatically determined, and therefore, they are consistent in their names. If the source domains are not consistent, source domains of various abstractions will probably be found. As a result, the metaphorical expressions will probably be scattered at different individual nodes. If this is the case, the building of a conceptual hierarchy will become more difficult. The results shown in this section will, thus, need additional evaluation in the future.

Research in this direction will not only contribute to the connectionist model (McClelland & Rumelhart 1986), but it will also help to establish knowledge domains based on a lexical approach through the categorization of metaphorical expressions and will further establish connectivity between metaphorical expressions. Furthermore, research in terms of concreteness using linguistic evidence has not been seen in previous work. If concreteness can be measured using computational methods, many of the problems faced when defining source domains in conceptual metaphors can be overcome. Therefore, questions answered in this book are essential to conceptual metaphor research.

The comparison between the conceptual knowledge approach (top-down) and the prototype approach (bottom-up) shows that humans seem to respond more positively to the prototype approach than the conceptual knowledge approach. Nevertheless, the real reasons for the responses garnered by human judgments may depend on their lack of understanding of the conceptual knowledge system when compared to the usage-based collocations with which they are familiar. Therefore, in terms of future research, in addition to the proposed research areas previously stated, it will also be possible to compare a folk taxonomy to a conceptual taxonomy, such as those from SUMO.

Finally, this book has important implications, not only in comparing languages used in different regions but also in providing empirical evidence that may help to prove how speakers of different communities differ in their cognition where metaphors are concerned. This book also contributes to the application of computational methods.
to linguistic analyses, which bridges the gap between scientific analyses with human language use. The methodologies used in this book can also contribute to natural language processing, specifically to the automatic extraction of conceptual metaphors in corpora, which has been attempted by many (see Chapter Three) but has not yet been successful because the source domains had not first been determined. Once the source domains have been determined, as we have done in this study, the automatic extraction of metaphors in corpora become more promising, as we now know what has been mapped in conceptual metaphors. Furthermore, understanding source domain categories constitutes a distinctive contribution to research in the categorization of human concepts. For example, we have shown in the hierarchical definition of source domains that source domains are derived from a hierarchical concept. The relationship between source domains, therefore, becomes important in understanding how humans categorize objects and events. In sum, the work in this book is significant in terms of its contributions, and in addition, it also leaves many avenues for future research open. In closing, we can say that the questions posed in this book provide a new direction for future metaphor research.
References


Blake, William. 1794. *Song of Experience*.


A Corpus-driven Approach to Source Domain Determination

References


A Corpus-driven Approach to Source Domain Determination


束定芳. 2000.《隱喻學研究》。上海：上海外語教育出版社。
黃居仁. 2004a. 《中文的詞義小辭典 1.0》[Meaning in Sense in Mandarin Chinese Version 1.0]。台北：中央研究院文獻語料庫與詞庫小組技術報告。
黃居仁. 2004b. 《詞類歧義的本質與解釋——以大量語料庫為本的分析研究》，收錄於石鋒、沈鍾偉主編《樂在其中——王士元教授七十華誕慶祝文集》，235-245。天津：南開大學出版社。
黃居仁. 2005. 《中文的詞義小辭典 2.0》[Meaning in Sense in Mandarin Chinese Version 2.0]。台北：中央研究院文獻語料庫與詞庫小組技術報告。
黃居仁. 2006. 《中文的詞義小辭典 3.0》[Meaning in Sense in Mandarin Chinese Version 3.0]。台北：中央研究院文獻語料庫與詞庫小組技術報告。
黃居仁. 2007. 《中文的詞義小辭典 4.0》[Meaning in Sense in Mandarin Chinese Version 4.0]。台北：中央研究院文獻語料庫與詞庫小組技術報告。
劉揚. 2004. 《雙語 WordNet 詞義知識庫的構造理論與工程實踐》，北京大學博士論文。
References

http://bow.sinica.edu.tw/ (SinicaBow) #
http://corpora.fi.muni.cz/chinese_all/ (Chinese Sketch Engine) #
http://wordsketch.ling.sinica.edu.tw/ (Chinese Sketch Engine) #
http://virtual.cvut.cz/ksmsaWeb/browser/title (KMASA Project) #
http://wordnet.princeton.edu/ (WordNet) #
http://www.touchgraph.com/ (TouchGraph)
http://sourceforge.net/projects/touchgraph/ (TouchGraph)
http://www.anu.edu.au/asianstudies/mcp/ (Malay Concordance Project)
http://www.ln.edu.hk/lle/cwd/project01/web/home.html (Metalue)
http://www.ontologyportal.org/ (Suggested Upper Merged Ontology; SUMO)
http://cogsci.berkeley.edu/lakoff/ (Conceptual Metaphor Homepage)

# Snapshots of these databases are given in the Appendices A4.1 through A4.4.
Appendices

Appendix A4.1: Procedures for Running the WordNet Database
(http://wordnet.princeton.edu/)
A Corpus-driven Approach to Source Domain Determination


Word to search for: notice

Display Options (Select option to change) Change

WordNet home page


Word to search for: notice

Display Options (Select option to change) Change

Key: "S." = Show subset (semantic) relations, "W." = Show Word (lexical) relations

Noun
- S. (of) politics, political relation (social relations involving intrigue to gain authority or power) "office politics is often counterproductive"
- S. (of) politics, political science, government (the study of government of states and other political units)
- S. (of) politics (the profession devoted to governing and to political affairs)
- S. (of) politics, political sympathies (the opinion you hold with respect to political questions)
- S. (of) politics (the activities and affairs involved in managing a state or a government) "unemployment dominated the politics of the inter-war years", "government agencies multiplied beyond the control of representative politics"

WordNet home page
Appendices


Word to search for: politics

Display Options: Change

Key: "S." = Show Synset (semantic) relations, "W." = Show Word (lexical) relations

Noun

- S: (n) politics, political relation (social relations involving intrigue to gain authority or power) "office politics is often counterproductive"
  - domain term category
  - direct hypernym / inherited hypernym / sister term
  - derivationally related form
- S: (n) politics, political science, government (the study of government of states and other political units)
  - S: (n) politics (the profession devoted to governing and to political affairs)
  - S: (n) politics, political sympathy (the opinion you hold with respect to political questions)
- S: (n) politics (the activities and affairs involved in managing a state or a government) "unemployment dominated the politics of the inter-war years", "government agencies multiplied beyond the control of representative politics"

WordNet home page
A Corpus-driven Approach to Source Domain Determination

WordNet Search - 3.0  WordNet home page  Glossary  Help

Word to search for: politics  Select Word

Display Options: [Select option to change]  Change

Key: "S." = Show Synset (semantic) relations, "W." = Show Word (lexical) relations

Noun
- S. (n) politics (social relations involving intrigue to gain authority or power) "office politics is often counterproductive"
  o domain term category
    o direct hypernym / inherited hypernym / sister term
      - S. (n) social relation (a relation between living organisms (especially between people))
        o S. (n) relation (an abstraction belonging to or characteristic of two entities or parts together)
          o S. (n) abstraction, abstract entity (a general concept formed by extracting common features from specific examples)
          o S. (n) entity (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))
    o derivationally related form
      - S. (n) politics, political science, government (the study of government of states and other political units)
      - S. (n) politics (the profession devoted to governing and to political affairs)

WordNet home page
Appendix A4.2: Procedures for Running the KSMSA Database
(http://virtual.cvut.cz/ksmsaWeb/browser/title)
A Corpus-driven Approach to Source Domain Determination
Synset politics

Root > wordnet > Nouns

the profession devoted to governing and to political affairs
Appendix A4.3: Procedures for Running the SinicaBow Database
(http://bow.sinica.edu.tw/)
Select any of these two
Appendix A4.4: Procedures for Running the Chinese Sketch Engine
(Registration Needed)
(http://corpora.fi.muni.cz/chinese_all/)
(http://wordsketch.ling.sinica.edu.tw/)
Select a sub-corpus and type in your keyword
CNA19910524.0254 重申：「政治的歸政治」理念。10月11日香港《大公报》一篇社評認為，政府當局對此觀點理解不足，唯實行政策時，往往衝破「法律的歸法律」界限，令行政機制運作失當。
CNA19910526.0270 事件。應讓司法的歸司法、政治的歸政治。唯實行政策時，往往衝破「法律的歸法律」界限，令行政機制運作失當。
CNA19910807.0182 一個個案代表整個社會，讓政治的歸政治。唯實行政策時，往往衝破「法律的歸法律」界限，令行政機制運作失當。
CNA19910807.0182 老百姓代表事前審查，讓政治的歸政治。唯實行政策時，往往衝破「法律的歸法律」界限，令行政機制運作失當。
CNA19910819.0244 政治並不等於藝術，政府當局應該把「政治的歸政治」理念，貫徹在政策的制定和執行上。
CNA19910924.0197 一、讓「法律的歸法律」理念，貫徹在政策的制定和執行上。
CNA19910924.0297 二、實行政策時，政府當局應該把「法律的歸法律」理念，貫徹在政策的制定和執行上。
CNA19911004.0347 呼籲讓「法律的歸法律、政治的歸政治」理念，貫徹在政策的制定和執行上。
CNA19911108.0098 院會中也提出了政府當局應該把「法律的歸法律、政治的歸政治」理念，貫徹在政策的制定和執行上。
CNA19911109.0170 院會中也提出了政府當局應該把「法律的歸法律、政治的歸政治」理念，貫徹在政策的制定和執行上。
CNA19911123.0193 專政體制下，讓政治的歸政治。唯實行政策時，往往衝破「法律的歸法律」界限，令行政機制運作失當。
CNA19911124.0067 在法律的制定和執行上，政府當局應該把「法律的歸法律」理念，貫徹在政策的制定和執行上。
CNA19920128.0076 政府當局應該把「法律的歸法律」理念，貫徹在政策的制定和執行上。
Appendix A5.1: Percentages of Coordinated Instances Deleted from the First 4 Per cent of the Single Target Domains (CNA)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Total First 4%</th>
<th>Coordinated Deleted</th>
<th>Remaining</th>
<th>Percentage Deleted</th>
</tr>
</thead>
<tbody>
<tr>
<td>政治 zhèngzhì ‘politics’</td>
<td>10,549</td>
<td>1,756</td>
<td>8,793</td>
<td>16.65%</td>
</tr>
<tr>
<td>外交 wàijiāo ‘foreign affairs’</td>
<td>2,493</td>
<td>468</td>
<td>2,025</td>
<td>18.77%</td>
</tr>
<tr>
<td>財政 cāizhèng ‘finance’</td>
<td>4,174</td>
<td>487</td>
<td>3,687</td>
<td>11.67%</td>
</tr>
<tr>
<td>教育 jiàoyù ‘education’</td>
<td>7,295</td>
<td>266</td>
<td>7,029</td>
<td>3.65%</td>
</tr>
<tr>
<td>經濟 jīngjī ‘economy’</td>
<td>24,951</td>
<td>2,890</td>
<td>22,061</td>
<td>11.58%</td>
</tr>
</tbody>
</table>

Appendix A5.2: Percentages of Coordinated Instances Deleted from the First 4 Per cent of the Single Target Domains (XIN)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Total First 4%</th>
<th>Coordinated Deleted</th>
<th>Remaining</th>
<th>Percentage Deleted</th>
</tr>
</thead>
<tbody>
<tr>
<td>政治 zhèngzhì ‘politics’</td>
<td>4,378</td>
<td>1,104</td>
<td>3,274</td>
<td>25.21%</td>
</tr>
<tr>
<td>外交 wàijiāo ‘foreign affairs’</td>
<td>2,213</td>
<td>235</td>
<td>1,978</td>
<td>10.62%</td>
</tr>
<tr>
<td>財政 cāizhèng ‘finance’</td>
<td>1,436</td>
<td>162</td>
<td>1,274</td>
<td>11.26%</td>
</tr>
<tr>
<td>教育 jiàoyù ‘education’</td>
<td>6,329</td>
<td>1,152</td>
<td>5,177</td>
<td>18.20%</td>
</tr>
<tr>
<td>經濟 jīngjī ‘economy’</td>
<td>26,888</td>
<td>4,744</td>
<td>22,144</td>
<td>17.64%</td>
</tr>
</tbody>
</table>
## Appendix A5.3: Instance Number per Year from 1991 through 2002 in CNA and XIN in Each Target Domain

<table>
<thead>
<tr>
<th>Years</th>
<th>外交 ‘foreign affairs’</th>
<th>政治 ‘politics’</th>
<th>財政 ‘finance’</th>
<th>教育 ‘education’</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CNA</td>
<td>XIN</td>
<td>CNA</td>
<td>XIN</td>
<td>CNA</td>
</tr>
<tr>
<td>1991</td>
<td>92</td>
<td>23</td>
<td>906</td>
<td>890</td>
<td>105</td>
</tr>
<tr>
<td>1992</td>
<td>94</td>
<td>40</td>
<td>877</td>
<td>1052</td>
<td>91</td>
</tr>
<tr>
<td>1993</td>
<td>88</td>
<td>24</td>
<td>860</td>
<td>901</td>
<td>110</td>
</tr>
<tr>
<td>1994</td>
<td>131</td>
<td>28</td>
<td>929</td>
<td>1055</td>
<td>166</td>
</tr>
<tr>
<td>1995</td>
<td>92</td>
<td>28</td>
<td>1196</td>
<td>922</td>
<td>107</td>
</tr>
<tr>
<td>1996</td>
<td>86</td>
<td>27</td>
<td>1093</td>
<td>945</td>
<td>107</td>
</tr>
<tr>
<td>1997</td>
<td>77</td>
<td>42</td>
<td>1032</td>
<td>1131</td>
<td>155</td>
</tr>
<tr>
<td>1998</td>
<td>88</td>
<td>39</td>
<td>1024</td>
<td>1033</td>
<td>133</td>
</tr>
<tr>
<td>1999</td>
<td>101</td>
<td>43</td>
<td>1002</td>
<td>1210</td>
<td>176</td>
</tr>
<tr>
<td>2000</td>
<td>91</td>
<td>34</td>
<td>1016</td>
<td>1072</td>
<td>143</td>
</tr>
<tr>
<td>2001</td>
<td>88</td>
<td>33</td>
<td>1013</td>
<td>1236</td>
<td>130</td>
</tr>
<tr>
<td>2002</td>
<td>91</td>
<td>42</td>
<td>750</td>
<td>699</td>
<td>118</td>
</tr>
<tr>
<td>Total</td>
<td>1097</td>
<td>423</td>
<td>11698</td>
<td>12236</td>
<td>1541</td>
</tr>
</tbody>
</table>

## Appendix A5.4: Chi-Square Analyses of Instances in Each Target Domain and in CNA and XIN

<table>
<thead>
<tr>
<th>外交 ‘foreign affairs’</th>
<th>政治 ‘politics’</th>
<th>財政 ‘finance’</th>
<th>教育 ‘education’</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNA/XIN versus years</td>
<td>CNA</td>
<td>XIN</td>
<td>CNA</td>
</tr>
<tr>
<td>11, N=1520</td>
<td>36.56</td>
<td>p&lt;.05</td>
<td>11, N=23934</td>
</tr>
<tr>
<td>Overall</td>
<td>298.87</td>
<td>p&lt;.05</td>
<td>12.09</td>
</tr>
<tr>
<td>CNA versus XIN</td>
<td>CNA</td>
<td>XIN</td>
<td>CNA</td>
</tr>
<tr>
<td>11, N=1097</td>
<td>27.27</td>
<td>p&lt;.05</td>
<td>11, N=423</td>
</tr>
<tr>
<td>Within Years 1991-2002</td>
<td>CNA</td>
<td>XIN</td>
<td>CNA</td>
</tr>
<tr>
<td>11, N=11689</td>
<td>156.54</td>
<td>p&lt;.05</td>
<td>11, N=12236</td>
</tr>
<tr>
<td>11, N=1341</td>
<td>61.78</td>
<td>p&lt;.05</td>
<td>11, N=814</td>
</tr>
<tr>
<td>11, N=60.33</td>
<td>p&lt;.05</td>
<td>11, N=47.00</td>
<td>p&lt;.05</td>
</tr>
</tbody>
</table>
**Appendix A6.1:** Types of Metaphorical Expressions, Tokens of Metaphorical Expressions and Tokens-per-Type Ratios in CNA, XIN, CNA+XIN and the Overlapped in CNA+XIN in the Single Target Domains and Coordinated Target Domains

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Total Instances Analyzed</th>
<th>Types of Metaphorical Expressions</th>
<th>Total Tokens of Metaphorical Expressions</th>
<th>Tokens-per-Type Ratio</th>
<th>Percentages of Tokens of Metaphorical Expressions from Total Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Target Domains</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CNA</td>
<td>43,595</td>
<td>1,126</td>
<td>14,075</td>
<td>12.50</td>
<td>32.29</td>
</tr>
<tr>
<td>XIN</td>
<td>33,847</td>
<td>677</td>
<td>10,799</td>
<td>15.95</td>
<td>31.91</td>
</tr>
<tr>
<td>CNA+XIN</td>
<td>77,442</td>
<td>1,428</td>
<td>24,878</td>
<td>17.42</td>
<td>32.12</td>
</tr>
<tr>
<td>Overlapped</td>
<td>77,442</td>
<td>375</td>
<td>21,696</td>
<td>57.86</td>
<td>28.02</td>
</tr>
<tr>
<td>Coordinated Target Domains</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CNA</td>
<td>14,258</td>
<td>452</td>
<td>4,218</td>
<td>9.33</td>
<td>29.58</td>
</tr>
<tr>
<td>XIN</td>
<td>14,765</td>
<td>221</td>
<td>3,409</td>
<td>15.43</td>
<td>23.09</td>
</tr>
<tr>
<td>CNA+XIN</td>
<td>29,023</td>
<td>534</td>
<td>7,627</td>
<td>14.28</td>
<td>26.28</td>
</tr>
<tr>
<td>Overlapped</td>
<td>29,023</td>
<td>139</td>
<td>6,378</td>
<td>45.88</td>
<td>21.98</td>
</tr>
</tbody>
</table>
## Appendix A6.2: Types of Metaphorical Expressions, Tokens of Metaphorical Expressions and Tokens-per-Type Ratios in CNA, XIN, CNA+XIN and the Overlapped in CNA+XIN in Different Target Domains

<table>
<thead>
<tr>
<th>Target Domains</th>
<th>Single/ Coordinated</th>
<th>CNA</th>
<th>XIN</th>
<th>CNA+XIN</th>
<th>Total Instances Analyzed</th>
<th>Types of Metaphorical Expressions</th>
<th>Tokens of Metaphorical Expressions</th>
<th>Tokens-per-Type Ratios</th>
<th>Percentages of Tokens of Metaphorical Expressions from Total Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>政治 zhengzhi ‘politics’</td>
<td>Coord</td>
<td>CNA</td>
<td>11,698</td>
<td>398</td>
<td>3,803</td>
<td>9.56</td>
<td>32.51%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>XIN</td>
<td>12,336</td>
<td>106</td>
<td>3,130</td>
<td>15.07</td>
<td>25.58%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CNA+XIN</td>
<td>23,034</td>
<td>472</td>
<td>6,933</td>
<td>14.69</td>
<td>28.07%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>封建 jingdian ‘finance’</td>
<td>Single</td>
<td>CNA</td>
<td>8,793</td>
<td>518</td>
<td>2,921</td>
<td>5.64</td>
<td>33.22%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>XIN</td>
<td>3,274</td>
<td>152</td>
<td>1,117</td>
<td>7.35</td>
<td>34.12%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CNA+XIN</td>
<td>12,067</td>
<td>565</td>
<td>4,038</td>
<td>7.15</td>
<td>33.46%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>封建 jingdian ‘finance’</td>
<td>Single+Coord</td>
<td>CNA</td>
<td>20,491</td>
<td>734</td>
<td>6,724</td>
<td>9.16</td>
<td>32.81%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>XIN</td>
<td>15,510</td>
<td>280</td>
<td>4,247</td>
<td>15.17</td>
<td>27.38%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CNA+XIN</td>
<td>36,001</td>
<td>825</td>
<td>10,971</td>
<td>13.30</td>
<td>30.47%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>外交 waijiao ‘foreign affairs’</td>
<td>Coord</td>
<td>CNA</td>
<td>814</td>
<td>60</td>
<td>109</td>
<td>1.82</td>
<td>13.39%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>XIN</td>
<td>1,541</td>
<td>31</td>
<td>116</td>
<td>3.74</td>
<td>7.53%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CNA+XIN</td>
<td>2,355</td>
<td>74</td>
<td>225</td>
<td>3.04</td>
<td>9.55%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>外交 waijiao ‘foreign affairs’</td>
<td>Single</td>
<td>CNA</td>
<td>2,025</td>
<td>99</td>
<td>623</td>
<td>6.29</td>
<td>30.77%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>XIN</td>
<td>1,978</td>
<td>88</td>
<td>338</td>
<td>3.84</td>
<td>17.09%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CNA+XIN</td>
<td>4,003</td>
<td>155</td>
<td>961</td>
<td>6.20</td>
<td>24.01%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>外交 waijiao ‘foreign affairs’</td>
<td>Single+Coord</td>
<td>CNA</td>
<td>2,839</td>
<td>133</td>
<td>732</td>
<td>5.50</td>
<td>25.78%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>XIN</td>
<td>3,519</td>
<td>104</td>
<td>454</td>
<td>4.37</td>
<td>12.90%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CNA+XIN</td>
<td>6,358</td>
<td>193</td>
<td>1,186</td>
<td>6.15</td>
<td>18.65%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>教育 jiaoyu ‘education’</td>
<td>Coord</td>
<td>CNA</td>
<td>1,097</td>
<td>80</td>
<td>216</td>
<td>2.70</td>
<td>19.69%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>XIN</td>
<td>423</td>
<td>22</td>
<td>77</td>
<td>3.59</td>
<td>18.20%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CNA+XIN</td>
<td>1,520</td>
<td>93</td>
<td>293</td>
<td>3.15</td>
<td>19.28%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>教育 jiaoyu ‘education’</td>
<td>Single</td>
<td>CNA</td>
<td>3,687</td>
<td>139</td>
<td>583</td>
<td>4.19</td>
<td>15.81%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>XIN</td>
<td>1,274</td>
<td>37</td>
<td>173</td>
<td>4.68</td>
<td>13.58%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CNA+XIN</td>
<td>4,961</td>
<td>157</td>
<td>756</td>
<td>4.82</td>
<td>15.24%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>教育 jiaoyu ‘education’</td>
<td>Single+Coord</td>
<td>CNA</td>
<td>4,764</td>
<td>190</td>
<td>799</td>
<td>4.21</td>
<td>16.79%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>XIN</td>
<td>1,697</td>
<td>56</td>
<td>250</td>
<td>4.46</td>
<td>14.73%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CNA+XIN</td>
<td>6,481</td>
<td>216</td>
<td>1,049</td>
<td>4.85</td>
<td>16.19%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>經濟 jingji ‘economy’</td>
<td>Coord</td>
<td>CNA</td>
<td>649</td>
<td>30</td>
<td>90</td>
<td>3.00</td>
<td>13.87%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>XIN</td>
<td>565</td>
<td>22</td>
<td>86</td>
<td>3.91</td>
<td>15.22%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CNA+XIN</td>
<td>1,214</td>
<td>46</td>
<td>176</td>
<td>3.83</td>
<td>14.50%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>經濟 jingji ‘economy’</td>
<td>Single</td>
<td>CNA</td>
<td>7,029</td>
<td>172</td>
<td>1,300</td>
<td>7.56</td>
<td>18.49%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>XIN</td>
<td>5,177</td>
<td>174</td>
<td>1,419</td>
<td>8.16</td>
<td>27.41%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CNA+XIN</td>
<td>12,206</td>
<td>281</td>
<td>2,791</td>
<td>9.93</td>
<td>22.87%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>經濟 jingji ‘economy’</td>
<td>Single+Coord</td>
<td>CNA</td>
<td>7,678</td>
<td>185</td>
<td>1,390</td>
<td>7.51</td>
<td>18.10%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>XIN</td>
<td>5,742</td>
<td>185</td>
<td>1,505</td>
<td>8.14</td>
<td>26.21%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CNA+XIN</td>
<td>13,420</td>
<td>302</td>
<td>2,895</td>
<td>9.59</td>
<td>21.57%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>經濟 jingji ‘economy’</td>
<td>Single</td>
<td>CNA</td>
<td>22,061</td>
<td>651</td>
<td>8,648</td>
<td>13.28</td>
<td>39.20%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>XIN</td>
<td>22,144</td>
<td>443</td>
<td>7,752</td>
<td>17.50</td>
<td>35.01%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CNA+XIN</td>
<td>44,205</td>
<td>864</td>
<td>16,400</td>
<td>15.51</td>
<td>37.10%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>經濟 jingji ‘economy’</td>
<td>Single+Coord</td>
<td>CNA</td>
<td>36,319</td>
<td>875</td>
<td>11,161</td>
<td>12.76</td>
<td>30.78%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>XIN</td>
<td>36,090</td>
<td>535</td>
<td>12,866</td>
<td>24.05</td>
<td>34.86%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CNA+XIN</td>
<td>73,228</td>
<td>1100</td>
<td>24,027</td>
<td>16.51</td>
<td>32.81%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Appendix A6.3: Comparisons of 政治 zhèngzhì ‘politics’ and 經濟 jīngjì ‘economy’ to other Single Target Domains through Kruskal-Wallis Tests

<table>
<thead>
<tr>
<th>Domain</th>
<th>Language</th>
<th>CNA</th>
<th>XIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>政治 zhèngzhì ‘politics’</td>
<td>CNA</td>
<td>Mean rank of A=717.31</td>
<td>Mean rank of B=476.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\chi^2(1)=151.51, p&lt;.01$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>XIN</td>
<td>Mean rank of A=442.68</td>
<td>Mean rank of B=248.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\chi^2(1)=148.65, p&lt;.01$</td>
<td></td>
</tr>
<tr>
<td>經濟 jīngjì ‘economy’</td>
<td>CNA</td>
<td>Mean rank of A=291.36</td>
<td>Mean rank of B=350.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\chi^2(1)=88.49, p&lt;.01$</td>
<td>Mean rank of C=603.58</td>
</tr>
<tr>
<td></td>
<td>XIN</td>
<td>Mean rank of A=87.93</td>
<td>Mean rank of B=225.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\chi^2(1)=67.67, p&lt;.01$</td>
<td>Mean rank of C=416.19</td>
</tr>
<tr>
<td>外交 wàijiāo ‘foreign affairs’</td>
<td>CNA</td>
<td>Mean rank of A=271.02</td>
<td>Mean rank of B=335.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\chi^2(1)=137.31, p&lt;.01$</td>
<td>Mean rank of D=637.10</td>
</tr>
<tr>
<td></td>
<td>XIN</td>
<td>Mean rank of A=115.81</td>
<td>Mean rank of B=235.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\chi^2(1)=110.21, p&lt;.01$</td>
<td>Mean rank of D=419.86</td>
</tr>
<tr>
<td>財政 cāizhèng ‘finance’</td>
<td>CNA</td>
<td>Mean rank of A=312.29</td>
<td>Mean rank of B=363.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\chi^2(1)=51.15, p&lt;.01$</td>
<td>Mean rank of E=593.19</td>
</tr>
<tr>
<td></td>
<td>XIN</td>
<td>Mean rank of A=198.43</td>
<td>Mean rank of B=265.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\chi^2(1)=41.75, p&lt;.01$</td>
<td>Mean rank of E=419.28</td>
</tr>
<tr>
<td>教育 jiàoyù ‘education’</td>
<td>CNA</td>
<td>Mean rank of A=433.13</td>
<td>Mean rank of B=593.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\chi^2(1)=131.76, p&lt;.01$</td>
<td>Mean rank of E=593.19</td>
</tr>
<tr>
<td></td>
<td>XIN</td>
<td>Mean rank of A=131.87</td>
<td>Mean rank of B=419.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\chi^2(1)=95.73, p&lt;.01$</td>
<td>Mean rank of E=419.28</td>
</tr>
</tbody>
</table>

---
**Appendix A6.4: Comparisons of Single Target Domains versus Coordinated Target Domains through Kruskal-Wallis Tests**

<table>
<thead>
<tr>
<th></th>
<th>Single Target Domains (I) versus Coordinated Target Domains (J)</th>
</tr>
</thead>
</table>
| **政治 zhèngzhì** | Mean rank of I=504.49  
| ‘politics’        | Mean rank of J=393.50  
|                   | $\chi^2(1)=41.33, p<.05$                                      |
|                   | Mean rank of I=231.83  
| XIN               | Mean rank of J=130.04  
|                   | $\chi^2(1)=89.47, p<.05$                                      |
| **外交 wàijiāo**  | Mean rank of I=82.33  
| ‘foreign affairs’ | Mean rank of J=155.58  
|                   | $\chi^2(1)=73.63, p<.05$                                      |
|                   | Mean rank of I=22.57  
| XIN               | Mean rank of J=42.50  
|                   | $\chi^2(1)=19.76, p<.05$                                      |
| **財政 cáizhèng** | Mean rank of I=68.38  
| ‘finance’         | Mean rank of J=97.67  
|                   | $\chi^2(1)=15.91, p<.05$                                      |
|                   | Mean rank of I=53.22  
| XIN               | Mean rank of J=79.26  
|                   | $\chi^2(1)=14.43, p<.05$                                      |
| **教育 jiàoyù**   | Mean rank of I=89.40  
| ‘education’       | Mean rank of J=167.10  
|                   | $\chi^2(1)=48.69, p<.05$                                      |
|                   | Mean rank of I=88.61  
| XIN               | Mean rank of J=171.82  
|                   | $\chi^2(1)=46.71, p<.05$                                      |
Appendix A7.1: The Percentages of Found Types and Tokens of Metaphorical Expressions in SinicaBow When Searched in the Merged Word List

<table>
<thead>
<tr>
<th>Percentages</th>
<th>Types</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNA</td>
<td>96.99</td>
<td>96.59</td>
</tr>
<tr>
<td>XIN</td>
<td>98.90</td>
<td>97.91</td>
</tr>
<tr>
<td>CNA and XIN</td>
<td>97.61</td>
<td>96.72</td>
</tr>
<tr>
<td>Overlapped</td>
<td>99.08</td>
<td>97.63</td>
</tr>
</tbody>
</table>
Appendices

Appendix A7.2: Questionnaire for the Concreteness Rating of the SUMO Nodes

語言調查問卷

您好！對不起佔用您寶貴的時間。我們要作一份有關人們使用語言的問卷，以進行語言方面的研究，竭誠邀請您的參與，協助我們完成本問卷，並請您據實回答。謝謝！

姓名：
學校系級：
年齡：
性別：  男  女

1. 您是僑生嗎？  是（請說明僑居地）：
否

2. 您會使用幾種語言？

<table>
<thead>
<tr>
<th>語言種類</th>
<th>何時開始學的？</th>
<th>在哪學的？</th>
<th>跟誰學的？</th>
<th>流利程度（只會一點---------------非常流利）</th>
</tr>
</thead>
<tbody>
<tr>
<td>國語</td>
<td>1 2 3 4 5 6 7</td>
<td>1 2 3 4 5 6 7</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>台語</td>
<td></td>
<td>1 2 3 4 5 6 7</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>英語</td>
<td></td>
<td>1 2 3 4 5 6 7</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. 您的哪一手比較靈活？ 左  右  一樣靈活

4. 您的直系親屬中有沒有人左手比較靈活？
有（請列出稱謂）：
沒有

5. 您腦部受過傷嗎？
是（請說明）：
否

6. 您曾因昏迷而送醫嗎？
是（多久？）：
否

7. 您有聽覺障礙未矯正嗎？
是（請說明）：
否

8. 您有視覺障礙未矯正嗎？
是（請說明）：
否
同學你好，

謝謝你參加本心理語言學的實驗。本實驗的目的是要了解人們對於不同概念之間的認知。這並不是智力測驗或語言遊戲，所以問題中並沒有陷阱，而且沒有標準答案，請各位用平常心以及你們的直覺來判斷作答就可以了。

問卷一共有 626 題，裡面都是英文的概念，填寫完畢後請交給負責人。

問卷中每個題目裡的一個英文概念，有些概念裡的英文詞會連在一起，那不是文法的錯誤，而是資料來源的關係。請你在看過這些英文概念之後，根據你的語感，判斷每題裡的概念是否具體。如果你覺得問題裡的英文概念是非常具體的話，請圈選 7，如果你覺得概念非常不具體，請圈選 1。
請必須每項都填寫完畢，舉例如下：

1  Cake    1 2 3 4 5 6 7
2  Soul    1 2 3 4 5 6 7

如果有任何問題請你現在提出來，等一下作答就不能發問。再次提醒你，等一下你要做的事就是根據你的語感，判斷每題裡的概念是否具體。如果你覺得問題裡的英文概念是非常具體的話，請圈選 7，如果你覺得概念非常不具體，請圈選 1。

最後謝謝你的協助和幫忙！
Appendix A8.1: List of General Keywords in the Top-down Approach

ABSTRACT
ACTIVITY
ASPECT
ATTRIBUTES
BASIS
CLASS
COMMON
CONCEPT
CORE
DOMAIN
ELEMENT
ENTITY
EVENT
FACT
FORM
INSTANCE
LIST
MASS
MATTER
MEAN
MEMBER
NAME
OBJECT
PART
PROCESS
RANGE
REASON
RELATION
SERIES
SET
SET_OR_CLASS
SHAPE
SIZE
SPECIFICATION
SUBCLASS
SUBSTANCE
THING
TYPE
UNIT
Appendix A9.1: Formula for the Calculation of Shortest Distance

Shortest Distance = \( d^2 = \left(\frac{i}{n1}\right)^2 + \left(\frac{f(i)}{n2}\right)^2 \)

Steps to derive this formula:

1. For method 1, see the following illustration:

2. The shortest distance is calculated using the following formula:

\[
\begin{align*}
    d^2 &= (i-0)^2 + (bi^n-0)^2 \\
    &= i^2 + b^2i^{2n}
\end{align*}
\]

Given \( \frac{\partial i^k}{\partial i} = k * i^{k-1} \)

Thus, to calculate the shortest distance is:

\[
\begin{align*}
    \frac{\partial d^2(i)}{\partial i} &= 2i + b^2 * 2ai^{2a-1} = 0 \\
    (i) + ab^2 i^{(2a-1)} &= 0 \\
    ab^2 &= -i^{(2-2a)} \\
    i &= (-ab^2)^{(1/2a)}
\end{align*}
\]
Formula for the Calculation of Slope (with value -1)

\[
\text{Slope} = \frac{\partial (bi^a)}{\partial i} = a * b * i^{a-1} = -1
\]

\[
i = (-a*b)^{1/1-a}
\]
Appendix A10.1: List of General Keywords in the Bottom-up Approach

ASPECT/局面
BASIS/基礎
BEHALF/方面
BRIEF/綱要
DESCRIPTION/性質
MATTER/問題
PECULIARITY/特性
POINT/特點
REQUISITENESS/必要
SETUP/格局
Appendices

Appendix A11.1: Questionnaire for the Evaluation of Source Domains

語言調查問卷

您好！對不起佔用您寶貴的時間。我們要作一份有關人們使用語言的問卷，以進行語言方面的研究，竭誠邀請您的參與，協助我們完成本問卷，並請您據實回答。謝謝！

姓名： 學校系級：
年齡： 性別： 男 女

1. 您是僑生嗎？ 是（請說明僑居地）： 否

2. 您會使用幾種語言？

<table>
<thead>
<tr>
<th>語言種類</th>
<th>何時開始學的？</th>
<th>在哪學的？</th>
<th>跟誰學的？</th>
<th>流利程度</th>
</tr>
</thead>
<tbody>
<tr>
<td>國語</td>
<td>1 2 3 4 5 6 7</td>
<td>1 2 3 4 5 6 7</td>
<td>1 2 3 4 5 6 7</td>
<td>1 2 3 4 5 6 7</td>
</tr>
<tr>
<td>台語</td>
<td>1 2 3 4 5 6 7</td>
<td>1 2 3 4 5 6 7</td>
<td>1 2 3 4 5 6 7</td>
<td>1 2 3 4 5 6 7</td>
</tr>
<tr>
<td>英語</td>
<td>1 2 3 4 5 6 7</td>
<td>1 2 3 4 5 6 7</td>
<td>1 2 3 4 5 6 7</td>
<td>1 2 3 4 5 6 7</td>
</tr>
</tbody>
</table>

3. 您的哪一手比較靈活？ 左 右 一樣靈活

4. 您的直系親屬中有沒有人左手比較靈活？

5. 您腦部受過傷嗎？ 是（請說明）： 否

6. 您曾因昏迷而送醫嗎？ 是（多久？）： 否

7. 您有聽覺障礙未矯正嗎？ 是（請說明）： 否

8. 您有視覺障礙未矯正嗎？ 是（請說明）： 否
同學你好，

謝謝你參加本實驗。本實驗為心理語言學實驗，目的是要了解人們對於不同概念之間的認知。這個實驗並不是智力測驗或語言遊戲，所以問題中沒有陷阱，而且沒有標準答案，請各位用平常心以及你們的直覺來判斷作答就可以了。

問卷共有 415 題，每個題目裡都有 A 和 B 兩個詞組。A 是詞義概念，B 是詞義的領域。請你在看過詞組之後，根據你的語感和直覺，決定每題裡 A 的詞義概念是否屬於 B 的領域，然後圈選作答。作答時，請在是或否兩個選項間圈選你覺得最適當的答案，每題都必須作答。舉例如下：

請問，A 的詞義概念是否屬於 B 的領域？

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>手錶</td>
<td>裝飾品</td>
<td>是</td>
<td>否</td>
</tr>
<tr>
<td>6</td>
<td>頭髮</td>
<td>文具</td>
<td>是</td>
<td>否</td>
</tr>
<tr>
<td>7</td>
<td>打球</td>
<td>運動</td>
<td>是</td>
<td>否</td>
</tr>
<tr>
<td>8</td>
<td>潛水</td>
<td>打架</td>
<td>是</td>
<td>否</td>
</tr>
</tbody>
</table>

如果有任何問題可以現在提出來，開始作答之後就不能發問。再次提醒你，等一下你要做的，是要根據你的語感和直覺，決定每題裡 A 的詞義概念是否屬於 B 的領域，然後圈選作答。請你在是或否之間圈選你覺得最適當的答案，並請依序作答。

最後，謝謝你的協助和幫忙！
指 導 語 (Bottom-up)

同學你好，

謝謝你參加本實驗。本實驗為心理語言學實驗，目的是要了解人們對於不同概念之間的認知。這個實驗並不是智力測驗或語言遊戲，所以問題中沒有陷阱，而且沒有標準答案，請各位用平常心以及你們的直覺來判斷作答就可以了。

問卷共有 413 題，每個題目裡都有 A 和 B 兩個詞組。請你在看過詞組之後，根據你的語感和直覺，決定每題裡 A 和 B 是否屬於相同的概念領域，然後圈選作答。作答時，請在是或否兩個選項間選你覺得最適當的答案，每題都必須作答。舉例如下:

請問，A 和 B 是否屬於相同的概念領域？

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.</td>
<td>手錶</td>
<td>戒指</td>
</tr>
<tr>
<td>10.</td>
<td>頭髮</td>
<td>文具</td>
</tr>
<tr>
<td>11.</td>
<td>打球</td>
<td>運動</td>
</tr>
<tr>
<td>12.</td>
<td>潛水</td>
<td>打架</td>
</tr>
</tbody>
</table>

如果有任何問題可以現在提出來，開始作答之後就不能發問。再次提醒你，等一下你要做的，是要根據你的語感和直覺，決定每題裡 A 和 B 是否屬於相同的概念領域，然後圈選作答。

請你在是或否之間圈選你覺得最適當的答案，並請依序作答。

最後，謝謝你的協助和幫忙！
Index

A
abstract, v, 5, 15, 115, 118, 203, 204, 205, 206, 207, 238, 264, 267, 281, 291, 296
abstraction, 207
algorithm, 119
automatic, vi, 10, 13, 15, 17, 21, 22, 28, 29, 37, 43, 46, 48, 51, 52, 73, 104, 107, 130, 168, 174, 185, 196, 200, 203, 229, 275, 281, 284, 289, 295, 296
axis, 178, 179, 180, 181, 182, 183

B
British National Corpus, 37, 175
BUILDING, 1, 10, 16, 17, 211, 274, 275, 294, 295, 302

C
Central News Agency of Taiwan. See CNA
Chinese Gigaword Corpus
Gigaword corpus, 40, 45, 46, 51, 52, 53, 65, 68, 77
clustering, 146, 169, 170, 189, 190, 194, 196, 197, 198, 208, 212, 232
CNA+XIN, xiv, xvi, 55, 68, 77, 83, 85, 107, 108, 142, 172, 197, 207
cognitive mechanism, 77, 288, 289
collocation, v, vi, 8, 24, 25, 28, 29, 39, 53, 165, 170, 191, 209, 211, 212, 227, 228, 229, 231, 266, 270, 277, 289, 297
computational method, 6, 7, 18, 74, 101, 151, 165, 171, 212, 229, 235, 269, 285, 288, 296, 297
computational tool, v, 7, 13
computational work, 18, 28
conceptual category, 3
conceptual density, 52, 194, 195, 201, 203, 204, 205, 206, 207, 212, 229
conceptual distance, 51, 122, 136
conceptual hierarchy, 22, 166, 285, 286, 295, 296
conceptual key, 263
conceptual knowledge, v, 4, 33, 101, 105, 107, 139, 166, 297
conceptual metaphor, v, vi, 1, 2, 3, 4, 5, 9, 10, 12, 13, 14, 15, 16, 17, 26, 72, 135, 136, 147, 149, 153, 161, 210, 228, 259, 263, 274, 283, 288, 289, 294, 295
Conceptual Metaphor Theory, 1, 2, 3, 5, 9, 17, 112, 113
conceptual similarity, 119, 121, 122, 123, 130, 143, 146, 167
conceptualization, vi, 2, 3, 6, 7, 205, 217, 289
concordance, 39, 46, 64, 75
concreteness, 1, 3, 5, 10, 15, 17, 51, 104, 111, 116, 118, 123, 125, 128, 130, 132, 137, 140, 143, 264, 290, 291, 292, 296
connectionist model, 296
conventional, 12, 13, 280, 281, 284, 285
co-occurrence, 12, 69, 75, 180
coordination, 66, 68, 90, 91, 101, 105, 158, 163, 223, 224, 288
corpus-based, 10, 17, 24, 100, 278, 303
corpus-driven, 281, 282, 283, 284, 285
correlation, 104, 251
cross-linguistic, 2, 234
cross-regional, 85, 287
curve, 178, 180, 181, 184, 185
Index

cut-off point, i, ii, 51, 177, 178, 180, 181,
182, 184, 185, 186, 187, 188, 190, 192,
194, 204, 206, 212, 227, 233, 234, 284
CWN, xxii, xxiii, 243, 244, 245, 246, 248,
252, 254, 257, 261, 268

D
database, 18, 19, 20, 29, 37, 39, 46, 50, 52,
143, 164, 167, 192, 206, 280, 281, 283,
284, 285, 307
dataset, 8, 31, 47, 48, 49, 51, 53, 67, 73, 77,
78, 79, 83, 84, 85, 91, 93, 102, 108,
109, 145, 146, 147, 157, 161, 170, 208,
209, 223, 225, 231, 232, 236, 250, 282
decision-making, 234, 285
definition, v, vi, 1, 4, 5, 8, 9, 10, 11, 14, 15,
16, 46, 47, 101, 104, 105, 109, 112,
113, 115, 117, 128, 129, 130, 131, 132,
136, 137, 138, 140, 209, 230, 231, 235,
239, 247, 263, 264, 268, 274, 275, 281,
282, 284, 285, 286, 287, 289, 296
DEVICE, xviii, 263, 269, 295, 296
dictionary, 20, 21, 22, 31, 168, 280
dictionaries, 36, 107, 111
distance, ii, iii, 50, 115, 118, 119, 120, 121,
122, 123, 124, 126, 127, 128, 136, 146,
166, 167, 180, 181, 183, 185, 188, 195
distributional pattern, 79, 81, 178
domain, v, 1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12,
13, 14, 15, 16, 17, 18, 19, 21, 23, 24,
25, 26, 28, 29, 32, 34, 43, 44, 46, 47,
48, 51, 52, 55, 56, 59, 62, 66, 68, 70,
71, 72, 73, 74, 75, 76, 79, 86, 87, 88,
89, 92, 93, 94, 95, 97, 98, 100, 101,
102, 103, 112, 113, 115, 129, 130, 131,
132, 135, 136, 138, 139, 140, 141, 142,
144, 145, 147, 148, 149, 151, 152, 154,
155, 157, 158, 159, 162, 164, 166, 167,
168, 169, 171, 172, 173, 190, 192, 198,
199, 202, 203, 204, 206, 207, 208, 210,
211, 212, 214, 215, 216, 217, 218, 219,
221, 222, 223, 224, 225, 229, 230, 231,
232, 234, 235, 236, 237, 238, 239, 240,
244, 245, 246, 247, 248, 251, 252, 255,

258, 260, 261, 262, 263, 264, 265, 266,
268, 269, 270, 271, 272, 273, 274, 275,
278, 279, 280, 281, 282, 283, 284, 285,
286, 287, 288, 289, 290, 291, 292, 293,
295, 296, 301
source domain, v, vi, 1, 2, 3, 4, 5, 6,
7, 8, 9, 10, 11, 12, 13, 14, 15, 16,
17, 18, 19, 20, 21, 22, 23, 24, 25,
26, 27, 28, 30, 31, 47, 48, 50, 51,
52, 53, 56, 57, 61, 62, 64, 73, 74,
75, 76, 77, 78, 92, 101, 102, 103,
104, 107, 108, 109, 113, 115, 116,
118, 130, 132, 136, 137, 138, 139,
140, 141, 142, 143, 144, 145, 146,
147, 148, 149, 150, 151, 152, 153,
154, 156, 157, 158, 159, 160, 161,
164, 165, 166, 167, 168, 169, 170,
171, 172, 174, 177, 184, 192, 193,
194, 198, 204, 205, 206, 207, 208,
209, 210, 211, 212, 213, 214, 216,
217, 218, 219, 220, 221, 222, 224,
225, 226, 227, 228, 229, 230, 231,
232, 234, 235, 236, 238, 239, 240,
241, 242, 244, 246, 247, 248, 249,
250, 251, 252, 254, 255, 256, 259,
260, 262, 263, 264, 265, 266, 267,
268, 269, 270, 271, 272, 273, 274,
275, 276, 277, 280, 281, 282, 283,
284, 285, 286, 287, 288, 289, 290,
291, 292, 293, 294, 295, 296, 297
target domain, i, v, 1, 2, 3, 8, 9, 14,
15, 18, 21, 22, 24, 47, 48, 52, 53,
55, 56, 57, 58, 60, 61, 62, 63, 64,
65, 66, 68, 69, 70, 71, 72, 73, 74,
75, 76, 77, 78, 80, 81, 89, 90, 91,
92, 93, 95, 96, 97, 98, 99, 100, 101,
102, 103, 104, 105, 106, 107, 108,
109, 115, 116, 146, 147, 152, 154,
158, 165, 167, 169, 170, 171, 173,
174, 175, 176, 177, 180, 181, 190,
195, 197, 206, 207, 208, 211, 212,
214, 218, 219, 222, 223, 242, 264,
266, 267, 268, 271, 287, 290

339


160, 165, 170, 212, 213, 214, 220, 221, 222, 224, 226, 230, 236, 251, 280, 283, 286
Kruskal-Wallis, xiv, xxiii, 65, 67, 79, 85, 88, 96, 98, 99, 102, 244, 247, 255, 257

L
lexical approach, 296
lexical item, 21, 22, 23, 30, 31, 48, 157, 170, 192, 239, 240, 246, 272, 288
linguistic pattern, 5, 43, 45, 49, 52

M
manual analysis, 53, 73, 116, 147
mapping, 1, 3, 11, 17, 25, 29, 35, 36, 37, 38, 50, 57, 101, 104, 105, 107, 109, 115, 118, 125, 128, 140, 141, 146, 154, 166, 192, 193, 194, 211
mean of means, 186, 188, 190, 194, 212
Merged Word List. See Chinese-English Merged Word List
metaphor identification, v, 8, 9, 10, 17, 18, 20, 22, 29, 30, 31, 47, 77, 78, 277, 282, 285, 288, 289
Metaphorical Pattern Analysis, 24, 26, 27, 76, 227
Mutual Information, 41, 177, 233, 300

N
naming, 6, 9, 14, 30, 101, 170, 263, 285, 287
newspaper, 4, 45, 50, 51, 69, 77, 78, 79, 93, 271, 288
node, 23, 32, 33, 35, 36, 50, 51, 52, 57, 60, 62, 104, 105, 115, 116, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 130, 132, 135, 136, 138, 139, 140, 143,
A Corpus-driven Approach to Source Domain Determination

271, 272, 274, 275, 285, 286, 287, 289, 290, 291
transformation, 121, 182, 186
TRANSPORTATION, 142, 269, 271, 296
TRANSPORTATION DEVICE, 142, 269, 271, 296
type of metaphorical expression, 50, 52, 78, 83, 84, 104, 105, 112, 130, 136, 139, 146, 156, 192, 218, 224, 235, 250, 252, 276
See type of metaphorical expression

V
vehicle, 191, 287
WordNet, vi, vii, ix, x, xii, xiii, xv, xvi, xix, xx, xxii, xxiii

W
war metaphor, 266, 272, 287

X
XIN, x, xii, xiv, xvi, xii, xxii, xxiii, 45, 46, 47, 51, 52, 53, 55, 56, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 89, 91, 92, 93, 95, 96, 97, 98, 99, 101, 102, 103, 105, 107, 109, 112, 113, 139, 142, 143, 147, 149, 156, 159, 165, 166, 170, 172, 173, 174, 175, 184, 192, 193, 197, 198, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 221, 222, 223, 226, 229, 230, 231, 232, 233, 234, 236, 237, 238, 244, 247, 252, 253, 254, 256, 257, 258, 259, 262, 266, 269, 273
Xinhua News Agency of Beijing. See XIN

Z
Zipf's law, 178
**Biodata**

Siaw-Fong Chung is now an Assistant Professor at the Department of English, National Chengchi University, Taiwan. After obtaining her Ph.D. degree from National Taiwan University, she worked as a post doctoral research fellow at National Taiwan University for about a year. Siaw-Fong Chung’s interests lie in corpus linguistics and lexical semantics. Her Ph.D. dissertation was awarded the 2007 Outstanding Doctoral Dissertation Award by the Association of Computational Linguistics and Chinese Language Processing (ACLCLP).

Mailing Address:
Siaw-Fong Chung
Department of English
National Chengchi University
No. 64, Sec. 2, ZhiNan Road
Wenshan District
Taipei City 11605, Taiwan
Email: sfchung@nccu.edu.tw
LANGUAGE AND LINGUISTICS

LANGUAGE AND LINGUISTICS is a department of the INSTITUTE OF LINGUISTICS, ACADEMIA SINICA. It furthers the Institute’s objective of excellence in research, scholarship, and education by publishing pioneering articles and monographs in Linguistics and other related fields. All publications have been reviewed rigorously according to academic standards.

Copyright © 2009 by INSTITUTE OF LINGUISTICS, ACADEMIA SINICA
128, Sec. 2, Academia Road, Nankang, Taipei 11529, Taiwan, R.O.C.
http://www.ling.sinica.edu.tw
All rights reserved.

LANGUAGE AND LINGUISTICS MONOGRAPH SERIES NUMBER A22

By Siaw-Fong Chung

A Corpus-driven Approach to Source Domain Determination

《語言暨語言學》
《語言暨語言學》隸屬於中央研究院語言學研究所，以出版語言學及相關領域之最新研究成果為宗旨。所有出版品均經過嚴格學術審查。

《語言暨語言學》專刊甲種之二十二
以語料庫驅動之隱喻源域界定研究
鍾曉芳 著

出版：中央研究院語言學研究所
發行：中央研究院語言學研究所
115 台北市 南港區 研究院路 2 段 128 號
http://www.ling.sinica.edu.tw
版權所有　翻印必究

印刷：文盛彩藝事業有限公司
初版：中華民國九十八年十二月 [2009]
定價：新台幣 700 元／US$40
GPN: 1009803179