Automatic lexicon construction for domain-specific sentiment analysis : a frame-based approach

Tan, Sang Sang

2020

Tan, S. S. (2020). Automatic lexicon construction for domain-specific sentiment analysis : a frame-based approach. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/143088>

<https://doi.org/10.32657/10356/143088>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Downloaded on 26 Apr 2023 08:56:39 SGT

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

**AUTOMATIC LEXICON CONSTRUCTION FOR
DOMAIN-SPECIFIC SENTIMENT ANALYSIS: A
FRAME-BASED APPROACH**

TAN SANG SANG

WEE KIM WEE SCHOOL OF COMMUNICATION AND INFORMATION

2020

AUTOMATIC LEXICON CONSTRUCTION FOR DOMAIN-SPECIFIC SENTIMENT ANALYSIS: A FRAME-BASED APPROACH

TAN SANG SANG

Wee Kim Wee School of Communication and Information

A thesis submitted to the Nanyang Technological University
in partial fulfilment of the requirement for the degree of
Doctor of Philosophy

2020

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

August 23, 2019

Date

Tan Sang Sang

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

August 23, 2019

Date

Assoc. Prof. Na Jin Cheon

Authorship Attribution Statement

This thesis contains material from three papers published in or under consideration by the following book chapters / conference proceedings. I am the first and corresponding author of the papers.

Some sentences / paragraphs in Chapter 1, 2, and 5 have been published in

Tan, S. S., & Na, J. C. (2016). Expanding Sentiment Lexicon with Multi-word Terms for Domain-Specific Sentiment Analysis. In: Morishima A., Rauber A., Liew C. (eds) Digital Libraries: Knowledge, Information, and Data in an Open Access Society. ICADL 2016. Lecture Notes in Computer Science, vol 10075. Springer, Cham.

Some sentences / paragraphs in Chapter 2 and 4 have been published in

Tan, S. S., & Na, J. C. (2017) Mining Semantic Patterns for Sentiment Analysis of Product Reviews. In: Kamps J., Tsakonas G., Manolopoulos Y., Iliadis L., Karydis I. (eds) Research and Advanced Technology for Digital Libraries. TPDL 2017. Lecture Notes in Computer Science, vol 10450. Springer, Cham.

I co-designed the studies in the above papers with Dr. Na Jin Cheon and prepared the manuscript drafts. He revised the manuscripts.

August 23, 2019

Date

Tan Sang Sang

ACKNOWLEDGMENTS

First and foremost, praises and thanks to God, the Almighty, for His abundant grace and showers of blessings along this journey.

I would like to express my deepest gratitude to my supervisor, Dr. Na Jin Cheon, for his guidance in the past four years. His kindness, sincerity, vision, and motivation have deeply inspired me. It has been a priceless experience working with him. I really appreciate him, both professionally and personally.

I am also grateful to my family and church members for their love, support, and sacrifices. Without their prayers, this challenging endeavor would not have been possible.

Finally, my thanks go to all the people who have supported me directly and indirectly throughout this research work.

TABLE OF CONTENTS

Statement of Originality	iii
Supervisor Declaration Statement	iv
Authorship Attribution Statement	v
ACKNOWLEDGMENTS	vi
TABLE OF CONTENTS	vii
LIST OF FIGURES	xi
LIST OF TABLES	xiii
SUMMARY	xvi
Chapter 1 Introduction	1
1.1 Background and Motivation	1
1.2 Research Questions and Objectives	4
1.3 Overview of the Research Process	10
1.3.1 Phase 1	11
1.3.2 Phase 2	13
1.3.3 Phase 3	15
1.4 Significance of Research	16
1.5 Outline of the Thesis	18
Chapter 2 Literature Review	20
2.1 Introduction	20

2.2 Automatic Lexicon Construction	21
2.2.1 Dictionary-Based Approach.....	21
2.2.2 Corpus-Based Approach	24
2.2.3 Constructing Phrasal Sentiment Lexicons	28
2.3 FrameNet and Frame Identification	34
2.4 Summary	41
Chapter 3 Frame Identification	42
3.1 Introduction.....	42
3.2 Background	43
3.3 The Proposed Frame Identification Model	45
3.3.1 Bidirectional Encoder Representations from Transformers (BERT)	46
3.3.2 Positional Attention-based Frame Identification with BERT (PAFIBERT)	48
3.3.3 Frame Filtering.....	55
3.3.4 Listing of Candidate Frames	56
3.4 Experiments	57
3.5 Results and Discussion	59
3.6 Summary	68
Chapter 4 GradFrameNet and Sentiment-Aware Semantic Abstraction	70
4.1 Introduction.....	70

4.2 Background	71
4.3 Method	73
4.3.1 GradFrameNet.....	74
4.3.2 Semantic Tagging for Semantic Abstraction	83
4.4 Experiment.....	87
4.5 Results and Discussion	90
4.5.1 Assessing the Semantic Abstraction Ability of the Proposed Method	108
4.5.2 Assessing the Information Discovery Ability of the Proposed Method	115
4.6 Summary	123
Chapter 5 Domain-Specific Lexicon Construction and Sentiment Classification	125
5.1 Introduction.....	125
5.2 Background	126
5.3 The Proposed Domain-Specific Lexicon Construction Method	128
5.3.1 Candidate Generation.....	130
5.3.2 Sentiment Score Calculation and Lexical Entry Selection	136
5.4 Experiment.....	138
5.4.1 Data	138
5.4.2 Sentiment Classifiers	139

5.4.3 Performance Measures	143
5.5 Results and Discussion	143
5.5.1 Effects of Frame-based Semantic Abstraction on Sparsity Degree, Lexicon Size, and Noise Level	144
5.5.2 Evaluation of Lexicons in Sentence-Level Sentiment Classification.....	150
5.6 Summary	154
Chapter 6 Conclusion.....	156
6.1 Introduction.....	156
6.2 Review of Research Objectives	157
6.3 Research Contributions	161
6.4 Limitations and Future Work.....	165
6.5 Concluding Remark	169
References.....	170
Appendix A.....	200

LIST OF FIGURES

Figure 1.1. Explicit (Set A) and implicit (Set B) sentiment expression	5
Figure 1.2. Overview of the Research Process	12
Figure 1.3. A simplified schematic view of the [Desirability] frame (FrameNet 1.7)	14
Figure 3.1. PAFIBERT without frame filtering	50
Figure 3.2. PAFIBERT with frame filtering	51
Figure 3.3. Simplified schematic representations for [Possibility], [Capability], [Possession], and [Have_associated]	62
Figure 3.4. Frame predictions for “Also There is so much product in it, so it is worth the \$\$”	63
Figure 3.5. Frame predictions for “It is tough to find in Toronto (it was at Walmart), but I'll stock up whenever I can!”	64
Figure 3.6. Frame predictions for “I do think it has a highlighting benefit (also a good shade), but it's just too dry and cakey to wear”	65
Figure 3.7. Frame predictions for “I used it for years, and I was sad to give it up”	65
Figure 4.1. The [Usefulness] frame (FrameNet 1.7)	72
Figure 4.2. The [Taking_time] frame (FrameNet 1.7)	73
Figure 4.3. Examples of frames and LUs that can be graduated based on Infused Intensification	77
Figure 4.4. The graduation taxonomy	79
Figure 4.5. The semantic tagging process	84

Figure 4.6. Frequencies of non-graduation and graduated frames in the product reviews	111
Figure 5.1. Overall process of the proposed domain-specific lexicon construction method.....	129
Figure 5.2. The frequency table and semantic lookup tables.....	133
Figure 5.3. Counts of low-frequency terms in eyeliner reviews.....	146
Figure 5.4. Counts of low-frequency terms in cleanser reviews.....	146
Figure 5.5. Examples of entries in the domain-specific lexicons (using hybrid semantic abstraction)	149
Figure 5.6. Figure 5.6. F-scores (Y-axis) for eyeliner data in relation to different relative pruning thresholds (X-axis). The threshold combinations for skip-4-grams, skip-3-grams, and skip-2-grams are displayed on the X-axis. The F-scores were obtained from 10-fold cross-validation using SVM classifiers trained on different feature sets.....	150
Figure 5.7. Figure 5.7. F-scores (Y-axis) for cleanser data in relation to different relative pruning thresholds (X-axis). The threshold combinations for skip-4-grams, skip-3-grams, and skip-2-grams are displayed on the X-axis. The F-scores were obtained from 10-fold cross-validation using SVM classifiers trained on different feature sets.....	151
Figure 6.1. Overview of Phase 1.....	158
Figure 6.2. Overview of Phase 2.....	160
Figure 6.3. Overview of Phase 3.....	161

LIST OF TABLES

Table 2.1. Syntactic categories of opinion words treated in previous works	29
Table 2.2. Concepts extracted by a concept parser (Poria et al., 2014)	32
Table 3.1. Examples of FrameNet’s annotations for frame identification. The target positions mark the beginning and ending characters of a target.	49
Table 3.2. The selected hyperparameters	58
Table 3.3. Accuracies with frame filtering	59
Table 3.4. Accuracies without frame filtering	60
Table 3.5. Most frequently confused pairs of frames	61
Table 3.6. Examples of misclassified instances	61
Table 4.1. Examples of LUs retrieved from FrameNet for the annotation task.	79
Table 4.2. Numbers of frames and LUs for each graduation type and level	83
Table 4.3. Detailed information about the data	88
Table 4.4. Collocation groups from negative eyeliner reviews (-E)	91
Table 4.5. Collocation groups from positive eyeliner reviews (+E)	93
Table 4.6. Collocation groups from negative cleanser reviews (-C)	95
Table 4.7. Collocation groups from positive cleanser reviews (+C).	97
Table 4.8. Collocation groups from negative moisturizer reviews (-M)	99
Table 4.9. Collocation groups from positive moisturizer reviews (+M)	101
Table 4.10. Collocation groups from negative television reviews (-T)	103
Table 4.11. Collocation groups from positive television reviews (+T).	105

Table 4.12. The frequencies of non-graduated frames and different graduation types. Examples are given for graduation types and graduation levels found in the extracted collocation groups.	110
Table 4.13. Top 15 semantic labels in the reviews of the four product types .	113
Table 4.14. Numbers of collocation groups that contain at least one low-frequency (< 5 occurrences) cases	115
Table 4.15. Statistics on the information types	117
Table 4.16. Collocations from negative eyeliner reviews (-E)	121
Table 4.17. Collocations from positive eyeliner reviews (+E)	121
Table 4.18. Collocations from negative cleanser reviews (-C).....	121
Table 4.19. Collocations from positive cleanser reviews (+C).....	122
Table 4.20. Collocations from negative moisturizer reviews (-M).....	122
Table 4.21. Collocations from positive moisturizer reviews (+M).....	122
Table 4.22. Collocations from negative television reviews (-T)	123
Table 4.23. Collocations from positive television reviews (+T)	123
Table 5.1. Semantic-level and lexical-level skip-grams extracted for the sentence “It has no staying power whatsoever and it is not that much cheaper than MAC.”	132
Table 5.2. Pseudo labeled data for domain-specific lexicon construction.....	139
Table 5.3. Manually labeled data for sentiment classification.....	139
Table 5.4. Classification models	140
Table 5.5. Statistics of the document-term matrices before and after abstraction	145

Table 5.6. Sizes of the domain-specific lexicons before and after applying frame-based semantic abstraction	148
Table 5.7. Sentiment classification results over 10-fold cross-validation (* $p \leq$.05, ** $p \leq .01$, *** $p \leq .001$).....	152
Table 5.8. The proportion of unseen targets in the development data	153

SUMMARY

The need to gather useful information efficiently from a vast amount of online data has given rise to the interdisciplinary field of automated sentiment analysis. In general, sentiment analysis researchers agree that detecting implicit sentiments is more challenging than recognizing explicit ones. While the latter can be recognized more easily through the presence of sentiment-laden words, the former are conveyed indirectly, usually by describing a situation that invokes or can be interpreted as related to a specific sentiment. Attempts to deal with the challenge of implicit sentiments have leveraged sentiment lexicons containing phrasal entries capable of capturing subtle clues for inferring sentiments.

The present research aims to establish an automated lexicon construction process to build phrasal sentiment lexicons by extracting regular phrasal patterns from domain-specific text corpora. The key challenge in detecting such patterns is that most phrasal units have low frequencies in text corpora and occur in different lexical forms even when they convey similar semantics. For instance, in beauty product reviews, ‘sting my face’ and ‘burn my skin’ are lexically different phrases that describe the same negative experience. Because of the existence of many phrasal variants, each with a low frequency, it is challenging to detect them from text corpora. This issue is related to the common language phenomenon of word sparsity. To address this problem, the present research set out to investigate the plausibility of a linguistically-motivated approach that uses semantic frames in FrameNet for frame-based semantic abstraction. The key idea is that, by creating abstract semantic representations for texts, semantically

similar phrasal units represented by the same frames will be grouped together, creating a more compact and less sparse textual space that facilitates the detection of regular phrasal patterns.

Two prerequisites need to be in place for this frame-based abstraction approach to work properly. First, the frame identification problem must be dealt with adequately. Frame identification is a common subtask in frame-semantic parsing. It disambiguates a marked target in a sentence (a word or a phrase) and associates it with the most likely frame in FrameNet. In this research, a robust frame identification model called Positional Attention-based Frame Identification with BERT (PAFIBERT) was developed, which outperformed existing solutions and was proven robust in handling unseen targets (i.e., out-of-dictionary targets not listed in FrameNet). The second prerequisite for frame-based semantic abstraction is related to the degree of abstraction provided by FrameNet’s frames. Since FrameNet was not initially created for sentiment-related tasks, some information that is essential for distinguishing positive opinions from negative ones is not encoded in the frames. The present research incorporated an appraisal concept called graduation into FrameNet to provide discriminative sentiment-related information, resulting in a sentiment-aware version of FrameNet known as GradFrameNet. Using GradFrameNet and PAFIBERT, this research demonstrated that the frame-based abstraction approach did not cause substantial loss of information but rather generalized in a sentiment-aware manner to extract valuable phrasal clues for sentiment analysis. With the above two requirements fulfilled, the final stage of this research established a domain-specific sentiment lexicon construction process that applied

frame-based semantic abstraction to increase the coverage of phrasal lexicons. From the performance gain obtained by incorporating the lexicons into sentiment classifiers, it was found that the integration of semantic abstraction was beneficial to the lexicon construction process despite the noise introduced by the imperfect frame identification step.

This work makes several original contributions. First, it contributes to the existing knowledge of corpus-based lexicon construction by addressing the issue of lexicon coverage and dealing specifically with the difficulty of extracting phrasal patterns from sparse textual corpora using a novel frame-based abstraction approach. As one of the first efforts to examine the potential of FrameNet in a rather novel setting and an early attempt to realize the graduation phenomenon on a large scale, this research enhances the understanding of how semantic frames can be combined with the graduation concept to provide abstract yet semantic-rich and sentiment-aware text representations. On a practical level, GradFrameNet and PAFIBERT also benefit the community of researchers and practitioners as useful linguistic resources or tools for sentiment analysis. In addition, the findings in this research shed light on the potential of the frame-based abstraction approach in confronting the fundamental problem of word sparsity, which is an inherent characteristic of textual data that poses additional challenges for various text mining tasks. Finally, since frame identification is a pivotal subtask in frame-semantic parsing, this research also makes an important contribution to the field by providing a model that achieved new state-of-the-art results.

CHAPTER 1

INTRODUCTION

1.1 Background and Motivation

Consumers' preferences and decisions are increasingly influenced by peer opinions from online reviews, social networks, blogs, and other user-generated content on the Web (Chen, Fay, & Wang, 2003; Constantinides & Fountain, 2008; Dviso, 2010; Ye, Law, Gu, & Chen, 2011; Zhang, Ye, Law, & Li, 2010). In particular, the impacts of online reviews on consumers' purchasing decisions and product sales have been well-established (Chen et al., 2003; Chevalier & Mayzlin, 2006; Dviso, 2010). Potential customers tend to regard online reviews as a relevant source of user-oriented information that describes the qualities of products or services from the perspectives of a consumer (Bickart & Schindler, 2001; Chen & Xie, 2008; Lee, Park, & Han, 2008). This growing reliance on online reviews opens up excellent opportunities but also poses new challenges for business stakeholders and marketers. While the information in online reviews is valuable in helping businesses generate innovative ideas, plan their product improvements, and manage marketing strategies to target consumers' concerns and requirements, acquiring such information can be time-consuming and resource-demanding because of the volume and dynamic nature of online data. Driven by the need to gather these data and turn them into actual business values, the interdisciplinary field of automated sentiment analysis has attracted significant research

interest in recent years. The primary goal of sentiment analysis is to classify the opinions expressed in texts into positive and negative categories with the aim of inferring people's attitudes toward specific targets and attributes. By providing an avenue to summarize and classify opinions automatically, sentiment analysis plays a pivotal role in supporting efficient decision-making in today's fast-paced business world.

It has become commonplace to classify sentiment analysis approaches into three categories: statistical, knowledge-based, and hybrid (Cambria, Das, Bandyopadhyay, & Feraco, 2017; Ravi & Ravi, 2015; Taboada, 2016). Statistical techniques in sentiment analysis leverage labeled data to train sentiment classifiers, usually in a supervised learning setting. In earlier studies, support vector machine (SVM) and naïve Bayes (NB) were among the most extensively studied statistical techniques in sentiment classification (Pang, Lee, & Vaithyanathan, 2002; Tan & Zhang, 2008; Wang & Manning, 2012; Zhang, Ye, Zhang, & Li, 2011). The performance of these statistical models depends on the features used in building the models. Commonly used features include word n-grams (Wang & Manning, 2012), negations (Jia, Yu, & Meng, 2009), and domain-specific features such as hashtags and emojis (Davidov, Tsur, & Rappoport, 2010), among others. In recent years, deep learning has emerged as a dominant machine learning technique in many application domains (e.g., He, Zhang, Ren, & Sun, 2016; Wang, Wang, & Yeung, 2015). In sentiment analysis, deep learning has also achieved state-of-the-art results with minimal feature engineering efforts (Devlin, Chang, Lee, & Toutanova, 2018; Howard & Ruder, 2018). However, as Cambria et al. (2017) pointed out, “[S]tatistical methods are generally semantically weak, i.e., lexical or co-occurrence elements in a statistical model have little predictive value individually” (p. 6). Therefore, a statistical technique usually works well

with larger text units such as documents and paragraphs but is less effective on its own in classifying smaller text units such as sentences and clauses (Cambria et al., 2017).

A large body of research has employed hybrid approaches combining statistical and knowledge-based techniques. In purely knowledge-based methods, sentiment lexicons are usually used in conjunction with rules to aggregate sentiment scores based on the prior sentiment values of words and phrases (e.g., Kennedy & Inkpen, 2006; Taboada, Brooke, Tofiloski, Voll, & Stede, 2011). In hybrid approaches, the rules are either replaced or complemented by statistical models that incorporate sentiment knowledge bases into model learning and inference processes (Kiritchenko, Zhu, & Mohammad, 2014a; Poria, Cambria, Winterstein, & Huang, 2014). In both cases, a high-quality sentiment lexicon is needed to achieve satisfactory classification rates. Therefore, a major interest within the field of sentiment analysis is the construction of sentiment lexicons—lexicons consisting of words or phrases associated with positive or negative sentiment scores—as instrumental resources for sentiment analysis.

Lexicon construction techniques in sentiment analysis can be categorized into dictionary-based (e.g., Blair-Goldensohn et al., 2008; Hu & Liu, 2004; Kamps, Marx, Mokken, & Rijke, 2004; Rao & Ravichandran, 2009; Wilson, Wiebe, & Hoffmann, 2005) and corpus-based approaches (e.g., Bross & Ehrig, 2013; Hamilton, Clark, Leskovec, & Jurafsky, 2016; Kaji & Kitsuregawa, 2007; Kiritchenko et al., 2014a; Velikovich, Blair-Goldensohn, Hannan, & McDonald, 2010). In general, the former usually results in generic lexicons consisting of domain-independent entries that are connotatively positive or negative such as “excellent,” “awful,” “beautiful,” and “ugly.” Apparently, not all opinions are expressed with this type of sentiment-laden or emotion-related terms that explicitly signify

certain sentiment orientations. For instance, consider the utterance “When the eyeliner dries, it forms a rubbery, latex finish material that comes off in flakes” that appears in a beauty product review. While it does not contain any sentiment-bearing terms, this sentence describes an undesirable fact about the beauty product. Without domain-specific knowledge, it is unlikely that this sentence can be associated correctly with negative sentiment by a sentiment classifier. Since it is extremely difficult, if not impossible, to construct a universal lexicon that caters to the needs of every domain, one of the most widely accepted alternatives is to automatically generate domain-specific lexicons to cover commonly used opinion words or phrases for each domain. This type of lexicon is usually produced using the corpus-based approach, which has been proven effective in domain-specific sentiment analysis studies including those of Bross and Ehrig (2013), Hamilton et al. (2016), Lu, Castellanos, Dayal, and Zhai (2011), and Qiu, Liu, Bu, and Chen (2009), among others.

1.2 Research Questions and Objectives

One of the many difficulties faced by the sentiment analysis research community is that sentiment can be expressed in many different ways, some of which involve the use of opinionated terms that convey attitude in a relatively straightforward manner while others contain no prominent sign of evaluative or emotion-related words but still express sentiment implicitly (Balahur, Hermida, Montoyo, & Muñoz, 2013; Cambria, Schuller, Xia, & Havasi, 2013; Liu, 2015; Noferesti & Shamsfard, 2015; Ofek et al., 2016). Dealing with implicit sentiment expression has been a lingering concern in sentiment analysis research. To illustrate the difficulty of handling this type of sentiment expression, Figure 1.1 shows two sets of sample sentences selected from beauty product reviews. Sentences in Set A are

Figure 1.1. Explicit (Set A) and implicit (Set B) sentiment expression

examples of explicit sentiment expression, in which strongly evaluative or emotion-related words like “glad,” “nice,” “irritate,” “gentle,” and “effective” are used. Sentences in Set B, on the other hand, are more subtle and implicit. Without the use of sentiment-bearing words, the reviewers convey their attitudes implicitly by describing usage scenarios, giving advice, and so forth. While explicit sentiment can be detected more easily from specific keywords, analyzing implicit sentiment often requires the interpretation of larger linguistic units like phrases or clauses. On a related but more general note, eminent researchers in linguistics have convincingly argued that meaning is not merely the attribute or property of individual words (Sinclair, 2004; Stubbs, 2001). Martin and White (2005), who proposed a comprehensive appraisal framework that describes evaluative language, also stated that although it is possible in some cases to pinpoint certain lexical items as carriers of assessment or sentiment in a sentence, a better approach would be to consider the complete evaluative meanings carried by utterances or propositions.

As researchers increasingly realize the importance of word compositions in sentiment analysis, several attempts have been made to enable the extraction and utilization of phrasal constructs in sentiment analysis. For example, Cambria and colleagues (Cambria, Havasi, & Hussain, 2012; Cambria, Olsher, & Rajagopal, 2014; Cambria, Poria, Bajpai, & Schuller, 2016; Cambria, Poria, Hazarika, & Kwok, 2018) developed SenticNet, a concept-level resource that provides affective information for common sense concepts. These concepts are mostly phrasal units that convey sentiment implicitly through context and common sense knowledge. For instance, “birthday party” and “celebrate special occasion” are concepts related to positive sentiment whereas “feel guilty” and “lose temper” are concepts pertaining to negative sentiment. Other studies that have explored concept-level knowledge include those of Agarwal, Mittal, Bansal, and Garg (2015a), Agarwal, Poria, Mittal, Gelbukh, and Hussain (2015b), Tsai, Wu, Tsai, and Hsu (2013), and Wu and Tsai (2014). Besides the concept-level knowledge-based approach, there has also been an increase in the attention given to phrasal sentiment compositions in other lexicon construction studies (Kiritchenko & Mohammad, 2016a, 2016b).

Although the importance of phrasal units in sentiment analysis is well-recognized in existing research, the discovery of useful phrasal clues could be hampered by word sparsity (Qu, Ifrim, & Weikum, 2010; Tan, Na, Theng, & Chang, 2012; Wilson, Wiebe, & Hwa, 2006). Sparsity is a common characteristic of text data: Based on the language phenomenon described by Zipf’s Law (Zipf, 1949), the frequency distribution of words in any text corpus is often defined approximately by a simple mathematical form such that there are few high-frequency words but a long-tail distribution of infrequent words. This characteristic of text data poses significant challenge not only to sentiment analysis but also to other natural language processing (NLP) problems dealing with textual data (Khan, Bashir, & Qamar,

2014; Liu, Chang, Wu, & Yang, 2017; Liu, Kang, Yu, & Wang, 2005; Saif, Fernández, He, & Alani, 2014; Saif, He, & Alani, 2012). So far, there has been little effort to specifically address the challenges of achieving extensive coverage for phrasal lexicons when phrasal units need to be extracted and be given sentiment scores automatically based on their occurrences in sparse text data. Wide coverage is a general requirement for all types of sentiment lexicons, but because of the compositional nature of phrases, this criterion is particularly difficult to fulfill for domain lexicons consisting of phrases aiming to capture domain-specific knowledge. Note that the phrasal patterns that are useful for domain-specific sentiment analysis should extend beyond fixed multi-word units carrying meanings that are not entirely obvious from those of their constituents. Fixed multi-word units may occur in the form of compounds (e.g., credit card), phrasal verbs (e.g., look after), fixed expressions (e.g., to a certain extent), and idioms (e.g., a piece of cake). Since these multi-word units have fixed meanings and most of them could be found in dictionaries and lexical resources, extracting them from text corpora may not be necessary. However, the phrasal patterns that this research is interested in are more variable combinations of words that capture domain knowledge. In the examples given in Figure 1.1, some potentially useful phrasal patterns may include “on third bottle,” “not move out of place,” and “clean face without drying.” Apparently, detecting these phrasal patterns from texts is challenging because phrases can vary in lexical form despite being semantically equivalent. This problem has also been pointed out and partially addressed in previous works using conceptual primitives obtained via hierarchical clustering and dimensionality reduction (Cambria et al., 2016; Cambria et al., 2018). To date, however, the plausibility of using a linguistically-motivated categorization approach to address the lexical variation and coverage problem has not been closely examined. The present research thus set out to explore the potential of

FrameNet (Baker, Fillmore, & Lowe, 1998) and semantic frames in producing abstract and semantically consistent representations to reduce data sparsity and increase the coverage of phrasal units in domain-specific lexicons.

FrameNet (Baker et al., 1998) is a lexical database that stemmed from frame semantics theory (Fillmore, 1976; Fillmore, Johnson, & Petruck, 2003). The idea underlying the creation of FrameNet is that semantic frames form semantic categories and contexts for the interpretation of meanings. The main goal of this research is to develop a sentiment lexicon construction method that extracts phrases from domain data and automatically assign sentiment scores to these phrasal lexicon entries. In line with this goal, the present research was conducted to determine whether mapping texts to FrameNet’s semantic frames based on frame-semantic similarity would reduce word sparsity and facilitate the extraction of interesting phrasal patterns for sentiment analysis. For instance, in describing a negative reaction to a beauty product, reviewers may use lexically-varied phrases such as “sting my face,” “sting my skin,” “hurt my eye,” and so forth, but at the more abstract semantic level, the reviewers’ negativity can be inferred on the basis that the product caused harm to their body. By mapping texts to semantic frames, the above phrases can be represented using a more abstract, frame-based semantic representation: {[Cause_harm], [Body_part]}. While this frame-based conceptualization seems plausible, two requirements need to be in place for this approach to be implemented successfully.

First, an essential prerequisite for using FrameNet’s frames in NLP is the availability of an automatic semantic parsing tool that processes text inputs and maps them to the most appropriate frames in FrameNet. The task of associating words with FrameNet’s frames is commonly known as frame identification (e.g., Das, Chen, Martins, Schneider, & Smith,

2014; Hermann, Das, Weston, & Ganchev, 2014; Hartmann, Kuznetsov, Martin, & Gurevych, 2017; Peng, Thomson, Swayamdipta, & Smith, 2018). Since a word such as “buy” or a fixed expression such as “take off” may be associated with multiple semantic frames because of homonymy and polysemy, the frame identification task is similar to a word sense disambiguation task that disambiguates words based on sentential contexts. Previous works on frame identification have mostly been conducted on an earlier version of FrameNet (FrameNet 1.5) and have not adequately addressed the issue related to unseen targets, i.e., words that are either not listed under any frame in FrameNet or do not appear in the hand-annotated data released along with FrameNet for the development of frame-semantic parsers. In the present research, since the frame identification process is intended for parsing domain-specific data, it is crucial for the frame identifier to properly identify frames for domain-specific words and jargons that are usually not listed in FrameNet. This issue leads to the first research question (RQ1): How can a robust frame identification model be developed to produce accurate semantic-level representations for domain-specific data?

The second requirement concerns the suitability of FrameNet as a resource for generating abstract semantic representations for sentiment analysis. My initial inspection of FrameNet’s frames revealed that some sentiment-related information that is essential for distinguishing positive opinions from negative opinions is missing from the semantic frames. Therefore, the second question (RQ2) of this research is: How can semantic frames in FrameNet be modified to produce semantic-level representations that not only reduce lexical variation but also generalize in a sentiment-aware manner to allow for the discovery of phrasal clues for sentiment analysis?

As described above, since the construction of a phrasal lexicon poses additional challenges on the coverage criterion, frame-based semantic abstraction is proposed as the solution to aid the extraction of useful phrasal sentiment clues from domain data. With the efforts devoted to adapt FrameNet to this sentiment-related task and improve the frame identification process to generate more accurate semantic-level representations, the third question (RQ3) posed in this research is: What are the effects of frame-based semantic abstraction on domain-specific lexicon construction and sentiment analysis?

To answer the three research questions, this research aims to achieve the following objectives:

- Develop a robust frame identification model that achieves state-of-the-art accuracy and decently handles unseen words in domain-specific data (RO1).
- Transform semantic frames in FrameNet to produce sentiment-aware semantic abstractions that retain discriminative sentiment-related information (RO2).
- Establish a domain-specific sentiment lexicon construction process that uses frame-based semantic abstraction to increase the coverage of phrasal lexicon entries, thereby improving the performance of sentiment classification models for domain-specific data (RO3).

1.3 Overview of the Research Process

This research aimed to improve the process of sentiment lexicon construction for phrasal entries. The proposed method focuses on solving the issue of lexicon coverage, which is one of the most crucial criteria in assessing the usefulness of sentiment lexicons. To this end, the

present research introduced a frame-based semantic abstraction method that associated words and phrases in textual contents with semantic frames, thus creating a representation layer from which semantically similar phrases could be easily extracted and added to sentiment lexicons. The overview of the research process is presented in Figure 1.2. As can be seen from the figure, the research process consists of three phases, and each phase serves to answer a research question and fulfills a research objective stated in the previous section. The three phases are described in detail below.

1.3.1 Phase 1

This research intends to exploit the frame-LU (**L**exical **U**nit) association defined in FrameNet for frame-based semantic abstraction. Phase 1 of this research focused on improving the frame identification process, which is an important subtask in frame-semantic parsing. The purpose of frame identification is to disambiguate words or phrases and associate them with the most likely frames based on sentential contexts. Previous works in frame identification (e.g., Das et al., 2014; Hermann et al., 2014; Peng et al., 2018) have tended to focus on improving accuracies on a commonly used benchmark data set (Das & Smith, 2011) without investigating the performance of frame identification models on unseen targets and domain-specific data rigorously. Also, far too little attention has been paid to the deployment of these models in more realistic setups that reflect the deployment scenarios in a real-world problem.

To provide a robust solution to the problem, a frame identification model was developed using deep learning neural networks on top of a state-of-the-art pretrained language representation model called Bidirectional Encoder Representations from

Figure 1.2. Overview of the Research Process

Transformers (BERT) (Devlin et al., 2018). The proposed frame identification model, known as PAFIBERT (Positional Attention-based Frame Identification with BERT), makes use of the positional information and contexts of words in a sentence to select the most suitable frames from FrameNet. Rigorous comparison with previous works was conducted to evaluate the performance of PAFIBERT in various experimental setups and examine its ability in generalizing to unseen targets in domain-specific data. The architecture of PAFIBERT and the experiments conducted to evaluate the model are described in detail in Chapter 3.

1.3.2 Phase 2

Semantic frames in FrameNet can be regarded as cognitive schemas that describe meaningful contexts and specify the arguments—commonly known as frame elements in the frame-semantic terminology—related to the frames. For instance, the [Desirability] frame shown in Figure 1.3 consists of arguments like *Evaluee*, *Parameter*, *Affected_party*, *Circumstances*, and so forth. The complete schematic representation defined in FrameNet for this frame is attached in Appendix A. Another essential component of a frame is its lexical units (LUs). The LUs associated with the same frame tend to share some semantic similarity, and the dimension of the similarity depends on the definition of the frame. In the example given in Figure 1.3, the LUs are similar in that they are words that describe different degrees of desirability. However, as also illustrated by this example, the frame-LU association defined by FrameNet does not take into account the distinction that people often make when expressing their attitude or sentiment. For instance, in a semantic-related sense,

“amazing.a” and “lame.a” are similar because they both describe desirability, but in a sentiment-related sense, they are distinct and should not be grouped into the same category.

Therefore, to benefit from the frame-based conceptualization provided by FrameNet while retaining the sentiment-related distinction necessary for telling apart positive opinions from negative opinions and vice versa, the LUs of each semantic frame were categorized manually based on the graduation concept defined in Martin and White’s (2005) appraisal

Desirability	
This frame concerns an Evaluatee being judged for its quality, i.e. how much it would probably be liked. In many cases, the Evaluatee is implicitly judged good or bad relative to other instances of its type. The Evaluatee's desirability is determined by one or more Parameters, which are scalar properties of the Evaluatee. The evaluation may also explicitly be relativized to a set of Circumstances, a Comparison_set of entities that belong to the same class as the Evaluatee, or an Affected_party. The Degree of goodness or badness may also be expressed. Note: With some targets, desirability is conventionally aligned with quantity, i.e., GOOD is MORE.	
Core Frame Elements	Evaluatee, Parameter
Non-Core Frame Elements	Affected_party, Circumstances, Comparison_set, Degree, Depictive
LU	amazing.a, appalling.a, astonishing.a, astounding.a, atrocious.a, average.a, awesome.a, awful.a, bad.a, best thing since sliced bread.n, cool.a, crap.n, crappy.a, decent.a, desirable.a, dreadful.a, excellence.n, excellent.a, execrable.a, extraordinary.a, fabulous.a, fair.a, fantastic.a, fine.a, first-rate.a, garbage.n, gem.n, gold.n, good.a, great.a, horrible.a, idyllic.a, incredible.a, inferior.a, junk.n, killer.a, lame.a, lamer.n, magnificent.a, marvel.n, marvellous.a, mediocre.a, miserable.a, nasty.a, okay.a, outstanding.a, pathetic.a, pitiful.a, poor.a, popular.a, rock.v, rotten.a, second-rate.a, sensational.a, shit.n, shitty.a, so-so.a, splendid.a, standout.n, stupendous.a, substandard.a, suck.v, super.a, superb.a, superlative.a, sweet.a, terrible.a, terrific.a, third-rate.a, tip-top.a, tolerable.a, top-notch.a, tremendous.a, uncool.a, unfortunate.a, well.adv, wonderful.a, wonderfully.adv, worthless.a
Exemplar Sentence	The view was <u>ASTONISHING</u> . On clear days, the view was <u>EXCELLENT</u> .

Figure 1.3. A simplified schematic view of the [Desirability] frame (FrameNet 1.7)

framework. For instance, the LUs of the [Desirability] frame can be grouped into instances that describe high desirability (e.g., “amazing.a” and “incredible.a”), moderate desirability

(e.g., “average.a” and “so-so.a”), and low desirability (e.g., “nasty.a” and “uncool.a”) based on a gradation type called Infused Intensification. This modification results in a sentiment-aware version of FrameNet known as GradFrameNet in this research. The usefulness of GradFrameNet was examined in the study described in Chapter 4, in which GradFrameNet was used with a semantic tagger to generate semantic-level representations for online reviews collected for three beauty products and one electronic product. Phrasal patterns in the form of collocations were extracted from the semantic-level representations and analyzed for semantic consistency and ability in uncovering important phrasal clues for inferring reviewers’ attitudes.

1.3.3 Phase 3

To achieve sufficient coverage for phrasal lexicon entries in automated domain lexicon construction, the final phase of this research leverages frame-based semantic abstraction as a means to reduce data sparsity and improve the process of finding discriminative phrasal patterns for building domain-specific sentiment lexicons. To this end, GradFrameNet and PAFIBERT were incorporated into the lexicon construction process described in Chapter 5 to generate wide-coverage domain-specific sentiment lexicons. This study also examined the effects of frame-based semantic abstraction on data sparsity degree, lexicon size, and noise level of the lexicons. A series of sentiment classification experiments were conducted to evaluate the effectiveness of the proposed lexicon construction method and investigate the role played by frame-based semantic abstraction in the process. Beauty product reviews were used in the study because reviewers in this domain often express their attitudes by describing

desirable and undesirable usage scenarios that provide plenty of examples for studying the more implicit kind of sentiment expression.

1.4 Significance of Research

It is well-established that combining sentiment lexicon knowledge into classification models tends to show promising improvements in sentiment classification (Bross & Ehrig, 2013; Cambria et al., 2012; Hamilton et al., 2016; Lu et al., 2011; Kiritchenko et al., 2014a; Kiritchenko & Mohammad, 2016a, 2016b). The study conducted by Mohammad, Kiritchenko, and Zhu (2013) has also highlighted the importance of sentiment lexicons through their experiments that compared the effects of various classification features by removing one feature set at a time in the classification process. Among the feature sets compared in the experiments (including n-grams, negation, part-of-speech, emoticons, punctuation, and hashtags, among others), sentiment lexicons were found to produce the most influential features for sentiment classification, to the extent that the removal of such features dropped the F-scores of the classification by more than 8.5%. Given the drastic impact demonstrated by sentiment lexicons, improving the process and outcome of lexicon construction is of considerable significance within the field. The present research contributes to this line of work by establishing a lexicon construction method that collects useful lexicon entries automatically from domain-specific data.

Unlike earlier research in lexicon construction that focused primarily on strongly evaluative or emotion-related terms like “good,” “bad,” “awful,” “excellent,” and so forth (Blair-Goldensohn et al., 2008; Hu & Liu, 2004; Kamps et al., 2004; Rao & Ravichandran, 2009; Wilson et al., 2005), this research emphasizes the importance of phrasal lexicon

entries, which play a vital role in detecting implicit sentiments and indirect opinions (Balahur et al., 2013; Cambria et al., 2012, 2014, 2016, 2018; Noferesti & Shamsfard, 2015; Ofek et al., 2016). By addressing the difficulties in extracting phrasal lexicon entries from text corpora, this research provides an important opportunity to advance our understanding of the complex problem of implicit sentiment analysis.

Furthermore, this research explores frame-based semantic abstraction as a solution to the sparsity problem that hinders the extraction of useful phrasal clues from text data. Word sparsity is a general issue not only in sentiment analysis and opinion mining but also in other text mining tasks (Khan et al., 2014; Liu et al., 2005, 2017; Saif et al., 2012, 2014). Therefore, although this research aims to tackle this sparsity problem in a lexicon construction setting, it will also shed light on the potential of this frame-based solution in dealing with word sparsity in general. Besides, part of the objectives of this research includes the development of GradFrameNet and PAFIBERT for the implementation of frame-based semantic abstraction. On a practical level, this research will benefit the community of researchers and practitioners in sentiment analysis by providing these useful resources or tools.

Finally, the cosmetics domain has been chosen as the primary domain to study the effectiveness of frame-based semantic abstraction and domain-specific lexicon construction method developed in this research, although electronic product reviews were also used in Chapter 4 for cross-domain comparison. Compared to other types of data such as movie reviews, electronic product reviews, and tweets, beauty product reviews have been relatively little explored by the research community of sentiment analysis. Reviewers of beauty products commonly express their sentiments through the sharing of experience and the

description of usage scenarios. Also, since beauty products are consumables, recurring purchases occur depending on customers' overall satisfaction with the products. Therefore, beauty product reviews not only contain information about consumers' positive and negative experiences and overall satisfaction but also provide insights into other consumer behavior like expectation and disconfirmation (Oliver, 1977; Oliver, 1980), retention intention (Ranaweera & Prabhu, 2003; Zeithaml, Berry, & Parasuraman, 1996), and word-of-mouth engagement (Mauri & Minazzi, 2013; Ranaweera & Prabhu, 2003; Zeithaml, Berry, & Parasuraman, 1993; Zeithaml et al., 1996). By exploring the process of mining phrasal sentiment clues from the beauty domain, this research provides an opportunity to explore a relatively new domain and allows researchers to discover previously unseen challenges and unsolved problems in sentiment analysis.

1.5 Outline of the Thesis

This thesis does not have a specific chapter on the research method, but specific methods and tools used in this research will be discussed alongside the three phases of the research, which are presented in Chapters 3, 4, and 5. This approach in structuring the thesis has been taken because each of the three phases focuses on solving a specific problem related to the main goal of the research. Presenting these phases, as well as the related methods and tools, in three self-contained chapters would most likely make the thesis more readable and comprehensible. In accordance with this structuring approach, the rest of the thesis is organized as follows. Chapter 2 summarizes previous works related to the construction of sentiment lexicons. Chapter 3 provides details on the development of PAFIBERT, a model that achieved new state-of-the-art performance on frame identification. Chapter 4 presents a

study in which GradFrameNet was created by incorporating the appraisal concept of graduation into FrameNet. Chapter 4 also describes how GradFrameNet and PAFIBERT were integrated into a frame-based semantic abstraction process for semantic tagging of product reviews and how this process allowed useful phrasal patterns to be extracted from the reviews. Chapter 5 focuses on the description of the fully-automated lexicon construction process established in this research. The effects of frame-based semantic abstraction on the text corpora and the evaluation of the resulted phrasal lexicons are presented in Chapter 5. Finally, the conclusion is given in Chapter 6 with a summary of the thesis, a review of the research objectives, and limitations that need to be examined more closely in future work.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The importance of sentiment lexicons in sentiment analysis is well-established in the literature. This chapter discusses two dominant approaches in sentiment lexicon construction: dictionary-based (or thesaurus-based) approach and corpus-based approach. Dictionary-based lexicon construction techniques usually create sentiment lexicons by exploiting word relations in dictionaries or similar linguistic resources. In contrast, the corpus-based approach utilizes text corpora such as online reviews and blog articles to generate lexicons automatically from textual contents. Also highlighted in this chapter are lexicon construction studies that focused on producing lexicons consisting of phrasal entries, which play a vital role in sentiment analysis because of their abilities to capture contexts, domain knowledge, and implicit sentiments (Agarwal et al., 2015a; Agarwal et al., 2015b; Cambria et al., 2012, 2014, 2016, 2018; Kiritchenko & Mohammad, 2016a, 2016b; Tsai et al., 2013; Wu & Tsai, 2014). Since frame identification is an integral part of the lexicon construction process proposed in this research, in addition to the lexicon construction literature, this chapter also provides a thorough review of previous works related to FrameNet and frame identification research.

2.2 Automatic Lexicon Construction

2.2.1 Dictionary-Based Approach

Constructing sentiment lexicons manually could be time-consuming and labor-intensive. Hence, fully-automated or semi-automated methods are commonly adopted to reduce manual efforts. In a dictionary-based approach, lexicon construction techniques usually follow the steps below (Liu, 2015):

- Prior to the lexicon construction process, an initial list of manually-labeled sentiment terms (also known as seeds) is collected.
- The lexicon construction algorithm exploits the relations between the seeds and their related terms in dictionaries. WordNet (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990) is commonly used for this purpose.
- Starting with a small set of seeds, the algorithm expands the lexicon iteratively, adding newly detected related terms to the pool of lexicon entries until no more entries can be found by the algorithm.
- Since the above steps often introduce some noise into the sentiment lexicon, it is a common practice to clean the lexicon either manually or automatically.

Examples of studies that follow the general workflow described above include those of Hu and Liu (2004), Kamps et al. (2004), and Blair-Goldensohn et al. (2008). Hu and Liu (2004) and Kamp et al. (2004) used word relations in WordNet to predict the sentiment orientations of adjectives. The two studies differed mainly in the way the seed set was expanded to find the lexicon entries. Hu and Liu (2004) predicted the sentiment orientations of lexicon entries by assuming that the synonyms of a term should have the same sentiment orientation as the term and its antonyms should have the opposite sentiment orientation. On

the other hand, Kamps et al. (2004) determined the sentiment orientations of adjectives using a distance-based measurement that computed the distance between two terms as the length of the shortest path connecting the terms. Based on the theory of semantics differentiation (Osgood, 1952), Kamps et al. (2004) predicted the sentiment orientations of lexicon entries by considering their relative distance to two nodes—“good” and “bad”—in WordNet.

In a similar vein, Blair-Goldensohn et al. (2008) proposed a lexicon construction method that also leveraged the synonym and antonym relations in WordNet but with several added constraints that aimed to improve the precision of the lexicon. An improvement introduced by Blair-Goldensohn et al. (2008) was the assignment of confidence scores to lexicon entries to indicate the likelihood of the sentiment orientations. Additionally, they appended part-of-speech (POS) tags to each candidate entry to handle word sense disambiguation for homonyms. The POS tags were taken into account when the sentiment scores were propagated through the synonymous and antonymous links in WordNet. This constraint ensured that the score propagation only affected candidate entries that had the same POS tags. Besides the POS constraint, a predefined list of neutral terms was used in their study to avoid the propagation of sentiment scores through neutral terms. Although dictionary-based techniques usually utilized WordNet to identify candidate entries for sentiment lexicons, similar procedures can be applied to build the lexicons from other linguistic resources. For instance, Rao and Ravichandran (2009) demonstrated the versatility of graph-based learning methods in constructing general-purpose sentiment lexicons from English WordNet, Hindi WordNet, and French OpenOffice thesaurus.

Dictionary-based lexicon construction often results in general-purpose sentiment lexicons, which have important research and practical values because of their reusability for

a wide range of sentiment analysis tasks. One of the commonly used resources for sentiment analysis is SentiWordNet (Baccianella, Esuli, & Sebastiani, 2010; Esuli & Sebastiani, 2005, 2006, 2007). For each synset in WordNet, SentiWordNet provides three scores that describe the level of objectivity, positivity, and negativity of the synset. A key assumption made in the construction of SentiWordNet is that terms with similar glosses (i.e., textual definitions) are likely to share the same sentiment orientations. Based on this assumption, the glosses of terms retrieved from WordNet were collated to generate vector representations, and a term classifier was trained to predict the sentiment orientations of the lexicon entries. Esuli and Sebastiani (2005) leveraged a binary classifier to classify the candidate terms as either positive or negative. The classification task was augmented by Esuli and Sebastiani (2006) with the inclusion of the “objective” category, which encompassed neutral terms such as “white” and “triangular.”

Although a generic sentiment lexicon of superior quality could be a useful resource in sentiment analysis, it is a context-free resource in the sense that the sentiment scores of its lexicon entries are domain-independent. Therefore, generic sentiment lexicons mainly consist of sentiment-bearing entries of which the scores or orientations are less ambiguous and can be determined with minimal uncertainty. When it comes to domain-specific sentiment analysis that often requires domain knowledge for interpretation of opinions, this kind of lexicon tends to suffer the problem of insufficient coverage and lack of specific knowledge. The corpus-based lexicon construction approach tackles this problem by deriving lexicon entries from text corpora.

2.2.2 Corpus-Based Approach

Corpus-based lexicon construction methods usually extract sentiment terms automatically from web data such as online reviews, tweets, and blog articles. Most studies in this line of research have adopted unsupervised or semi-supervised strategies that make use of unlabeled text corpora for automated lexicon construction. Without the information provided by human annotators, the sentiment scores of lexicon entries are usually inferred from various sources of information such as consumer ratings and common website structures in product reviews, hashtags and emoticons in tweets, and so forth.

Hatzivassiloglou and McKeown (1997) were among the first to adopt a corpus-based lexicon construction approach. They postulated that adjectives joined by cumulative conjunctions (e.g., “and” and “or”) were likely to share the same sentiment orientations, whereas adjectives connected by adversative conjunctions (e.g., “but”) were likely to have opposite sentiment orientations. Based on this simple postulation, they demonstrated that a sentiment lexicon could be built from unlabeled text data using only a small seed set and the conjunctions between adjectives as inference rules to expand the seed set. Similarly, Velikovich et al. (2010) examined the viability of building a sentiment lexicon semi-automatically without WordNet and other linguistic resources. Instead, they built a large-scale sentiment lexicon from unlabelled web documents. Taking as input the positive and negative seed sets, their lexicon construction method derived a similarity graph based on statistical co-occurrence information in the web documents. A graph propagation algorithm was used to compute the sentiment scores of candidate entries based on the distance between the candidates and the preselected seeds. In addition to capturing domain-specific knowledge, another advantage of employing the corpus-based approach is the acquisition of

non-standard entries, most of which are not documented in formal dictionaries. As demonstrated by Velikovich et al. (2010), the constructed lexicon also extracted non-standard entries, including commonly misspelled words, multi-word expressions, slang, and word variations.

In addition to the use of seed words, sentiment scores or orientations of lexicon entries can also be inferred from other clues. Kaji and Kitsuregawa (2007), for example, developed a Japanese sentiment lexicon automatically by exploiting multiple types of sentiment-revealing clues in web documents. They assumed that cue words such as “pros” and “advantage” were indicators of positive sentences, whereas words such as “cons” and “disadvantage” signified negativity. These keyword indicators were used in conjunction with website layout structural clues—itemization structure and table structure—to collect opinionated sentences from the web documents. Chi-square and Pointwise Mutual Information (PMI) were used to score the candidate entries extracted from the opinionated sentences based on their co-occurrences with the cue words.

Each type of web data has its unique characteristics that could be useful for inferring sentiment scores of lexicon entries. For example, Feng, Song, Wang, and Yu (2015) assumed that emoticons were indicative of sentiments expressed in tweet messages. They proposed a graph-based mutual reinforcement model to represent the co-occurrence relations between emoticons and candidate words. The emoticons and candidate words formed the vertices of the graph, and the weights of the edges—which were indicative of the strengths of association between the vertices—were calculated using PMI. Lexicon entries were selected based on the ranking generated by a random walk algorithm. Song et al. (2015) extended the work of Feng et al. (2015) by also taking into account the co-occurrence relations between

the candidate words and the preselected seed words. The graph model thus contained three types of vertices: emoticons, seeds, and candidates.

In recent years, there has been a growing trend towards fine-grained sentiment analysis, which aims to classify aspect-specific opinions (Kiritchenko, Zhu, Cherry, & Mohammad, 2014b; Poria, Chaturvedi, Cambria, & Bisio, 2016; Wang, Huang, & Zhao, 2016). This trend has also led to the studies on aspect-specific sentiment lexicon construction. The motivation for creating an aspect-specific lexicon is that the same opinion word may have different sentiment scores when associated with various aspects in the same domain. For instance, in cellphone reviews, “long” may have a positive sentiment orientation when associated with “battery life” but a negative sentiment orientation when describing “reboot time.” The lexicon entries produced by this method are usually in the form of triples that associate sentiment scores or orientations with aspect-opinion pairs (e.g., (ROOM: huge: positive) and (NOISE: high: negative)). Examples of studies that fall into this line of work include those of Bross and Ehrig (2013), Ding, Liu, and Yu (2008), Qiu et al. (2009), and Lu et al. (2011), among many others.

To generate a domain-specific and aspect-aware sentiment lexicon for product reviews, Bross and Ehrig (2013) utilized a predefined product ontology and a string matching algorithm to identify potential product aspects from the product reviews. Opinion words related to the identified aspects were extracted and associated with the aspects. The likelihood ratio test was used to determine the sentiment scores of the aspect-specific lexicon entries. They also increased the lexicon coverage via label propagation through word relations defined in WordNet.

Ding et al. (2008) also aimed to overcome the shortcomings of sentiment lexicons in aspect-level sentiment analysis by generating lexicons consisting of aspect-opinion pairs and their sentiment orientations. To infer the sentiment scores of the lexicon entries, they applied various linguistic rules and leveraged contextual information obtained from multiple sources. With respect to the use of linguistic rules, their study was similar to those of Hatzivassiloglou and McKeown (1997) and Kanayama and Nasukawa (2006) that also leveraged conjunction rules in lexicon construction. However, Ding et al. (2008) extended the rules to handle inter-sentence conjunctions such as “however” that signified the shift of sentiments across neighboring sentences.

Similarly, Qiu et al. (2009) built aspect-specific sentiment lexicons by applying conjunction rules and making inferences from other contextual clues such as overall ratings in product reviews. In product reviews, opinion words often co-occur with the aspects they describe. Hence, Qiu et al. (2009) exploited this co-occurrence relation to extract new opinion words and new aspects in an alternating and iterative manner. Specifically, using a set of seed words and the dependency relations between aspects and their opinion words, Qiu et al. (2009) designed dependency rules to perform the following extraction tasks:

- Extract opinion words using opinion words
- Extract aspects using opinion words
- Extract opinion words using aspects
- Extract aspects using aspects

New opinion words and aspects were detected from this iterative process and used to find other opinion words and aspects in the subsequent iterations. This process continued until no more opinion words and aspects could be found.

Lu et al. (2011) proposed an aspect-specific lexicon construction method that combined different sources of information to infer the sentiment scores of lexicon entries. Their method gathered information from four sources: generic sentiment lexicons, review ratings, synonyms and antonyms from a thesaurus, and linguistic rules such as conjunction and negations rules. Although these sources of information had been used in earlier studies, their work combined numerous sources of information into a unified framework. The biggest challenge in combining these multiple sources was that the signals might contain contradictory information. They treated the conflicts as an optimization problem and demonstrated that the problem could be transformed into an equivalent linear programming problem to be solved efficiently with existing optimization methods.

Although aspect-specific sentiment lexicons could be very useful for aspect-level sentiment analysis, most studies in this line of research have restricted the syntactic categories of aspects to nouns or noun phrases, and have assumed that the aspect-specific opinion words consist of adjectives or adverbs (e.g., Bross & Ehrig, 2013; Lu et al., 2011; Qiu et al., 2009). This assumption may hold in most cases for explicit aspects and opinions. However, implicit aspect detection (e.g., Chen & Chen, 2016; Cruz, Gelbukh, & Sidorov, 2014; Schouten et al., 2015) is much more complex and has not been studied extensively in aspect-specific lexicon construction.

2.2.3 Constructing Phrasal Sentiment Lexicons

As mentioned above, previous works in corpus-based lexicon construction have often assumed that certain syntactic categories such as adjectives and adverbs are more relevant to the expression of sentiment. Hence, many studies have confined the extraction of lexicon

entries to these more relevant syntactic categories. Table 2.1 provides a summarized view of the syntactic categories of opinion words treated in existing lexicon construction studies.

Table 2.1. Syntactic categories of opinion words treated in previous works

Study	Syntactic Categories of Opinion Words
Hatzivassiloglou and McKeown (1997)	adjective
Kamps et al. (2004)	adjective
Blair-Goldensohn et al. (2008)	adjective, adverb, noun, and verb
Qiu et al. (2009)	adjective
Rao and Ravichandran (2009)	adjective, noun, and verb
Lu et al. (2011)	adjective and adverb
Bross and Ehrig (2013)	adjective
Feng et al. (2015)	adjective, verb, noun, and adverb
Song et al. (2015)	adjective, verb, noun, and adverb
Ofek et al. (2016)	adjective

Although it has become an acceptable practice and a logical choice in lexicon construction studies to assume that lexicon entries consist of opinion words under the syntactic types shown in Table 2.1, such assumption has somehow undermined the interrelations between words and overlooked the importance of phrasal units, which play a vital role in capturing contexts, domain knowledge, and implicit sentiments (Agarwal et al., 2015a; Agarwal et al., 2015b; Cambria et al., 2012, 2014, 2016, 2018; Kiritchenko & Mohammad, 2016a, 2016b; Tsai et al., 2013; Wu & Tsai, 2014).

According to Cambria et al. (2012), “[O]pinions and sentiments are expressed implicitly through context and domain-dependent concepts, which make purely syntactical approaches ineffective. To overcome such a problem, we need to use NLP techniques that rely on semantics and sentics, rather than syntactics” (p. 203). To this end, SenticNet

(Cambria et al., 2012, 2014, 2016, 2018) was produced to provide a concept-level resource for a more effective interpretation and processing of subtle and implicit sentiments. The focus of their studies is to create a sentiment knowledge base that provides polarity and other affective information for what they called the “common-sense concepts.” Basically, these common-sense concepts consist mostly of phrasal units that capture subtle sentiment information related to various life scenarios. For instance, “accomplish goal,” “birthday party,” and “celebrate special occasion” are positive concepts, whereas “feel guilty,” “lose temper,” and “shed tear” are negative concepts. Their sentiment knowledge base was built by means of sentic computing—which integrated the use of various techniques in artificial intelligence, semantic web, and data mining, as well as supporting theory and philosophy in linguistics, psychology, and sociology—into a unified framework. Earlier versions of SenticNet were created by assigning sentiment and other emotion-related scores to concepts obtained from ConceptNet (Speer & Havasi, 2012), a semantic network of common-sense knowledge. SenticNet 3.0 (Cambria et al., 2014) integrated other resources such as DBPedia (Bizer et al., 2009) and Probase (Wu, Li, Wang, & Zhu, 2012) to build a more comprehensive resource.

Other studies that explored concept-level lexicon construction stemmed mainly from the work of Cambria and his colleagues. The effectiveness of concept-level lexicon was also exemplified in a study undertaken by Tsai et al. (2013), in which a method was proposed to propagate sentiment scores through ConceptNet to build a concept-level sentiment lexicon. The first step of the method was an iterative step that predicted the initial value of each concept on ConceptNet using a regression model. They combined features such as relation types and relation directions in ConceptNet, as well as prior scores obtained from two other resources—SenticNet and ANEW (Affective Norms for English Words) (Bradley & Lang,

1999) —to generate initial values for concepts. In the second step, the predicted values were propagated through the links in ConceptNet using a random-walk technique. In the extended work proposed by Wu and Tsai (2014), relation filtering and bias correction were integrated into the score propagation step. Specifically, Tsai et al. (2013) found that some relations in ConceptNet were non-beneficial to sentiment score propagation. Therefore, Wu and Tsai (2014) limited the score propagation to only sentiment-relevant links in ConceptNet. The best combinations of relations were determined using a sequential forward search by removing relation types one at a time. Another issue observed by Tsai et al. (2013) was the bias introduced in the random-walk step that caused values to be shifted and wrong sentiment scores to be assigned in the iterative process. Improvement was introduced by Wu and Tsai (2014) to correct this bias. Other studies along the same lines have been undertaken by Rajagopal, Cambria, and Olsher, and Kwok (2013) and Agarwal et al. (2015a, 2015b).

An additional challenge in the acquisition and utilization of phrasal lexicon entries is that phrases must first be extracted from texts. In the case of a concept-based lexicon, the reliance on concept extraction tools to deconstruct sentences into concepts also contributes to the complexity of the problem. As far as this research is concerned, a highly reliable solution for concept extraction has yet to exist. Particularly challenging is the extraction of concepts from utterances that convey similar meanings but are lexically varied because of word choices, word forms, word orders, and so forth. Table 2.2 provides several examples of concepts extracted by a concept parsing tool (Poria, Agarwal, Gelbukh, Hussain, & Howard, 2014).

Table 2.2. Concepts extracted by a concept parser (Poria et al., 2014)

Sentence	Extracted Concepts
We won't purchase this product again.	purchase product, product
No, I won't buy this product ever again.	buy product, product
This product is not something I'll buy again.	product, something product, buy something

As can be seen from the table, given three sentences with similar semantics, the concept parsing tool returned different concepts. This drawback in the concept parsing process complicates the use of concept-based lexicons because the extracted concepts may contain too many variations, thus aggravating the word sparsity problem. As a result, it is difficult to achieve sufficient coverage with concept-based sentiment lexicons. This issue of insufficient coverage has also been pointed out and partially addressed by Cambria et al. (2016) and Cambria et al. (2018). To increase the generalizability of common sense concepts, Cambria et al. (2016) used hierarchical clustering and dimensionality reduction techniques to link (verb+noun) concepts in SenticNet to what they called the conceptual primitives. To further extend the comprehensiveness of SenticNet, Cambria et al. (2018) used recurrent neural networks to find related concepts for (verb+noun) and (adjective+noun) concepts retrieved from ConceptNet. The present research also stemmed from the similar idea that lexical-level representations can be generalized to the more abstract semantic-level representations to achieve broader coverage for phrasal lexicon entries. However, instead of relying on statistical techniques, this research seeks to explore the potential of a linguistically-motivated approach that performs generalization based on semantic categories suggested by frame semantics theory (Fillmore, 1976). Another key difference between this research and Cambria et al.'s work is that SenticNet is a general lexicon, which was not constructed using a corpus-based approach, it thus does not focus on

capturing domain-specific knowledge. Although Ofek et al. (2016) proposed a method to enrich SenticNet with domain-specific concepts, their method was limited to adjective opinion words.

Besides the concept-based approach, there are other ways to deconstruct texts into phrasal units. A simple phrase extraction approach that has been extensively used in NLP tasks is the n -gram approach, which relies on the sequential order of words to extract sequences of n tokens from sentences. Mohammad et al. (2013) used hashtags and emoticons to produce pseudo-labeled tweets and built n -gram sentiment lexicons automatically for sentiment analysis of tweets and SMS. Their lexicons included not only unigrams and bigrams, but also contiguous and non-contiguous pairs of unigram-unigram, unigram-bigram, and bigram-bigram entries. By combining the features generated from the n -gram lexicons with other features to train an SVM model, Mohammad et al. (2013) created a classifier that achieved the best performance in a sentiment analysis shared task in SemEval-2013. Their ablation analysis of the classifier’s features showed that lexicon-related features contributed substantially to the classifier’s performance. Kiritchenko et al. (2014a) also adopted the same approach but further improved the lexicon construction process with special handling of negated contexts. Along the same lines of research, Kiritchenko and Mohammad (2016a) created a sentiment composition lexicon that consists of phrases containing negators, degree adverbs, and modals. Another composition lexicon was generated for opposing polarity phrases such as “crazy talented” and “great loss” that are composed of words with opposing polarities (Kiritchenko & Mohammad, 2016b). However, the two phrasal lexicons provided by Kiritchenko and Mohammad (2016a, 2016b) were created via manual or semi-manual processes instead of learned automatically from text corpora.

Another common approach for extracting phrasal units from sentences is by exploiting the syntactic relations or dependencies between words. Syntactic relations can be obtained via syntactic or dependency parsing. Examples of studies that adopted this approach include those of Kaji and Kitsuregawa (2007) and Kanayama and Nasukawa (2006). Kaji and Kitsuregawa used dependency parsing to extract adjective phrases of the form (noun + postpositional particle + adjective) as candidate phrases for lexicon construction. Likewise, Kanayama and Nasukawa applied two types of syntactic patterns, i.e., modality patterns and conjunctive patterns, to the structures of syntactic trees for the detection of candidate clauses in lexicon construction.

While previous works in this line of research have presented some plausible solutions and demonstrated the usefulness of phrasal sentiment lexicons, the fundamental problems of word sparsity and insufficient lexicon coverage remain. The multitude of phrasal expressions that could possibly be formed by different combinations of words renders the construction of a wide-coverage phrasal sentiment lexicon extremely difficult.

2.3 FrameNet and Frame Identification

FrameNet has been identified as the linguistic resource suitable for semantic abstraction in this research. As a well-known linguistic resource, FrameNet is often discussed and considered alongside two other popular resources, i.e., WordNet (Miller et al., 1990) and PropBank (Palmer, Gildea, & Kingsbury, 2005). WordNet is a lexical database that organizes words and phrases into lexical categories called synsets. Synsets may consist of only one lemma or multiple synonymous lemmas. For example, Synset(buy.v.01) consists of two synonymous words: “buy” and “purchase.” Synsets are interconnected by various

semantic relations such as hypernym, hyponym, meronym, holonym, and so on. For instance, Synset(get.v.01) is the hypernym of Synset(buy.v.01). WordNet version 3.0 contains 155,287 unique words and a total of 206,941 word-sense pairs, organized into 117,659 synsets (wordnet.princeton.edu/documentation/wnstats7wn).

The second resource frequently compared to FrameNet is PropBank (Palmer et al., 2005), which provides predicate-argument information on top of Penn Treebank's phrase structure annotation (Marcus, Santorini, & Marcinkiewicz, 1993). PropBank has known to be a widely used resource in semantic role labeling (SRL). Unlike frame-based SRL, which is frame-oriented, Propbank-based SRL is verb-oriented, i.e., it identifies verbs as predicates and marks the arguments or semantic roles related to the verb predicates. The annotation of FrameNet was guided by frame semantics theory (Fillmore, 1976; Fillmore et al., 2003), but PropBank's annotation is theory-neutral because the goal of the project is to develop a large-scale annotated corpus to provide consistent and generic labeling (e.g., Arg0, Arg1, etc.) of semantic roles. Detailed comparisons between PropBank and FrameNet can be found in Kshirsagar et al. (2015) and Palmer, Gildea, and Xue (2010).

For semantic abstraction in the present research, FrameNet has been chosen over other potential resources because it not only provides semantically-sound categorization for lexical items but also does not limit its entries to certain syntactic categories. In contrast, PropBank provides annotations only for verb predicates, and the annotation process did not place emphasis on defining any semantic categories for its verb entries (Palmer et al., 2010). Besides, unlike the categorization approach employed in WordNet to group words or phrases into synsets, the approach adopted by FrameNet extends beyond synonymous relations such that the LUs grouped under the same frame are not necessarily synonymous, nor do they

have to share the same syntactic categories. The frame-based categorization approach enriches word relations with new similarity dimensions so that words or phrases are not only perceived as similar when they are synonymous but also when they share the same semantic criteria defined by a frame.

However, the implementation of frame-based semantic abstraction inevitably entails the undertaking of creating a robust frame identification tool, which is crucial to accurately associate lexical items with the right semantic frames in FrameNet. In particular, to work well in real-world applications, frame identification models should be able to make sensible predictions for both seen and unseen targets. The solutions proposed in existing works, as well as challenges and gaps in frame identification, are discussed in more detail in the next subsection.

FrameNet has generated wide interest in other lines of research (e.g., Ofoghi, Yearwood, & Ma, 2009; Sikos & Padó, 2018), but the field of sentiment analysis has hardly benefited from this resource because the categorization in FrameNet is not sentiment-aware. Motivated by the need for more in-depth sentiment analysis, Ruppenhofer and Rehbein (2012) proposed SentiFrameNet as an extension to FrameNet by identifying the multi-dimensional opinions, event structures, and presuppositions associated with each FrameNet entry. Although I agree with the researchers' view that linguistically motivated approaches are worth exploring in sentiment analysis, the complex relations and nuances that they tried to encode in SentiFrameNet render the use of this resource arduous, especially in automated sentiment analysis. Other studies that applied FrameNet to sentiment-related tasks include those of Kim and Hovy (2004) and Sharma, Gupta, Agarwal, and Bhattacharyya (2015). The former used FrameNet to identify opinion holders and sentiment topics in news articles; the

latter utilized FrameNet to find polar adjectives that could be useful for differentiating sentiment intensity. Although these studies have shown promising results, the potential of FrameNet as a resource that provides immediate access to salient linguistic categories has not been fully explored. Since the present research aimed to apply frame-based semantic abstraction to the problem of sentiment analysis, the incorporation of sentiment-related information into FrameNet's frames is essential to ensure the relevance of these frames to sentiment analysis. Therefore, part of the objectives of this research is to complement semantic frames with useful information for distinguishing different sentiments, thereby extending the potential of FrameNet to sentiment analysis. More descriptions about FrameNet's components and the procedure conducted in the present research to transform FrameNet into a sentiment-aware resource are presented in Chapter 4.

Frame-semantic parsing refers to the SRL process in which sentences are tagged with semantic frames and roles based on the lexical information provided by FrameNet. This parsing process involves two major tasks: frame identification and semantic role assignment. The present research focuses only on the first task, which identifies the most appropriate semantic frame for a marked lexical predicate (also known as "target") in a given sentence. Frame-semantic parsing has started to attract more interest after the introduction of the SemEval 2007 shared task on frame semantic structure extraction, which used FrameNet 1.3 for the subtasks of identifying frames and assigning semantic roles (Baker, Ellsworth, & Erk, 2007). The best performing system in SemEval 2007 was presented by Johansson and Nugues (2007). For the frame identification subtask, they employed Support Vector Machines (SVMs) to learn a collection of classifiers. The features they used for training the classifiers included lemmas and word forms of the targets, as well as features generated from dependency parsing such as dependencies of the targets, parents of the targets, and so forth.

The system proposed by Johansson and Nugues was then outperformed by an earlier version of SEMAFOR (Das, Schneider, Chen, & Smith, 2010), a frame-semantic parser which used a conditional log-linear model for frame identification. One of the challenges in frame identification is the handling of unseen targets, i.e., targets that are not found in FrameNet lexicon and the hand-annotated data released along with FrameNet. While Johansson and Nugues (2007) tried to mitigate this issue by expanding the list of frame-evoking words with the help of WordNet synsets, Das et al. (2010) proposed an approach that involved the use of a latent variable to incorporate lexical-semantic relations into their model. These lexical-semantic features allowed the model to make predictions for unseen targets based on their relations with known LUs. Further improvement was suggested by Das and Smith (2011) to address the problem of unseen targets. They adopted a semi-supervised learning approach to build a graph from a large amount of unlabeled data and FrameNet annotated data. The graph vertices corresponded to targets, whereas the edges represented frame similarity and distributional similarity. For the known targets (a subset of the graph vertices), their weights were initialized based on the distribution of frames available in FrameNet annotated data. Through label propagation, the frames (i.e., labels) were transferred to unlabeled examples, which were vertices representing unknown targets in the graph. The constructed graph was then used to suggest candidate frames for unseen targets.

Earlier work on frame identification mainly focused on feature-based approaches that made use of hand-engineered features. Since the popularization of distributed representations (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013b), more recent work has gradually shifted toward the use of word embedding or low-dimensional vector representations. This approach not only reduces the need for expensive feature engineering

of semantic structure but also leads to improved accuracy due to better generalization (FitzGerald, Täckström, Ganchev, & Das, 2015). Hermann et al. (2014) were among the earliest to leverage distributed representation to represent targets and their syntactic contexts as inputs for frame identification. Given a target in a sentence, Hermann et al.’s (2014) technique extracted the syntactic contexts of the target and generated an initial high-dimensional vector space using a set of general word embeddings learned from a large corpus. The WSABIE algorithm (Weston, Bengio, & Usunier, 2011) was then used to determine—from the FrameNet labeled data—a transformation function that mapped the high-dimensional vectors to low-dimensional representations. Embeddings were also learned for frame labels in this low-dimensional space such that for each target, the embedding of its matching frame was the nearest in the vector space compared to those of the competing frames. At inference time, after a target and its context were projected using the learned function to the low dimensional space, the nearest frame label was chosen as the prediction. The accuracy achieved by Hermann et al. (2014) was highly promising, surpassing the results of previous studies by a large margin. However, for the generation of the initial high-dimensional vectors, their technique still involved a great deal of feature engineering to extract features through syntactic parsing.

In light of Hermann et al.’s (2014) work, Hartmann et al. (2017) experimented with pretrained word embeddings to represent targets and their contexts as inputs for learning two classification models: a two-layer neural network and a WSABIE model as proposed by Hermann et al. (2014). Although their models did not outperform Hermann et al.’s (2014) system on Das and Smith’s (2011) benchmark test set, the results reported on out-of-domain data showed that their models generalized better to other domains.

Following the success of deep learning in many natural language processing tasks, researchers in frame-semantic parsing have also explored different types of neural networks for frame identification. Swayamdipta, Thomson, Dyer, and Smith (2017) combined pretrained word embedding, learned word embedding, and part-of-speech (POS) embedding for frame identification using a long short-term memory (LSTM) neural network and achieved an accuracy similar to that of Hartmann et al. (2017). Yang and Mitchell (2017) examined the effectiveness of a multi-layer neural network for frame identification and found that the model’s overall performance was competitive, and it outperformed other models in predicting frames for ambiguous LUs (i.e., LUs associated with multiple frames).

Peng et al. (2018) adopted a joint decoding approach that could make use of disjoint data to improve model performance in a multi-task scenario. This approach allowed the sharing of information across tasks to benefit all the tasks. For the frame-semantic parsing problem, which consists of the frame identification and semantic role assignment subtasks, the joint decoding approach learned model parameters jointly for both tasks by optimizing a multi-task objective. Their models showed improvement over previous works, achieving a new state-of-the-art accuracy of 90% for frame identification on FrameNet 1.5 and 89.1% on FrameNet 1.7.

Although extensive research has been carried out in this area, most studies have focused on improving accuracies on the benchmark test set (Das & Smith, 2011) without investigating rigorously how well frame identification models perform on unseen targets and domain-specific data. Since the present research intends to use frame identification for tagging product reviews, which contain tons of domain-specific words and jargons that are not listed in FrameNet, a generalizable model that can handle these cases is essential.

Therefore, this research set out to develop a model that not only outperforms other models on the benchmark data but also works equally well under realistic deployment scenarios.

2.4 Summary

In this chapter, the two dominant approaches for automatic lexicon construction—dictionary-based and corpus-based—were discussed. While the dictionary-based approach builds lexicons by exploiting the relations between words in dictionaries, the corpus-based approach derives lexicon entries from textual data. This chapter has identified several gaps in previous works. In particular, most studies have tended to focus on lexicon entries pertaining to certain syntactic categories such as adjectives and adverbs. Because of this common assumption, little attention has been paid to the construction of phrasal sentiment lexicons, especially in a corpus-based setting that aims to extract phrasal entries automatically from text corpora. As discussed in Section 2.2.3, corpus-based phrasal lexicon construction often suffers the problems of word sparsity and insufficient lexicon coverage. The present research thus set out to tackle these major issues with detail and to deal specifically with the difficulty of extracting phrasal patterns from sparse textual corpora. To this end, a lexicon construction process was proposed. As will be presented in the next three chapters, this process involves the modification of FrameNet to produce semantic frames that are suitable for sentiment analysis and the development of a robust model for frame identification. The current chapter also provided an in-depth discussion of existing frame identification solutions.

CHAPTER 3

FRAME IDENTIFICATION

3.1 Introduction

To address RQ1 and RO1, this chapter presents a frame identification model called **Positional Attention-based Frame Identification with BERT (PAFIBERT)**. This frame identifier was developed using deep learning neural networks on top of the state-of-the-art Bidirectional Encoder Representations from Transformers (BERT) language representation model (Devlin et al., 2018). Previous works on frame identification have been mostly limited to model evaluation on the benchmark test set suggested by Das and Smith (2011). Also, little attention has been paid to evaluate these models with more realistic experimental setups that reflect the performance of the models in a real-world problem. Since this research intends to use the frame identifier for tagging product reviews, the proposed model was also evaluated qualitatively on product reviews using an experimental setup that reflected accurately the intended deployment scenarios in this research. PAFIBERT not only outperformed existing models but also generalized well on domain-specific data and handled unseen targets decently.

3.2 Background

The use of FrameNet in natural language processing (NLP) tasks like question answering (Shen & Lapata, 2007) and textual entailment (Aharon, Szpektor, & Dagan, 2010) cannot be isolated from semantic role labeling (SRL). SRL is an automatic semantic parsing task in which “for each predicate in a sentence, the goal is to identify all constituents that fill a semantic role, and to determine their roles (Agent, Patient, Instrument, etc.) and their adjuncts (Locative, Temporal, Manner, etc.)” (p. 257, Punyakanok, Roth, & Yih, 2008). By detecting key events described in a given sentence and identifying participants and their roles in the events, SRL allows for semantic interpretation of the sentence to answer the question “Who did What to Whom, and How, When and Where?” (p. 1, Palmer et al., 2010). The development of most SRL tools relies on syntactic parsing, machine learning, and manually labeled data or lexical resources that provide hand-labeled annotations for predicates and their arguments. Recent years have seen increasingly rapid advances in SRL due to the continuing efforts devoted to the development and maintenance of high-quality resources such as PropBank (Palmer et al., 2005) and FrameNet (Baker et al., 1998).

SRL using FrameNet is also known as frame-semantic parsing. It involves tagging sentences with semantic frames and roles based on the lexical information provided by FrameNet. The majority of approaches to frame-semantic parsing has focused on the following subtasks:

- Frame identification. This task identifies the most likely frame for a marked lexical predicate, often referred to as “target,” which can be any word or phrase occurring in a sentence. Because of homonymy and polysemy, a target may be associated with

multiple semantic frames. Hence, this frame identification task is similar to a word sense disambiguation task that disambiguates the target based on sentential contexts.

- Semantic role assignment. Given the target and its identified frame, the second subtask aims to associate constituents of a sentence with semantic roles, which are defined as frame elements in FrameNet.

An example of frame-semantic parsing for the sentence “Coke raised the price of their soda by 2%.” is shown below. The [Cause_change_of_position_on_a_scale] frame is evoked by the word “raised,” along with four semantic roles (i.e., Agent, Attribute, Item, and Difference) associated with the frame:

(Coke)_{Agent} raised [Cause_change_of_position_on_a_scale] (the price)_{Attribute} (of their soda)_{Item} (by 2%)_{Difference}.

This research intends to exploit the frame-LU relations defined in FrameNet for frame-based semantic abstraction. An essential prerequisite for this task is the availability of a tool that processes text inputs and maps them to the most likely semantic frames in FrameNet. As described above, the task of associating a target word with a frame is traditionally known as frame identification in the field of frame-semantic parsing. The present study only deals with the first subtask of frame-semantic parsing since the frame elements or semantic roles in FrameNet are not essential for frame-based semantic abstraction. Apparently, this first subtask plays a vital role in frame-semantic parsing because accurate frame identification is a crucial precondition for the subsequent step of semantic role assignment. It limits the search scope for the most suitable semantic roles from the hundreds of roles defined in FrameNet to at most 30 potential roles per frame (Hartmann et al., 2017). Hartmann et al. (2017) evaluated a widely used parser called SEMAFOR (Das

et al., 2014) on a benchmark dataset (Das & Smith, 2011) and showed that errors in frame identification could cause drastic degradation in the accuracy of semantic role assignment—from 55.20% accuracy when semantic roles were assigned based on gold frames to 30.08% when assigned based on predicted (and therefore error-prone) frames.

3.3 The Proposed Frame Identification Model

This research proposes an attention-based neural network model called **Positional Attention-based Frame Identification with BERT** (PAFIBERT) as a solution to the frame identification problem. In NLP, the use of attention mechanism has its root in machine translation (Bahdanau, Cho, & Bengio, 2014). It has since become an essential component of neural network models for many other NLP problems such as text summarization (e.g., Nallapati, Zhou, Gulcehre, & Xiang, 2016; Rush, Chopra, & Weston, 2015), question answering (e.g., Xiong, Merity, & Socher, 2016; Yang, Ai, Guo, & Croft, 2016), and aspect-based sentiment analysis (e.g., Ma, Li, Zhang, & Wang, 2017; Wang et al., 2016). In general, the attention mechanism in deep learning leverages a vector consisting of weights that represent the importance of other elements in the model. The PAFIBERT model in this study makes use of the positional information of a target to attend to parts of the given sentence that are most useful for disambiguating the target. The proposed model was built on top of the state-of-the-art Bidirectional Encoder Representations from Transformers (BERT) language representation model (Devlin et al., 2018). The following subsection provides background information about BERT.

3.3.1 Bidirectional Encoder Representations from Transformers (BERT)

There are various ways to convert textual data into numerical representations that are processable by machine learning algorithms. Local Representations in the form of sparse, binary vectors can be obtained easily through one-hot encoding, but this kind of representation suffers several problems such as the curse of dimensionality and lack of semantics (Bengio & Bengio, 1999; Zhang, Du, & Wang, 2016). To overcome the limitations of local representations, research in distributional semantics (e.g., Bruni, Tran, & Baroni, 2014; Mikolov, Chen, Corrado, & Dean, 2013a; Mikolov et al., 2013b; Mitchell & Lapata, 2010) aims to produce numerical representations for linguistic items based on their distributions in textual data. This line of research has led to the popularization of word embedding, which has become a standard representation technique in NLP. In contrast to the local representations generated from one-hot encoding, distributed representations are low-dimensional vectors of real numbers that are capable of capturing the semantic and syntactic relations between linguistic items (Mikolov et al., 2013a, 2013b). Therefore, developing methods that can generate high-quality word embeddings efficiently is of significant interest to the research community. In particular, generating pretrained word embeddings has been an active research area (e.g., Mikolov, Grave, Bojanowski, Puhersch, & Joulin, 2017; Pennington, Socher, & Manning, 2014). Pretrained word embeddings can be obtained through unsupervised learning on unlabeled data. Since the pretraining process is task-agnostic and domain-independent, the resulted word vectors are applicable to a wide range of downstream tasks.

Although pretrained word vectors obtained via algorithms such as GloVe (Pennington et al., 2014) and Word2Vec (Mikolov et al., 2013a) have brought NLP a long

way, this type of pretraining approach is considered shallow because the word vectors usually serve as input features to initialize the first layer of neural network models, but the rest of the models are trained from scratch using task-specific data (Howard & Ruder, 2018). There have been some attempts to incorporate pretrained embeddings into other layers of neural networks. However, since the embeddings are still treated as fixed weights in those models, their effectiveness and usefulness are somewhat limited (Howard & Ruder, 2018).

Outside of NLP, transfer learning in computer vision research has shown revolutionary improvements in image classification task (He et al., 2016; Krizhevsky, Sutskever, & Hinton, 2012; Simonyan & Zisserman, 2014). It is thus not surprising that transfer learning has also attracted tremendous research interest among the NLP community in recent years. Transfer learning usually involves fine-tuning a pretrained model, i.e., a model that was trained in an unsupervised setting on a massive unlabeled dataset. Instead of learning from scratch, the learning process for a downstream task can then benefit from the weights of the pretrained models and only needs to fine-tune the weights to adapt to the specific task, usually with supervised learning on a relatively smaller set of labeled data.

Compared to pretrained word embeddings that only serve as input features to a neural network model, employing a pretrained model with the fine-tuning approach allows all model weights—including weights in the pretrained layers and the task-specific layers—to be jointly updated for a downstream task. Recent advances in NLP have obtained convincing empirical proof that supports the shift of paradigm from shallow to deep pretraining (Devlin et al., 2018; Radford, Narasimhan, Salimans, & Sutskever, 2018; Howard & Ruder, 2018). New state-of-the-art results have been achieved via pretrained language models such as Universal Language Model Fine-tuning (ULMFiT) (Howard & Ruder, 2018), OpenAI’s

Generative Pretrained Transformer (Radford et al., 2018), and Google’s Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2018) on a wide range of NLP tasks.

BERT was introduced by Devlin et al. (2018) to overcome some limitations in the implementation of language models. They argued that the unidirectional architecture of language models like OpenAI’s GTP (Radford et al., 2018) could be suboptimal. BERT is a deep bidirectional model that was trained using two novel pretraining tasks that allowed the language model to capture: (1) contextual information of tokens from both left-to-right and right-to-left directions and (2) relationship between sentences.

Although the concept underlying the development of BERT is simple, it has been empirically proven as more potent than other pretrained language models, rewriting the records of state-of-the-art results in eleven NLP tasks on benchmark datasets. The new achievements were overwhelming, including more than 15% absolute improvements on a linguistic acceptability task and a textual entailment recognition task when compared to OpenAI’s GPT. The model architecture used by BERT is a multi-layer transformer encoder, implemented based on the original transformer architecture proposed by Vaswani et al. (2017).

3.3.2 Positional Attention-based Frame Identification with BERT (PAFIBERT)

This section defines the frame identification problem in detail and describes the proposed solution. Several examples of FrameNet’s annotations for frame identification are given in Table 3.1. This example is revealing in several ways: First, it shows that a target is marked

by indices that indicate its position in the sentence; second, a target may consist of non-contiguous stretches of texts; third, the same word or phrase occurring at different positions in a sentence may be associated with different frames.

Table 3.1. Examples of FrameNet’s annotations for frame identification. The target positions mark the beginning and ending characters of a target.

Sentence	Target Positions	Lexical Unit (LU)	Frame
Because these tales were not - as far as we know - collected in the Renaissance and because , like ballads and songs , they must have been subject to endless variation and mutual interference , we can not now know the Pedro de Urdemalas who endeared himself to the folk " (218) .	(44, 47)	know.v	[Awareness]
	(209, 212)	know.v	[Familiarity]
A half - hour tour costs HK\$50 ; pay at the end , or the driver may cut your trip short .	(68, 70), (82, 86)	cut short.v	[Change_event_duration]

Given a sentence s and a set of indices $x = \{(a_1, b_1), (a_2, b_2), \dots\}$ indicating the position of a target, the goal of frame identification is to determine the most suitable frame for the target based on its usage in s . Two variants of PAFIBERT—without frame filtering and with frame filtering—are presented in this study. The two neural network designs are depicted in Figure 3.1 and 3.2. The key difference between these two variants will be discussed in the next subsection, but in general, both variants of PAFIBERT need two inputs: sentence s and a target indicated by its position indices x . The model will then identify a frame for each target of the sentence one target at a time.

Figure 3.1. PAFIBERT without frame filtering

Figure 3.2. PAFIBERT with frame filtering

The input sequence generated from s is fed through the transformer layers of BERT to produce H , an $(n \times d)$ matrix of hidden states where n is the input sequence length, and d is the hidden dimension. In the representation of input sequences, BERT uses special tokens such as [CLS] and [SEP] to indicate the start and end of sequences. To inform the model about a given target, which is required for the frame identification problem, this research proposes a network architecture that uses the target’s position for position-based attention. The attention mechanism transforms the hidden states H produced by BERT into a pooled attentional hidden state $\tilde{h}$, which is a vector that captures information about the input sentence and the target. Using the default configuration in BERT_{BASE}, the hidden dimension is set to 768. As suggested by Luong, Pham, and Manning (2015), the attentional hidden state can be computed as:

$$\tilde{h} = \tanh(W_C[c; t]) \quad (1)$$

where the context vector c and target vector t are combined via concatenation and passed through a neural network layer. W_C is the weight matrix of the layer, and the context vector c is a d -dimensional vector derived as follows:

$$c = H^T \alpha \quad (2)$$

In Equation (2), the context vector c captures the relations between the hidden states and the n -dimensional alignment vector α , which defines how much each hidden state contributes to the prediction of an output. Various alternatives exist for computing the alignment vector α in the above equation. This study adopted a local attention mechanism (Luong et al., 2015) that restricted the attention to only a subset of the input hidden states. Compared to a global attention mechanism that computes the alignment scores for all hidden

states of the input sequence, a local attention mechanism is less expensive in terms of computational costs. The intuition of using a local attention mechanism for the frame identification problem is that, given a target in a lengthy sentence, the model may not need the whole sentence to predict the frame of the target correctly. Instead, the context words adjacent to the target are likely to be sufficient for this task. Hence, the possible scope for localized attention was limited to a fixed window of w hidden states before and after the target positions. In other words, the model only computes the alignment scores for hidden states within the range of $[\beta_1, \beta_2]$ where:

$$\beta_1 = \begin{cases} p_{start} - w & (p_{start} - w) > 1 \\ 1 & (p_{start} - w) \leq 1 \end{cases} \quad (3)$$

$$\beta_2 = \begin{cases} p_{end} + w & (p_{end} + w) < n \\ n & (p_{end} + w) \geq n \end{cases} \quad (4)$$

In the above formulas, p_{start} and p_{end} denote the smallest and largest target positions ($p_{start} = p_{end}$ for single-word targets). By observing the typical scope needed to infer the frame of a given target in FrameNet’s hand-annotated sentences, the size of the fixed window was selected as $w = 10$ in this study. For each input hidden state at position i (i.e., for the i^{th} input token) within the range of $[\beta_1, \beta_2]$, its alignment score (α_i) is calculated as follows:

$$\alpha_i = \frac{\exp(h_i t)}{\sum_{j=\beta_1}^{\beta_2} \exp(h_j t)} \quad (5)$$

Note that Equation (5) describes a softmax function such that α represents a probability distribution that defines the importance of each input hidden state. The alignment scores for input hidden states outside the fixed window are set to zeros. The derivation of t , the target vector, in Equation (1) and (5) is very much problem-specific and varies across

different attention-based models. In the proposed model, t is extracted from the input hidden states at the target positions. To this end, target positions were represented using an n -dimensional position vector (p), which consists of binary values indicating whether a particular position of the input sequence corresponds to the target positions (1 for a target position, 0 for a non-target position). The d -dimensional target vector is then derived from the input hidden states based on the target positions:

$$t = H^T p \quad (6)$$

3.3.3 Frame Filtering

Hartmann et al. (2017) pointed out that most frame-semantic parsing studies have adopted the standard experimental setup in which FrameNet lexicon was used to list a set of frames as candidates for frame identification. Under this setup, if a target is unambiguous—it has only one candidate frame—the frame is selected directly. On the other hand, if the target is ambiguous, this frame filtering step limits the search to a small list of candidate frames instead of all frames in FrameNet. This setup makes strong assumptions about the comprehensiveness of FrameNet: It assumes that FrameNet specifies every possible sense for the target and that the list of LUs under each frame is exhaustive. These assumptions may be acceptable when applied to the development data released by FrameNet, but they certainly do not hold for real-world application data. Therefore, following the suggestion of Hartmann et al. (2017), this study experimented with two model architectures: with frame filtering and without frame filtering.

In the no-filtering model, the attentional hidden state ($\tilde{h}$) obtained from Equation (1) is simply fed through a softmax layer to generate the final predictive distribution y using Equation (7). Each element of y indicates the probability of a frame given a sentence s and a target's position indices x . W_s is the weight matrix of the softmax layer.

$$y = \text{softmax}(W_s \tilde{h}) \quad (7)$$

In the model with frame filtering, the filtering step is added before the softmax layer so that the probability values for the non-candidate frames are set to zeros, whereas the probability values for the candidate frames sum up to 1. This filtering step generates the filtered prediction vector o , which is a k -dimensional vector calculated using the formula below:

$$o = W_k \tilde{h} \odot v \quad (8)$$

In the above equation, $\tilde{h}$ is passed through a fully-connected layer consisting of k nodes, where k corresponds to the number of frames in FrameNet. W_k is the weight matrix of the fully-connected layer. Element-wise multiplications are then performed between the output of the fully-connected layer and vector v , which is a k -dimensional vector consisting of binary values that specify whether a particular frame is a candidate frame. For unseen targets, this step assumes that every frame in FrameNet is a potential candidate (i.e., all elements in v have value 1). Because of the element-wise multiplications, only the candidate frames will have non-zero probability values. Finally, the model produces the k -dimensional frame probability distribution as follows:

$$y = \text{softmax}(W_s o) \quad (9)$$

3.3.4 Listing of Candidate Frames

As shown in Table 3.1, FrameNet’s annotations associate each target with the gold LU and frame. Therefore, for frame identification models with frame filtering, existing studies including that of Hermann et al. (2014), Swayamdipta et al. (2017), and Peng et al. (2018) have usually assumed the gold LUs are known. In other words, the models could use the gold LUs to retrieve a list of candidate frames from FrameNet.

Apparently, this simplified setup does not reflect a real-world deployment scenario, in which the gold LUs should be unknown and the model is expected to predict based on the given target itself, which may occur in any word forms (e.g., singular or plural form, past tense or present tense, comparative or superlative form, and so forth). When applying frame filtering to the proposed model, this study experimented with the following two approaches:

- *Frame filtering by LUs.* This is the approach described above, in which the model is allowed to use the gold LUs to retrieve candidate frames for frame filtering.
- *Frame filtering by targets.* This approach assumes that the gold LUs are unknown. Additional pre-processing steps are thus needed to enumerate the potential candidate frames for frame filtering. For this purpose, an NLP toolkit called Pattern (<https://www.clips.uantwerpen.be/pattern>) was utilized to obtain various derived forms for all LUs in FrameNet, including their lemmatized forms, comparatives, superlatives, conjugated verbs, and so forth. A lookup table was then created to link these inflections to potential frames. During model training and testing, the candidate frames were selected using the target as the lookup key. For instance, for the LU “know.v,” this step generated “know,” “knows,” “knowing,” and “knew” as the derived forms. When one of these derived forms was marked as the target, the lookup

table returned [Certainty], [Differentiation], [Awareness], and [Familiarity] as the candidate frames.

3.4 Experiments

The latest version of FrameNet is 1.7, but FrameNet 1.5 has been the version used almost exclusively in frame-semantic research. Compared to FrameNet 1.5, FrameNet 1.7 has a substantially larger number of frames and LUs, thus presenting a more challenging stage for frame identification. To date, Peng et al. (2018) was the first and only study that reported evaluation results on FrameNet 1.7. However, to facilitate comparison with related work, experiments were conducted in this study for both FrameNet versions. FrameNet provides the following types of hand-annotated data with each release:

- Exemplary data containing sentences annotated only for a single LU.
- Full-text data consisting of sentences with multiple annotations pertaining to multiple targets.

In this study, both the exemplary data and the full-text data were used for model training, with the same train/test split as in previous works. Following Hermann et al. (2014), 16 documents from the full-text data were used as the development data for hyperparameter tuning. With this standard setup, the models were tested on the DAS test set, i.e., the benchmark test data suggested by Das and Smith (2011). In FrameNet 1.5, the DAS test set contains 4,458 targets from 23 documents, whereas, in FrameNet 1.7, the number of annotations in these documents has been increased to 6,728 targets. Each target-sentence pair was treated as an instance to train and evaluate the proposed models. It is worth noting

that some sentences in the DAS test set also occur in FrameNet’s exemplary data, but this study removed all the duplicated sentences to ensure that the test data were completely unseen during the model training stage.

PAFIBERT was built on the smaller BERT model (BERT_{BASE} uncased), which consists of 12 transformer blocks, 768 hidden dimensions, and 12 self-attention heads. All hyperparameters were tuned on the development data based on the range of hyperparameter values suggested by Devlin et al. (2018). Specifically, the hyperparameter values were selected for the batch size and learning rate from {16, 32} and {2e-5, 3e-5, 5e-5} respectively. However, it was found that the range suggested for the number of epochs (i.e., {3, 4}) did not produce optimal models in the initial experiments. The range was thus increased to {3, 4, 5, 6, 7, 8} for the search. The maximum input sequence length (n) was set based on the length of the longest sentence in the training data. The value of n was set to 260 and 320 for FrameNet 1.5 and 1.7 respectively. The selected hyperparameters are shown in Table 3.2.

Table 3.2. The selected hyperparameters

Hyperparameter	FrameNet 1.5			FrameNet 1.7		
	Filtered by LUs	Filter by Targets	No Frame Filtering	Filtered by LUs	Filter by Targets	No Frame Filtering
Batch size	32	32	32	16	16	16
Learning rate	3e-5	3e-5	3e-5	3e-5	3e-5	3e-5
Epoch	8	6	8	7	6	8

The difficulty in frame identification lies in the ambiguous LUs that may link to multiple frames. Hence, previous works have assessed these more challenging and

interesting cases separately to provide a more representative measurement of model performance. This study followed the same experimental setup to evaluate the proposed models for all targets and ambiguous targets (Presented as All and Ambiguous in the results below).

3.5 Results and Discussion

Table 3.3 and 3.4 compare the classification rates of PAFIBERT to the results reported in several representative studies in frame-semantic parsing. As can be seen from the table, PAFIBERT outperformed other models on both versions of FrameNet, yielding new state-of-the-art results for frame identification. As expected, the accuracies obtained with frame filtering by gold LUs were slightly better than the accuracies produced with target-based frame filtering, suggesting that a realistic setup is crucial in measuring model performance more precisely.

Table 3.3. Accuracies with frame filtering

Model	FrameNet 1.5		FrameNet 1.7	
	All	Ambiguous	All	Ambiguous
Semafor (Das et al., 2014)	83.60	69.19	-	-
Hermann et al. (2014)	88.73	73.67	-	-
Yang and Mitchell (2017)	88.20	75.70	-	-
Hartmann et al. (2017)	87.63	73.80	-	-
Botschen et al. (2018)	88.82	75.28	-	-
Peng et al. (2018)	90.00	78.00	89.10	77.50
PAFIBERT with frame filtering by LUs	92.04	82.50	91.44	82.55
PAFIBERT with frame filtering by targets	91.39	82.80	90.15	81.92

Table 3.4. Accuracies without frame filtering

Model	FrameNet 1.5		FrameNet 1.7	
	All	Ambiguous	All	Ambiguous
Hartmann et al. (2017)	77.49	-	-	-
Botschen et al. (2018)	81.21	72.51	-	-
PAFIBERT without frame filtering	89.57	81.86	88.97	81.83

Table 3.5 shows the top five most frequently confused pairs of frames from the experiments. There is a substantial overlap in the lists of the most confusing frames across models. In particular, PAFIBERT seemed to have trouble distinguishing the following frames:

- Possibility and Capability
- Possession and Have_associated

Some examples of misclassified instances are shown in Table 3.6. Figure 3.3 gives the simplified schematic view of the above frames to aid the interpretation of the misclassified examples. Through the comparison of the ambiguous LUs, it was found that recognizing the distinction between these confusing senses is indeed quite tricky and requires a deep understanding of the underlying sentential contexts.

Despite the difficulty in distinguishing nuanced senses, the overall improvement made by PAFIBERT is promising. The boost in accuracies is especially notable for ambiguous targets, gaining 4-5% absolute improvements with frame filtering (Table 3.3) and a significant increase of 9.35% in the absence of frame filtering (Table 3.4). Interestingly, in the more challenging setup that involved no frame filtering, the accuracies obtained by PAFIBERT were still on par with prior state-of-the-art accuracies achieved with

Table 3.5. Most frequently confused pairs of frames

Gold Frame	Predicted Frame	Number of Misclassified Instances	Potentially Confusing LUs
PAFIBERT without frame filtering			
Possibility	Capability	13	can.v
Possession	Have_associated	12	have.v, have got.v
Capability	Possibility	10	can.v
Locative_relation	Spatial_co_location	9	at.prep, here.adv, there.adv
Visiting	Arriving	9	visit.v, visit.n
PAFIBERT with frame filtering by targets			
Possession	Have_associated	14	have.v, have got.v
Possibility	Capability	13	can.v
Measure_duration	Calendric_unit	12	century.n, day.n, decade.n, fortnight.n, hour.n, millennium.n, minute.n, month.n, second.n, week.n, year.n
Political_locales	Locale_by_use	9	country.n, city.n, village.n
Visiting	Arriving	9	visit.v, visit.n

Table 3.6. Examples of misclassified instances

Test Sentence	Gold Frame	Frame Predicted by PAFIBERT with Frame Filtering
Macau 's own currency , the pataca , is pegged to the Hong Kong dollar , and you <u>can</u> use your Hong Kong currency freely in Macau .	Possibility	Capability
The people who <u>can</u> benefit most directly from your generosity have no time to waste .	Possibility	Capability
I mean , what <u>can</u> i say ?	Capability	Possibility
You <u>can</u> help them to know that feeling .	Capability	Possibility
Hong Kong is crowded - it <u>has</u> one of the world 's greatest population densities .	Possession	Have_associated
Though much smaller than Hong Kong Island , Kowloon <u>has</u> almost twice the population .	Possession	Have_associated
The casinos <u>have</u> no admission charge and formal dress is optional , though long pants for men are required .	Have_associated	Possession
Macau <u>has</u> an ample supply of Portuguese wines .	Have_associated	Possession

Possibility	
A Possible_event is deemed to have all pre-conditions satisfied; in other words, it has some probability of occurrence, if some (generally implicit) further Condition pertains...	
LU	can.v, could.v, may.v, might.v
Exemplar Sentence	This sense of lack <u>CAN</u> never be satisfied .

Capability	
An Entity meets the pre-conditions for participating in an Event. A Degree modifier may be included to indicate by how much the Entity exceeds or falls short of the minimum requirements.	
LU	ability.n, able.a, can.v, capability.n, capable.a, capacity.n, inability.n, incapable.a, incapacity.n, potential.a, potential.n, power (statistical).n, power.n, powerful.a, powerless.a, powerlessness.n, unable.a
Exemplar Sentence	The GOP is afraid that it might not be <u>CAPABLE</u> of keeping the seat vacated by Dunn.

Possession	
An Owner has (or lacks) a Possession.	
LU	assets.n, belong.v, belongings.n, custody.n, effects.n, got.v, have got.v, have.v, lack.n, lack.v, lacking.a, own.v, owner.n, ownership.n, possess.v, possession (definite noun).n, possession (of goods).n, possession.n, possessor.n, property.n, want.n, want.v, wanting.a
Exemplar Sentence	The river's part of my manor, but of course it also <u>BELONGS</u> to the River Police.

Have_associated	
A Topical_entity has properties which are affected by the existence and association of an Entity.	
LU	have got.v, have.v, with.prep
Exemplar Sentence	By the 1920s a characteristic book <u>HAD</u> black and white photographic illustrations only .

Figure 3.3. Simplified schematic representations for [Possibility], [Capability], [Possession], and [Have_associated]

frame filtering. This significant performance gain demonstrates the generalizability and robustness of the proposed solution, especially for real-world frame-semantic parsing problems. As will be demonstrated in the discussion below, a model built without frame filtering has the ability to generalize well in other domains and make reasonable inferences for unseen targets.

Since this study intends to use the frame identifiers on product reviews, it is crucial to make sure the models not only perform satisfactorily on FrameNet’s data but also predict accurately when given product reviews, which contain domain-specific information. However, it is time-consuming and resource-intensive to create a set of hand-annotated data from this domain to assess the models quantitatively. This study thus resolved to qualitative evaluation using selected sample sentences from the beauty domain. This section presents interesting cases that highlight the strengths and weaknesses of the models. All tokens in the sentences, except certain stopwords, were passed as targets to the models. Sentences that contain unseen targets and domain-specific terms were selected to examine the robustness of the models in handling these more complicated cases. The predictions made by PAFIBERT with and without frame filtering were compared. Figure 3.4 to 3.7 show the sentences, targets, and predicted frames. Unseen targets are highlighted in blue, bolded font.

<i>Also There is so much product in it, so it is worth the \$\$</i>		
Target	With Lexicon-Filtering (by Target)	Without Lexicon-Filtering
Also	Duplication	Increment
There is	Existence	Existence
so	Degree	Degree
much	Proportional_quantity	Quantified_mass
product	Manufacturing	Manufacturing
in	Interior_profile_relation	Interior_profile_relation
it		
,		
so	Causation	Causation
it		
is		
worth	Deserving	Earnings_and_losses
the		
\$\$		

Figure 3.4. Frame predictions for “*Also There is so much product in it, so it is worth the \$\$*”

<i>It is tough to find in Toronto (it was at Walmart), but I'll stock up whenever I can!</i>			
Target	With Lexicon-Filtering (by Target)		Without Lexicon-Filtering
It			
is			
tough	Difficulty		Difficulty
to			
find	Locating		Locating
in	Interior_profile_relation		Interior_profile_relation
Toronto	Duplication		Political_locales
(			
it			
was			
at	Locative_relation		Locative_relation
Walmart	Prison		Businesses
)			
,			
but	Concessive		Concessive
I			
'll	Increment		Likelihood
stock	Carry_goods		Carry_goods
up	Direction		Amassing
whenever	Duplication		Frequency
I			
can	Capability		Capability
!			

Figure 3.5. Frame predictions for “It is tough to find in Toronto (it was at Walmart), but I’ll stock up whenever I can!”

<i>I do think it has a highlighting benefit (also a good shade), but it's just too dry and cakey to wear.</i>		
Target	With Lexicon-Filtering (by Target)	Without Lexicon-Filtering
I		
do		
think	Opinion	Opinion
it		
has		
a		
highlighting	Duplication	Body_decoration
benefit	Conferring_benefit	Conferring_benefit
(		
also	Increment	Increment
a		
good	Desirability	Desirability
shade	Type	Color
)		
,		
but	Concessive	Concessive
it		
's		
just	Increment	Degree
too	Sufficiency	Sufficiency
dry	Being_dry	Being_dry
and		
cakey	Duplication	Ground_up
to		
wear	Wearing	Wearing

Figure 3.6. Frame predictions for “I do think it has a highlighting benefit (also a good shade), but it's just too dry and cakey to wear”

<i>I used it for years, and I was sad to give it up</i>		
Target	With Lexicon-Filtering (by Target)	Without Lexicon-Filtering
I		
used	Using	Using
it		
for	Taking_sides	Duration_relation
years	Measure_duration	Measure_duration
,		
and		
I		
was		
sad	Emotion_directed	Emotion_directed
to		
give	Giving	Surrendering_possession
it		
up	Direction	Surrendering_possession

Figure 3.7. Frame predictions for “I used it for years, and I was sad to give it up”

First and foremost, because of the way PAFIBERT was designed to rely on target positions for frame identification, both models were capable of handling ambiguous targets that should be associated with different frames based on their positions in the same sentence. This scenario is exemplified by the utterance in Figure 3.4 in which the target “so” occurs twice but is used to express [Degree] in the first usage and [Causation] in the second.

What stands out the most from this comparison is that, although PAFIBERT with frame filtering achieved high accuracy on the benchmark DAS test set, it did not handle unseen targets satisfactorily. On the other hand, the PAFIBERT model without frame filtering seemed to generalize more decently in the beauty product domain. It is generally agreed that inadequate coverage of frame-evoking LUs is a major shortcoming of FrameNet (Botschen, Gurevych, Klie, Mousselly-Sergie, & Roth, 2018; Hartmann et al., 2017). Previous studies have tried to address this issue in many ways; some extended FrameNet using other resources such as WordNet, VerbNet, and so forth (Johansson & Nugues, 2007; Shi & Mihalcea, 2005; Tonelli & Giuliano, 2009), others adopted statistical or graph-based strategies to generalize to unseen targets (Das et al., 2010; Das & Smith, 2011). Nevertheless, unseen targets remain a major obstacle in applying frame-semantic parsing to real-world problems.

From the sample sentences and the predicted frames, it was observed that PAFIBERT without frame filtering was able to select suitable frames for most of the unseen targets, including “cakey” (Figure 3.6), “highlighting” (Figure 3.6), “Toronto” (Figure 3.5), and “Walmart” (Figure 3.5), among others. Particularly encouraging is the ability of the model to recognize “Toronto” as [Political_locales] and “Walmart” as [Businesses]. This behavior of the model resembles that of a named entity recognizer (e.g., Finkel, Grenager, & Manning,

2005) but encompasses a broader range of categories in comparison to the latter. Also, since the usage of words is domain-specific, it is encouraging that the model was able to identify frames in accordance with the domain context. For instance, identifying the word “highlight” as [Body_decoration] is appropriate in this context because a highlighter is a cosmetic product that is often applied to some regions of the face to reflect light and create contours. Another domain-sensitive identification is exemplified by the word “shade” (Figure 3.6). Although “shade.n” is listed under the [Type] frame in FrameNet, it is tagged as [Color] by the PAFIBERT model without frame filtering. Compared to the frame identified by PAFIBERT with frame filtering, [Color] seems to be a better match for “shade” in this context.

Furthermore, PAFIBERT without frame filtering worked better on non-contiguous multiword LUs containing insertions. For example, when the constituents of the “give up.v” LU are separated by insertions (Figure 3.7), PAFIBERT with frame filtering failed to associate “up” with the [Surrendering_possession] frame because this frame was not listed as one of the candidate frames for “up.” However, without frame filtering, the PAFIBERT model was able to recognize the word “up” as part of the phrasal verb “give up.”

So far, PAFIBERT without frame filtering is in the lead in the above comparison of model generalizability. However, there is a disadvantage in performing frame identification without FrameNet frame filtering. Specifically, since the model depends solely on the sentential contexts to make its predictions, it is highly sensitive to contextual variations, and it is more prone to identification errors caused by contexts that differ significantly from the normal contexts seen in the training data. In contrast, PAFIBERT with frame filtering can benefit from the listing of candidate frames provided by FrameNet to limit its search scope,

thus reducing the risk of being misled by the contexts and straying from the correct frame. For example, with frame filtering, the LU “compete.v” is unambiguous because FrameNet only associates this LU with the [Competition] frame. However, for the sentences below, PAFIBERT without frame filtering did not select the [Competition] frame:

- These are good enough to compete with those at the department store. → Prediction: [Evaluative_comparison]
- If I find a good sale on dove sensitive bars that can compete with jergen's mild soap (33 cents a bar) then I will buy it. → Prediction: [Evaluative_comparison]
- Though this can compete with NYX. → Prediction: [Compatibility]

Although the predicted frames still make sense, this characteristic of the model may introduce too many variations at the semantic-level, thus defeating the purpose of using semantic abstraction, i.e., to reduce word variation and sparsity. This issue creates a dilemma because while frame filtering can remedy the problem caused by contextual variations, enforcing the candidate frames would reduce the generalizability of the model.

3.6 Summary

Frame identification is an integral part of frame-semantic parsing and an indispensable step in the semantic abstraction framework proposed in this research. This chapter presented a novel neural network solution called **P**ositional **A**ttention-based **F**rame **I**dentification with BERT (PAFIBERT) that leveraged targets’ positions for position-based attention. Although the design of the proposed solution is conceptually simple, rigorous comparison with existing models revealed that the proposed models are empirically powerful and superior to

other models under various experimental setups. However, through the analysis of the most frequently confused pairs of frames, it was found that the identifiers failed to recognize more subtle differences that required a more profound understanding of sentence contexts and semantics. Although PAFIBERT was built on top of a state-of-the-art language representation model (i.e., BERT) which, based on recent empirical evidence, is the answer to rich semantic representations for NLP tasks, it is still unclear as to how far this latest advancement can take us in dealing with nuanced sense distinction. Since the top frequently confused frames caused a substantial loss in model accuracies, future research in frame identification should consider placing emphasis on solving these more challenging cases.

CHAPTER 4

GRADFRAMENET AND SENTIMENT-AWARE SEMANTIC ABSTRACTION

4.1 Introduction

This chapter addresses RQ2 and RO2. The study presented in this chapter aims to incorporate an appraisal concept called graduation (Martin & White, 2005) into FrameNet to produce a sentiment-aware version of the resource, named as GradFrameNet in this research. This effort to extend FrameNet is necessary to make the resource more suitable for the task of sentiment analysis. Several steps were performed in this study to examine the usefulness of GradFrameNet. First, GradFrameNet was used in conjunction with the frame identifier (PAFIBERT) developed in Phase 1 of this research to perform frame-based semantic abstraction and generate semantic-level representations for online reviews. Second, phrasal patterns were extracted from the representations to assess the semantic abstraction ability and information discovery ability of the abstraction process. The semantic abstraction ability refers to whether the abstraction process results in semantically-consistent representations that generalize in a sentiment-aware manner, whereas the information discovery ability refers to whether the process facilitates the discovery of phrasal patterns that reveal interesting clues for inferring reviewers' sentiments.

4.2 Background

Cognitive linguistics is a field that studies language with emphasis on its cognitive function for organizing information in humans' minds and for forming informational structures that mediate humans' interactions with the world. Generally, there is a consensus among cognitive linguists that meaning is a linguistic phenomenon of primary importance in linguistic analysis, and that the study of meaning should be governed by cognitive principles of categorization (Geeraerts, 2006). Cognitive linguists think of language as a collection of meaningful categories that serve as the instrument to help humans organize knowledge about the world, thus allowing humans to convey, process, and understand meanings with respect to the structures imposed by these categories (Geeraerts, 2006; Geeraerts & Cuyckens, 2007).

FrameNet is a linguistic resource that has been created as a realization for frame semantics theory (Fillmore, 1976), which has a far-reaching impact in cognitive linguistics. According to this theory, word meanings should be described and organized in relation to semantic frames (Fillmore et al., 2003). The latest version of FrameNet (version 1.7) consists of more than 1,000 frames. These semantic frames aim to encode real-world knowledge or common scenarios in a schematized form (Lowe, 1997). Figure 4.1 shows a simplified representation of an example frame called [Usefulness]. This schematic representation includes the description of the frame, the frame's elements (FEs), and a list of lexical units (LUs). This study only focused on the encapsulation relation between frames and their LUs. This relation defines the association between lemmas and frames, which provides a means for representing texts using a more abstract and less diverse semantic form. For example, [Usefulness] can be used as an abstraction or a semantic-level representation for "effective," "work," "useful," and so forth. The potential benefit that we can gain from this kind of

semantic generalization is tremendous. By associating words with a predefined set of semantic frames, a denser word space consisting of a more manageable set of linguistic categories instead of countless individual terms can be created. As a result, the word sparsity problem can be mitigated, and common phrasal patterns become more discernible because semantically similar phrases are likely to share the same semantic forms.

Usefulness	
An Entity aids in the successful completion of a Purpose . Its Degree of utility may also be expressed...	
Core FE	Entity, Purpose
Non-core FE	Degree, Domain, Time
LU	effective.a, excellent.a, fantastic.a, fine.a, good.a, great.a, ideal.a, ineffective.a, marvellous.a, outstanding.a, perfect.a, splendid.a, strong.a, super.a, superb.a, terrific.a, tremendous.a, useful.a, utility.n, valuable.a, value.n, wonderful.a, work.v

Figure 4.1. The [Usefulness] frame (FrameNet 1.7)

In relation to linguistic categories, some intriguing questions may be raised regarding the level of semantic abstraction that facilitates the discovery of evaluative patterns in sentiment analysis. A closer look at FrameNet’s frames and their LUs should quickly reveal that, although FrameNet provides a means for transforming lexical-level representations into abstract representations, the information essential for distinguishing positive and negative opinions is missing. To understand this problem, consider the utterance “The delivery was slow,” which expresses a negative attitude towards the delivery service. The word “slow” in this utterance belongs to the [Taking_time] frame in FrameNet (Figure 4.2), which also contains other LUs such as “fast,” “rapid,” and “speedy.” For sentiment analysis and opinion mining, coalescing antonymic terms into the same category is usually undesirable because this information is needed to tell apart positive opinions from the negative ones and vice

versa. In other words, [Taking_time] cannot be used as a common abstraction for “slow,” “fast,” “rapid,” and so forth because this will cause a substantial loss of sentiment-related information. To benefit from the frame-based abstraction provided by FrameNet while retaining the sentiment-related information expressed in the original text inputs, this study set out to complement FrameNet’s frames with an appraisal concept called graduation (Martin & White, 2005). The process to incorporate this concept into FrameNet is described in the next section.

Taking_time	
An Activity takes some Time_length to complete. If an Agent is involved they may stand in metonymically for the Activity ...	
Core FE	Activity, Agent, Degree, Time_length
Non-core FE	Attribute, Circumstances, Duration, Time
LU	fast.a, gradually.adv, in.prep, quick.a, quickly.adv, rapid.a, slow.a, slowly.adv, speedy.a, swift.a, take.v

Figure 4.2. The [Taking_time] frame (FrameNet 1.7)

4.3 Method

This study involved two principal stages. First, a sentiment-aware FrameNet was created by incorporating the graduation concept defined by Martin and White (2005) into FrameNet. Second, a semantic tagger was used for semantic abstraction, i.e., transforming input sentences to their semantic-level representations. These two stages are described below.

4.3.1 GradFrameNet

GradFrameNet was created in this study by categorizing manually the LUs of each semantic frame based on the graduation concept described in Martin and White's (2005) appraisal framework. This framework covers the evaluative aspects of natural language comprehensively. According to the appraisal framework, the linguistic elements of appraisals divide into three categories—attitude, engagement, and graduation—that operate interactively.

The attitude category encompasses words or phrases that express feelings, emotions, and assessments. This category has traditionally been dealt with extensively in sentiment analysis, especially in lexicon-based sentiment analysis, which builds on the concept of sentiment polarity associated with words and phrases. Linguistic elements in this category are used to express positive and negative feelings or emotional reactions, as well as favorable or unfavorable assessments towards some attitudinal targets. In the appraisal framework, the attitude category is further classified into the following subtypes:

- Affect, which is related to emotional reactions (e.g., “happy” and “sad”)
- Judgment, which describes assessments of human behavior based on social norms like ethics and morality (e.g., “honest” and “deceptive”)
- Appreciation, which describes aesthetical assessments by reference to social valuation (e.g., “beautiful” and “ugly”)

The engagement category consists of locutions that provide a means for dialogistic positioning. Using these locutions, writers and speakers may position themselves in various ways—as affirming, reporting, acknowledging, denying, countering, and so forth—towards

a proposition. There are four engagement approaches—proclamation, disclamation, probabilization, and attribution—that may be adopted by a writer or speaker in forming a dialogistically positioned statement. These approaches represent different types of engagement with a proposition, which indicates a writer’s or speaker’s degree of approval, commitment, and certainty in the proposition (White, 2015). For instance, in the sentences below, “will definitely be” signifies a strong engagement whereas “could be” shows a weaker engagement.

- This product will definitely be a game-changer for you. (Proclamation)
- This product could be a game-changer for you. (Probabilization)

The third appraisal category is the graduation category, which provides a means for writers or speakers to adjust the strength of their appraisal (Read & Carroll, 2012) by intensifying or mitigating the strength of the opinions they express (Carretero & Taboada, 2014; White, 2015). Martin and White (2005) stated that graduation is a ubiquitous phenomenon in natural language, and its members exert upscaling and downscaling effects on other language elements, including elements under the attitude and engagement categories. For instance, an attitude expression such as “upset” can be scaled by preceding it with an intensifier (e.g., “very” and “extremely”) or a mitigator (e.g., “a little” and “slightly”). The appraisal framework classifies graduation into the following subtypes:

- *Focus*. This type of graduation is used to graduate inherently unrankable terms. Focus-based graduation can adjust the “membership” of other terms with respect to the notion of prototypicality. A member of a language category is considered prototypical if it is a representative member of the category. Focus-based graduation

adjusts this membership to draw an instance closer to or away from the core of a category. For example, in the utterance “he plays real jazz,” the word “real” has an upscaling effect, which contrasts the downscaling effect of “of sorts” in “he made a curtain of sorts.”

- *Force*. This type of graduation operates on rankable language categories, differentiating the members of these categories based on Quantification and Intensity. Language categories that can be graduated based on Quantification include those related to quantity (e.g., “a few” and “many”), mass (e.g., “large” and “tiny”), proximity (e.g., “far” and “close”), and distribution (e.g., “brief” and “prolonged”). Graduation with respect to Intensity usually operates on qualities and processes. This type of graduation operates in two modes: isolated and infused. The former refers to cases where isolated intensifiers such as “slightly” and “very” are used to adjust the strengths or degrees of other terms (e.g., “slightly disappointed” and “very disappointed”). The infused mode refers to cases where the intensification or mitigation mechanism is an integral part of a term, i.e., the graduation is “inbuilt” alongside the meaning of the term. This graduation mode is exemplified by terms like “pleased,” “happy,” and “overjoyed” that express different degrees of happiness.

While isolated intensifiers are easy to identify, pinpointing infused intensifiers is less straightforward because of the absence of well-defined semantic realms within which comparable terms can be graduated. Specifically, for two terms to be comparable, they should have some common properties that place them under the same semantic category. Hence, the task of identifying infused intensifiers must be preceded by the tasks of creating and defining such semantic categories and enumerating the members belonging to each

category. Although Martin and White’s (2005) appraisal framework has been applied to many studies in sentiment analysis and opinion mining (Argamon, Bloom, Esuli, & Sebastiani, 2007; Korenek & Šimko, 2014; Whitelaw, Garg, & Argamon, 2005), Infused Intensification has not received much attention in prior works. This research gap is filled in the present study by using FrameNet’s frames as the semantic realms. Frame members (i.e., LUs) can then be graduated with respect to some common semantic properties defined by the frames. The examples of frames and LUs in Figure 4.3 illustrate the phenomenon of Infused Intensification.

	[Degree_of_processing]	[Sufficiency]	[Temperature]
High	enriched.a, processed.a	plenty.a, ample.a, too.adv, so.adv, etc.	hot.a, scalding.a, warm.a
Medium		enough.a, suffice.v, sufficient.a, adequate.a, etc.	lukewarm.a, tepid.a
Low	raw.a, unprocessed.a	insufficient.a, insufficiency.n, inadequate.a, inadequacy.n, etc.	cool.a, freezing.a, cold.a, frigid.a

Figure 4.3. Examples of frames and LUs that can be graduated based on Infused Intensification

Figure 4.4 shows the graduation taxonomy deployed in this study to transform FrameNet (version 1.7) into GradFrameNet. Although the appraisal framework consists of three appraisal categories, the present research only incorporated the graduation concept into FrameNet because as stated by Martin and White (2005), “attitude and engagement are domains of graduation which differ according to the nature of the meanings being scaled” (p. 136). Read and Carroll (2012) also described graduation as a system that “addresses how

language functions to amplify or diminish the attitude and engagement conveyed by a text” (p. 423). These descriptions suggest that graduation does not stand alone as an evaluative category, but is often inextricably intertwined with other language elements, including elements of attitude and engagement. In other words, gradability can be perceived as a general property of attitude and engagement, and most attitude and engagement expressions can be associated with one of the subtypes of graduation. For example, although most LUs of the [Usefulness] frame (Figure 4.1) fall under the attitude category, these LUs can also be associated with Infused Intensification, which allows graduation to be applied with respect to the meaning (i.e., usefulness) being scaled. Other frames in FrameNet that contain attitude expressions include [Desirability], [Emotion_directed], [Aesthetics], [Candidness], [Morality_evaluation], and so forth. Likewise, frames such as [Certainty] and [Opinion] contain engagement expressions (e.g., “believe,” “sure,” “think,” and “suppose”) that can be scaled based on Infused Intensification. Therefore, deploying only the graduation taxonomy on FrameNet while excluding attitude and engagement would most likely ensure the simplicity and usability of the resulted resource without sacrificing crucial information for sentiment analysis.

The deployment of the graduation taxonomy was conducted by the author and an undergraduate student at Nanyang Technological University. Given the LUs retrieved from FrameNet, the purpose of the annotation process was to decide whether an LU could be associated with any graduation type and to determine the level or value of graduation if applicable. Table 4.1 shows some examples of LUs retrieved from FrameNet for the annotation task.

Figure 4.4. The graduation taxonomy

Table 4.1. Examples of LUs retrieved from FrameNet for the annotation task

Frame	Lexical Unit	Definition
Abandonment	abandon.v	FN: leave behind
Abandonment	leave.v	FN: let remain in another place as abandoned
Abandonment	abandonment.n	FN: the act of relinquishing to no one
Abandonment	abandoned.a	FN: the state of being left behind
Abandonment	forget.v	FN: leave something behind unintentionally
Abounding_with	teem.v	COD: be full of or swarming with.
Abounding_with	crawling.a	FN: unpleasantly covered or crowded with.
Abounding_with	crawl.v	COD: be unpleasantly covered or crowded with.
Abounding_with	throng.v	FN: be filled (with), teem (with)
Abounding_with	thronged.a	FN: filled (with)
Abounding_with	crowded.a	FN: filled so far as to inhibit movement.
Abounding_with	swarming.a	FN: moving in or form a large group of people or things
...	...	...

As described earlier in this section, the taxonomy derived from the appraisal framework defines the following graduation types:

- Quantity: Graduation type for words related to amount or quantity (e.g., “a few,” “many,” “galore,” and “millions”)
- Mass: Graduation type for words related to mass and size (e.g., “huge,” “tiny,” “tall,” and “gigantic”)
- Distribution: Graduation type for words related to distribution in time and space (e.g., “brief,” “lengthy,” and “wide-spread”)
- Proximity: Graduation type for words related to distance in time and space (e.g., “near,” “far,” and “distant”)
- Isolated Intensification: Graduation type for independent modifiers that can be used to adjust the degrees of other words (e.g., “slightly,” “very,” “extremely,” and “a little bit”)
- Infused Intensification: This graduation type contains words that can be ranked by their semantics. For example, “devastated,” “sad,” “pleased,” “happy,” and “overjoyed” are examples of Infused Intensification that express different degrees of happiness.
- Focus: This graduation type contains words that can adjust the “membership” of other words or the “precision” of other words. For instance, “real” has an upscaling effect on other words (e.g., “He is a gentleman” versus “He is a real gentleman”), whereas “about” has a downscaling effect on other words (e.g., “You have two hours” versus “You have about two hours”).

For each LU that could be associated with one of the graduation types, the coders also selected a level of graduation. A three-point scale was used to define the levels of graduation. For instance, when Mass was selected as the graduation type for an LU, the coders should choose from one of the following graduation levels: Small, Medium, or Large. As can be seen in Table 4.1, each LU in the list consists of a word or phrase and its syntactic category (“v” for verb, “n” for noun, “a” for adjective, and so forth), and is associated with a semantic frame and a definition. The coders interpreted the meanings of the LUs based on the frames and the LUs’ definitions.

The deployment differentiated between frame-dependent and frame-independent graduation. Six of the graduation types, i.e., Quantity, Mass, Distribution, Proximity, Isolated Intensification, and Focus, are collectively referred to as frame-independent graduation in the present research because the graduation is neither interpreted nor applied within the context of a semantic frame. The frame-dependent graduation consists only of the Infused Intensification type, which is different from other graduation types in that the graduation should be interpreted with respect to the specific context defined by a semantic frame. To obtain more consistent annotations, the coders used the following thought process:

- Step 1: Can the LU be associated with one of the frame-independent graduation types (i.e., Quantity, Mass, Distribution, Proximity, Isolated Intensification, or Focus)?
 - YES → Select the graduation type and determine the level of graduation.
 - NO → Go to the next step.
- Step 2: Does the frame of the LU under consideration contain any antonymous terms (e.g., happy and sad, honest and deceitful, and so forth)?
 - YES → Select Infused Intensification as the graduation type and determine the level of graduation.

- NO → Compare the LU under consideration with other LUs in the same frame. If they cannot be ranked based on any common criteria, this LU should not be assigned any graduation type.

The intuition underlying Step 2 in the annotation process is that since Infused Intensification utilizes the semantic frames as the contexts or realms of graduation, the existence of antonymous LUs (e.g., “cold” and “hot”) within a frame (e.g., [Temperature]) suggests that the LUs of the frame can be ranked or graduated based on certain criteria.

Cohen's κ was run to determine intercoder consistency. There was substantial agreement between the two coders' annotations ($\kappa = .768$). An annotation was considered consistent only when both coders selected the same graduation type and graduation level. The conflicting annotations were reviewed and resolved by the author. Out of 13,689 entries and 1,073 frames, graduation was applied to 4,537 LUs from 255 frames. In addition to applying graduation to the current LUs in FrameNet, the lists of negators and intensifiers under the [Negation] and [Degree] frames were also extended. As mentioned earlier, the valence shifting effects of negation and intensification have been studied extensively in sentiment analysis. However, it was noticed that the lists of negators and intensifiers in FrameNet are rather thin (6 LUs in [Negation] and 21 LUs in [Degree]). Hence, the lists were extended with manually collected negators and intensifiers from various sources, increasing the sizes of the [Negation] and [Degree] frames to 12 and 61 respectively. Table 4.2 shows the numbers of frames and LUs for each graduation type and level.

Table 4.2. Numbers of frames and LUs for each graduation type and level

Graduation Type	Graduation Level	Number of Frames	Number of LUs
Quantity	Large	7	69
	Medium	1	4
	Small	5	23
Mass	Large	5	60
	Medium	3	16
	Small	4	29
Distribution	Wide	17	92
	Medium	5	37
	Narrow	9	57
Proximity	Far	6	39
	Medium	6	21
	Close	5	15
Isolated Intensification	High	2	44
	Medium	1	13
	Low	1	8
Focus	Upscale	3	12
	Downscale	3	23
Infused Intensification	High	211	2,039
	Medium	60	364
	Low	208	1,572

4.3.2 Semantic Tagging for Semantic Abstraction

Figure 4.5 gives an overview of the semantic tagging process used in this study. An example is also given in the figure to provide a better illustration of the process. Each sentence from the input documents (i.e., product reviews) went through standard pre-processing steps, including cleaning, case folding, part-of-speech (POS) tagging, and lemmatization. Stanford CoreNLP was utilized for these pre-processing steps to generate a sequence of tokens. Each token is a tuple of five values: the positions of the first and last characters of the token, as well as the word form, lemma, and POS of the token.

Figure 4.5. The semantic tagging process

The brand tagging and product keyword tagging steps were included in the process to provide further generalization by incorporating domain-specific information. The former identified product brands and replaced them with a [Product_brand] tag, whereas the latter identified keywords referring to the reviewed products and replaced them with a [Product] tag. Depending on the types of input documents, these two steps can be skipped or replaced by other similar processes. Product brands can be scraped from review websites that allow users to search by brands. Product keywords can be listed manually because there are only a handful of keywords that reviewers frequently use for each product type. For instance, in camera reviews, common product keywords include “camera,” “digital camera,” “digital cam,” and “device.” In product reviews, these keywords are often preceded by the word

“this.” In the example given in Figure 4.5, the reviewer used “this product” to refer to the reviewed product. To identify product brands and product keywords in sentences, this study performed text pattern matching using a set of regular expressions defined from lists of product brands and product keywords.

As described in Chapter 3, in Phase 1 of this research, progress has been made on frame identification, i.e., the process of associating text tokens or targets with the most appropriate semantic frames. The semantic tagging process presented in this section incorporated the PAFIBERT model described in Chapter 3 for the frame tagging step. Since PAFIBERT without frame-filtering worked better for unseen targets, this variant was chosen for the experiments presented in this chapter. For the example given in Figure 4.5, the following inputs would be fed into PAFIBERT for frame identification:

$(S_I, 14, 17), (S_I, 19, 28), (S_I, 31, 33), (S_I, 58, 71), (S_I, 82, 84)$

where S_I is the sample sentence in Figure 4.5. The second and third components in these tuples mark the positions of tokens in S_I . Note that stopwords and tokens that had been tagged in the brand tagging and product keyword tagging steps were omitted from frame tagging. Besides, since GradFrameNet contains multi-word lexical units (MWLUs) like idioms, compound nouns, and other forms of phrasal LUs, multi-word tokens such as “well known” in S_I were passed as single tokens instead of splitting them into their constituents. The MWLUs listed in GradFrameNet were used to identify these multi-word tokens via string matching.

As depicted in Figure 4.5, the frame tagging step generated a sequence of frame tags, $F = \{f_1, f_2, \dots, f_n\}$, where f_j is a tuple of six values, denoted as (frStartIndex, frEndIndex, targetLemma, predictedFrame, graduationCategory, categoryValue). Each frame tag linked

a stretch of text at positions [frStartIndex, frEndIndex] of the input sentence to a frame and a graduation type in GradFrameNet. The graduationCategory and categoryValue elements in the frame tags held the graduation types (e.g., Quantity and Infused_intensification) and levels (e.g., Large and Low). Let L be the set of LUs listed under the frame predicted by PAFIBERT. Determining the graduation type and its value should be straightforward if the target at the specified positions can be associated directly with an LU that exists in L , i.e., $lu_t \in L$. However, if t is an unseen target or the frame predicted by PAFIBERT does not contain lu_t , the graduation type and value of t were inferred based on the similarity between t and each LU listed in L . For instance, “beautiful.a” falls under the [Aesthetics] frame. When PAFIBERT returns [Aesthetics] as the predicted frame for “beautiful,” the graduation type and value of the LU (i.e., [Aesthetics: Infused_intensification = High]) can be retrieved easily from GradFrameNet. However, for an unseen target such as “gorgeous” that is associated by PAFIBERT with the [Aesthetics] frame, the LUs of the [Aesthetics] frame are used to infer the graduation type and value of “gorgeous.” In this case, since “gorgeous” has a high similarity with “beautiful,” it will be assigned the same graduation type and value, i.e., [Aesthetics: Infused_intensification = High]. Formally, the graduation type and value of t is inferred from that of LU z for

$$z = \underset{l \in L}{\operatorname{argmax}}(\operatorname{similarity}(t, l)) \quad (1)$$

For the measurement of similarity, cosine similarity (e.g., Larsen & Aone, 1999) was computed from the GloVe embeddings (Pennington et al., 2014) of t and l . However, since product reviews often contain out-of-dictionary words such as informal abbreviations or misspelled words, t may not have a match in the GloVe embeddings. In that case,

Levenshtein distance is used as the similarity measure to compute $\text{similarity}(t, l)$ in the above equation.

As can be seen in Figure 4.5, the final output of the semantic tagging process is the semantic-level representations of the input sentence. There are three semantic forms. For a non-graduated frame such as [Concessive], its semantic form only consists of the frame name. The semantic form for frame-dependent graduation differs from that of other graduation types. As discussed in the previous section, frame-dependent graduation must be realized within the context defined by a semantic frame. Therefore, in the semantic-level representations, while other graduation types can stand on their own (e.g., [Isolated_intensification = High] and [Mass = Small]), frame-dependent graduation only makes sense when attached to a semantic frame (e.g., [Desirability: Infused_intensification = Low] and [Fame: Infused_intensification = High]).

4.4 Experiment

Two properties of the proposed abstraction method were examined to assess its usefulness in sentiment analysis: first, its semantic abstraction ability in grouping together semantically similar phrases in a sentiment-aware manner; second, its information discovery ability in uncovering information related to reviewers' sentiments. In the present study, this assessment was accomplished through a careful analysis of the phrasal patterns extracted from the semantic-level representations described earlier. The data used in this study consisted mainly of online reviews collected from the beauty domain. Despite being a multimillion-dollar industry, the beauty domain has been little explored in sentiment analysis. Beauty product reviews provide good examples to study the challenging problem

of implicit sentiments and indirect opinions because reviewers in this domain often express their sentiments indirectly through descriptions of usage experiences. Web scraping was used to acquire online reviews for eyeliner, cleanser, and moisturizer products from MakeupAlley (www.makeupalley.com), a popular review site in the beauty domain. For each product type, reviews with five-star ratings (positive reviews) and one-star ratings (negative reviews) were collected from the top 30 most reviewed products. In addition to the beauty product reviews, this study also included online reviews for an electronic product, i.e., television, for cross-domain comparison. The television reviews were obtained from a publicly available Amazon dataset (Wang, Lu, & Zhai, 2010). After removing duplicate reviews and brief reviews containing less than 20 words, 500 positive and negative reviews were sampled randomly for each product type. Table 4.3 shows detailed information about the data.

Table 4.3. Detailed information about the data

	Data Subset	Number of Reviews	Number of Sentences	Number of Words
Eyeliner	Negative (-E)	500	2,989	42,730
	Positive (+E)	500	2,935	44,833
Cleanser	Negative (-C)	500	2,682	42,432
	Positive (+C)	500	2,936	45,199
Moisturizer	Negative (-M)	500	2,800	43,343
	Positive (+M)	500	2,951	46,609
Television	Negative (-T)	500	4,145	79,870
	Positive (+T)	500	4,636	92,815

This study extended the collocation finders in Natural Language Toolkit (NLTK) to extract phrasal patterns from semantic-level representations. The phrasal patterns were extracted in the form of trigram collocations using Pointwise Mutual Information (PMI) as

association measure to select collocations based on the scores given by the following formula (Manning & Schütze, 1999):

$$PMI(w_1, w_2, w_3) = \log \frac{N * c(w_1, w_2, w_3)}{c(w_1, w_2) * c(w_3)} \quad (2)$$

where N is the total number of tokens, $c(w_3)$ is the number of times w_3 occurs in the data, $c(w_1, w_2)$ is the number of times w_1 and w_2 co-occur, and likewise, $c(w_1, w_2, w_3)$ denotes the co-occurrence of w_1 , w_2 , and w_3 . Co-occurrences are usually counted within a window of fixed size. This window size was set to five in this study. In other words, the collocations consisted of sequences of three tokens that were not necessarily adjacent within a five-token window. Often the candidate n-grams obtained from calculating the above frequencies are filtered to remove unuseful terms. This study filtered out tokens consisting of less than two characters and all the words found in a list of manually selected stopwords. Examples of filtered tokens include punctuations, pronouns, articles, and common verbs such as “do” and “have,” among others. A minimum frequency threshold was set so that the collocation finders only considered candidates that appeared at least five times in the data.

The enhancement made to the collocation finders allowed PMI scores to be calculated based on token frequencies in the semantic-level representations, but a lookup table was maintained to store the lexical-level n-grams corresponding to each semantic-level candidate n-gram. These lexical-level collocations are referred to as collocation groups. Based on the PMI scores, the top 25 collocation groups in each data subset (200 in total) were selected and presented in the discussion below. Since the data subsets contain only five-star and one-star reviews, the collocation groups extracted from +E/+C/+M/+T and -E/-C/-M/-T consist mostly of positive phrasal patterns and negative phrasal patterns respectively. There was,

however, a small number of collocation groups that were returned as top collocation groups for both positive and negative reviews. These ambiguous collocation groups were removed from the results. Finally, to assess how useful and informative the extracted phrasal patterns are, the information types of the collocation groups were also identified manually. Each collocation was assigned one of the following information types: General / Ambiguous, Experiential, Dis/satisfaction, Expectation, Retention, and Word of Mouth. Except for the first information type, other information types are related to concepts frequently discussed in the field of sentiment analysis and the related fields of consumers' behavior. Therefore, linking the collocation groups to these information types helps to provide theoretical reasoning for how the extracted collocation groups relate to the reviewers' sentiments. These information types were determined based on the sentences in the product reviews from which the collocations were extracted. The sentence examples in the tables should provide sufficient contexts for readers to make sense of the collocations. In cases where the collocations seem more ambiguous, multiple examples are given to clarify the contexts. These information types will be discussed in more depth in the next section.

4.5 Results and Discussion

The collocation groups obtained for the four product types are shown in Table 4.4 to 4.11. In general, the findings showed that the following properties and characteristics of the reviewed products were typically perceived as important by consumers.

- The eyeliner reviewers were concerned about the staying power, change of consistency, smoothness, color saturation, and ease of use of the products.

Table 4.4. Collocation groups from negative eyeliner reviews (-E)

Ref.	Collocation Group	Examples from the -E Subset	Info. Type
-E1	{complete/waste/money}: 4, {total/waste/money}: 2	• <i>This was a <u>complete waste of money</u>.</i>	Dis/satisfaction
-E2	{if/look/for}: 6	• <i>If your <u>looking for</u> a good liquid liner ...</i>	General / Ambiguous
-E3	{stay/far/from}: 3, {stay/away/from}: 2	• <i>... <u>stay far away</u> from this product!</i>	Dis/satisfaction
-E4	{have to/go/over}: 5, {im/go/over}: 1	• <i>I find I always <u>have to go over</u> it with a powder ...</i>	Experiential
-E5	{not like/at/all}: 7	• <i>I did <u>not like</u> this <u>at all</u>.</i>	Dis/satisfaction
-E6	{high/hope/for}: 5	• <i>very dissapointing because I had <u>high hopes for</u> this one.</i>	Expectation
-E7	{look/like/racoon}: 4, {look/like/raccoon}: 2	• <i>I <u>look like</u> a <u>raccoon</u> within ...</i>	Experiential
-E8	{only/thing/about}: 7	• <i>... <u>only good thing about</u> it ...</i> • <i>... <u>only thing I liked about</u> it ...</i>	Dis/satisfaction
-E9	{smudge/like/crazy}: 7	• <i>... cream <u>smudges like crazy</u> ...</i>	Experiential
-E10	{tug/at/lid}: 3, {tug/at/skin}: 1, {pull/at/skin}: 1	• <i>... it <u>tugged at</u> my <u>lids</u>.</i> • <i>... so it was <u>pulling at</u> my <u>skin</u>.</i>	Experiential
-E11	{dry up/so/fast}: 4, {dry out/so/fast}: 1	• <i>I was really disappointed that it <u>dried up so fast</u> ...</i>	Experiential
-E12	{will/not buy/again}: 4, {will/not purchase/again}: 1, {will/never buy/again}: 1	• <i>I guess I <u>will never buy</u> it <u>again</u>.</i>	Retention
-E13	{low/than/30}: 1, {less/than/7}: 1, {less/than/5}: 1, {less/than/30}: 1, {less/than/2}: 1	• <i>... lasted for <u>less than 2</u> hours ...</i> • <i>... icky mess in <u>less than 5</u> minutes?</i>	Experiential
-E14	{run/down/face}: 2, {stream/down/face}: 1, {come/stream/face}: 1, {come/down/face}: 1	• <i>... <u>BLACK STREAKS RUNNING DOWN YOUR FACE!!!!</u></i>	Experiential
-E15	{after/read/all}: 5, {after/read/many}: 1	• <i><u>After reading all</u> the great reviews I thought this product would be great.</i>	Expectation
-E16	{first/time/use}: 5, {first/time/apply}: 1	• <i>This is my <u>first time using</u> liquid eyeliner ...</i>	General / Ambiguous
-E17	{for/two/day}: 2, {for/2/hour}: 2, {for/two/hour}: 1, {for/couple/hour}: 1, {for/8/hour}: 1	• <i>I worked in retail <u>for an 8 hour</u> shift ...</i> • <i>... I lost my vision temporarily <u>for two days</u>.</i>	General / Ambiguous
-E18	{take/five/minute}: 1, {less/30/minute}: 1, {in/30/minute}: 1, {in/30/min}: 1, {in/20/minute}: 1	• <i>First of all, it <u>takes almost five</u> <u>minutes</u> to fully dry.</i> • <i>It melted off my eyes in <u>less than 30 mins</u>.</i>	Experiential

Table 4.4. (continued)

Ref.	Collocation Group	Examples from the -E Subset	Info. Type
-E19	{chuck/in/trash}: 2, {toss/in/trash}: 1, {throw/in/garbage}: 1, {throw/in/bin}: 1	<ul style="list-style-type: none"> • <i>I <u>tossed</u> it <u>in</u> the <u>trash</u>.</i> • <i>I used it a few times and actually <u>threw</u> it <u>in</u> the <u>bin</u>.</i> 	Dis/satisfaction
-E20	{all/good/review}: 6, {many/great/review}: 1, {many/good/review}: 1, {many/favorable/review}: 1, {majority/great/review}: 1, {all/positive/review}: 1, {all/great/review}: 1, {a lot/good/review}: 1	<ul style="list-style-type: none"> • <i>I have no idea why this eyeliner has so <u>many favorable reviews</u>.</i> • <i>Going by <u>all</u> the <u>positive reviews</u> here, i went and bought it - bad idea.</i> 	Expectation
-E21	{about/5/time}: 2, {around/7/time}: 1, {about/six/time}: 1, {about/20/time}: 1	<ul style="list-style-type: none"> • <i>I had to dip my brush in the gel <u>around 7 times</u> ...</i> • <i>I had to layer it <u>about six times</u> for it to show up ...</i> 	Experiential
-E22	{not know/if/just}: 4, {dont know/if/just}: 1	<ul style="list-style-type: none"> • <i>I <u>don't know if</u> its <u>just</u> a crap product or crap packaging...</i> 	Dis/satisfaction
-E23	{real/want/like}: 11, {real/want/love}: 8	<ul style="list-style-type: none"> • <i>I <u>really wanted</u> to <u>love</u> this eyeliner ...</i> 	Dis/satisfaction
-E24	{try/put/on}: 4, {attempt/put/on}: 1	<ul style="list-style-type: none"> • <i>... when I <u>try</u> to <u>put it on</u> my eyes it clumps</i> 	Experiential
-E25	{eyeliner/out/there}: 2, {products/out/there}: 1, {pencil/out/there}: 1, {eyeliners/out/there}: 1	<ul style="list-style-type: none"> • <i>I've heard a lot of good reviews for this product as one of the best eyeliner <u>pencil out there</u>.</i> • <i>There are much better <u>products out there</u>.</i> 	General / Ambiguous

Table 4.5. Collocation groups from positive eyeliner reviews (+E)

Ref.	Collocation Group	Examples from the +E Subset	Info. Type
+E1	{skin/dark-eyed/girl}: 1, {pale/skin/girl}: 1, {pale/dark-eyed/girl}: 1, {brown/eye/gal}: 1, {blue/eye/girl}: 1	<ul style="list-style-type: none"> • <i>This is a must have for <u>brown eyed gals</u>.</i> 	General / Ambiguous
+E2	{until/take/off}: 9, {until/wash/off}: 1, {till/take/off}: 1	<ul style="list-style-type: none"> • <i>... won't budge or fade <u>until</u> you <u>take it off</u>!</i> 	Experiential
+E3	{go/long/way}: 5	<ul style="list-style-type: none"> • <i>... the product <u>goes a long way</u> ...</i> 	Experiential
+E4	{buy/over/over}: 4, {re-purchase/over/over}: 1, {purchase/over/over}: 1	<ul style="list-style-type: none"> • <i>I would <u>buy</u> this <u>over</u> and <u>over</u> again!</i> 	Retention
+E5	{not tug/pull on/eye}: 1, {not tug/on/lid}: 1, {not tug/at/lid}: 1, {not tug/at/eye}: 1, {no pull/at/skin}: 1, {less tug/at/lid}: 1	<ul style="list-style-type: none"> • <i>Glides on very easily, <u>no pulling at</u> the <u>skin</u> ...</i> 	Experiential
+E6	{many/different/type}: 3, {many/ different/brand}: 1, {a lot/different/ type}: 1	<ul style="list-style-type: none"> • <i>I tried <u>a lot of different types</u> of liners, including dept store brands and this kind stays on the best.</i> 	Dis/satisfaction
+E7	{put/on/in}: 3, {put/on/at}: 2	<ul style="list-style-type: none"> • <i>I <u>put it on in</u> the morning with an angled eyeliner brush ...</i> 	General / Ambiguous
+E8	{off/at/night}: 7, {off/at/pm}: 1	<ul style="list-style-type: none"> • <i>until I took it <u>off</u> around 9 <u>at night</u>.</i> • <i>... looks new until I take it <u>off at</u> 9 <u>PM</u>.</i> 	Experiential
+E9	{for/four/now}: 1, {for/7/now}: 1, {for/5/now}: 1, {for/4/now}: 1, {for/2/now}: 1, {for/10/now}: 1	<ul style="list-style-type: none"> • <i>I have used it <u>for</u> about <u>2</u> years <u>now</u> ...</i> • <i>... been using <u>for</u> about <u>10</u> years <u>now</u>.</i> 	Retention
+E10	{stay/all/day}: 6	<ul style="list-style-type: none"> • <i>It glides on and <u>stays all day</u> ...</i> 	Experiential
+E11	{a bit/pricey/but}: 5, {a little/pricey/ but}: 1	<ul style="list-style-type: none"> • <i>It is <u>a bit pricey</u> but I pay for the quality.</i> 	Dis/satisfaction
+E12	{easy/work/with}: 5	<ul style="list-style-type: none"> • <i>... the tip is very <u>easy to work with</u> ...</i> 	Experiential
+E13	{will/repurchase/again}: 5, {will/def/again}: 1	<ul style="list-style-type: none"> • <i><u>Will repurchase again</u> and again!</i> 	Retention
+E14	{rich/dark/brown}: 2, {deep/dark/color}: 2, {deep/dark/brown}: 2, {rich/reddish/brown}: 1, {rich/deep/brown}: 1, {rich/deep/black}: 1, {rich/dark/color}: 1, {rich/dark/black}: 1, {rich/deep/color}: 1, {deep/dark/black}: 1, {dark/reddish/brown}: 1	<ul style="list-style-type: none"> • <i>It's a <u>rich, dark reddish brown</u> with just a hint of gold.</i> 	Experiential
+E15	{thing/about/this product}: 2, {thing/about/this liner}: 1, {thing/about/this eyeliner}: 1, {thing/about/gel liner}: 1	<ul style="list-style-type: none"> • <i>One of my fave <u>thing about this liner</u> ...</i> 	General / Ambiguous
+E16	{glide/on/smooth}: 6, {glide/on/quick}: 1	<ul style="list-style-type: none"> • <i>It <u>glides on smoothly</u> and is very black.</i> 	Experiential
+E17	{use/angle/brush}: 3, {use/slant/brush}: 1, {apply/angle/brush}: 1	<ul style="list-style-type: none"> • <i>... I <u>use an angled eyeliner brush</u> ...</i> 	General / Ambiguous

Table 4.5. (continued)

Ref.	Collocation Group	Examples from the +E Subset	Info. Type
+E18	{with/makeup/remover}: 6	<ul style="list-style-type: none"> ... <i>come off</i> <u>with</u> <u>makeup remover</u> so you dont have to tug ... 	Experiential
+E19	{make/eye/pop}: 5	<ul style="list-style-type: none"> ... and <u>made</u> my <u>eyes pop</u>. 	Experiential
+E20	{thin/thick/line}: 5	<ul style="list-style-type: none"> ... allows you to draw very <u>thin</u> to very <u>thick lines</u> ... 	Experiential
+E21	{for/long/time}: 10	<ul style="list-style-type: none"> ... lasts <u>for</u> a <u>long time</u>. 	Experiential
+E22	{on/oily/lid}: 4, {on/oily/eyelid}: 4, {on/oily/skin}: 1	<ul style="list-style-type: none"> ... great even <u>on</u> my sometimes <u>oily eyelids</u>. 	General / Ambiguous
+E23	{around/8/hour}: 1, {around/5-6/hour}: 1, {about/8/hour}: 1, {about/4-6/hour}: 1, {about/12/hour}: 1	<ul style="list-style-type: none"> It lasts pretty well on my waterline, <u>about 4-6 hours</u>. it lasts me <u>about 8 hours</u> ... 	Experiential
+E24	{nice/dark/black}: 3, {nice/dark/brown}: 2, {nice/dark/colour}: 1, {nice/dark/color}: 1	<ul style="list-style-type: none"> It's color is just right, <u>nice dark brown</u>. 	Experiential
+E25	{in/small/purse}: 1, {in/small/bottle}: 1, {in/little/pot}: 1, {in/a little/jar}: 1, {in/a little/glass}: 1	<ul style="list-style-type: none"> It comes <u>in</u> a very <u>small bottle</u> ... 	General / Ambiguous

Table 4.6. Collocation groups from negative cleanser reviews (-C)

Ref.	Collocation Group	Examples from the -C Subset	Info. Type
-C1	{same/thing/happen}: 4, {exact/thing/happen}: 2	<ul style="list-style-type: none"> • ... <u>returned to using it within a week, <u>same thing happened.</u></u> 	Experiential
-C2	{at/same/time}: 5	<ul style="list-style-type: none"> • .. <u>same thing to my husband <u>at exactly the <u>same time!</u></u></u> • .. <u>dry, and oily <u>at the <u>same time.</u></u></u> 	General / Ambiguous
-C3	{wake up/next/morning}: 5	<ul style="list-style-type: none"> • <u>when i <u>woke up the <u>next morning</u> i had tine white bumps ...</u></u> • ... <u>when I <u>woke up the <u>next morning</u> my face was horrible.</u></u> 	Experiential
-C4	{read/many/review}: 3, {read/sooo/review}: 1, {read/million/review}: 1, {read/a lot/review}: 1	<ul style="list-style-type: none"> • I had <u>read so <u>many</u> good <u>reviews</u> of this face wash, and was really excited to try it.</u> 	Expectation
-C5	{hard/wash/off}: 2, {hard/rinse/off}: 1, {hard/rid/off}: 1, {difficult/rinse/off}: 1	<ul style="list-style-type: none"> • Plus it is very <u>hard to <u>wash off,</u> it sticks ...</u> 	Experiential
-C6	{only/thing/about}: 9	<ul style="list-style-type: none"> • <u>the <u>only thing</u> i like <u>about it ...</u></u> • <u>The <u>only good thing</u> <u>about the product ...</u></u> 	Dis/satisfaction
-C7	{hear/great/thing}: 4, {read/good/thing}: 1, {hear/good/thing}: 1	<ul style="list-style-type: none"> • I was so disappointed as I've <u>heard <u>great things.</u></u> 	Expectation
-C8	{not like/at/all}: 5	<ul style="list-style-type: none"> • I really <u><u>don't like this <u>at all.</u></u></u> 	Dis/satisfaction
-C9	{itch/like/crazy}: 4, {sting/like/crazy}: 2	<ul style="list-style-type: none"> • ... it make me <u><u>itch like crazy,</u> so I stopped using it.</u> 	Experiential
-C10	{will/never buy/again}: 3, {will/not buy/again}: 2, {will/not purchase/again}: 1	<ul style="list-style-type: none"> • <u><u>Would not purchase <u>again.</u></u></u> 	Retention
-C11	{not take/makeup/off}: 2, {not get/makeup/off}: 1, {not clean/makeup/off}: 1, {never take/make up/off}: 1	<ul style="list-style-type: none"> • <u><u>Didn't take <u>makeup off</u> very well either.</u></u> 	Experiential
-C12	{someone/with/sensitive}: 2, {people/with/sensitive}: 2, {people/with/non-sensitive}: 1, {lady/with/sensitive}: 1, {for/with/non-sensitive}: 1, {anyone/with/sensitive}: 1	<ul style="list-style-type: none"> • <u>attention <u>ladies with sensitive skin:</u> this product is not for you!</u> 	General / Ambiguous
-C13	{at/first/think}: 5	<ul style="list-style-type: none"> • <u><u>At first I thought</u> this was a great product ...</u> 	Experiential
-C14	{throw/in/trash}: 2, {throw/in/bin}: 2, {toss/in/trash}: 1, {throw/in/garbage}: 1, {chuck/in/bin}: 1, {throw/in/sink}: 1	<ul style="list-style-type: none"> • ... I <u><u>threw it in the garbage.</u></u> 	Dis/satisfaction
-C15	{high/hope/this product}: 3, {high/for/this product}: 2, {high/for/this cleanser}: 2, {high/hope/this cleanser}: 1	<ul style="list-style-type: none"> • ... <u><u>high hopes for this product...but those dreams were crushed.</u></u> 	Expectation

Table 4.6. (continued)

Ref.	Collocation Group	Examples from the -C Subset	Info. Type
-C16	{little/red/bump}: 6, {tiny/red/bump}: 2, {little/red/pimple}: 2, {tiny/flesh-coloured/bump}: 1, {tiny/color/bump}: 1, {little/red/itchy}: 1, {little/flesh-coloured/bump}: 1	<ul style="list-style-type: none"> • <i>But after about a week my skin felt dry....and then the <u>little red bumps</u> came.</i> 	Experiential
-C17	{not remove/makeup/well}: 4, {not take/makeup/well}: 1, {dose not remove/mascara/well}: 1	<ul style="list-style-type: none"> • <i>It does <u>not remove makeup</u> very <u>well</u> ...</i> 	Experiential
-C18	{may/work/for}: 2, {will/work/someone}: 1, {probable/work/people}: 1, {may/work/someone}: 1, {may/work/people}: 1	<ul style="list-style-type: none"> • <i>I do have sensitive skin though so this <u>may work for</u> others.</i> 	Dis/satisfaction
-C19	{make/feel/tight}: 2, {leave/feel/tight}: 2, {make it/feel/tight}: 1, {make/feel/stiff}: 1, {make/feel/smooth}: 1, {leave/feel/gritty}: 1	<ul style="list-style-type: none"> • <i>... left a film on my skin that <u>made it feel tight</u> all day.</i> 	Experiential
-C20	{dry/skin/out}: 7, {dry/face/out}: 4	<ul style="list-style-type: none"> • <i>... just <u>dried</u> my <u>skin out</u>.</i> 	Experiential
-C21	{face/like/no other}: 2, {skin/like/nothing else}: 1, {skin/like/no other}: 1, {pore/like/no other}: 1	<ul style="list-style-type: none"> • <i>... dried my <u>skin out like no other!</u></i> • <i>... clogged my <u>pores like no other</u>.</i> 	Experiential
-C22	{little/red/all over}: 3, {tiny/red/all over}: 1, {little/white/all over}: 1	<ul style="list-style-type: none"> • <i>and <u>little RED</u> bumps <u>all over</u> my face ...</i> 	Experiential
-C23	{2/week/now}: 4, {3/week/now}: 1	<ul style="list-style-type: none"> • <i>... after stopping use for <u>2 weeks now</u> ...</i> 	General / Ambiguous
-C24	{after/a few/use}: 3, {after/one/use}: 2	<ul style="list-style-type: none"> • <i><u>After a few uses</u> I don't find any difference in acne ...</i> 	Experiential
-C25	{real/want/like}: 9, {real/want/love}: 2, {real/hope/like}: 2	<ul style="list-style-type: none"> • <i>I was <u>really hoping</u> I would <u>like</u> it, too, so I am very disappointed.</i> 	Dis/satisfaction

Table 4.7. Collocation groups from positive cleanser reviews (+C)

Ref.	Collocation Group	Examples from the +C Subset	Info. Type
+C1	{go/long/way}: 7	<ul style="list-style-type: none"> • <i>And a little <u>goes</u> a <u>long way</u> too!</i> 	Experiential
+C2	{more/expensive/than}: 5, {even/expensive/than}: 1	<ul style="list-style-type: none"> • <i>... now its even <u>more expensive</u> <u>than</u> MAC.</i> • <i>It is <u>more expensive</u> <u>than</u> I'd like, but it lasts ...</i> 	General / Ambiguous
+C3	{will/repurchase/again}: 5	<ul style="list-style-type: none"> • <i>... <u>will repurchase</u> <u>again</u> and again.</i> 	Retention
+C4	{only/need/tiny}: 5, {only/need/small}: 1, {only/need/pea-size}: 1, {only/need/little}: 1	<ul style="list-style-type: none"> • <i>You <u>only need</u> a <u>pea-size</u> to wash your face.</i> 	Experiential
+C5	{follow/up/with}: 5	<ul style="list-style-type: none"> • <i>I <u>follow up</u> <u>with</u> a toner and moisterizer.</i> 	General / Ambiguous
+C6	{without/dry/out}: 5	<ul style="list-style-type: none"> • <i>cleans my combo skin <u>without</u> <u>drying</u> it <u>out</u> or leaving a residue.</i> 	Experiential
+C7	{not irritate/at/all}: 6	<ul style="list-style-type: none"> • <i><u>Doesn't</u> irritate my skin <u>at all</u>.</i> 	Experiential
+C8	{on/second/bottle}: 2, {on/5th/bottle}: 2, {on/2nd/bottle}: 2, {onto/second/tub}: 1, {on/third/tub}: 1, {on/third/bottle}: 1, {on/4th/tube}: 1, {on/30th/bottle}: 1, {on/12th/bottle}: 1	<ul style="list-style-type: none"> • <i>This has become my hg cleanser, I'm <u>on</u> my <u>second</u> <u>bottle</u>.</i> • <i>... every night and I think I am <u>on</u> my <u>4th</u> <u>tube</u>.</i> 	Retention
+C9	{look/for/cleanser}: 3, {look/for/face wash}: 1, {look/for/face cleanser}: 1	<ul style="list-style-type: none"> • <i>I was cruising the skincare isle <u>looking for</u> a new <u>face wash</u> ...</i> 	General / Ambiguous
+C10	{not make/feel/tight}: 3, {not leave/feel/tight}: 1, {never make/feel/tight}: 1	<ul style="list-style-type: none"> • <i><u>doesn't</u> make my face <u>feel</u> <u>tight</u> or anything, which is good.</i> 	Experiential
+C11	{for/6/now}: 2, {for/two/now}: 1, {for/8/now}: 1, {for/4/now}: 1, {for/2/now}: 1, {for/17/now}: 1, {for/15/now}: 1s	<ul style="list-style-type: none"> • <i>... I've had it <u>for</u> over <u>6</u> months <u>now</u> i believe.</i> • <i>I've been using it <u>for</u> <u>15</u> years <u>now</u>.</i> 	General / Ambiguous
+C12	{separate/makeup/remover}: 4, {separate/make-up/remover}: 1	<ul style="list-style-type: none"> • <i>... I don't use <u>separate</u> eye <u>make-up</u> <u>remover</u>.</i> • <i>I do use a <u>separate</u> eye <u>makeup</u> <u>remover</u> to get mascara off ...</i> 	General / Ambiguous
+C13	{will/recommend/anyone}: 4, {will/recommend/everyone}: 1	<ul style="list-style-type: none"> • <i>I <u>would</u> <u>recommend</u> this to <u>anyone</u>.</i> 	Word of Mouth
+C14	{for/long/time}: 11	<ul style="list-style-type: none"> • <i>... will be sticking with this one <u>for</u> a <u>long</u> <u>time</u>.</i> • <i>... been using this <u>for</u> a <u>long</u> <u>time</u> & will continue to use it.</i> 	Retention
+C15	{cleansers/out/there}: 3, {products/out/there}: 2, {product/out/there}: 1	<ul style="list-style-type: none"> • <i>This is one of the few <u>products</u> <u>out</u> <u>there</u> that ...</i> 	General / Ambiguous
+C16	{come/in/pump}: 3, {come/in/bottle}: 2, {come/in/container}: 1, {come/in/tube}: 1	<ul style="list-style-type: none"> • <i>These towelettes <u>come</u> <u>in</u> a plastic <u>container</u> ...</i> 	General / Ambiguous

Table 4.7. (continued)

Ref.	Collocation Group	Examples from the +C Subset	Info. Type
+C17	{mascara/without/have to}: 2, {makeup/without/have to}: 1, {make-up/without/have to}: 1, {liner/without/have to}: 1	<ul style="list-style-type: none"> • <i>it takes off all of my eye <u>makeup</u> <u>without having to scrub</u> around ...</i> 	Experiential
+C18	{great/job/remove}: 4, {good/job/remove}: 1, {good/job/clean}: 1, {fabulous/job/remove}: 1	<ul style="list-style-type: none"> • <i>It's really gentle, but does a <u>great job removing</u> all my makeup.</i> 	Experiential
+C19	{since/start/use}: 3, {just/start/use}: 3	<ul style="list-style-type: none"> • <i>... haven't had any breakouts <u>since</u> I <u>started using</u> this product.</i> • <i>I <u>just started using</u> it (its been about a month) and could not be happier!</i> 	Experiential
+C20	{try/many/other}: 3, {try/hundred/other}: 1, {try/a lot/other}: 1	<ul style="list-style-type: none"> • <i>I've <u>tried many other</u> brands of face wash, towelettes, pads, etc and I have yet to find anything better.</i> 	Dis/satisfaction
+C21	{makeup/include/mascara}: 2, {makeup/include/makeup}: 1, {makeup/include/liner}: 1, {makeup/include/eyemakeup}: 1, {makeup/except/mascara}: 1	<ul style="list-style-type: none"> • <i>Removes all of my <u>makeup</u>, <u>including mascara</u></i> 	Experiential
+C22	{take/makeup/off}: 7, {take/make up/off}: 2, {wipe/makeup/off}: 1, {wash/make-up/off}: 1, {take/mascara/off}: 1, {remover/mascara/off}: 1, {get/make up/off}: 1, {take/make-up/off}: 1, {remove/ makeup/from}: 1, {get/mascara/off}: 1, {get/makeup/off}: 1,	<ul style="list-style-type: none"> • <i>This cleaner <u>takes</u> my <u>make-up off</u> without me having to ...</i> • <i>It <u>takes</u> heavy duty <u>mascara off</u> really easily ...</i> 	Experiential
+C23	{will/continue/use}: 8	<ul style="list-style-type: none"> • <i>I <u>will continue</u> to <u>use</u> and repurchase.</i> 	Retention
+C24	{love/love/love}: 33, {love/love/lovee}: 1	<ul style="list-style-type: none"> • <i>I <u>love love LOVE</u> this face wash!</i> 	Dis/satisfaction
+C25	{up/very/well}: 2, {up/real/well}: 1, {up/real/nice}: 1, {lather/very/well}: 1, {lather/very/nice}: 1, {lather/so/well}: 1	<ul style="list-style-type: none"> • <i>You only need a very little bit for it to lather <u>up really nicely</u>.</i> 	Experiential

Table 4.8. Collocation groups from negative moisturizer reviews (-M)

Ref.	Collocation Group	Examples from the -M Subset	Info. Type
-M1	{read/many/review}: 3, {read/all/review}: 3	• <i>I was surprised, since I had <u>read</u> so <u>many</u> rave <u>reviews</u> ...</i>	Expectation
-M2	{high/hope/for}: 9	• <i>I had <u>high</u> <u>hopes</u> <u>for</u> this ...</i>	Expectation
-M3	{will/not use/again}: 3, {will/never use/again}: 2	• <i>I'll <u>never</u> <u>use</u> this <u>again</u>.</i>	Retention
-M4	{break out/like/crazy}: 8, {start/like/crazy}: 1, {out/like/crazy}: 1, {break out/like/crazyyy}: 1, {break/like/crazy}: 1	• <i>This made me <u>break out</u> <u>like</u> <u>crazy</u>!</i> • <i>... eyes felt like they were on fire and <u>started</u> watering <u>like</u> <u>crazy</u>.</i>	Experiential
-M5	{not moisturize/at/all}: 5, {not hydrate/at/all}: 3, {hard moisturize/at/all}: 1	• <i>This is <u>not</u> <u>hydrating</u> <u>at</u> <u>all</u>.</i>	Experiential
-M6	{for/2/day}: 3, {for/two/day}: 2, {for/5/day}: 2, {for/3/day}: 2, {for/three/day}: 1, {for/8/day}: 1, {for/7/day}: 1, {for/4/day}: 1, {for/30/day}: 1, {for/20/min}: 1, {for/2-3/day}: 1	• <i>... only used it <u>for</u> <u>2</u> <u>days</u> to see ...</i> • <i>... decided to buy the positively radiant moisturizer and have used it <u>for</u> <u>4</u> <u>days</u>.</i>	General / Ambiguous
-M7	{absorb/into/skin}: 6	• <i>It doesn't even get <u>absorbed</u> <u>into</u> the <u>skin</u> ...</i>	Experiential
-M8	{there be/products/there}: 2, {there be/products/out}: 2, {there/products/out}: 1 {there/products/there}: 1,	• <i>... <u>there</u> <u>are</u> better <u>products</u> out <u>there</u>.</i>	General / Ambiguous
-M9	{little/red/bump}: 3, {small/red/bump}: 2, {tiny/red/pimple}: 1, {little/white/bump}: 1	• <i>... got <u>little</u> <u>red</u> irritation <u>bumps</u> after each use.</i>	Experiential
-M10	{6/new/pimple}: 2, {8/new/pimple}: 1, {5/new/blemish}: 1, {10/new/zit}: 1	• <i>... 2 days of using it I got <u>8</u> <u>new</u> <u>pimples</u> for my troubles.</i>	Experiential
-M11	{skin/turn/red}: 4, {face/turn/red}: 3, {cheek/turn/red}: 1	• <i>It burned my <u>skin</u> and I <u>turned</u> <u>red</u>.</i>	Experiential
-M12	{break out/in/rash}: 4, {break out/in/pimple}: 2, {break out/in/cystic}: 1, {break out/in/cyst}: 1, {break out/in/bump}: 1, {break out/in/blotchy}: 1, {break out/in/blotch}: 1	• <i>This lotion made me <u>break out</u> <u>in</u> a <u>rash</u>.</i>	Experiential
-M13	{someone/with/skin}: 4, {people/with/skin}: 2, {anyone/with/skin}: 2, {for/with/skin}: 1, {somebody/with/skin}: 1, {everyone/with/skin}: 1, {anyone/with/pore}: 1	• <i>I have dry skin so I can only imagine how <u>somebody</u> <u>with</u> <u>oily</u> <u>skin</u> would cope!</i> • <i>All this is from <u>someone</u> <u>with</u> <u>dry</u> <u>skin</u>.</i>	General / Ambiguous
-M14	{use/small/amount}: 4, {use/tiny/amount}: 1, {apply/tiny/amount}: 1	• <i>I <u>used</u> a <u>small</u> <u>amount</u>, applied all over my face ...</i>	General / Ambiguous
-M15	{make/feel/greasy}: 2, {leave/feel/greasy}: 2, {make it/feel/gross}: 1, {make/feel/oily}: 1	• <i>It <u>leaves</u> my skin <u>feeling</u> <u>greasy</u> ...</i>	Experiential

Table 4.8. (continued)

Ref.	Collocation Group	Examples from the -M Subset	Info. Type
-M16	{sit/on top of/skin}: 4, {sit/on/skin}: 4, {sit/on top of/face}: 1	• <i>This moisturizer <u>sits on top of</u> my <u>skin</u> ...</i>	Experiential
-M17	{many/people/love}: 1, {many/people/ like}: 1, {a lot/woman/like}: 1, {a lot/ people/love}: 1, {a lot/people/like}: 1	• <i>Apparently <u>a lot of women</u> really <u>like</u> it, but I'm not ...</i>	Expectation
-M18	{real/want/like}: 11, {real/want/love}: 4	• <i>I <u>really wanted to like</u> this after reading all the great reviews.</i>	Dis/satisfaction
-M19	{face/feel/tight}: 3, {skin/feel/tight}: 1, {skin/feel/bumpy}: 1	• <i>... my <u>skin</u> was <u>feeling bumpy</u>.i realized this after using ...</i>	Experiential
-M20	{look/like/oil}: 3, {smell/like/plastic}: 1, {look/like/grease}: 1	• <i>... made my face <u>look like</u> an <u>oil slick</u></i>	Experiential
-M21	{for/sensitive/skin}: 5	• <i>Prefer: Olay lotion <u>for sensitive skin</u></i>	General / Ambiguous
-M22	{use/at/night}: 4, {use/in/morning}: 2, {apply/at/night}: 1	• <i>I <u>applied</u> MSE <u>at night</u> ...</i>	General / Ambiguous
-M23	{break/so/bad}: 4, {break/soooo/bad}: 1	• <i>It <u>broke</u> me out <u>soooo Bad</u> ...</i>	Experiential
-M24	{in/giant/pimple}: 2, {in/large/pimple}: 1, {in/large/cystic}: 1, {in/deep/pimple}: 1	• <i>my skin broke out <u>in giant</u> giant painful <u>pimples</u> ...</i>	Experiential
-M25	{someone/with/prone}: 1, {someone/acne/prone}: 1, {people/acne/prone}: 1, {for/with/prone}: 1, {for/acne/prone}: 1	• <i>... for <u>people</u> with <u>acne prone</u> skin.</i>	General / Ambiguous

Table 4.9. Collocation groups from positive moisturizer reviews (+M)

Ref.	Collocation Group	Examples from the +M Subset	Info. Type
+M1	{go/long/way}: 17	• <i>A little bit goes a <u>long way</u> ...</i>	Experiential
+M2	{will/repurchase/again}: 4, {definite/repurchase/again}: 3	• <i><u>Will</u> definitely <u>repurchase</u> <u>again</u>.</i>	Retention
+M3	{soak/into/skin}: 10	• <i>... it <u>soaked into</u> my <u>skin</u> and didn't feel wet, but silky smooth.</i>	Experiential
+M4	{for/about/now}: 4, {for/some/now}: 1, {for/almost/now}: 1	• <i>... been using Josie Maran's Argan oil <u>for about</u> a month <u>now</u> and I love it.</i>	Retention
+M5	{on/third/tube}: 1, {on/third/bottle}: 1, {on/second/jar}: 1, {on/second/bottle}: 1, {on/5th/jar}: 1, {on/4th/bottle}: 1, {on/3rd/jar}: 1, {on/3rd/bottle}: 1	• <i>I think I am <u>on</u> my <u>5th jar</u>.</i>	Retention
+M6	{last/5/year}: 2, {past/2/year}: 1, {past/2/month}: 1, {last/three/year}: 1, {last/couple/month}: 1, {last/4/year}: 1, {last/30/year}: 1	• <i>I've been using it for the <u>past 2 years</u> on my face, body, and hair.</i> • <i>I will keep purchasing as I have been for the <u>last 4</u> or <u>5 years</u>.</i>	Retention
+M7	{someone/with/dry}: 2, {anyone/with/dry}: 2, {people/with/dry}: 1	• <i>I recommend this to <u>anyone with dry</u> skin!</i> • <i>This is definitenly not for <u>someone with dry</u> skin though!</i>	General / Ambiguous
+M8	{absorb/very/quick}: 2, {absorb/so/fast}: 2, {absorb/real/quick}: 1, {absorb/pretty/fast}: 1	• <i>It <u>absorbs</u> really <u>quick</u> and doesn't feel greasy ...</i>	Experiential
+M9	{in/with/this product}: 2, {in/love/this product}: 2, {in/love/this lotion}: 1, {in/love/this cream}: 1	• <i>I am <u>in love with this product</u> and i have yet to find a better moisturizer ...</i>	Dis/satisfaction
+M10	{in/cold/month}: 2, {in/coole/month}: 1, {during/cold/season}: 1, {during/cold/month}: 1	• <i>... feeling nice and moisturized <u>during the cold seasons</u>!</i> • <i>... Number 1 moisturiser <u>in the cooler months</u> ...</i>	Experiential
+M11	{skin/feel/supple}: 1, {skin/feel/soft}: 1, {skin/feel/smooth}: 1, {skin/feel/silky}: 1, {hair/feel/soft}: 1	• <i>I did notice my <u>skin feeling smoother</u> ...</i>	Experiential
+M12	{well/under/foundation}: 2, {great/under/makeup}: 2, {wonderful/under/makeup}: 1, {well/underneath/makeup}: 1, {well/under/makeup}: 1, {well/under/make up}: 1, {good/under/makeup}: 1	• <i>It's <u>wonderful under</u> my <u>makeup</u>.</i>	Experiential
+M13	{about/two/ago}: 2, {about/2/ago}: 2, {almost/2/ago}: 1, {about/five/ago}: 1	• <i>... bought it <u>about two weeks ago</u></i> • <i>Then <u>about 2 years ago</u> I received this as a deluxe sample ...</i>	General / Ambiguous
+M14	{so/many/different}: 5	• <i>I have tried <u>so many different</u> types and this is by far my favourite.</i>	Dis/satisfaction

Table 4.9. (continued)

Ref.	Collocation Group	Examples from the +M Subset	Info. Type
+M15	{will/continue/purchase}: 2, {will/keep on/purchase}: 1, {will/keep/purchase}: 1, {will/keep/buy}: 1	<ul style="list-style-type: none"> • <i>Will <u>keep on purchasing</u>.</i> 	Retention
+M16	{sink/in/well}: 4, {sink/into/well}: 1	<ul style="list-style-type: none"> • <i>It <u>sinks in well</u> and moisturizes perfectly.</i> 	Experiential
+M17	{night/after/cleanse}: 3, {tonight/after/shower}: 1, {night/after/wash}: 1, {morning/after/cleanser}: 1	<ul style="list-style-type: none"> • <i>At <u>night, after I wash</u> my face, I put it on before ...</i> 	General / Ambiguous
+M18	{by/far/good}: 2, {by/far/favorite}: 2, {by/far/favourite}: 1	<ul style="list-style-type: none"> • <i>This is <u>by far the best</u> moisturizer I have every used.</i> 	Dis/satisfaction
+M19	{over/3/year}: 2, {over/six/year}: 1, {over/6/month}: 1, {over/5/year}: 1, {over/2/year}: 1, {over/15/year}: 1, {over/10/year}: 1	<ul style="list-style-type: none"> • <i>... been using this face cream for <u>over 5 years</u> ...</i> • <i>... been using this product for <u>over 10 years</u>.</i> 	Retention
+M20	{oily/acne/prone}: 4, {dry/acne/prone}: 2	<ul style="list-style-type: none"> • <i>... recommended this to me because of my <u>dry, acne prone</u> skin.</i> 	General / Ambiguous
+M21	{only/can/use}: 5	<ul style="list-style-type: none"> • <i>This is the <u>ONLY</u> moisturizer i <u>can use</u> in the world.</i> • <i>... it was the <u>only</u> mositurizer I <u>could use</u> that wouldn't irrate ...</i> 	Dis/satisfaction
+M22	{since/start/use}: 4, {once/start/use}: 1	<ul style="list-style-type: none"> • <i>... softer & radiant looking <u>since</u> I've <u>started using</u> this.</i> 	Experiential
+M23	{moisturizer/with/spf}: 2, {products/with/spf}: 1, {product/with/spf}: 1, {lotions/with/spf}: 1, {cream/with/spf}: 1	<ul style="list-style-type: none"> • <i>I also discovered that a <u>moisturizer with SPF 15</u> was needed to ...</i> 	General / Ambiguous
+M24	{dry/in/winter}: 7, {oily/in/daytime}: 1, {hydrate/during/winter}: 1, {dry/during/winter}: 1, {dry/by/midday}: 1, {dry/at/night}: 1, {dry/at/evening}: 1, {dehydrate/during/winter}: 1	<ul style="list-style-type: none"> • <i>I have oily skin in the daytime which is <u>dry as hell at evenings</u>.</i> 	General / Ambiguous
+M25	{but/always/come back}: 5, {but/keep/come back}: 1, {but/alway/come back}: 1	<ul style="list-style-type: none"> • <i>I've actually tried other things just to see if there's anything better out there, <u>but I keep coming back</u> to this.</i> 	Retention

Table 4.10. Collocation groups from negative television reviews (-T)

Ref.	Collocation Group	Examples from the -T Subset	Info. Type
-T1	{class/action/suit}: 4, {class/action/lawsuit}: 2	<ul style="list-style-type: none"> • <i>I wish someone would get a <u>class action suit</u> against LG.</i> 	Dis/satisfaction
-T2	{make/buzz/noise}: 2, {make/pop/noise}: 1, {make/click/sound}: 1, {make/buzz/sound}: 1	<ul style="list-style-type: none"> • <i>... it would just <u>make</u> this weird <u>buzzing noise</u>.</i> 	Experiential
-T3	{not stand/behind/products}: 3, {not stand/behind/product}: 3	<ul style="list-style-type: none"> • <i>... Sony would <u>not stand behind</u> their <u>product</u> ...</i> 	Dis/satisfaction
-T4	{white/glove/delivery}: 7	<ul style="list-style-type: none"> • <i>... because of their <u>white glove delivery</u> service promise.</i> 	General / Ambiguous
-T5	{power/supply/board}: 6	<ul style="list-style-type: none"> • <i>... the <u>power supply board</u> needs replacement!!!</i> • <i>... seriously wrong in the <u>power supply board</u> of this TV.</i> 	Experiential
-T6	{authorize/repair/cent}: 5	<ul style="list-style-type: none"> • <i>I had an <u>authorized Toshiba repair center</u> come out to look ...</i> • <i>... no response from <u>authorized repair center</u> after 3 calls.</i> 	Experiential
-T7	{sound/but/no picture}: 7	<ul style="list-style-type: none"> • <i><u>Sound</u> still works <u>but no picture</u> at all.</i> 	Experiential
-T8	{white/line/appear}: 1, {white/dot/appear}: 1, {red/line/develop}: 1, {green/line/appear}: 1, {color/shape/appear}: 1, {color/line/appear}: 1	<ul style="list-style-type: none"> • <i>Then last week two vertical <u>white lines</u> <u>appeared</u> on either side ...</i> 	Experiential
-T9	{in/dec./2005}: 2, {on/dec./28th}: 1, {on/dec./27}: 1, {on/dec./2006}: 1	<ul style="list-style-type: none"> • <i>I purchased this tv for \$1,300.00 <u>on Dec. 27</u>, 2006, on Feb. 28 ...</i> 	General / Ambiguous
-T10	{still/under/warranty}: 8	<ul style="list-style-type: none"> • <i>It was <u>still under warranty</u> and called customer service.</i> 	Experiential
-T11	{happy/with/until}: 5	<ul style="list-style-type: none"> • <i>... <u>happy with this TV until</u> yesterday when it failed.</i> 	Dis/satisfaction
-T12	{as/mention/someone}: 2, {as/by/someone}: 2, {as/note/person}: 1	<ul style="list-style-type: none"> • <i>BTW, <u>as mentioned by someone</u> before ...</i> • <i>I bought this tv from sears and <u>as noted</u> by another <u>person</u> ...</i> 	General / Ambiguous
-T13	{up/leave/corner}: 1, {top/leave/leave}: 1, {top/leave/corner}: 1, {top/leave/bottom}: 1, {top/bottom/leave}: 1, {side/top/bottom}: 1, {leave/right/side}: 1, {leave/bottom/right}: 1, {leave/bottom/leave}: 1, {bottom/leave/bottom}: 1	<ul style="list-style-type: none"> • <i>... high-pitched hum emanating from the back of the unit in the <u>upper left corner</u>.</i> • <i>... huge black bar missing from the picture in the <u>top left corner</u>.</i> 	Experiential
-T14	{screen/go/black}: 8, {screen/go/green}: 1	<ul style="list-style-type: none"> • <i>The <u>screen</u> just <u>went green</u> when using HDMI.</i> 	Experiential
-T15	{after/warranty/expire}: 6, {right/warranty/expire}: 1, {after/warrantee/expire}: 1	<ul style="list-style-type: none"> • <i>This TV worked great until shortly <u>after the warranty expired</u>.</i> 	Experiential

Table 4.10. (continued)

Ref.	Collocation Group	Examples from the -T Subset	Info. Type
-T16	{work/fine/for}: 5, {work/great/for}: 4, {work/perfect/for}: 1	<ul style="list-style-type: none"> • <i>It <u>worked fine</u> for about 6 months then ...</i> 	Experiential
-T17	{bad/customer/service}: 4, {horrible/ customer/service}: 3, {poor/customer/ service}: 3, {lousy/customer/ service}: 1, {superior/customer/ service}: 1, {terrible/customer/ service}: 1, {crappy/customer/service}: 1	<ul style="list-style-type: none"> • <i>... <u>crappy</u> products and <u>horrible</u> <u>customer service</u>.</i> 	Experiential
-T18	{vertical/white/line}: 1, {vertical/red/ line}: 1, {vertical/green/line}: 1, {vertical/color/band}: 1, {horizontal/ blurry/bar}: 1	<ul style="list-style-type: none"> • <i>... two <u>vertical white lines</u> appeared on either side ...</i> 	Experiential
-T19	{black/at/top}: 3, {red/at/top}: 1, {blue/at/bottom}: 1	<ul style="list-style-type: none"> • <i>Tv then had a <u>red band at top</u> of screen.</i> 	Experiential
-T20	{brand/new/xbr6}: 1, {brand/new/xbr}: 1, {brand/new/ln-t4661f}: 1, {brand/ new/lg}: 1, {brand/new/bravia}: 1	<ul style="list-style-type: none"> • <i>We bought a <u>brand new</u> Sony <u>Bravia XBR</u> this past May ...</i> 	General / Ambiguous
-T21	{spot/on/screen}: 5	<ul style="list-style-type: none"> • <i>... came with dark <u>spots on the screen</u> that will not go away.</i> 	Experiential
-T22	{start/blue/dot}: 3, {start/red/line}: 1, {start/black/line}: 1	<ul style="list-style-type: none"> • <i>... <u>started</u> to get <u>black lines</u> across the screen.</i> 	Experiential
-T23	{call/customer/support}: 5, {contact/customer/support}: 1	<ul style="list-style-type: none"> • <i>... <u>called customer support</u> and they were NOT willing to ...</i> 	Experiential
-T24	{tv/stop/work}: 7, {this tv/stop/work}: 1	<ul style="list-style-type: none"> • <i>... the <u>TV stopped working</u>, and could not be repaired.</i> 	Experiential
-T25	{picture/quality/great}: 2, {picture/quality/good}: 2, {image/quality/great}: 2	<ul style="list-style-type: none"> • <i>The <u>image quality</u> was <u>great</u>, but only for one year and two days.</i> • <i>It looks like a good deal (and the <u>image quality is great</u>) ...</i> 	Experiential

Table 4.11. Collocation groups from positive television reviews (+T)

Ref.	Collocation Group	Examples from the +T Subset	Info. Type
+T1	{in/perfect/condition}: 7, {in/excellent/condition}: 1 {in/good/condition}: 1,	• ... <i>arrived earlier than expected and <u>in perfect condition</u>.</i>	Experiential
+T2	{progressive/scan/dvd}: 5	• ... <i>for hi-def TV and <u>progressive scan DVD</u>.</i>	General / Ambiguous
+T3	{if/look/for}: 12	• <i>If you are <u>looking for</u> a smaller flat panel LCD ...</i>	General / Ambiguous
+T4	{can/not say/about}: 2, {can/not comment/about}: 2, {can/not comment/on}: 1	• <i>I <u>can not comment on</u> the sound as I have never used it ...</i>	General / Ambiguous
+T5	{at/same/time}: 6	• ... <i>new LCD TV's that came out <u>at the same time</u> ...</i>	General / Ambiguous
+T6	{for/month/now}: 12, {for/week/now}: 6, {for/year/now}: 2	• <i>So far no problems and I have had this one <u>for a few months now</u>.</i>	Experiential
+T7	{sit/foot/away}: 3, {stand/foot/away}: 1, {sit/foot/from}: 1	• <i>I <u>sit</u> about 9 <u>feet away</u>.</i> • <i>I <u>sit</u> about 12 <u>feet from</u> the TV ...</i>	General / Ambiguous
+T8	{build/in/speaker}: 7	• <i>I don't use the <u>built in speakers</u> much ...</i>	Experiential
+T9	{spend/month/research}: 3, {spend/week/research}: 1, {spend/long/research}: 1	• <i>I <u>spent months researching</u> these televisions, and finally ...</i>	Expectation
+T10	{no problem/at/all}: 4, {no trouble/at/all}: 1	• <i>I have had <u>no troubles at all</u>.</i>	Dis/satisfaction
+T11	{home/theat/system}: 5, {home/theatre/system}: 2, {home/theatre/setup}: 1, {home/theatre/optical}: 1, {home/theat/setup}: 1	• <i>If you do not have a <u>home theater</u> audio system ...</i> • ... <i><u>home theater systems</u> and home electronics.</i>	General / Ambiguous
+T12	{will/recommend/anyone}: 10	• <i>I <u>would recommend</u> samsung to <u>anyone</u> ...</i>	Word of Mouth
+T13	{black/bar/side}: 2, {black/bar/leave}: 2, {black/bar/either}: 2, {black/bar/top}: 1	• ... <i>watch them in 4:3 with <u>black bars</u> on either <u>side</u>.</i>	General / Ambiguous
+T14	{order/on/november}: 2, {order/on/thursday}: 1, {order/on/sunday}: 1, {order/on/octobe}: 1, {order/on/january}: 1, {order/on/friday}: 1, {order/on/8th}: 1, {order/on/28th}: 1	• <i>I <u>ordered</u> this <u>on January 3rd</u> ...</i>	General / Ambiguous
+T15	{hang/on/wall}: 5, {put/on/wall}: 2, {lean/against/wall}: 1	• ... <i><u>put it on</u> the wall and attached a local HD antenna.</i>	General / Ambiguous
+T16	{decide/go/with}: 7, {decision/go/with}: 2	• <i>So, I <u>decided to go with</u> a flat panel LCD HDTV.</i>	General / Ambiguous
+T17	{within/price/range}: 3, {in/price/range}: 3	• ... <i>superior to all others <u>in the price range</u>.</i>	Dis/satisfaction

Table 4.11. (continued)

Ref.	Collocation Group	Examples from the +T Subset	Info. Type
+T18	{side/by/side}: 5	<ul style="list-style-type: none"> ... <u>side by side</u> comparisons are not always so cut and dried. 	General / Ambiguous
+T19	{buy/circuit/city}: 5	<ul style="list-style-type: none"> If you do <u>buy</u> from a <u>Circuit City</u>, Best Buy, etc ... 	General / Ambiguous
+T20	{use/component/cable}: 6, {use/component/connector}: 1, {use/component/connection}: 1	<ul style="list-style-type: none"> I hooked up my Xbox to this TV <u>using the component cables</u> and the graphics are incredible. 	General / Ambiguous
+T21	{good/value/money}: 2, {terrific/value/money}: 1, {standout/value/money}: 1, {great/value/money}: 1	<ul style="list-style-type: none"> Overall it's a <u>terrific value</u> for the <u>money</u>. 	Dis/satisfaction
+T22	{save/bundle/by}: 1, {save/500/through}: 1, {save/500/by}: 1, {save/450/by}: 1, {save/400/by}: 1, {save/250/by}: 1	<ul style="list-style-type: none"> I <u>saved</u> over <u>\$500</u> <u>through</u> Amazon. 	Dis/satisfaction
+T23	{research/before/buy}: 4, {research/before/purchase}: 2	<ul style="list-style-type: none"> I did hours and hours of <u>research</u> <u>before</u> <u>buying</u> ... 	Expectation
+T24	{standard/cable/look}: 3, {regular/cable/look}: 2, {standard/signal/look}: 1, {regular/cable/feel like}: 1	<ul style="list-style-type: none"> ..., <u>regular cable TV</u> <u>looks</u> very nice and the settings ... 	General / Ambiguous
+T25	{other/reviewer/sug}: 1, {other/reviewer/note}: 1, {other/reviewer/mention}: 1, {another/reviewer/report}: 1, {another/reviewer/mention}: 1	<ul style="list-style-type: none"> As <u>another reviewer mentioned</u>, this unit is ... 	General / Ambiguous

- The cleanser reviewers cared about cleaning ability, after-use sensation, comedogenicity, and sensitivity reaction.
- After-use sensation, comedogenicity, and sensitivity reaction were also key determinants that affect consumers' attitudes towards moisturizers. Additionally, the reviewers were concerned about whether a moisturizer could be absorbed easily by their skin and whether it paired well with makeup products.
- The television reviewers evaluated a television based on durability, value for money, quality of customer service, and whether the unit functioned properly without any noise in sound and picture.

The collocation groups were sorted in descending order by their PMI scores. Identifiers were assigned to the collocation groups based on their data subsets for easier reference in the discussion below. The extracted collocations are presented as tuples of three lemmatized elements, followed by their frequencies in the data. Each element may be formed by one lemma (e.g., “very”) or multiple lemmas (e.g., “dry up”). Each extracted group consists of one or more trigram collocations sharing the same semantic form. For instance, {[Completeness: Infused_intensification = High] / [Frugality: Infused_intensification = High] / [Money]} is the semantic form for {complete/waste/money} and {total/waste/money} in -E1 (Table 4.4). Likewise, the followings are examples of semantic forms for other product types:

- +E24: {[Stimulus_focus: Infused_intensification = High] / [Color_qualities: Infused_intensification = Low] / [Color]}

- -C9: {[Perception_body: Infused_intensification = Low] / [Similarity: Infused_intensification = High] / [Mental_property: Infused_intensification = Low]}
- +C18: {[Desirability: Infused_intensification = High] / [Work] / [Removing]}
- -M9: {[Mass = Low] / [Color] / [Body_mark]}
- +M8: {[Soaking_up] / [Isolated_intensification = High] / [Speed_description: Infused_intensification=High]}
- -T24: {[Product] / [Activity_stop] / [Being_operational: Infused_intensification = High]}
- +T21: {[Desirability: Infused_intensification = High] / [Usefulness] / [Money]}

Since this semantic abstraction approach relied on automated frame identification and other pre-processing techniques in NLP, some of the extracted collocation groups may contain noise or instances that are inconsistent with the majority in a collocation group. For example, in -C19, {make/feel/smooth} is an error introduced by the frame identifier since “smooth” is not semantically similar to “tight,” “stiff,” and “gritty.” This issue, however, was observed only in a very small number of collocation groups.

4.5.1 Assessing the Semantic Abstraction Ability of the Proposed Method

In regard to semantic abstraction ability, the proposed method relies on GradFrameNet’s frames for generalization. GradFrameNet consists of frames inherited directly from FrameNet (i.e., non-graduated) and frames that were further categorized based on graduation types and graduation levels (i.e., graduated). The semantic abstraction process relied on both the non-graduated and graduated frames to produce semantic representations that generalize

in a sentiment-aware manner. From the extracted results, it seemed the non-graduated frames provided semantically consistent abstraction for words or phrases that were hardly rankable such as:

- “chuck,” “toss,” and “throw” in -E19 and -C14 (abstracted as [Cause_motion])
- “bottle,” “tub,” “tube,” etc. in +C8, +C16, and +M5 (abstracted as [Containers])
- “bump,” “pimple,” “rash,” etc. in -C16, -M9, -M10, -M12, -M24, -M25, and +M20 (abstracted as [Body_mark])
- “line,” “dot,” “shape,” etc. in -T8, -T18, -T22 (abstracted as [Shapes])

On the other hand, the graduated frames played an indispensable role in maintaining semantic and sentiment consistency when grouping rankable words or phrases. In other words, when FrameNet’s frames were extended using graduation, their abstraction degrees were modified to become more consistent with the distinctions that people typically make when expressing their attitudes or opinions. The examples in Table 4.12 help to elucidate the effects of graduation. Without graduation, the semantic representations generated by the semantic abstraction process and the phrasal patterns extracted from these representations would become too general and noisy to be useful for sentiment analysis. For example, if the [Desirability] frame were not extended with graduation, all words and phrases describing different levels of desirability would be grouped under one overly general frame. Apparently, in sentiment analysis, it is necessary to distinguish between desirable (e.g., “terrific”) and undesirable (e.g., “horrible”) situations. As demonstrated by the examples in Table 4.12, graduation provided a means to achieve this goal. The frequencies of non-graduated frames and different types of graduated frames are also presented in Table 4.12.

Table 4.12. The frequencies of non-graduated frames and different graduation types. Examples are given for graduation types and graduation levels found in the extracted collocation groups.

Graduation	% of Frames in the Product Reviews				Example of Collocation for Different Graduation Levels
	Eyeliner	Cleanser	Moisturizer	Television	
No Graduation	57.35	60.83	60.53	63.65	-
Quantity					
Small	1.32	1.03	1.12	1.00	-C24: {after/a <u>few</u> /use}
Medium	0.21	0.11	0.16	0.26	-
Large	1.45	1.56	1.32	1.76	-E20: { <u>majority</u> /great/review}
Mass					
Small	0.48	0.35	0.44	0.27	-C16: { <u>little</u> /flesh-coloured/bump}
Medium	0.01	0.01	0.01	0.06	-
Large	0.53	0.47	0.73	0.57	-M24: {in/ <u>giant</u> /pimple}
Distribution					
Narrow	0.83	0.86	0.77	0.79	-E18: {less/30/ <u>minute</u> }
Medium	1.22	0.68	0.78	0.61	+E23: {around/8/ <u>hour</u> }
Wide	2.93	1.96	2.12	1.93	+M6: {last/5/ <u>year</u> }
Proximity					
Close	0.23	0.18	0.24	0.52	+T18: {side/ <u>by</u> /side}
Medium	0.09	0.11	0.12	0.20	+T7: {stand/ <u>foot</u> /away}
Far	0.25	0.20	0.22	0.42	-E3: {stay/ <u>far</u> /from}
Isolated Intensification					
Low	0.13	0.16	0.15	0.08	+E11: {a <u>bit</u> /pricey/but}
Medium	0.16	0.15	0.14	0.13	-
High	4.08	3.70	3.92	2.07	+M8: {absorb/ <u>very</u> /quick}
Focus					
Downscale	0.53	0.51	0.59	0.73	-E21: { <u>about</u> /5/time}
Upscale	0.02	0.01	0.01	0.02	-C25: { <u>real</u> /hope/like}
Infused Intensification					
Low	6.55	5.46	5.75	4.90	-T17: { <u>horrible</u> /customer/service} (infused with [Desirability])
Medium	3.17	4.44	3.93	3.35	-C18: { <u>probable</u> /work/people} (infused with [Likelihood])
High	18.47	17.22	16.93	16.70	+T21: { <u>terrific</u> /value/money} (infused with [Desirability])

The frame frequencies are grouped by graduation types in Figure 4.6 to allow for easier comparison across different product categories. It can be seen that although the non-graduated frames still dominate in frequency, there is a substantial number of graduated frames in the product reviews. In the beauty domain, the three products showed similar

Figure 4.6. Frequencies of non-graduation and graduated frames in the product reviews

frequencies, but the proportion of graduated frames in the eyeliner reviews was slightly higher for most graduation types. In particular, the Distribution graduation type was noticeably more frequent in eyeliner reviews because the reviewers frequently commented

on the staying power of the products, which is an aspect related to distribution in time. In comparison, the television reviews seemed to have a higher percentage of non-graduated frames. The frames frequencies for the Isolated Intensification and Infused Intensification categories were also much lower in the television reviews.

Table 4.13 shows the top 15 most frequent semantic labels (may consist of frames, graduation types, or the combinations of both) and their top matching tokens found in the four types of product reviews. Some general semantic labels such as [Negation], [Isolated_intensification = High], and [Concessive] are common for all product types. As expected, the [Desirability] and [Experiencer_focus] frames, which are closely related to consumers' likes and dislikes, also occur frequently in the four types of reviews. Different domains seem to have their own domain-specific frames. For instance, [Body_parts] is probably specific to beauty product reviews, whereas [Gizmo] and [Physical_artworks] are more common in electronic reviews. Since different types of products in the same domain serve specific purposes, they are also associated with different customer requirements. From the top frames of the beauty products, it seems these frames are somewhat reflective of the major requirements for the products. For example, both cleansers and moisturizers have a high frequency of the [Body_mark] frame, suggesting that the comedogenicity of cleansers and moisturizers has a huge impact on consumers' attitudes towards the products. Having [Removing] and [Body_decoration] among the most frequent frames for cleansers may be evident that consumers care about the ability of cleanser products in removing traces of makeup. Likewise, the [Color] frame in the eyeliner reviews may suggest that color is an important purchasing criterion for eyeliner users.

Table 4.13. Top 15 semantic labels in the reviews of the four product types

Negative Reviews		Positive Reviews	
Semantic Label	Top Matching Token	Semantic Label	Top Matching Token
<u>Eyeliner</u>			
Negation	n't	Isolated_intensification = High	really
Isolated_intensification = High	so	Distribution = Wide	lasts
Concessive	but	Negation	n't
Body_parts	eyes	Desirability: Infused_intensification = High	great
Distribution = Wide	stay	Color	black
Color	black	Using	use
Causation	because	Concessive	but
Using	use	Body_parts	eyes
Desirability: Infused_intensification = High	good	Experiencer_focus: Infused_intensification = High	love
Spatial_contact	on	Causation	so
Desirability: Infused_intensification = Low	bad	Spatial_contact	on
Quantity = Large	all	Quantity = Large	all
Quantity = Small	one	Distribution = Medium	day
Commerce_buy	bought	Quantity = Small	one
Likelihood: Infused_intensification = Medium	would	Containers	brush
<u>Cleanser</u>			
Body_parts	skin	Body_parts	skin
Negation	n't	Isolated_intensification = High	really
Isolated_intensification = High	really	Negation	n't
Concessive	but	Using	use
Using	using	Desirability: Infused_intensification = High	great
Causation	so	Concessive	but
Distribution = Wide	week	Distribution = Wide	years
Time_vector: Infused_intensification = High	after	Body_decoration	makeup
Quantity = Large	all	Causation	so
Desirability: Infused_intensification = High	good	Experiencer_focus: Infused_intensification = High	love
Body_mark	acne	Removing	remove
Likelihood: Infused_intensification = Medium	would	Quantity = Large	all
Desirability: Infused_intensification = Low	bad	Grooming	wash
Spatial_contact	on	Accompaniment: Infused_intensification = High	with
Being_dry: Infused_intensification = Medium	dry	Level_of_force_resistance: Infused_intensification = High	gentle

Table 4.13. (continued)

Negative Reviews		Positive Reviews	
Semantic Label	Top Matching Token	Semantic Label	Top Matching Token
<u>Moisturizer</u>			
[Body_parts]	skin	[Body_parts]	skin
[Negation]	n't	[Isolated_intensification = High]	very
[Isolated_intensification = High]	really	[Negation]	n't
[Causation]	so	[Using]	use
[Concessive]	but	[Distribution = Wide]	years
[Using]	use	[Desirability: Infused_intensification = High]	great
[Distribution = Wide]	week	[Causation]	so
[Desirability: Infused_intensification = High]	good	[Concessive]	but
[Quantity = Large]	all	[Experiencer_focus: Infused_intensification = High]	love
[Spatial_contact]	on	[Temporal_collocation]	in
[Time_vector: Infused_intensification = High]	after	[Being_dry: Infused_intensification = Medium]	dry
[Likelihood: Infused_intensification = Medium]	would	[Spatial_contact]	on
[Body_mark]	acne	[Quantity = Large]	all
[Desirability: Infused_intensification = Low]	bad	[Calendric_unit]	night
[Give_impression]	smells	[Quantity = Small]	one
<u>Television</u>			
Negation	not	Desirability: Infused_intensification = High	great
Distribution = Wide	months	Isolated_intensification = High	very
Cardinal_numbers	2	Negation	not
Commerce_buy	buy	Cardinal_numbers	2
Quantity = Large	all	Quantity = Large	all
Temporal_collocation	in	Concessive	but
Concessive	but	Physical_artworks	picture
Isolated_intensification = High	very	Gizmo	player
Increment	another	Interior_profile_relation: Infused_intensification = High	in
Time_vector: Infused_intensification = High	after	Weapon	dlp
Desirability: Infused_intensification = High	great	Spatial_contact	on
Spatial_contact	on	Accompaniment: Infused_intensification = High	with
Contacting	called	Commerce_buy	buy
Distribution = Narrow	again	Distribution = Wide	years
Gizmo	unit	Increment	more

4.5.2 Assessing the Information Discovery Ability of the Proposed Method

This study also examined the ability of the proposed method in uncovering useful information for sentiment analysis and opinion mining. In most cases, the collocations seem to reflect the word frequency distribution described by Zipf's Law (Zipf, 1949), which states that given a text collection, the frequency of any word can be approximated by a power function of its rank in the word frequency table. Therefore, only a handful of words or word compositions tend to occur frequently, but there is a long tail of low-frequency instances (Zipf, 2016). However, the frame-based abstraction method grouped together low-frequency instances to generate uniform representations that facilitated the discovery of regular patterns. As a result, some collocation groups consisting of collocations with low individual frequencies were detected. This is evident in cases such as -E10, +E5, -C5, +C8, -M10, +M8, -T8, +T21, and so forth. Table 4.14 gives the numbers of collocation groups that contain at least one low-frequency (less than five occurrences) collocation. Recall that one of the main challenges in generating a high-coverage phrasal sentiment lexicon is the difficulty in identifying lexically-varied but semantically similar phrasal units. The ability to detect these low-frequency instances is therefore valuable to such effort.

Table 4.14. Numbers of collocation groups that contain at least one low-frequency (< 5 occurrences) cases

Dataset	Number of Collocation Groups Containing Low-frequency (< 5 Occurrences) Collocations
Eyeliner	37 (74%)
Cleanser	38 (76%)
Moisturizer	42 (84%)
Television	36 (72%)

Not all of the extracted collocations are useful for sentiment analysis. Some phrasal patterns just happen to occur more frequently in certain product domains, but they might not carry useful information for sentiment analysis. These collocation groups were labeled as “General / Ambiguous” in Table 4.4 to 4.11. It can be tricky to tell apart the “General / Ambiguous” collocations from the “Experiential” collocations, which describe positive and negative usage experiences. A collocation group was labeled as “General / Ambiguous” if it is describing some general facts about the consumers or products and the consumers’ routines or habits when using the products instead of their reactions or consequences from using the products. For instance, {someone/with/sensitive} (-C12) and {oily/acne/prone} (+M20) may look like negative collocations on the surface but they are actually general collocations that describe the consumers’ skin types and conditions. Likewise, {use/angle/brush} (+E17) and {hang/on/wall} (+T15) are general collocations that describe the consumers’ routines or habits when using the products. As summarized in Table 4.15, in total, 26% of the collocation groups fall under the “General / Ambiguous” type. The percentage is higher for television reviews (38%) in comparison to beauty product reviews (20% to 26%). Other collocation groups have revealed useful sentiment-clues that could be categorized into five information types, i.e., consumers’ positive or negative experiences, their overall satisfaction with the products, their expectations as shaped by various factors, and customers’ retention and intention of word-of-mouth (WOM) engagement. As can be seen from Table 4.15, the distribution of these information types did not seem to differ considerably across different products in the beauty domain. However, for the television reviews, the results showed that it was uncommon for the reviewers to express their retention intention. A highly possible explanation for this might be that unlike beauty products, which are consumables that people purchase recurrently, televisions—or electronic products in

general—are capital goods or durable goods. The identified information types are described in detail with supporting examples below.

Table 4.15. Statistics on the information types

Dataset	General / Ambiguous	Experiential	Dis/satisfaction	Expectation	Retention	WOM
Eyeliner	10 (20%)	24 (48%)	9 (18%)	3 (6%)	4 (8%)	0 (0%)
Cleanser	10 (20%)	24 (48%)	7 (14%)	3 (6%)	5 (10%)	1 (2%)
Moisturizer	13 (26%)	21 (42%)	5 (10%)	3 (6%)	8 (16%)	0 (0%)
Television	19 (38%)	21 (42%)	7 (14%)	2 (4%)	0 (0%)	1 (2%)
Total:	52 (26%)	90 (45%)	28 (14%)	11 (5.5%)	17 (8.5%)	2 (1%)

- *Experiential opinion.* A large proportion of the collocation groups contains the reviewers' descriptions of their usage experiences, and this was observed for all product types. As expected, most of the collocation groups extracted from the positive reviews are related to positive experiences, and vice versa. The experiential opinions found in the eyeliner reviews, for example, indicate that consumers are fond of eyeliners that have excellent staying power and long-lastingness (+E2, +E3, +E8, +E10, +E21, and +E23), high color saturation (+E14 and +E24), versatility (+E20), ease of use (+E12), beautiful effect (+E19), and smooth application (+E5 and +E16). In contrast, insufficient staying power and excessive smudging (-E7, -E9, -E13, -E14, and -E18), troublesome application (-E4, -E10, -E18, -E21, and -E24), and very low durability (-E11) are some unfavorable properties of eyeliners. Interestingly, a slightly expensive price (+E11) was not regarded as deal-breakers by the reviewers since they still rated the products highly. These experiential opinions provide valuable clues for inferring consumers' sentiments based on the desirable and undesirable properties of the products.

- *Customer satisfaction / dissatisfaction.* Customer satisfaction is a central concept in consumer marketing (Churchill Jr & Surprenant, 1982). It represents a customer's overall positive or negative feeling about his or her purchase based on a perceived net value derived from comparing the costs and gains of the purchase (Churchill Jr & Surprenant, 1982; Yang & Peterson, 2004). In online reviews, customer satisfaction and dissatisfaction can be regarded as equivalent to positive and negative sentiments in sentiment analysis. The collocation groups detected signals of customer satisfaction and dissatisfaction, even though some signals were conveyed in a rather indirect manner. For example, in -E23 and -T11, although the utterances contain words such as "love," "like," and "happy," the reviewers were actually dissatisfied with the products (e.g., "I really wanted to love this eyeliner" and "I was happy with this TV until yesterday"). Conversely, in +E11, collocations such as {a bit/pricey/but} actually conveyed the reviewers' satisfaction indirectly. Although "pricey" is a negative word, this collocation suggested that the reviewers' overall satisfaction was high because they perceived their purchases as "worth the money" (i.e., the perceived net values were high).
- *Customer expectation.* Research in consumer marketing has suggested that customer expectation and disconfirmation are potent determinants of customer satisfaction (Oliver, 1980; Tse & Wilton, 1988). According to the expectation disconfirmation theory (Oliver, 1977; Oliver, 1980), disconfirmation occurs based on the discrepancy between perceived performance and customer expectation. When a product or service performs better than expected, there is a positive disconfirmation, which contributes to customer satisfaction and vice versa. Word-of-mouth (WOM) plays a pivotal role in shaping customers' expectations when they gather information prior

to making purchase decisions (Mauri & Minazzi, 2013; Zeithaml et al., 1993). In the collocation groups extracted from the negative beauty product reviews (e.g., -E6, -E15, -E20, -C4, -C7, -C15, -M1, -M2, and -M17), the reviewers frequently expressed a high pre-consumption expectation (e.g., {many/favorable/review}, {hear/great/thing}, and {high/hope/for}) as a contrast to a post-consumption disconfirmation of expectation. Since negative disconfirmation often results in customer dissatisfaction, these collocation groups may also serve as subtle clues to separate negative opinions from positive opinions. However, the opposite was observed for the electronic product, in which collocations related to customers' expectations such as {spend/month/research} (+T9) and {research/before/purchase} (+T23) appeared in positive reviews instead of negative reviews.

- *Customer retention and Word-of-Mouth (WOM) engagement.* Previous research has established the influence of customer satisfaction on customer loyalty (Ranaweera & Prabhu, 2003; Zeithaml et al., 1996). The behavioral dimension of loyalty is measured by consumers' retention (i.e., repurchase or continue a relationship with a company or brand), whereas the affective dimension is measured as consumers' intention to engage in positive WOM (Ranaweera & Prabhu, 2003; Zeithaml et al., 1996). The patterns discovered from the collocation groups are consistent with the findings in the consumer behavior literature. Specifically, the reviewers showed higher loyalty in positive reviews (+E4, +E9, +E13, +C3, +C8, +C13, +C14, +C23, +M2, +M4, +M5, +M6, +M15, +M19, +M25 and +T12) but lower loyalty in negative reviews (-E12, -C10, and -M3). Since customer satisfaction is a precondition for retention and WOM engagement, expressions related to the reviewers' retention (e.g., {purchase/over/over}, {on/third/bottle}, {over/six/year}, {but/always/come

back}, {will/never buy/again}, and {will/not use/again})) and their willingness or reluctance to engage in positive WOM (e.g., {will/recommend/anyone})) are important clues for recognizing consumers' sentiments.

For comparison, Table 4.16 to 4.23 show the collocations obtained without applying the semantic abstraction process described in Section 4.3.2. The NLTK collocation finders were used to extract collocations from the pre-processed review sentences. The same datasets and experimental settings described in Section 4.4 were used. It was found that these collocations also extracted some interesting phrasal patterns that revealed the reviewers' sentiments. In particular, there are some informative collocations such as {too/harsh/for} (Table 4.18, #24), {feel/clean/refresh} (Table 4.19, #24), {not/stand/smell} (Table 4.20, #25), {lamp/burn/out} (Table 4.22, #14), {no/dead/pixel} (Table 4.23, #3), {bang/for/buck} (Table 4.23, #7), {can/tell/difference} (Table 4.23, #24) that are missing from the collocation groups in Table 4.4 to 4.11. Although a possible explanation for this might simply be that these collocations have lower significance than the top 25 collocation groups in Table 4.4 to 4.11, the possibility of this being caused by an imperfect semantic tagging process cannot be ruled out. Conversely, as discussed earlier, low-frequency collocations like those in -E10, +E5, -C5, +C8, and so forth were only detected when abstraction was applied. Furthermore, with abstraction, similar collocations such as {real/want/love} and {real/want/like} could be organized into the same collocation group. In practice, this may also simplify the analyses of common phrasal patterns and facilitate the identification and generation of rules to use these patterns for sentiment analysis.

Table 4.16. Collocations from negative eyeliner reviews (-E)

#	Collocation	#	Collocation	#	Collocation
1	{ew/ew/ew}	11	{dry/so/fast}	21	{after/read/all}
2	{low/lash/line}	12	{after/few/hour}	22	{never/buy/again}
3	{not/understand/hype}	13	{no/stay/power}	23	{good/thing/about}
4	{high/hope/for}	14	{only/thing/about}	24	{flake/like/crazy}
5	{stay/away/from}	15	{solid/black/line}	25	{look/like/raccoon}
6	{revlon/colorstay/liquid}	16	{glide/on/smooth}		
7	{all/over/place}	17	{real/want/love}		
8	{revlon/colorstay/liner}	18	{give/raccoon/eye}		
9	{if/make/mistake}	19	{smudge/like/crazy}		
10	{as/soon/as}	20	{so/many/review}		

Table 4.17. Collocations from positive eyeliner reviews (+E)

#	Collocation	#	Collocation	#	Collocation
1	{bobbi/brown/gel}	11	{thin/thick/line}	21	{will/never/back}
2	{until/take/off}	12	{never/go/back}	22	{will/buy/again}
3	{in/makeup/bag}	13	{bit/pricey/but}	23	{on/low/lash}
4	{hand/down/good}	14	{will/definite/repurchase}	24	{will/repurchase/again}
5	{low/lash/line}	15	{go/long/way}	25	{last/long/time}
6	{make/eye/pop}	16	{dry/out/quick}		
7	{by/far/good}	17	{any/problem/with}		
8	{with/makeup/remover}	18	{stay/in/place}		
9	{top/lash/line}	19	{definite/buy/again}		
10	{off/at/night}	20	{just/make/sure}		

Table 4.18. Collocations from negative cleanser reviews (-C)

#	Collocation	#	Collocation	#	Collocation
1	{wake/next/morning}	11	{little/all/over}	21	{only/good/thing}
2	{wake/up/morning}	12	{little/bump/all}	22	{good/thing/about}
3	{stay/away/from}	13	{bump/all/over}	23	{remove/eye/makeup}
4	{wake/up/next}	14	{bad/than/ever}	24	{too/harsh/for}
5	{up/next/morning}	15	{take/off/makeup}	25	{acne/prone/skin}
6	{little/red/bump}	16	{leave/film/on}		
7	{little/bump/over}	17	{at/same/time}		
8	{as/soon/as}	18	{remove/makeup/well}		
9	{so/many/review}	19	{all/over/cheek}		
10	{only/thing/about}	20	{sensitive/acne/prone}		

Table 4.19. Collocations from positive cleanser reviews (+C)

#	Collocation	#	Collocation	#	Collocation
1	{always/come/back}	11	{will/definite/repurchase}	21	{for/long/time}
2	{on/cotton/pad}	12	{leave/any/residue}	22	{follow/up/with}
3	{only/need/tiny}	13	{love/love/purity}	23	{try/so/many}
4	{little/long/way}	14	{will/buy/again}	24	{feel/clean/refresh}
5	{go/long/way}	15	{take/off/mascara}	25	{irritate/at/all}
6	{little/go/way}	16	{will/definite/again}		
7	{hg/hg/hg}	17	{will/repurchase/again}		
8	{more/expensive/than}	18	{not/live/without}		
9	{little/go/long}	19	{oily/acne/prone}		
10	{only/need/amount}	20	{one/bottle/last}		

Table 4.20. Collocations from negative moisturizer reviews (-M)

#	Collocation	#	Collocation	#	Collocation
1	{stay/away/from}	11	{acne/prone/skin}	21	{after/read/all}
2	{just/sit/top}	12	{stay/from/product}	22	{just/sit/on}
3	{sit/on/top}	13	{end/up/with}	23	{out/like/crazy}
4	{wake/up/with}	14	{hydrate/at/all}	24	{not/absorb/into}
5	{after/read/review}	15	{break/out/crazy}	25	{not/stand/smell}
6	{read/all/review}	16	{can/not/stand}		
7	{high/hope/for}	17	{absorb/into/skin}		
8	{bump/all/over}	18	{little/all/over}		
9	{leave/film/on}	19	{break/like/crazy}		
10	{not/enough/moisture}	20	{never/buy/again}		

Table 4.21. Collocations from positive moisturizer reviews (+M)

#	Collocation	#	Collocation	#	Collocation
1	{extra/virgin/oil}	11	{mix/with/foundation}	21	{with/spf/15}
2	{always/come/back}	12	{so/many/different}	22	{but/come/back}
3	{before/go/bed}	13	{all/over/body}	23	{will/buy/again}
4	{little/long/way}	14	{soak/into/skin}	24	{but/always/come}
5	{go/long/way}	15	{wake/up/with}	25	{after/wash/face}
6	{will/repurchase/again}	16	{oily/acne/prone}		
7	{little/go/long}	17	{sink/in/quick}		
8	{not/clog/pore}	18	{will/definite/again}		
9	{little/go/way}	19	{cause/break/out}		
10	{last/long/time}	20	{base/for/makeup}		

Table 4.22. Collocations from negative television reviews (-T)

#	Collocation	#	Collocation	#	Collocation
1	{class/action/suit}	11	{give/5/star}	21	{happy/with/until}
2	{class/action/lawsuit}	12	{authorize/repair/cent}	22	{few/day/ago}
3	{class/action/against}	13	{not/stand/behind}	23	{more/than/hour}
4	{white/glove/delivery}	14	{lamp/burn/out}	24	{not/waste/money}
5	{stand/behind/product}	15	{still/under/warranty}	25	{less/than/month}
6	{there/class/action}	16	{will/stand/behind}		
7	{power/supply/board}	17	{less/than/2}		
8	{at/circuit/city}	18	{less/than/year}		
9	{from/circuit/city}	19	{after/warranty/expire}		
10	{stay/away/from}	20	{little/over/year}		

Table 4.23. Collocations from positive television reviews (+T)

#	Collocation	#	Collocation	#	Collocation
1	{brick/mortar/store}	11	{upconvert/dvd/player}	21	{progressive/dvd/player}
2	{progressive/scan/player}	12	{side/by/side}	22	{twice/as/much}
3	{no/dead/pixel}	13	{scan/dvd/player}	23	{black/bar/on}
4	{progressive/scan/dvd}	14	{hang/on/wall}	24	{can/tell/difference}
5	{no/sale/tax}	15	{in/perfect/condition}	25	{about/month/ago}
6	{home/theat/system}	16	{surround/sound/system}		
7	{bang/for/buck}	17	{with/progressive/scan}		
8	{buy/circuit/city}	18	{lcd/rear/projection}		
9	{local/electronic/store}	19	{play/video/game}		
10	{sony/rear/projection}	20	{keep/in/mind}		

4.6 Summary

This chapter described GradFrameNet and its use in a frame-based semantic abstraction process to facilitate the discovery of interesting phrasal patterns for sentiment analysis and opinion mining. As an early attempt to examine the potential of frame-based semantic abstraction in a novel setting, this study has gone some way towards enhancing the understanding of how FrameNet’s semantic frames plus graduation can provide abstract yet semantic-rich and sentiment-aware representations. This study used GradFrameNet and the

PAFIBERT frame identifier developed in Phase 1 of this research to generate semantic-level representations for online reviews. Phrasal patterns in the form of trigram collocations were then extracted from the representations. These phrasal patterns were evaluated qualitatively based on two qualities: semantic abstraction ability and information discovery ability. The qualitative assessment showed that GradFrameNet resulted in semantically-consistent representations that generalized in a sentiment-aware manner. The extracted phrasal patterns also demonstrated the effectiveness of the proposed frame-based abstraction process in facilitating the discovery of interesting sentiment clues related to reviewers' positive and negative experiences, as well as customers' satisfaction, expectation, retention, and word-of-mouth engagement.

CHAPTER 5

DOMAIN-SPECIFIC LEXICON CONSTRUCTION AND SENTIMENT CLASSIFICATION

5.1 Introduction

The first two phases of this research have addressed the problem of frame identification adequately and demonstrated that the proposed frame-based abstraction solution performed satisfactorily on domain-specific data. To address RQ3 and RO3, this chapter presents a fully-automated lexicon construction method that exploits the semantic abstraction process to create wide-coverage domain-specific lexicons. Using GradFrameNet and PAFIBERT developed in the first and second phases, the present study provides empirical support for the usefulness of frame-based semantic abstraction in creating a more compact but still discriminative data space that facilitates the discovery of phrasal sentiment clues for lexicon construction. To this end, frame-based semantic abstraction was applied to beauty product reviews to aid the extraction of phrases that formed the candidate pool for the domain-specific lexicons. The evaluation of the sentiment lexicons was conducted on sentence-level sentiment classification tasks in the beauty domain. With the features generated from the domain-specific lexicons, a series of experiments were conducted using statistical sentiment classifiers. While all the lexicon-based models outperformed the baseline models, the results

showed that the lexicons obtained via frame-based semantic abstraction improved the models' performance to a greater extent.

5.2 Background

In general, manually created sentiment lexicons are less noisy compared to automatically generated lexicons like SentiWordNet (Baccianella et al., 2010). However, the coverage of the latter is usually much more comprehensive. More importantly, since sentiment scores of lexicon entries are domain-dependent, creating sentiment lexicons manually for different domains are too time-consuming and labor-intensive. Previous research has established that domain-adapted lexicons are superior to general-purpose lexicons in domain-specific sentiment analysis (Bross & Ehrig, 2013; Choi & Cardie, 2009; Hamilton et al., 2016; Kanayama & Nasukawa, 2006; Lu et al., 2011; Ofek et al., 2016; Zhang & Singh, 2014). To minimize the manual efforts needed in building domain lexicons, existing studies have primarily focused on unsupervised or semi-supervised techniques to eliminate the need for manually labeled data. This approach is commonly known as the corpus-based approach, of which the main idea is to assign sentiment scores automatically to lexicon entries based on some measurements of relatedness such as co-occurrence statistics, structural clues, linguistic clues, and context similarity (Bross & Ehrig, 2013).

Without the sentiment labels provided by human annotators, sentiment scores of lexicon entries can be inferred from various information sources available in domain-specific corpora such as consumer ratings or common structures in product reviews, emoticons in tweets, and so forth. For example, Bross and Ehrig (2013), Huang, Niu, and Shi (2014), and Kaji and Kitsuregawa (2007) leveraged the pros and cons listed in product reviews to derive

positive and negative sentiment terms respectively. As demonstrated by Kaji and Kitsuregawa (2007), certain HTML layout structures on online review sites are also useful for this purpose because the pros and cons in reviews are usually listed in point form. Since the use of such an approach depends on the availability of specific information and structures in the domain corpora, another common approach avoids this constraint by using seed words (usually domain-independent sentiment words such as “good” and “bad”) to discover domain-specific sentiment terms automatically. The work of Qiu et al. (2009) falls under this category of research. Starting from a small list of seeds, their algorithm propagated through a large domain corpus, searching for pairs of aspects and opinion words recursively by exploiting a set of manually defined syntactic relations between the aspects and opinion words. Their work is closely related to that of Kanayama and Nasukawa (2006), who introduced intra-sentential and inter-sentential context coherency rules for detecting sentiment words.

Although extensive research has been carried out to improve the process of automatic domain lexicon construction, little attention has been paid to the challenges in creating a wide-coverage lexicon consisting of phrasal units or sentiment compositions. Sentiment polarities of lexicon entries are usually more ambiguous when words are isolated from their neighbors or contexts. For instance, it is difficult to assign a sentiment polarity to the word “thoroughly,” but when it is associated with other words in “dry out thoroughly” (negative) and “clean thoroughly” (positive), the polarities of the phrases can be determined with less uncertainty. The usefulness of phrases is thus well-supported in sentiment analysis (Cambria et al., 2012, 2014, 2016, 2018; Kiritchenko & Mohammad, 2016a, 2016b; Ofek et al., 2016; Polanyi & Zaenen, 2006) because they capture contextual information that, when omitted, will change or even reverse the intended meanings of sentiment expressions.

Wide coverage or recall is an essential criterion for sentiment lexicons. However, because of the inherent richness of natural language and the language phenomenon described by Zipf's Law (Zipf, 1949), even when a large domain corpus is used to learn a sentiment lexicon, it is almost certain that there will be some terms in the unseen data that the lexicon fails to cover. This problem is amplified when dealing with phrasal lexicon entries because the number of terms that need to be covered would increase exponentially because of the composition of words. Therefore, this study aims to develop a fully-automated lexicon construction method that focuses on achieving sufficient coverage for phrasal lexicon entries that capture domain-specific knowledge in product reviews. The novelty of the proposed method lies in its use of frame-based semantic abstraction to mitigate the problem of sparsity and lexical variation that hinder the discovery of useful phrasal sentiment clues.

5.3 The Proposed Domain-Specific Lexicon Construction Method

Figure 5.1 depicts the overall process of the proposed domain-specific lexicon construction method. After basic pre-processing like data cleaning and sentence splitting, pseudo labels were assigned to the review sentences based on the consumers' ratings. However, these document-level ratings may not be indicative of the sentiment expressed in individual sentences. To minimize the amount of noise introduced in this step, although the online review website (www.makeupalley.com) from which the product reviews were collected adopts a five-point scale system, this study only used reviews of one-star (negative) and five-star (positive) ratings for lexicon construction because these reviews are relatively less noisy. Specifically, reviewers tend to express one-sided opinions in one-star and five-star reviews,

so the pseudo labels inferred from the ratings and the actual sentiments expressed in the sentences extracted from the reviews are more likely to coincide.

Figure 5.1. The overall process of the proposed domain-specific lexicon construction method

The pre-processed and pseudo-labeled sentences were passed through the frame-based semantic abstraction process, which is the key to achieve sufficient coverage in the proposed lexicon building method. This abstraction process made use of GradFrameNet and the PAFIBERT model developed in the first two phases of this research to transform texts into semantic-level representations. The outputs of the semantic tagger were fed into a

frequency-based candidate generation process that extracted frequent phrasal candidates as potential entries for the lexicon. Finally, each candidate was scored based on its association with the pseudo labels. Since the frame-based semantic abstraction process has been presented in detail in Section 4.3.2, the following subsections only describe the other crucial steps of the lexicon construction method.

5.3.1 Candidate Generation

With the semantic-level representations generated from the preceding step, the candidate generation procedure aims to provide a pool of candidates for building the lexicon. Since the proposed lexicon building method focuses on sentiment phrases, the extraction of phrasal forms from sentences also plays a vital role in the coverage and quality of the final lexicon. Some studies made use of sequential relations between words to generate n-grams as candidate phrases for lexicon construction (e.g., Dey, Jenamani, & Thakkar, 2018; Mohammad et al., 2013), others extracted syntactically related terms from syntactic or dependency parsing (e.g., Huang et al., 2014; Kaji & Kitsuregawa, 2007; Kanayama & Nasukawa, 2006; Zhang & Singh, 2014).

Most existing studies in domain lexicon construction have assumed that sentiment terms consist of adjectives or adverbs that can be associated with certain aspects (Bross & Ehrig, 2013; Lu et al., 2011; Ofek et al., 2016; Qiu et al., 2009). Hence, in the candidate generation step, these studies focused on extracting nouns (e.g., battery, room, etc.) or noun phrases (e.g., battery life, room service, etc.) as aspects, and leveraged dependency parsing to identify corresponding opinion words. The resulted lexicon entries are usually in the form of relation triples such as (battery life, long, +1) and (room, small, -1). While this approach

is appealing because it can detect aspect-specific sentiment terms, its usage is still somewhat limited to certain domains such as mobile phone reviews and hotel reviews, where aspects are relatively well-defined and commonly stated explicitly by reviewers. However, in the beauty product domain, reviewers often express their likes and dislikes through descriptions of usage scenarios without explicitly mentioning the product aspects. For instance, consider the sentence “When I returned to the hotel, this liner was still on in its place.” Although the reviewer was describing a positive experience related to the long-lastingness of the makeup product, it is difficult to pinpoint the aspect term in this sentence. This issue calls for more complex techniques that can detect implicit aspects (e.g., Chen & Chen, 2016; Cruz et al., 2014; Schouten et al., 2015). Unfortunately, the accuracies of state-of-the-art implicit aspect detection techniques are still quite limited (Tubishat, Idris, & Abushariah, 2018).

Therefore, this study adopted the skip-gram approach (Guthrie, Allison, Liu, Guthrie, & Wilks, 2006), which is commonly used in language modeling as a more generalizable variant of the n-gram approach. Like n-grams, skip-grams consist of sequences extracted based on the linear structure of texts. However, the components in skip-grams need not follow the consecutive order of words in the texts under consideration but may instead skip certain number of words. The number of words that can be skipped depends on a predefined window size (w) used for the extraction. For each sentence in the pseudo-labeled data, skip-n-grams were extracted for $n = 2, 3$, and 4 with $w = (n + 2)$. The Ngrams module in Natural Language Toolkit (NLTK) was extended to leverage the semantic-level representations for extracting semantic-level skip-grams and their corresponding lexical-level skip-grams. A small list of stopwords was applied to remove less meaningful skip-grams in the extraction process. Table 5.1 shows the semantic-level and lexical-level skip-grams derived for the sentence “It has no staying power whatsoever and it is not that much cheaper than MAC.”

Table 5.1. Semantic-level and lexical-level skip-grams extracted for the sentence “It has no staying power whatsoever and it is not that much cheaper than MAC.”

Item No.	Semantic-Level	Lexical-Level
<u>Skip-2-grams</u>		
1	([Negation] / [Distribution=Wide])	(no / stay)
2	([Negation] / [Level_of_force_exertion:Infused_intensification=High])	(no / power)
3	([Negation] / [Spatial_co_location])	(no / whatsoever)
4	([Distribution=Wide] / [Level_of_force_exertion:Infused_intensification=High])	(stay / power)
5	([Distribution=Wide] / [Spatial_co_location])	(stay / whatsoever)
6	([Level_of_force_exertion:Infused_intensification=High] / [Spatial_co_location])	(power / whatsoever)
7	([Negation] / [Proportional_quantity])	(not / much)
8	([Negation] / [Expensiveness:Infused_intensification=Low])	(not / cheap)
9	([Proportional_quantity] / [Expensiveness:Infused_intensification=Low])	(much / cheap)
10	([Proportional_quantity] / than)	(much / than)
11	([Proportional_quantity] / [Product_brand])	(much / mac)
12	([Expensiveness:Infused_intensification=Low] / than)	(cheap / than)
13	([Expensiveness:Infused_intensification=Low] / [Product_brand])	(cheap / mac)
14	(than / [Product_brand])	(than / mac)
<u>Skip-3-grams</u>		
15	([Negation] / [Distribution=Wide] / [Level_of_force_exertion:Infused_intensification=High])	(no / stay / power)
16	([Negation] / [Distribution=Wide] / [Spatial_co_location])	(no / stay / whatsoever)
17	([Negation] / [Level_of_force_exertion:Infused_intensification=High] / [Spatial_co_location])	(no / power / whatsoever)
18	([Distribution=Wide] / [Level_of_force_exertion:Infused_intensification=High] / [Spatial_co_location])	(stay / power / whatsoever)
19	([Negation] / [Proportional_quantity] / [Expensiveness:Infused_intensification=Low])	(not / much / cheap)
20	([Negation] / [Proportional_quantity] / than)	(not / much / than)
21	([Negation] / [Expensiveness:Infused_intensification=Low] / than)	(not / cheap / than)
22	([Proportional_quantity] / [Expensiveness:Infused_intensification=Low] / than)	(much / cheap / than)
23	([Proportional_quantity] / [Expensiveness:Infused_intensification=Low] / [Product_brand])	(much / cheap / mac)
24	([Proportional_quantity] / than / [Product_brand])	(much / than / mac)
25	([Expensiveness:Infused_intensification=Low] / than / [Product_brand])	(cheap / than / mac)
<u>Skip-4-grams</u>		
26	([Negation] / [Distribution=Wide] / [Level_of_force_exertion:Infused_intensification=High] / [Spatial_co_location])	(no / stay / power / whatsoever)
27	([Negation] / [Proportional_quantity] / [Expensiveness:Infused_intensification=Low] / than)	(not / much / cheap / than)
28	([Negation] / [Proportional_quantity] / [Expensiveness:Infused_intensification=Low] / [Product_brand])	(not / much / cheap / mac)
29	([Negation] / [Proportional_quantity] / than / [Product_brand])	(not / much / than / mac)
30	([Negation] / [Expensiveness:Infused_intensification=Low] / than / [Product_brand])	(not / cheap / than / mac)
31	([Proportional_quantity] / [Expensiveness:Infused_intensification=Low] / than / [Product_brand])	(much / cheap / than / mac)

As depicted in Figure 5.2, a frequency table was created to store the positive and negative counts of each semantic-level skip-gram, which were obtained by counting the skip-gram's occurrences in positive and negative reviews respectively. Two semantic lookup tables (one for positive and one for negative reviews) were maintained during the skip-gram generation process to keep track of the association between the semantic-level skip-grams and their corresponding lexical-level skip-grams.

Frequency Table		
Semantic-level Skip-gram	Positive Frequency	Negative Frequency
(([Negation] / [Distribution=Wide]))	2	15
:	:	:
:	:	:

Positive Semantic Lookup Table	
Semantic-level Skip-gram	Lexical-level Skip-gram
(([Negation] / [Distribution=Wide]))	(not / stay) (not / a while)
:	:
:	:

Negative Semantic Lookup Table	
Semantic-level Skip-gram	Lexical-level Skip-gram
(([Negation] / [Distribution=Wide]))	(no / stay) (not / stay) (not / last) (no / last) (without / last) (not / go on)
:	:
:	:

Figure 5.2. The frequency table and semantic lookup tables

A problem that arises from using the skip-gram approach for lexicon candidate generation is that the extracted skip-grams may erroneously exclude negation terms, thus reversing the sentiments intended by the reviewers. Consider the skip-gram examples given in Table 5.1. Because of the skipping of negation terms, skip- n -grams such as (stay / power), (much / cheap / than), (cheap / than / mac), and so forth do not reflect the original opinions conveyed by the utterance. As a result, candidate phrases that frequently co-occur with negation terms would have distorted frequencies. In other words, the frequencies of these skip-grams might not be indicative of their association with positive or negative sentiments. In the above example, since the sentence was pseudo-labeled as a negative sentence, the negative counts of (stay / power), (much / cheap / than), (cheap / than / mac), and so forth would be erroneously incremented in the frequency table. These spurious increments would incorrectly strengthen the association between these phrases and the negative sentiment, thus resulting in an erroneous calculation of sentiment scores in the subsequent step.

As a remedy to this problem, this study handled negation by excluding—from the frequency count—skip-grams that inappropriately omitted any necessary negation terms. To determine whether a skip-gram should be excluded, the following rule was applied:

Given a skip-gram w of n elements, if we can extract from the same sentence a skip-gram of $(n+1)$ elements with a [Negation] frame inserted in front of any element of w , then w should be excluded from the frequency count.

For example, item #25: (cheap / than / mac) in Table 5.1 should not be counted because a [Negation] frame can be inserted in front of its semantic-level skip-gram to form a skip-4-gram that matches (not / cheap / than / mac), which can also be extracted from the

same sentence. Following the above rule, other items such as item #4: (stay / power), item #5: (stay / whatsoever), item #9: (much / cheap), and so forth should also be excluded.

When deciding whether a term should be included in an automatically constructed lexicon, previous works have commonly applied a pruning threshold to remove low-frequency terms. Low-frequency terms that fall below the selected threshold should be excluded because their observed occurrences—either more frequent in positive or negative reviews—could be purely coincidental. As far as this research is concerned, there is no definitive guideline or criterion that specifies how an optimal value can be selected for this threshold. However, it was found that in the research line of Latent Dirichlet Allocation (LDA) topic modeling (Blei, Ng, & Jordan, 2003), which is also an NLP research problem related to text mining and pattern extraction, Maier et al. (2018) provided a comprehensive guide on several issues and best practice related to the problem. According to Maier et al. (2018), relative pruning of infrequent terms can be applied with a threshold of 0.5%. This approach is also commonly used in topic modeling studies, including those of Grimmer and Stewart (2013) and Denny and Spirling (2017). This study followed a similar approach of relative pruning but with much lower values because phrasal expressions are inherently infrequent. Since higher-order skip-grams are generally less frequent than lower-order skip-grams, the thresholds were also varied according to skip-gram orders. This study started with a low threshold combination of (0.05%, 0.08%, 0.10%) for skip-4-grams, skip-3-grams, and skip-2-grams respectively, and experimented with threshold combinations within the range of (0.05%, 0.08%, 0.10%) and (0.10%, 0.13%, 0.15%).

5.3.2 Sentiment Score Calculation and Lexical Entry Selection

The sentiment scores of the phrasal candidates were calculated using Pointwise Mutual Information (PMI) (Turney & Littman, 2002). PMI has been used in many studies to compute the sentiment scores of lexicon entries (e.g., Kiritchenko et al., 2014a; Li, Guo, Shi, Zhu, & Zheng, 2018; Turney & Littman, 2002). Another popular method for obtaining sentiment scores is the Chi-square method. For instance, this method was used by Zhang, Ghosh, Dekhil, Hsu, and Liu (2011) in their research to build sentiment lexicons for tweets. However, in a study that explored the use of PMI and Chi-square for selecting polar phrases from a massive collection of HTML documents, PMI was found to be superior to Chi-square (Kaji & Kitsuregawa, 2007). The latter method not only produced lower precision in most cases but also suffered low recall rates.

Among a number of PMI variants, the implementation adopted in this study is the one used by Kiritchenko et al. (2014a), in which no seed words are needed to infer the sentiment scores. Based on this implementation, a term (w) is more likely to contain positive sentiment if it is frequently associated with positive sentences compared to negative sentences, and vice versa. The sentiment score of the term is thus derived as:

$$score(w) = PMI(w, positive) - PMI(w, negative) \quad (1)$$

$$PMI(w, positive) = \log_2 \frac{f(w, positive) * N}{f(w) * f(positive)} \quad (2)$$

$$PMI(w, negative) = \log_2 \frac{f(w, negative) * N}{f(w) * f(negative)} \quad (3)$$

which can be further simplified to:

$$score(w) = \log_2 \frac{f(w, positive) * f(negative)}{f(w, negative) * f(positive)} \quad (4)$$

where $f(w, \text{positive})$ and $f(w, \text{negative})$ denote the frequencies of w in positive sentences and negative sentences, whereas $f(\text{positive})$ and $f(\text{negative})$ are the total numbers of tokens in positive and negative sentences respectively.

When adapted to the proposed method, the score was computed for each semantic-level skip-gram using the frequencies stored in the frequency table shown in Figure 5.2. The score was then assigned to lexical-level skip-grams that shared the same semantic-level representation. The lexical-level skip-grams formed the entries of the domain-specific lexicons. To minimize the amount of noise, if the score of a semantic-level skip-gram was positive, this step only kept its lexical-level skip-grams stored in the positive semantic lookup table, and vice versa. From the example in Figure 5.2, the lexical skip-grams for ([Negation] / [Distribution=Wide]) in the negative semantic lookup table will be added to the domain-specific lexicons because the score of this semantic-level skip-gram is negative.

Another factor that usually affects the selection of lexicon entries is the sentiment score threshold, θ . While the frequency thresholds described in Section 5.3.1 aim to filter out low-frequency terms, the sentiment score threshold removes ambiguous candidates that do not exhibit a strong association with any sentiment polarity. These candidates usually have rather balanced frequencies in both positive and negative sentences, so their sentiment scores are small, indicating a rather weak association with either positive or negative sentiment. In this study, θ was set to 0.5 so that candidates with absolute scores below 0.5 were ignored. Examples of these ambiguous candidates include (draw / line): -0.48, (purchase / color / brown): 0.11, (wash / face / with): -0.38, and (smell / mint): 0.10, among others.

5.4 Experiment

5.4.1 Data

Web scraping was used to collect beauty product reviews from www.makeupalley.com. The acquired data were used for two purposes: automatic construction of domain-specific lexicons and development of sentiment classifiers. The experiments were conducted for two product types—eyeliner and cleanser—that were also examined in the study in Chapter 4. As shown in Chapter 4, the consumers of these two types of beauty products have different requirements: The eyeliner users were concerned about the staying power, smoothness, color saturation, change of consistency, and ease of use of the products, whereas the cleanser users usually evaluated a cleanser based on its cleaning ability, comedogenicity, triggering of sensitivity reaction, and after-use sensation. Therefore, this study created two product-specific lexicons and conducted the classification experiments separately for eyeliner and cleanser reviews. Table 5.2 and 5.3 provide detailed information about the data.

The data collection and preparation process ensured that the lexicon construction data and the sentiment classification data were mutually exclusive, i.e., the product reviews collected for lexicon construction were not included in the labeled data used for sentiment classification. For the preparation of the labeled data, the author and an undergraduate student at Nanyang Technological University manually labeled the pre-processed sentences as positive and negative. Cohen's κ was run to determine intercoder reliability. There was high agreement between the two coders' labels ($\kappa = .897$).

Table 5.2. Pseudo labeled data for domain-specific lexicon construction

Product Type	Sentiment Polarity	Number of Reviews	Number of sentences	Number of tokens
Eyeliner	Positive	1,000	5,897	88,124
	Negative	1,000	5,816	82,792
Cleanser	Positive	1,000	5,928	91,464
	Negative	1,000	5,397	84,130

Table 5.3. Manually labeled data for sentiment classification

Product Type	Number of Sentences		
	Positive	Negative	Total
Eyeliner	750	750	1,500
Cleanser	777	723	1,500

5.4.2 Sentiment Classifiers

To evaluate the usefulness and effectiveness of the proposed lexicon construction method, a series of sentiment classification experiments were conducted by incorporating lexicon-related information into statistical sentiment classification models. Support Vector Machine (SVM) (Cortes & Vapnik, 1995) was adopted as the machine learning algorithm to build the classifiers. SVM is a state-of-the-art machine learning algorithm that has been proven powerful in a wide range of text classification problems including sentiment analysis (Bhaskar, Sruthi, & Nedungadi, 2015; Khan, Qamar, & Bashir, 2017; Singh & Husain, 2014; Wang & Manning, 2012; Xia, Zong, & Li, 2011). Although in recent years, deep learning has achieved remarkable results in sentiment analysis (Devlin et al., 2018; Howard & Ruder, 2018), SVM is still widely adopted when it comes to integration with sentiment lexicons (Chikersal, Poria, & Cambria, 2015; Khan et al., 2017; Mohammad, Sobhani, & Kiritchenko, 2017; Mohammad, Zhu, Kiritchenko, & Martin, 2015). Besides, feature engineering for

SVM is well-studied, and previous works have also provided abundant information about the use of lexicon-related features to improve the performance of SVM classifiers.

The experiments were conducted using SVM classifiers on a linear kernel ($C = 1.0$). For each experiment, the classification models were evaluated over a 10-fold cross-validation. As shown in Table 5.4, six classification models were compared: a baseline model, a simple n-gram model, and four lexicon-based models created using slightly different domain-specific lexicons.

Table 5.4. Classification models

Classification Model	Features
Baseline	Unigrams
Simple N-gram	Unigrams + Bigrams
No Abstraction (NoA)	Unigrams + Lexicon (no semantic abstraction)
Abstraction without Frame Filtering (ANoFil)	Unigrams + Lexicon (frame-based semantic abstraction using PAFIBERT without frame filtering)
Abstraction with Frame Filtering (AFil)	Unigrams + Lexicon (frame-based semantic abstraction using PAFIBERT with frame filtering using targets)
Hybrid Abstraction (AHybrid)	Unigrams + Lexicon (frame-based semantic abstraction using both PAFIBERT variants)

The baseline classifiers were learned from bag-of-words features (i.e., unigrams) consisting of binary values that indicated the presence or absence of a unigram. Other alternatives for bag-of-words implementation exist, but simple binary values have been shown to outperform frequency-based implementation in sentiment classification (Pang et al., 2002). The simple n-gram classifiers used bag-of-unigrams and bag-of-bigrams as features. According to Wang and Manning (2012), although the use of bigrams has yielded mixed results in topical text classification tasks, including bigrams often led to remarkable

improvements in sentiment classification, presumably because bigrams are capable of capturing modified and negated terms. All unigrams and bigrams used in this study were lemmatized.

The effectiveness of the proposed lexicon construction method was evaluated based on the impact of the lexicon-related features on the sentiment classifiers' performance. For a more rigorous evaluation of the effects of frame-based semantic abstraction, several comparable sets of domain-specific lexicons were created. The procedures that generated these lexicons differed mainly in the semantic abstraction and candidate generation steps. The NoA model was created following the steps of the proposed lexicon construction method described in Section 5.3 and illustrated in Figure 5.1, except that the NoA model did not apply frame-based semantic abstraction before the candidate generation step. In other words, the lexicons created for the NoA model were constructed directly from pre-processed texts without semantic abstraction. Besides, since the negation handling process described in Section 5.3.1 utilized the [Negation] semantic frame from GradFrameNet, for a fair comparison, this study also used the LUs of this frame for negation handling in the NoA model.

In Chapter 3, two variants of PAFIBERT were implemented and compared. It was found that each implementation has its pros and cons. Without frame filtering, the frame identifier generalizes better for unseen targets. However, this generalizability comes at a price, as the model tends to become over-sensitive to the sentential contexts. In the present chapter, this study included both models for comparison in the sentiment classification experiments. Furthermore, since the strength of PAFIBERT without frame filtering is mainly on the identification of frames for unseen targets, whereas PAFIBERT with frame filtering

may be more reliable for known targets found in FrameNet, this study also experimented with a hybrid semantic abstraction process that used the predictions of the former for unseen targets and the predictions of the latter for other cases. The lexicons used for the ANoFil, AFil, and AHybrid models were constructed using these three different variants of frame identification (i.e., without frame filtering, with frame filtering, and hybrid approach).

Previous research has provided plenty of examples on how lexicon-related features could be derived and incorporated into SVM classifiers (Chikersal et al., 2015; Khan et al., 2017; Mohammad et al., 2017; Mohammad et al., 2015). Examples of commonly used features include the sum of sentiment scores, counts, maximum scores, minimum scores, and so forth. In this study, given a sentence, the following features were derived from the domain-specific lexicon for each skip-gram order:

- *sum_positive, sum_negative*: the sum of sentiment scores for the positive and negative skip-grams occurring in the sentence.
- *count_positive, count_negative, count_neutral*: the number of positive, negative, and neutral skip-grams in the sentence. Skip-grams that did not appear in the lexicon were considered neutral.
- *max_positive, min_positive, max_negative, min_negative*: the most positive (or negative) and least positive (or negative) skip-grams in the sentence.
- *bag-of-lexicon-entries*: These features were generated as an intersection set between the lexicon entries and the skip-grams extracted from the training sentences. Binary values were used to indicate the presence or absence of skip-grams.

5.4.3 Performance Measures

Since the labeled data contain slightly different numbers of sentences in each class, the weighted means of precision (P), recall (R), and F-score (F) were used as the performance metrics. For each of the k classes ($k = 2$), the precision, recall, and F-score of the class were weighted by the number of instances in the class, and N is the total number of instances.

$$\text{WeightedAverage}(P) = \frac{\sum_{i=1}^k n_i P_i}{N} \quad (5)$$

$$\text{WeightedAverage}(R) = \frac{\sum_{i=1}^k n_i R_i}{N} \quad (6)$$

$$\text{WeightedAverage}(F) = \frac{\sum_{i=1}^k n_i F_i}{N} \quad (7)$$

where the values of P , R , and F are given by the following equations:

$$P = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalsePositive}} \quad (8)$$

$$R = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalseNegative}} \quad (9)$$

$$F = \frac{2PR}{P + R} \quad (10)$$

5.5 Results and Discussion

This research has focused on developing a frame-based semantic abstraction method that is applicable to opinionated texts to mitigate the problem of word sparsity and facilitate the discovery of phrasal patterns from texts. This section starts by discussing the effects of this semantic abstraction process on the sparsity degrees of the development data. As will be shown in the following analysis, the impact of semantic abstraction on the proposed lexicon construction method is reflected in the drastic increase of lexicon sizes. Issues regarding the source of noise in the lexicon construction process are discussed, followed by a comparison

of the sentiment classification results obtained using the baseline models and the lexicon-based models.

5.5.1 Effects of Frame-based Semantic Abstraction on Sparsity Degree, Lexicon Size, and Noise Level

To examine the effects of frame-based semantic abstraction, this study compared the document-term matrices of the development data (pseudo-labeled data for lexicon construction) based on vocabulary size, number of zero elements, average term frequency, and sparsity degree before and after abstraction. The sparsity degree of a text corpus can be calculated as follows (Makrehchi & Kamel, 2008):

$$S_D = \frac{(n \times m) - \sum_{i=1}^n \sum_{j=1}^m d_{i,j}}{n \times m} \quad (11)$$

This calculation assumes that the text corpus is represented as a document-term matrix of $(n \times m)$ dimension, where n is the number of documents, and m is the vocabulary size of the data, i.e., the number of unique terms. Let D be the document-term matrix of the corpus. An element $d_{i,j} \in D$ will have value 1 if term j occurs in document i , and vice versa. The sparsity degree, S_D , is thus a measurement of the proportion of zero-valued elements in matrix D . The value of S_D falls within the range of 0 and 1. In general, most text corpora have a sparsity degree close to 1 (i.e., the number of zero-valued elements is usually high in text data) because the majority of terms only occur in a small number of documents.

Table 5.5 presents the statistics of the document-term matrices before and after applying different variants of frame-based semantic abstraction. It can be seen that the vocabulary sizes, numbers of zero elements, and sparsity degrees were reduced after

Table 5.5. Statistics of the document-term matrices before and after abstraction

Abstraction Variant	Vocabulary Size	Number of Zero Elements	Average Term Frequency	Sparsity Degree
Eyeliners				
No Abstraction (NoA)	4,526	52,587,295	34	0.99708
Abstraction without Frame Filtering (ANoFil)	1,526	17,637,765	95	0.99186
Abstraction with Frame Filtering (AFil)	1,455	16,812,674	98	0.99160
Hybrid Abstraction (AHybrid)	1,549	17,904,899	94	0.99193
Cleanser				
No Abstraction (NoA)	4,917	55,222,095	32	0.99715
Abstraction without Frame Filtering (ANoFil)	1,644	18,368,002	90	0.99199
Abstraction with Frame Filtering (AFil)	1,557	17,390,615	94	0.99168
Hybrid Abstraction (AHybrid)	1,658	18,524,773	90	0.99201

abstraction, and the average term frequencies increased significantly. On average, each term only occurs in about 30 sentences in the original data, but after the semantic abstraction process, the average frequencies were increased to roughly 90 occurrences. As noted in earlier work (Saif et al., 2014), the change in sparsity degree (as measured by Equation (11)) is usually small because sparsity reduction may affect both the number of zero elements (i.e., the numerator in the equation) and the total number of elements (i.e., the denominator) at a similar ratio.

To highlight the effects of semantic abstraction on the document-term matrices, the counts of low-frequency terms are presented in Figure 5.3 and 5.4. From these figures, it is obvious that the counts of low-frequency terms decreased drastically with frame-based semantic abstraction. In particular, the numbers of singletons (i.e., terms that appear only

Figure 5.3. Counts of low-frequency terms in eyeliner reviews

Figure 5.4. Counts of low-frequency terms in cleanser reviews

once in a corpus) reduced by approximately a factor of four, from more than 2,000 singletons to only a few hundreds. Since singletons and low-frequency terms are the main culprits of high dimensionality and sparsity in textual data, the significant reduction of low-frequency terms should mitigate these problems.

All the frame-based semantic abstraction variants seem to have positive impacts on the document-term matrices. However, as discussed in Chapter 3, when the PAFIBERT model is used without frame filtering, the predictions made by the model are not restricted to the frame-LU association defined by FrameNet. Specifically, given a target word, it is possible for the model to associate the target with a frame in which none of the LUs matches the target. Therefore, as demonstrated in Chapter 3, the data space generated by PAFIBERT without frame filtering (ANoFil) may contain more variations compared to that acquired with frame filtering (AFil). For the same reason, the data space after applying hybrid abstraction (AHybrid) showed slightly higher sparsity than ANoFil. Similar trends were observed for both eyeliner and cleanser reviews.

From the comparison of the document-term matrices, it was found that the data space became denser after the abstraction process. This denser space facilitated the discovery of semantic-level phrasal patterns, which were then used to retrieve the lexical-level skip-grams that constituted the entries of the lexicons. Consequently, the sizes of the domain-specific lexicons increased significantly (Table 5.6).

Generally speaking, automatically constructed sentiment lexicons are noisier than manually created lexicons. The domain-specific lexicons generated in this study are of no exception. Indeed, by inspecting the entries of the lexicons generated from the proposed lexicon construction method, it was found that the errors introduced by the frame

Table 5.6. Sizes of the domain-specific lexicons before and after applying frame-based semantic abstraction

Abstraction Variant	Skip-2-gram	Skip-3-gram	Skip-4-gram	Total
Eyeliner				
No Abstraction (NoA)	956	250	76	1,282
Abstraction without Frame Filtering (ANoFil)	9,857	2,852	537	13,246
Abstraction with Frame Filtering (AFil)	10,652	5,667	867	17,186
Hybrid Abstraction (AHybrid)	9,044	2,573	540	12,157
Cleanser				
No Abstraction (NoA)	1129	376	141	1,646
Abstraction without Frame Filtering (ANoFil)	11,700	3,743	953	16,396
Abstraction with Frame Filtering (AFil)	11,736	6,677	1,412	19,825
Hybrid Abstraction (AHybrid)	10,578	3,342	799	14,719

identification step constituted the primary source of the noise. Several examples of entries in the lexicons are given in Figure 5.5. As can be seen from the figure, frame-based semantic abstraction facilitated the extraction of uncommon phrases and allowed the sentiment scores of these low-frequency phrases to be inferred from other phrases of similar meanings. Consequently, singletons or relatively uncommon phrases in the data such as (cease / use), (sting / nasal), (welt / face), and (no / pull / at / skin) were assigned the same sentiment scores as their semantically equivalent phrases. However, because of the frame identification errors, sentiments scores could also be propagated incorrectly among semantically distinct phrases that were clustered into the same semantic category by mistake. Examples of such phrases include (either / for / dry), (not / make / oily), and (not / make / greasy) that were grouped under the same semantic-level representation ([Negation] / [Causation] / [Being_dry]) as other more semantically consistent phrases such as (not / leave / dry), (never / leave / dry),

and (not / make / flaky). The semantically-inconsistent phrases are shown in boldface in Figure 5.5.

Figure 5.5. Examples of entries in the domain-specific lexicons (using hybrid semantic abstraction)

5.5.2 Evaluation of Lexicons in Sentence-Level Sentiment Classification

This section describes the evaluation of the automatically constructed lexicons in sentiment classification experiments. As mentioned earlier in this chapter, there is no definitive guideline that specifies the optimal threshold for the relative pruning of lexicon candidates. Therefore, this study experimented with varying thresholds within a reasonable range. The changes in F-scores in relation to different thresholds are depicted in Figure 5.6 and 5.7.

Figure 5.6. F-scores (Y-axis) for eyeliner data in relation to different relative pruning thresholds (X-axis). The threshold combinations for skip-4-grams, skip-3-grams, and skip-2-grams are displayed on the X-axis. The F-scores were obtained from 10-fold cross-validation using SVM classifiers trained on different feature sets.

Figure 5.7. F-scores (Y-axis) for cleanser data in relation to different relative pruning thresholds (X-axis). The threshold combinations for skip-4-grams, skip-3-grams, and skip-2-grams are displayed on the X-axis. The F-scores were obtained from 10-fold cross-validation using SVM classifiers trained on different feature sets.

The best classification results for each model are presented in Table 5.7. From the table, t-tests showed that most of the lexicon-based classifiers achieved statistically significant improvements over the baseline classifiers in terms of F-score, precision, and recall. The highest F-scores were achieved by ANoFil and AFil in the classification of eyeliner and cleanser data respectively. The margin of improvement is slightly greater in the classification of cleanser reviews, with AFil yielded more than an 8% absolute increase in F-score over the baseline. Consistent with what has been established in the literature (Wang

& Manning, 2012), adding bigrams also improved the performance of the classifiers considerably. The improvements, however, were still less than those obtained using sentiment lexicons. It should also be noted that while the bag-of-n-grams (either unigrams or bigrams) were generated as features from the manually labeled data, which is essential for supervised learning, the sentiment lexicons were constructed from the relatively noisier unlabeled data for different purposes. For starters, sentiment lexicon construction is a one-off effort of creating useful resources that could be shared and reused by researchers and practitioners in the field. Although sentiment lexicons can also be incorporated as additional features for supervised learning, they are valuable resources for unsupervised learning and rule-based methods in sentiment analysis. When paired with unsupervised learning, automated lexicon construction would completely eliminate the need for manually labeled data.

Table 5.7. Sentiment classification results over 10-fold cross-validation (* $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$)

Classification Model	F-Score	Precision	Recall
Eyelineer			
Baseline	73.23	73.44	73.27
Simple N-gram	77.63 ***	77.84 ***	77.67 ***
No Abstraction (NoA)	76.03 *	76.24 *	76.06 *
Abstraction without Frame Filtering (ANoFil)	78.25 ***	78.34 ***	78.26 ***
Abstraction with Frame Filtering (AFil)	75.79	75.85	75.81
Hybrid Abstraction (AHybrid)	77.25 **	77.38 **	77.26 **
Cleanser			
Baseline	70.81	71.13	70.79
Simple N-gram	75.57 ***	75.94 ***	75.53 ***
No Abstraction (NoA)	75.99 ***	76.43 ***	76.00 ***
Abstraction without Frame Filtering (ANoFil)	77.48 ***	77.80 ***	77.47 ***
Abstraction with Frame Filtering (AFil)	78.94 ***	79.17 ***	78.95 ***
Hybrid Abstraction (AHybrid)	78.40 ***	78.66 ***	78.41 ***

Overall, the lexicons constructed with frame-based semantic abstraction (AFil, ANoFil, and AHybrid) resulted in more accurate classification models than those generated using the NoA method, with the exception of AFil that produced lower F-score than that of NoA in classifying eyeliner reviews. Contrary to expectations, none of these models consistently outperformed other models in the experiments, although all the best results in the comparison were achieved with the help of frame-based semantic abstraction. In particular, the AFil model achieved a high classification rate for cleanser data, but its performance in classifying eyeliner data is disappointing. Further investigation revealed that there is a higher proportion of unseen targets in the eyeliner data than in the cleanser data (Table 5.8). Therefore, this inconsistency in the performance of the AFil classifier is likely to be related to the drawback of using candidate frame filtering in the PAFIBERT model, which restricted the model’s ability to make more sensible frame predictions for unseen targets.

Table 5.8. The proportion of unseen targets in the development data

Data	Number of Targets	Number of Unseen Targets
Eyeliner	86,294	14,299 (16.57%)
Cleanser	90,683	14,103 (15.55%)

Also surprising are the classification rates obtained by ANoFil, which used the domain-specific lexicon built from PAFIBERT without frame filtering. In Chapter 3, this variant of PAFIBERT demonstrated high generalizability in the beauty domain and was proven effective in associating unseen targets with appropriate semantic frames. Nevertheless, although ANoFil achieved better F-scores than the baseline and NoA models,

it did not consistently outperform AFil in the experiments. While the inconsistency observed in the AFil model could be attributed to its lack of generalizability, the fluctuation in ANoFil’s performance may be due to its mere reliance on sentential contexts when making frame predictions. As discussed in Chapter 3, this characteristic of the model gives it better generalizability to unseen targets and the ability to identify frames based on contexts, but the model would also become more prone to errors caused by context sensitivity since it is not bounded by lists of candidate frames retrieved from FrameNet.

In comparison, the AHybrid model that integrated two types of frame identification approaches reported more stable performance gain. However, it should be noted that this hybrid abstraction approach does not represent the most optimal frame identification solution because it still partially relies on the frame candidate filtering strategy in PAFIBERT, which makes a strong assumption about the comprehensiveness and exhaustiveness of FrameNet. Taken together, the above observations in the classifiers’ performance suggest that although frame-based semantic abstraction has shown promising potential in sentiment analysis, its effectiveness depends heavily on the outcome of frame identification.

5.6 Summary

The study in this chapter developed a fully automated lexicon construction method that made use of frame-based semantic abstraction to achieve better coverage on phrasal sentiment clues. The role of semantic abstraction in the proposed lexicon construction method was to reveal the relatedness between phrases that—although constituted of distinct lexical terms—convey similar meanings in a sentiment-aware sense. With this approach, useful phrasal singletons can be detected automatically from the collected domain corpora, and

semantically related phrases can share the same sentiment scores in the constructed lexicons. The constructed domain-specific lexicons were evaluated via incorporation of lexicon-related features into SVM sentiment classifiers. The classification results showed that the lexicon-based models benefited from the use of semantic abstraction in lexicon construction and achieved the best classification rates in the experiments. However, the comparison also revealed that while frame-based semantic abstraction has the potential of improving sentiment analysis, the effectiveness of such approach highly relies on the accuracy of frame identification, which is itself an open research problem. Therefore, there is ample room for further progress in exploring the full potential of this approach in sentiment analysis. Despite the limitations of frame identification, the overall achievements of the proposed method in boosting the coverage of phrasal sentiment lexicons and improving sentiment classification rates are still promising.

CHAPTER 6

CONCLUSION

6.1 Introduction

Sentiment lexicons consisting of entries extracted from domain corpora allow sentiment classification models to benefit from domain knowledge. The development of lexicon construction techniques in sentiment analysis contributes to the field by minimizing repetitive and manual efforts in creating these lexicons. By undertaking the challenge of addressing a fundamental problem in building comprehensive phrasal lexicons, this research makes several contributions to the current literature of sentiment analysis and corpus-based sentiment lexicon construction. This final chapter reviews the research objectives and findings of each phase of the research process. The research contributions are highlighted in Section 6.3. The limitations of the research are discussed in Section 6.4, along with the recommendations for further investigation and experimentation in future work. Finally, this chapter ends with a concluding section that sums up this research.

6.2 Review of Research Objectives

The primary goal of the present research is to establish a fully-automated lexicon construction method that generates wide-coverage lexicons consisting of phrasal sentiment clues extracted from domain corpora. To this end, this research was carried out in three phases pertaining to three research objectives.

An overview of the first phase of this research is shown in Figure 6.1. Phase 1 investigated the problem of frame identification, which is a common subtask in frame-semantic parsing and an essential step in utilizing frame semantics to produce accurate semantic-level representations for texts. This task can be regarded as a word sense disambiguation task that associates a marked target in a sentence with the most appropriate frame in FrameNet. The study conducted in Phase 1 set out to achieve the first research objective (RO1): Develop a robust frame identification model that achieves state-of-the-art accuracy and decently handles unseen words in domain-specific data. To this end, this research designed a frame identification model called Positional Attention-based Frame Identification with BERT (PAFIBERT) that outperformed existing models in various experimental setups and achieved new state-of-the-art performance in frame identification. With the exception of very nuanced senses, the model showed satisfactory performance in identifying frames for ambiguous cases. In particular, the model was proven robust in handling domain-specific data, which often contains jargons and domain-specific words that occur neither in FrameNet nor the development data used for model training.

Figure 6.1. Overview of Phase 1

As part of the efforts to apply frame-based semantic abstraction to lexicon construction, an appraisal concept called graduation (Martin & White, 2005) was incorporated into FrameNet to achieve the second research objective (RO2): Transform semantic frames in FrameNet to produce sentiment-aware semantic abstractions that retain discriminative sentiment-related information. An overview of Phase 2 is depicted in Figure

6.2. The output of this research phase is a sentiment-aware version of FrameNet known as GradFrameNet. To examine the effectiveness of GradFrameNet, a study was conducted to mine phrasal patterns from the semantic-level representations produced by GradFrameNet and a semantic tagger built on PAFIBERT. The phrasal patterns were analyzed to assess the ability of GradFrameNet to generate semantic-level representations that are abstract yet discriminative in a sentiment-related sense. The findings showed that semantic abstraction using GradFrameNet and PAFIBERT facilitated the discovery of interesting phrasal patterns, including sentiment-revealing clues related to consumers' positive and negative experiences, satisfaction, expectation, retention, and word-of-mouth engagement.

Finally, the last phase of the research achieved the third research objective (RO3): Establish a domain-specific sentiment lexicon construction process that uses frame-based semantic abstraction to increase the coverage of phrasal lexicon entries, thereby improving the performance of sentiment classification models for domain-specific data. An overview of Phase 3 is given in Figure 6.3. At this final research stage, the fully-automated lexicon construction method leveraged GradFrameNet and PAFIBERT developed in the preceding phases to create wide-coverage phrasal lexicons from beauty product reviews. The effectiveness of the lexicon construction method was assessed based on the performance gain obtained from incorporating the domain-specific lexicons into sentiment classification models. The study also examined the effects of frame-based semantic abstraction on the sparsity degrees of the development data, as well as its impacts on lexicon size and noise. It was found that although the lexicon size increased drastically with the help of frame-based semantic abstraction, some noise was also introduced because of the frame identification errors. Overall, the classification results showed that the integration of semantic abstraction was beneficial to the lexicon construction process and the performance of sentiment

classification models, but the performance gain relied heavily on the accuracy of the frame identifiers.

Figure 6.2. Overview of Phase 2

Figure 6.3. Overview of Phase 3

6.3 Research Contributions

This research contributes to existing knowledge of corpus-based lexicon construction by treating the issue of word sparsity and lexicon coverage with detail and dealing specifically with the difficulty of extracting phrasal patterns from sparse textual corpora. In this regard, this work makes several original contributions. First and foremost, in addressing the challenges related to the extraction of phrasal clues that capture domain-specific knowledge,

this research investigated and demonstrated the usefulness of semantic-level representations empirically. The idea of extracting phrasal patterns at the semantic-level offers some new insights into how domain lexicon construction can benefit from the semantic similarity that transcends lexical-level variation. Since phrasal units play a crucial role in detecting implicit sentiments and indirect opinions (Balahur et al., 2013; Cambria et al., 2012, 2014, 2016, 2018; Noferesti & Shamsfard, 2015; Ofek et al., 2016), new understanding and avenue in dealing with the extraction of phrasal units should help to foster progress on this complex problem of sentiment analysis.

Compared to the original FrameNet, GradFrameNet produced in this research consists of frames that maintain the discriminative power necessary to distinguish positive opinions from negative opinions and vice versa. This research has been one of the first efforts to examine the plausibility of using FrameNet as a means of semantic abstraction in sentiment-related tasks. Although FrameNet has generated broad interest in various lines of research including question and answering, textual entailment, and other NLP tasks (Aharon et al., 2010; Nuzzolese, Gangemi, & Presutti, 2011; Shen & Lapata, 2007; Shi & Mihalcea, 2005), its potential has not been extensively explored in sentiment analysis because it is not specifically tailored for this purpose. As an early attempt to examine the potential of FrameNet in a rather novel setting in sentiment analysis, this research has enhanced the understanding of how semantic frames in FrameNet can be combined with an important sentiment-related concept defined in the appraisal framework (Martin & White, 2005) to provide abstract yet semantic-rich and sentiment-aware representations. This research demonstrated that the proposed abstraction method is capable of balancing the trade-off between semantic abstraction and sentiment distinction. In other words, the abstraction

process did not cause substantial loss of information but rather generalized in a sentiment-aware manner to extract valuable information for sentiment-related tasks.

Studying interrelations between words—in particular, the effects of negators, intensifiers, mitigators, and other contextual valence shifters—is of significant interest among the research community of sentiment analysis (Jia et al., 2009; Kennedy & Inkpen, 2006; Kiritchenko & Mohammad, 2016a; Polanyi & Zaenen, 2006). However, lists of contextual valence shifters are usually collected on a rather ad hoc manner because of the absence of a comprehensive linguistic resource that can be used for this purpose (e.g., Balahur & Perea-Ortega, 2015; Kiritchenko & Mohammad, 2016a; Qian, Huang, Lei, & Zhu, 2016). This lack of alignment may hinder direct comparisons of research findings across studies. Therefore, on a practical level, GradFrameNet, as well as the frame identifier developed in this research, will benefit the community of researchers and practitioners in sentiment analysis as a useful linguistic resource or tool that provides easy access to sentiment-related semantic categories. By making this resource or tool available for the greater benefits of research and real-world applications, this contribution may also open up opportunities to explore the interplay between other sentiment-related semantic frames, especially those related to the graduation concept defined in the appraisal framework (Martin & White, 2005). Besides, as the first endeavor to realize Martin and White’s (2005) graduation concept on a large scale, this study has also created a useful resource for exploring the phenomenon of graduation in evaluative language.

Furthermore, since sparsity is an inherent characteristic of textual data and is known to affect not only sentiment analysis but also other text mining tasks (Khan et al., 2014; Liu et al., 2005, 2017; Saif et al., 2012, 2014), the findings in this research may also have a

bearing on this fundamental problem in NLP. This research lays the groundwork for future research into frame-based semantic abstraction not only by demonstrating the plausibility of reducing text sparsity using semantic frames but also by addressing various issues related to the implementation of such an approach. Compared to other strategies for data sparsity reduction (Colace, De Santo, Greco, & Napoletano, 2014; Makrehchi & Kamel, 2008; Saif et al., 2012, 2014; Yan, Guo, Lan, & Cheng, 2013; Zhou, Zhang, & Hu, 2008) that mainly rely on statistical methods such as smoothing, dynamic stopword removal, topic extraction, and other dimension reduction techniques, the findings of this research shed new light on the potential of a linguistically-motivated and theoretically-sound approach to this problem.

Finally, this research created a novel attention-based neural network architecture on top of BERT (Devlin et al., 2018) for solving the problem of frame identification. The contributions of this frame identification solution are two-fold. First, frame identification is a crucial step in frame-semantic parsing, and errors in frame identification could cause a significant drop in the overall accuracy of the parsing process (Hartmann et al., 2017). By providing a robust solution for frame identification, this research has made a solid contribution to the field of frame-semantic parsing. Second, in addition to achieving new state-of-the-art results in frame identification, the proposed neural network design also extends our knowledge about the use of BERT (Devlin et al., 2018) in this specific problem of word sense disambiguation. The findings and achievements in this work provide convincing empirical proof that adds to the rapidly expanding body of research related to transfer learning and fine-tuning using pretrained language model in NLP (Devlin et al., 2018; Radford et al., 2018; Howard & Ruder, 2018).

6.4 Limitations and Future Work

The main limitation of this research is that the effectiveness of the lexicon construction process relies on an imperfect frame identification step. From the lexicon construction and sentiment classification experiments in Chapter 5, it seems the accuracy of frame identification is a determining factor that affects the effectiveness of the lexicon construction process. Hence, improving the frame identification accuracy would be a fruitful area for further work. Although the frame identification model (PAFIBERT) developed in this research has achieved new state-of-the-art performance, the analysis of the most frequently confused frames in Chapter 3 revealed that PAFIBERT still failed to recognize more nuanced differences that required a more in-depth understanding of sentence contexts and meanings. Closer inspection of the ambiguous LUs in these frequently confused frames found that distinguishing these frames was indeed extremely challenging. Therefore, further research may want to focus specifically on recognizing the most frequently confused frames since they accounted for a big part of the frame identification errors. Potential ways to improve the model’s accuracy for the frequently confused frames include collecting more annotations for these challenging cases to be used for model training and generating useful features that may help differentiate these frames based on knowledge of linguistic experts. Besides, given that transfer learning using pretrained language models is an emerging area in NLP, the task of frame identification may also benefit from more recent breakthroughs such as ELECTRA (Clark, Luong, Le, & Manning, 2020) and Reformer (Kitaev, Kaiser, & Levskaya, 2020) in the field.

This research focused on extending the coverage or recall of sentiment lexicons. However, both precision and recall are important criteria for high-quality lexicons.

Unfortunately, while it is straightforward to assess the improvements in lexicon coverage quantitatively, because of the size of automatically constructed lexicons, it is not possible to evaluate the precision of the lexicons directly without involving tedious and laborious manual efforts. Therefore, in this research, the automatically constructed lexicons were evaluated indirectly by incorporating lexicon knowledge into sentiment classification models and assessing the performance gain. Although the improvements induced by the lexicon knowledge on the classification models provided evidence for the usefulness of the lexicons, comprehensive evaluation that takes into account both the recall and precision of the lexicons is highly recommended. A plausible way to carry out such evaluation is through crowdsourcing, which breaks down a large amount of manual work into small, independent chunks. These micro-tasks can then be performed in parallel by a large group of people over crowdsourcing platforms such as Amazon Mechanical Turk. The crowdsourcing model has been employed for more efficient and inexpensive human annotations in many linguistic tasks, including those of Callison-Burch (2009), Mohammad and Turney (2013), and Snow, O'Connor, Jurafsky, and Ng (2008). Although crowdsourcing provides a fast and low-cost way to obtain human annotations, one of the major challenges is the quality control of the annotations. As pointed out by Mohammad and Turney (2013), since the host of an annotation task has minimal control over the educational background of the annotators, it is almost certain that some of the participants will provide random and incorrect inputs, either intentionally or because of inability to follow complex instructions. This issue, however, can be mitigated by carefully defining the task, giving clear and brief instructions, and implementing proper post-processing to remove random and erroneous annotations.

With regard to the generalizability of the proposed approach, it is still not completely clear how generalizable this approach is for other types of reviews. In the experiments

presented in Chapter 4, four types of product reviews were examined, including three beauty products and one electronic product. These experiments only required raw data collected from online review websites. However, in the study presented in Chapter 5, since manually labeled data were needed for model development and evaluation, the experiments were only conducted on eyeliner and cleanser products. From the results produced in Chapter 4, it seems the frame-based abstraction approach is quite generalizable across different categories of beauty products, although their frequencies of graduation types and frames may differ. For example, the study in Chapter 4 found that in the eyeliner reviews, the proportion of the Distribution graduation type was noticeably higher because the consumers were often concerned about how long the products could last after they put them on, and this aspect of the eyeliner products is related to distribution in time. Despite the distinction in graduation type frequencies and frame frequencies, valuable information—especially in the form of experiential opinions—was uncovered from the three beauty product types. In contrast, applying the frame-based abstraction approach to the electronic product in Chapter 4 seemed to result in smaller gains in detecting useful phrasal patterns. In particular, phrasal patterns related to retention (e.g., “will buy again”) were not detected in the electronic product reviews. Nevertheless, the proposed approach was still proven effective in revealing information related to customer experiences, satisfaction, expectations, and word-of-mouth. While the sentiment classification experiments in Chapter 5 demonstrated the effectiveness of frame-based semantic abstraction in creating phrasal lexicons that improved the performance of sentiment classifiers, the scope of the experiments was limited to the beauty domain. Although the experiments explored two types of beauty products that required distinct product-specific knowledge for inferring consumers’ likes and dislikes, these experiments are still somewhat inadequate in supporting the generalizability of the proposed

approach. For data in other domains such as hotel reviews, inferring the positivity and negativity of the reviewers' attitudes may require distinction to be drawn within the LUs of certain semantic frames. For example, in hotel reviews, travelers are generally concerned about the hotel location, room size, cleanliness, service quality, and so forth. Hence, for the proposed lexicon construction method to be useful for hotel reviews, the semantic abstraction should not cause loss of information related to the above aspects. In other words, the abstraction process must retain the necessary information for distinguishing between desirable and undesirable facts in the domain (e.g., "close to the airport" versus "far from the airport," "spacious bathroom" versus "small bathroom," etc.). It seems the graduation taxonomy in GradFrameNet might be able to handle these opinions with graduation types such as Proximity (Far, Medium, and Close) and Mass (Large, Medium, and Small). However, replicating the experiments on benchmark datasets or data collected from other domains will provide more conclusive information on the effectiveness and generalizability of the proposed approach.

Another limitation of this research is on the granularity of graduation in GradFrameNet. This research employed a rather coarse-grained strategy when incorporating the graduation concept into FrameNet. In the taxonomy derived from the graduation concept, a three-point scale was used to categorize LUs into three levels of graduation. For instance, the LUs in the [Temperature] frame were assigned Infused_intensification as the graduation type, and three graduation levels grouped the LUs into the following subcategories:

- [Temperature: Infused_intensification = Low]: cool.a, cold.a, freezing.a, frigid.a
- [Temperature: Infused_intensification = Medium]: lukewarm.a, tepid.a
- [Temperature: Infused_intensification = High]: warm.a, hot.a, scalding.a

Although this taxonomy worked well in most cases, more fine-grained graduation based on a five-point scale or even continuous values could be beneficial for sentiment analysis tasks that seek to distinguish between different sentiment strengths or intensities (e.g., Benamara, Cesarano, Picariello, Recupero, & Subrahmanian, 2007; Thelwall, Buckley, & Paltoglou, 2012; Thelwall, Buckley, Paltoglou, Cai, & Kappas, 2010). In other words, this kind of fine-grained graduation will allow GradFrameNet to draw a distinction between an extreme opinion such as “incredible stay power” and a less extreme one such as “stay pretty well.”

6.5 Concluding Remarks

Given the interdisciplinary nature of sentiment analysis, an important principle implied by the present research is that researchers should take advantage of the advancements in other lines of work to bring the field forward. To this end, this research has undertaken the task of examining the usefulness of FrameNet—a linguistically-motivated resource—in improving sentiment lexicon construction. Despite the above limitations, the efforts of this research in enhancing the automatic construction of phrasal sentiment lexicons via frame-based semantic abstraction have shown promising results that reveal the potential of such an approach conceptually and empirically. Further studies in exploring the full potential of this approach in sentiment analysis would be worthwhile.

References

- Agarwal, B., Mittal, N., Bansal, P., & Garg, S. (2015). Sentiment analysis using common-sense and context information. *Computational Intelligence and Neuroscience*, 2015, 30-38. doi:10.1155/2015/715730
- Agarwal, B., Poria, S., Mittal, N., Gelbukh, A., & Hussain, A. (2015). Concept-level sentiment analysis with dependency-based semantic parsing: A novel approach. *Cognitive Computation*, 7, 487-499. doi:10.1007/s12559-014-9316-6
- Aharon, R. B., Szpektor, I., & Dagan, I. (2010). Generating entailment rules from framenet. *Proceedings of the ACL 2010 Conference*, 241-246. Retrieved from <https://www.aclweb.org/anthology/P10-2045>
- Argamon, S., Bloom, K., Esuli, A., & Sebastiani, F. (2007). Automatically determining attitude type and force for sentiment analysis. *Proceedings of the Language and Technology Conference 2007*, 218-231. doi:10.1007/978-3-642-04235-5_19
- Baccianella, S., Esuli, A., & Sebastiani, F. (2010). Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. *Proceedings of LREC 2010*, 2200-2204. Retrieved from <https://www.aclweb.org/anthology/L10-1531>
- Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. Retrieved from arXiv preprint arXiv:1409.0473.
- Baker, C. F., Ellsworth, M., & Erk, K. (2007). SemEval'07 task 19: Frame semantic structure extraction. *Proceedings of the 4th International Workshop on Semantic*

Evaluations, 99-104. Retrieved from <https://aclweb.org/anthology/papers/S/S07/S07-1018>

Baker, C. F., Fillmore, C. J., & Lowe, J. B. (1998). The Berkeley FrameNet project. *Proceedings of the 17th international conference on Computational linguistics*, 86-90. doi:10.3115/980845.980860

Balahur, A., & Perea-Ortega, J. M. (2015). Sentiment analysis system adaptation for multilingual processing: The case of tweets. *Information Processing & Management*, 51, 547-556. doi:10.1016/j.ipm.2014.10.004

Balahur, A., Hermida, J. M., Montoyo, A., & Muñoz, R. (2013). Detecting implicit expressions of affect in text using EmotiNet and its extensions. *Data & Knowledge Engineering*, 88, 113-125. doi:10.1016/j.datak.2013.08.002

Benamara, F., Cesarano, C., Picariello, A., Recupero, D. R., & Subrahmanian, V. S. (2007). Sentiment analysis: Adjectives and adverbs are better than adjectives alone. *Proceedings of ICWSM 2007*, 1-7. Retrieved from <https://scinapse.io/papers/2293633461>

Bengio, Y., & Bengio, S. (1999). Modeling high-dimensional discrete data with multi-layer neural networks. *Proceedings of the 12th International Conference on Neural Information Processing Systems*, 400-406. Retrieved from <https://dl.acm.org/citation.cfm?id=3009714>

Bhaskar, J., Sruthi, K., & Nedungadi, P. (2015). Hybrid approach for emotion classification of audio conversation based on text and speech mining. *Procedia Computer Science*, 46, 635-643. doi:10.1016/j.procs.2015.02.112

- Bickart, B., & Schindler, R. M. (2001). Internet forums as influential sources of consumer information. *Journal of Interactive Marketing*, 15, 31-40. doi:10.1002/dir.1014
- Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., & Hellmann, S. (2009). Dbpedia: A crystallization point for the Web of Data. *Journal of Web Semantics*, 7, 154-165. doi:10.1016/j.websem.2009.07.002
- Blair-Goldensohn, S., Hannan, K., McDonald, R., Neylon, T., Reis, G. A., & Reynar, J. (2008). Building a sentiment summarizer for local service reviews. *Proceedings of NLPIX 2008*. Retrieved from <https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/34368.pdf>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993-1022. Retrieved from <http://papers.nips.cc/paper/2070-latent-dirichlet-allocation.pdf>
- Botschen, T., Gurevych, I., Klie, J.-C., Mousselly-Sergie, H., & Roth, S. (2018). Multimodal frame identification with multilingual evaluation. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1481-1491. doi:10.18653/v1/N18-1134
- Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings*. Technical Report C-1, University of Florida.
- Bross, J., & Ehrig, H. (2013). Automatic construction of domain and aspect specific sentiment lexicons for customer review mining. *Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management - CIKM '13*, 1077-1086. doi:10.1145/2505515.2505574

- Bruni, E., Tran, N. K., & Baroni, M. (2014). Multimodal distributional semantics. *Journal of Artificial Intelligence Research*, 49, 1-47. doi:10.1613/jair.4135
- Callison-Burch, C. (2009). Fast, cheap and creative: Evaluating translation quality using amazon's mechanical turk. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2009)*, 286-295.
doi:10.5555/1699510.1699548
- Cambria, E., Das, D., Bandyopadhyay, S. and Feraco, A. (2017). *A Practical Guide to Sentiment Analysis*. Cham, Switzerland: Springer.
- Cambria, E., Havasi, C., & Hussain, A. (2012). SenticNet 2: A semantic and affective resource for opinion mining and sentiment analysis. *Proceedings of the 25th International Florida Artificial Intelligence Research Society Conference*.
doi:10.1.1.392.6279
- Cambria, E., Olsher, D., & Rajagopal, D. (2014). SenticNet 3: A common and common-sense knowledge base for cognition-driven sentiment analysis. *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, 1515-1521. doi:10.5555/2892753.2892763
- Cambria, E., Poria, S., Bajpai, R., & Schuller, B. (2016). SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives. *Proceedings of COLING 2016*, 2666-2677. Retrieved from <https://www.aclweb.org/anthology/C16-1251>
- Cambria, E., Poria, S., Hazarika, D., & Kwok, K. (2018). SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings. *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 1795-1802. Retrieved from <https://aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16839>

- Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent systems*, 28, 15-21. doi:10.1109/MIS.2013.30
- Carretero, M., & Taboada, M. (2014). Graduation within the scope of attitude in English and Spanish consumer reviews of books and movies. *Evaluation in Context*, 221-239. doi:10.1075/pbns.242.11car
- Chen, H. Y., & Chen, H. H. (2016). Implicit polarity and implicit aspect recognition in opinion mining. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 20-25. doi:10.18653/v1/P16-2004
- Chen, Y., & Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science*, 54, 477-491. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=618782.
- Chen, Y., Fay, S., & Wang, Q. (2003). Marketing implications of online consumer product reviews. *Business Week*, 7150, 1-36. Retrieved from <http://plaza.ufl.edu/faysa/review.pdf>
- Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43, 345-354. doi:10.1509/jmkr.43.3.345
- Chikersal, P., Poria, S., & Cambria, E. (2015). SeNTU: Sentiment analysis of tweets by combining a rule-based classifier with supervised learning. *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, 647-651. doi:10.18653/v1/S15-2108

- Choi, Y., & Cardie, C. (2009). Adapting a Polarity Lexicon using Integer Linear Programming for Domain-Specific Sentiment Classification. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, 590-598.
Retrieved from <https://www.aclweb.org/anthology/D09-1062>
- Churchill Jr, G. A., & Surprenant, C. (1982). An investigation into the determinants of customer satisfaction. *Journal of Marketing Research*, 19, 491-504.
doi:10.2307/3151722
- Clark, K., Luong, M. T., Le, Q. V., & Manning, C. D. (2020). Electra: Pre-training text encoders as discriminators rather than generators. *Proceedings of the 2020 International Conference on Learning Representations (ICLR 2020)*. Retrieved from <https://openreview.net/pdf?id=r1xMH1BtvB>
- Colace, F., De Santo, M., Greco, L., & Napoletano, P. (2014). Text classification using a few labeled examples. *Computers in Human Behavior*, 30, 689-697.
doi:10.1016/j.chb.2013.07.043
- Constantinides, E., & Fountain, S. J. (2008). Web 2.0: Conceptual foundations and marketing issues. *Journal of Direct, Data and Digital Marketing Practice*, 9, 231-244.
doi:10.1057/palgrave.dddmp.4350098
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273-297. doi:10.1007/BF00994018
- Cruz, I., Gelbukh, A. F., & Sidorov, G. (2014). Implicit Aspect Indicator Extraction for Aspect based Opinion Mining. *International Journal of Computational Linguistics and*

- Applications*, 5, 135-152. Retrieved from <https://www.ijcla.org/2014-2/IJCLA-2014-2-pp-135-152-Implicit.pdf>
- Das, D., & Smith, N. A. (2011). Semi-supervised frame-semantic parsing for unknown predicates. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 1435-1444. Retrieved from <https://www.aclweb.org/anthology/P11-1144>
- Das, D., Chen, D., Martins, A. F., Schneider, N., & Smith, N. A. (2014). Frame-semantic parsing. *Computational Linguistics*, 40, 9-56. Retrieved from <http://repository.cmu.edu/cgi/viewcontent.cgi?article=1173&context=lti>
- Das, D., Schneider, N., Chen, D., & Smith, N. A. (2010). *SEMAFOR 1.0: A probabilistic frame-semantic parser*. Language Technologies Institute, School of Computer Science, Carnegie Mellon University.
- Davidov, D., Tsur, O., & Rappoport, A. (2010). Enhanced sentiment learning using twitter hashtags and smileys. *Proceedings of the 23rd International Conference on Computational Linguistics*, 241-249. Retrieved from <https://dl.acm.org/citation.cfm?id=1944594>
- Denny, M. J., & Spirling, A. (2018). Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it. *Political Analysis*, 26, 168-189. doi:10.1017/pan.2017.44
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. Retrieved from arXiv preprint arXiv:1810.04805.

- Dey, A., Jenamani, M., & Thakkar, J. J. (2018). Senti-N-Gram : An n-gram lexicon for sentiment analysis. *Expert Systems with Applications*, 103, 92-105.
doi:10.1016/j.eswa.2018.03.004
- Ding, X., Liu, B., & Yu, P. S. (2008). A holistic lexicon-based approach to opinion mining. *Proceedings of the 2008 International Conference on Web Search and Data Mining*, 231-240. doi:10.1145/1341531.1341561
- Dviso, A. (2010). Through the eyes of the consumer: Consumer shopping habit survey. *Channel Advisor Corporation*, 2010, 5-9. Retrieved from <http://go.channeladvisor.com/rs/channeladvisor/images/us-wp-consumer-survey-2010.pdf>
- Esuli, A., & Sebastiani, F. (2005). Determining the semantic orientation of terms through gloss classification. *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, 617-624. doi:10.1145/1099554.1099713
- Esuli, A., & Sebastiani, F. (2006). Determining term subjectivity and term orientation for opinion mining. *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*. doi:10.1.1.60.8645
- Esuli, A., & Sebastiani, F. (2007). SentiWordNet: A high-coverage lexical resource for opinion mining. *Evaluation*, 1-26. Retrieved from <http://ontotext.fbk.eu/Publications/sentiWN-TR.pdf>
- Feng, S., Song, K., Wang, D., & Yu, G. (2015). A word-emoticon mutual reinforcement ranking model for building sentiment lexicon from massive collection of microblogs. *World Wide Web*, 18, 949-967. doi:10.1007/s11280-014-0289-x

- Fillmore, C. J. (1976). Frame semantics and the nature of language. *Proceedings of the Annals of the New York Academy of Sciences: Annals of the New York Academy of Sciences*, 20-32. Retrieved from <http://www.icsi.berkeley.edu/pubs/ai/framesemantics76.pdf>
- Fillmore, C. J., Johnson, C. R., & Petruck, M. R. (2003). Background to FrameNet. *International journal of lexicography*, 16, 235-250. doi:10.1093/ijl/16.3.235
- Finkel, J. R., Grenager, T., & Manning, C. (2005). Incorporating non-local information into information extraction systems by gibbs sampling. *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, 363-370. doi:10.3115/1219840.1219885
- FitzGerald, N., Täckström, O., Ganchev, K., & Das, D. (2015). Semantic role labeling with neural network factors. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 960-970. doi:10.18653/v1/D15-1112
- Geeraerts, D. (2006). *Cognitive linguistics: Basic readings*. Berlin, New York: Mouton de Gruyter.
- Geeraerts, D., & Cuyckens, H. (2007). Introducing cognitive linguistics. *The Oxford Handbook of Cognitive Linguistics*. doi:10.1093/oxfordhb/9780199738632.013.0001
- Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21, 267-297. doi:10.1093/pan/mps028P

- Guthrie, D., Allison, B., Liu, W., Guthrie, L., & Wilks, Y. (2006). A closer look at skip-gram modelling. *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, 1222-1225. Retrieved from <https://www.aclweb.org/anthology/papers/L/L06/L06-1210>
- Hamilton, W. L., Clark, K., Leskovec, J., & Jurafsky, D. (2016). Inducing Domain-Specific Sentiment Lexicons from Unlabeled Corpora. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 595-605. doi:10.18653/v1/D16-1057
- Hartmann, S., Kuznetsov, I., Martin, T., & Gurevych, I. (2017). Out-of-domain FrameNet semantic role labeling. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, 471-482. Retrieved from <https://www.aclweb.org/anthology/E17-1045>
- Hatzivassiloglou, V., & McKeown, K. R. (1997). Predicting the semantic orientation of adjectives. *Proceedings of the 8th Conference on European Chapter of the Association for Computational Linguistics*, 174-181. doi:10.3115/976909.979640
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770-778. doi:10.1109/CVPR.2016.90
- Hermann, K. M., Das, D., Weston, J., & Ganchev, K. (2014). Semantic Frame Identification with Distributed Word Representations. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 1448-1458. doi:10.3115/v1/P14-1136

- Howard, J., & Ruder, S. (2018). Universal Language Model Fine-tuning for Text Classification. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, 328-339. doi:10.18653/v1/P18-1031
- Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 168-177. doi:10.1145/1014052.1014073
- Huang, S., Niu, Z., & Shi, C. (2014). Automatic construction of domain-specific sentiment lexicon based on constrained label propagation. *Knowledge-Based Systems*, 56, 191-200. doi:10.1016/j.knosys.2013.11.009
- Jia, L., Yu, C., & Meng, W. (2009). The effect of negation on sentiment analysis and retrieval effectiveness. *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, 1827-1830. doi:10.1145/1645953.1646241
- Johansson, R., & Nugues, P. (2007). Johansson and Nugues - SEMEVAL 2007. *Proceedings of the 4th International Workshop on Semantic Evaluations - SemEval'07*, 227-230. doi:10.3115/1621474.1621522
- Kaji, N., & Kitsuregawa, M. (2007). Building lexicon for sentiment analysis from massive collection of HTML documents. *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 1075-1083. Retrieved from <http://www.aclweb.org/anthology/D07-1115>

- Kamps, J., Marx, M., Mokken, R. J., & De Rijke, M. (2004). Using WordNet to measure semantic orientations of adjectives. *Proceedings of LREC 2004*, 1115-1118. Retrieved from <http://humanities.uva.nl/~kamps/publications/2004/kamp:usin04.pdf>
- Kanayama, H., & Nasukawa, T. (2006). Fully automatic lexicon expansion for domain-oriented sentiment analysis. *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, 355-363. Retrieved from <https://www.aclweb.org/anthology/W06-1642>
- Kennedy, A., & Inkpen, D. (2006). Sentiment classification of movie reviews using contextual valence shifters. *Computational Intelligence*, 22, 110-125.
doi:10.1111/j.1467-8640.2006.00277.x
- Khan, F. H., Bashir, S., & Qamar, U. (2014). TOM: Twitter opinion mining framework using hybrid classification scheme. *Decision Support Systems*, 57, 245-257.
doi:10.1016/j.dss.2013.09.004
- Khan, F. H., Qamar, U., & Bashir, S. (2017). A semi-supervised approach to sentiment analysis using revised sentiment strength based on SentiWordNet. *Knowledge and Information Systems*, 51, 851-872. doi:10.1007/s10115-016-0993-1
- Kim, S. M., & Hovy, E. (2004). Determining the sentiment of opinions. *Proceedings of the 20th international conference on Computational Linguistics*, 1367-1373.
Retrieved from <https://www.aclweb.org/anthology/papers/C/C04/C04-1200/>
- Kiritchenko, S. & Mohammad, S. M. (2016a), The effect of negators, modals, and degree adverbs on sentiment composition. *Proceedings of the 2016 Workshop on*

- Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 43-52. doi:10.18653/v1/W16-0410
- Kiritchenko, S. & Mohammad, S. M. (2016b), Sentiment composition of words with opposing polarities. *Proceedings of NAACL-HLT 2016*, 1102-1108.
doi:10.18653/v1/N16-1128
- Kiritchenko, S., Zhu, X., & Mohammad, S. M. (2014a). Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research*, 50, 723-762.
doi:10.1613/jair.4272
- Kiritchenko, S., Zhu, X., Cherry, C., & Mohammad, S. (2014b). NRC-Canada-2014: Detecting aspects and sentiment in customer reviews. *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, 437-442.
doi:10.3115/v1/S14-2076
- Kitaev, N., Kaiser, Ł., & Levskaya, A. (2020). Reformer: The efficient transformer. *Proceedings of the 2020 International Conference on Learning Representations (ICLR 2020)*. Retrieved from <https://openreview.net/pdf?id=rkgNKkHtvB>
- Korenek, P., & Šimko, M. (2014). Sentiment analysis on microblog utilizing appraisal theory. *World Wide Web*, 17, 847-867. doi:10.1007/s11280-013-0247-z
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097-1105. doi:10.1145/3065386

Kshirsagar, M., Thomson, S., Schneider, N., Carbonell, J., Smith, N. A., & Dyer, C.

(2015). Frame-semantic role labeling with heterogeneous annotations. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, 218-224.

doi:10.3115/v1/P15-2036

Larsen, B., & Aone, C. (1999). Fast and effective text mining using linear-time document clustering. *Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 16-22. doi:10.1145/312129.312186

Lee, J., Park, D. H., & Han, I. (2008). The effect of negative online consumer reviews on product attitude: An information processing view. *Electronic Commerce Research and Applications*, 7, 341-352. doi:10.1016/j.elerap.2007.05.004

Li, W., Guo, K., Shi, Y., Zhu, L., & Zheng, Y. (2018). DWWP: Domain-specific new words detection and word propagation system for sentiment analysis in the tourism domain. *Knowledge-Based Systems*, 146, 203–214. doi:10.1016/j.knosys.2018.02.004

Liu, B. (2015), *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. New York, NY: Cambridge University Press.

Liu, J., Chang, W. C., Wu, Y., & Yang, Y. (2017). Deep learning for extreme multi-label text classification. *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 115-124.

doi:10.1145/3077136.3080834

Liu, L., Kang, J., Yu, J., & Wang, Z. (2005). A comparative study on unsupervised feature selection methods for text clustering. *Proceedings of the 2005 International*

- Conference on Natural Language Processing and Knowledge Engineering*, 597-601.
doi:10.1109/NLPKE.2005.1598807
- Lowe, J. B. (1997). A frame-semantic approach to semantic annotation. *Proceedings of the 1997 Siglex Workshop/ANLP97*. doi:10.1.1.11.4657
- Lu, Y., Castellanos, M., Dayal, U., & Zhai, C. (2011). Automatic construction of a context-aware sentiment lexicon: An optimization approach. *Proceedings of the 20th International Conference on World Wide Web*, 347-356.
doi:10.1145/1963405.1963456
- Luong, M. T., Pham, H., & Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 1412-1421. doi:10.18653/v1/D15-1166
- Ma, D., Li, S., Zhang, X., & Wang, H. (2017). Interactive attention networks for aspect-level sentiment classification. *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, 4068-4074. Retrieved from
<https://dl.acm.org/citation.cfm?id=3171854>
- Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., & Schmid-Petri, H. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, 12, 93-118.
doi:10.1080/19312458.2018.1430754
- Makrehchi, M., & Kamel, M. S. (2008). Automatic extraction of domain-specific stopwords from labeled documents. *Proceedings of the European Conference on Information Retrieval*, 222-233. doi:10.1007/978-3-540-78646-7_22

- Manning, C.D. and Schütze, H. (1999), *Foundations of Statistical Natural Language Processing*. Cambridge, Massachusetts: MIT Press.
- Marcus, M., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19, 313–330.
doi:10.5555/972470.972475
- Martin, J.R. and White, P.R. (2005), *The Language of Evaluation*. Basingstoke, UK: Palgrave Macmillan.
- Mauri, A. G., & Minazzi, R. (2013). Web reviews influence on expectations and purchasing intentions of hotel potential customers. *International Journal of Hospitality Management*, 34, 99-107. doi:10.1016/j.ijhm.2013.02.012
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013a). Efficient estimation of word representations in vector space. *Proceedings of ICLR 2013*. Retrieved from <https://arxiv.org/abs/1301.3781>
- Mikolov, T., Grave, E., Bojanowski, P., Puhersch, C., & Joulin, A. (2017). Advances in pre-training distributed word representations. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*, 52-55. Retrieved from <https://www.aclweb.org/anthology/L18-1008>
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. *Proceedings of the 26th International Conference on Neural Information Processing Systems*, 3111-3119. Retrieved from <https://dl.acm.org/citation.cfm?id=2999959>

- Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, 3, 235-244. doi:10.1093/ijl/3.4.235
- Mitchell, J., & Lapata, M. (2010). Composition in distributional models of semantics. *Cognitive science*, 34, 1388-1429. doi:10.1111/j.1551-6709.2010.01106.x
- Mohammad, S. M., Kiritchenko, S., & Zhu, X. (2013). NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets. *Proceedings of the Second Joint Conference on Lexical and Computational Semantics*, 321–327. Retrieved from <https://pdfs.semanticscholar.org/5020/80d69b7c4827d20f437b9a2ecd62f6d7cb2c.pdf>
- Mohammad, S. M., Sobhani, P., & Kiritchenko, S. (2017). Stance and sentiment in tweets. *ACM Transactions on Internet Technology*, 17. doi:10.1145/3003433
- Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29, 436-465. doi:10.1111/j.1467-8640.2012.00460.x
- Mohammad, S. M., Zhu, X., Kiritchenko, S., & Martin, J. (2015). Sentiment, emotion, purpose, and style in electoral tweets. *Information Processing & Management*, 51, 480-499. doi:10.1016/j.ipm.2014.09.003
- Nallapati, R., Zhou, B., Gulcehre, C., & Xiang, B. (2016). Abstractive text summarization using sequence-to-sequence RNNs and beyond. *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning*, 280-290. Retrieved from <https://www.aclweb.org/anthology/K16-1028>

- Noferesti, S., & Shamsfard, M. (2015). Resource construction and evaluation for indirect opinion mining of drug reviews. *PloS one*, *10*. doi:10.1371/journal.pone.0124993
- Nuzzolese, A. G., Gangemi, A., & Presutti, V. (2011). Gathering lexical linked data and knowledge patterns from FrameNet. *Proceedings of the 6th International Conference on Knowledge Capture*, 41-48. doi:10.1145/1999676.1999685
- Ofek, N., Poria, S., Rokach, L., Cambria, E., Hussain, A., & Shabtai, A. (2016). Unsupervised commonsense knowledge enrichment for domain-specific sentiment analysis. *Cognitive Computation*, *8*, 467-477. doi:10.1007/s12559-015-9375-3
- Ofoghi, B., Yearwood, J., & Ma, L. (2009). The impact of frame semantic annotation levels, frame-alignment techniques, and fusion methods on factoid answer processing. *Journal of the American Society for Information Science and Technology*, *60*, 247-263. doi:10.1002/asi.20989
- Oliver, R. L. (1977). Effect of expectation and disconfirmation on postexposure product evaluations: An alternative interpretation. *Journal of Applied Psychology*, *62*, 480-486. doi:10.1037/0021-9010.62.4.480
- Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, *17*, 460-469. doi:10.2307/3150499
- Osgood, C. E. (1952). The nature and measurement of meaning. *Psychological Bulletin*, *49*, 197-237. doi:10.1037/h0055737

Palmer, M., Gildea, D., & Kingsbury, P. (2005). The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31, 71-106.

doi:10.1162/0891201053630264

Palmer, M., Gildea, D., & Xue, N. (2010). Semantic role labeling. *Synthesis Lectures on Human Language Technologies*, 3, 1-103.

doi:10.2200/S00239ED1V01Y200912HLT006

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: Sentiment classification using machine learning techniques. *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, 79-86. doi:10.3115/1118693.1118704

Peng, H., Thomson, S., Swayamdipta, S., & Smith, N. A. (2018). Learning joint semantic parsers from disjoint data. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1492-1502. doi:10.18653/v1/N18-1135

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532-1543. doi:10.3115/v1/D14-1162

Polanyi, L., & Zaenen, A. (2006). Contextual valence shifters. *Computing Attitude and Affect in Text: Theory and Applications*, 20, 1-10. doi:10.1007/1-4020-4102-0_1

Poria, S., Agarwal, B., Gelbukh, A., Hussain, A., & Howard, N. (2014). Dependency-based semantic parsing for concept-level text analysis. *Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics*, 113-127.

doi:10.1007/978-3-642-54906-9_10

- Poria, S., Cambria, E., Winterstein, G., & Huang, G. B. (2014). Sentic patterns: Dependency-based rules for concept-level sentiment analysis. *Knowledge-Based Systems*, 69, 45-63. doi:10.1016/j.knosys.2014.05.005
- Poria, S., Chaturvedi, I., Cambria, E., & Bisio, F. (2016). Sentic LDA: Improving on LDA with semantic similarity for aspect-based sentiment analysis. *Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN)*, 4465-4473. doi:10.1109/IJCNN.2016.7727784
- Punyakank, V., Roth, D., & Yih, W. T. (2008). The importance of syntactic parsing and inference in semantic role labeling. *Computational Linguistics*, 34, 257-287. doi:10.1162/coli.2008.34.2.257
- Qian, Q., Huang, M., Lei, J., & Zhu, X. (2016). Linguistically regularized LSTMs for sentiment classification. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, 1679-1689. doi:10.18653/v1/P17-1154
- Qiu, G., Liu, B., Bu, J., & Chen, C. (2009). Expanding domain sentiment lexicon through double propagation. *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence*, 1199-1204. doi:10.5555/1661445.1661637
- Qu, L., Ifrim, G., & Weikum, G. (2010). The bag-of-opinions method for review rating prediction from sparse text patterns. *Proceedings of the 23rd International Conference on Computational Linguistics*, 913-921. doi:10.5555/1873781.1873884
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. Retrieved from <https://s3-us-west->

2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf

Rajagopal, D., Cambria, E., Olsher, D., & Kwok, K. (2013). A graph-based approach to commonsense concept extraction and semantic similarity detection. *Proceedings of the 22nd International Conference on World Wide Web*, 565-570.

doi:10.1145/2487788.2487995

Ranaweera, C., & Prabhu, J. (2003). On the relative importance of customer satisfaction and trust as determinants of customer retention and positive word of mouth. *Journal of Targeting, Measurement and Analysis for marketing*, 12, 82-90.

doi:10.1057/palgrave.jt.5740100

Rao, D., & Ravichandran, D. (2009). Semi-supervised polarity lexicon induction. *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, 675-682. doi:10.3115/1609067.1609142

Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based Systems*, 89, 14-46.

doi:10.1016/j.knosys.2015.06.015

Read, J., & Carroll, J. (2012). Annotating expressions of appraisal in English. *Language Resources and Evaluation*, 46, 421-447. doi:10.1007/s10579-010-9135-7

Ruppenhofer, J., & Rehbein, I. (2012). Semantic frames as an anchor representation for sentiment analysis. *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*, 104-109. Retrieved from

<https://www.aclweb.org/anthology/W12-3716/>

- Rush, A. M., Chopra, S., & Weston, J. (2015). A neural attention model for abstractive sentence summarization. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 379-389. doi:10.18653/v1/D15-1044
- Saif, H., Fernández, M., He, Y., & Alani, H. (2014). On stopwords, filtering and data sparsity for sentiment analysis of twitter. *Proceedings of LREC 2014, Ninth International Conference on Language Resources and Evaluation*, 810-817. Retrieved from <http://oro.open.ac.uk/40666/>
- Saif, H., He, Y., & Alani, H. (2012). Alleviating data sparsity for twitter sentiment analysis. *Proceedings of the 2nd Workshop on Making Sense of Microposts at the 21st International Conference on the World Wide Web (WWW'12)*, 2-9. Retrieved from <http://oro.open.ac.uk/38501/>
- Schouten, K., De Boer, N., Lam, T., Van Leeuwen, M., Van Luijk, R., & Frasincar, F. (2015). Semantics-driven implicit aspect detection in consumer reviews. *Proceedings of the 24th International Conference on World Wide Web*, 109-110. doi:10.1145/2740908.2742734
- Sharma, R., Gupta, M., Agarwal, A., & Bhattacharyya, P. (2015). Adjective intensity and sentiment analysis. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2520-2526. doi:10.18653/v1/D15-1300
- Shen, D., & Lapata, M. (2007). Using semantic roles to improve question answering. *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 12-21. Retrieved from <https://www.aclweb.org/anthology/D07-1002/>

- Shi, L., & Mihalcea, R. (2005). Putting pieces together: Combining FrameNet, VerbNet and WordNet for robust semantic parsing. *Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics*, 100-111.
doi:10.1007/978-3-540-30586-6_9
- Sikos, J., & Padó, S. (2018). FrameNet's Using relation as a source of concept-based paraphrases. *Constructions and Frames*, 10, 38-60. doi:10.1075/cf.00010.sik
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. Retrieved from <https://arxiv.org/abs/1409.1556>
- Sinclair, J. (2004). The search for units of meaning. *Trust the Text*, 24-48.
doi:10.1093/applin/amn052
- Singh, P. K., & Husain, M. S. (2014). Methodological study of opinion mining and sentiment analysis techniques. *International Journal on Soft Computing*, 5, 11-21.
doi:10.5121/ijsc.2014.5102
- Snow, R., O'Connor, B., Jurafsky, D., and Ng, A. (2008). Cheap and fast - but is it good? Evaluating nonexpert annotations for natural language tasks. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2008)*, 254-263. doi:10.5555/1613715.1613751
- Song, K., Feng, S., Gao, W., Wang, D., Chen, L., & Zhang, C. (2015). Build emotion lexicon from microblogs by combining effects of seed words and emoticons in a heterogeneous graph. *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, 171-195. doi:10.1145/2700171.2791035

- Speer, R., & Havasi, C. (2012). Representing general relational knowledge in ConceptNet5. *Proceedings of LREC 2012*, 3679-3686. Retrieved from <https://www.aclweb.org/anthology/L12-1639/>
- Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell Publishers.
- Swayamdipta, S., Thomson, S., Dyer, C., & Smith, N. A. (2017). Frame-semantic parsing with softmax-margin segmental RNNs and a syntactic scaffold. Retrieved from <https://arxiv.org/abs/1706.09528>
- Taboada, M. (2016). Sentiment analysis: An overview from linguistics. *Annual Review of Linguistics*, 2, 325-347. doi:10.1146/annurev-linguistics-011415-040518
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). *Lexicon-based methods for sentiment analysis*. *Computational linguistics*, 37, 267-307. doi:10.1162/COLI_a_00049
- Tan, L. K. W., Na, J. C., Theng, Y. L., & Chang, K. (2012). Phrase-level sentiment polarity classification using rule-based typed dependencies and additional complex phrases consideration. *Journal of Computer Science and Technology*, 27, 650-666. doi:10.1007/s11390-012-1251-y
- Tan, S., & Zhang, J. (2008). An empirical study of sentiment analysis for Chinese documents. *Expert Systems with applications*, 34, 2622-2629. doi:10.1016/j.eswa.2007.05.028

- Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63, 163-173. doi:10.1002/asi.21662
- Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61, 2544-2558. doi:10.1002/asi.21416
- Tonelli, S., & Giuliano, C. (2009). Wikipedia as frame information repository. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, 276-285. Retrieved from <https://www.aclweb.org/anthology/papers/D/D09/D09-1029/>
- Tsai, A. C. R., Wu, C. E., Tsai, R. T. H., & Hsu, J. Y. J. (2013). Building a concept-level sentiment dictionary based on commonsense knowledge. *IEEE Intelligent Systems*, 28, 22-30. doi:10.1109/MIS.2013.25
- Tse, D. K., & Wilton, P. C. (1988). Models of consumer satisfaction formation: An extension. *Journal of Marketing Research*, 25, 204-212. doi:10.2307/3172652
- Tubishat, M., Idris, N., & Abushariah, M. A. (2018). Implicit aspect extraction in sentiment analysis: Review, taxonomy, opportunities, and open challenges. *Information Processing & Management*, 54, 545-563. doi:10.1016/j.ipm.2018.03.008
- Turney, P. D., & Littman, M. L. (2002). Unsupervised learning of semantic orientation from a hundred-billion-word corpus. Retrieved from <https://arxiv.org/ftp/cs/papers/0212/0212012.pdf>

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Proceedings of the 31st Conference on Neural Information Processing Systems*, 5998-6008. Retrieved from <https://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf>
- Velikovich, L., Blair-Goldensohn, S., Hannan, K., & McDonald, R. (2010). The viability of web-derived polarity lexicons. *Proceedings of the 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 777-785. doi:10.1.1.182.3213
- Wang, H., Lu, Y., & Zhai, C. (2010). Latent aspect rating analysis on review text data: A rating regression approach. *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 783-792. doi:10.1145/1835804.1835903
- Wang, H., Wang, N., & Yeung, D. Y. (2015). Collaborative deep learning for recommender systems. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1235-1244. doi:10.1145/2783258.2783273
- Wang, S., & Manning, C. D. (2012). Baselines and bigrams: Simple, good sentiment and topic classification. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, 90-94. doi:10.5555/2390665.2390688
- Wang, Y., Huang, M., & Zhao, L. (2016). Attention-based LSTM for aspect-level sentiment classification. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 606-615. doi:10.18653/v1/D16-1058

- Weston, J., Bengio, S., & Usunier, N. (2011). Wsabie: Scaling up to large vocabulary image annotation. *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, 2764–2770. doi:10.5555/2283696.2283856
- White, P. R. (2015). Appraisal theory. *The International Encyclopedia of Language and Social Interaction*, 1, 1-7. doi:10.1002/9781118611463/wbielsi041
- Whitelaw, C., Garg, N., & Argamon, S. (2005). Using appraisal groups for sentiment analysis. *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, 625-631. doi:10.1145/1099554.1099714
- Wilson, T., Wiebe, J., & Hwa, R. (2006). Recognizing strong and weak opinion clauses. *Computational Intelligence*, 22, 73-99. doi:10.1111/j.1467-8640.2006.00275.x
- Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. *Proceedings of the 2005 Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, 347-354. doi:10.3115/1220575.1220619
- Wu, C. E., & Tsai, R. T. H. (2014). Using relation selection to improve value propagation in a ConceptNet-based sentiment dictionary. *Knowledge-Based Systems*, 69, 100-107. doi:10.1016/j.knosys.2014.04.043
- Wu, W., Li, H., Wang, H., & Zhu, K. Q. (2012). Probase: A probabilistic taxonomy for text understanding. *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, 481-492. doi:10.1145/2213836.2213891

- Xia, R., Zong, C., & Li, S. (2011). Ensemble of feature sets and classification algorithms for sentiment classification. *Information Sciences*, 181, 1138-1152.
doi:10.1016/j.ins.2010.11.023
- Xiong, C., Merity, S., & Socher, R. (2016). Dynamic memory networks for visual and textual question answering. *Proceedings of the International Conference on Machine Learning*, 2397-2406. doi:10.5555/3045390.3045643
- Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013). A biterm topic model for short texts. *Proceedings of the 22nd International Conference on World Wide Web*, 1445-1456.
doi:10.1145/2488388.2488514
- Yang, B., & Mitchell, T. (2017). A joint sequential and relational model for frame-semantic parsing. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 1247-1256. doi:10.18653/v1/D17-1128
- Yang, L., Ai, Q., Guo, J., & Croft, W. B. (2016). aNMM: Ranking short answer texts with attention-based neural matching model. *Proceedings of the 25th ACM International Conference on Information and Knowledge Management*, 287-296. Retrieved from <https://arxiv.org/abs/1801.01641>
- Yang, Z., & Peterson, R. T. (2004). Customer perceived value, satisfaction, and loyalty: The role of switching costs. *Psychology & Marketing*, 21, 799-822.
doi:10.1002/mar.20030
- Ye, Q., Law, R., Gu, B., & Chen, W. (2011). The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to

hotel online bookings. *Computers in Human Behavior*, 27, 634-639.

doi:10.1016/j.chb.2010.04.014

Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1993). The nature and determinants of customer expectations of service. *Journal of the Academy of Marketing Science*, 21, 1-12. doi:10.1177/0092070393211001

Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1996). The behavioral consequences of service quality. *Journal of Marketing*, 60, 31-46. doi:10.2307/1251929

Zhang, L., Ghosh, R., Dekhil, M., Hsu, M., & Liu, B. (2011). *Combining lexicon-based and learning-based methods for Twitter sentiment analysis*. HP Laboratories, Technical Report HPL-2011, 89.

Zhang, W., Du, T., & Wang, J. (2016). Deep learning over multi-field categorical data. *Proceedings of the European Conference on Information Retrieval*, 45-57. doi:10.1145/3233770

Zhang, Z., & Singh, M. P. (2014). Renew: A semi-supervised framework for generating domain-specific lexicons and sentiment analysis. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 542-551. doi:10.3115/v1/P14-1051

Zhang, Z., Ye, Q., Law, R., & Li, Y. (2010). The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews. *International Journal of Hospitality Management*, 29, 694-700. doi:10.1016/j.ijhm.2010.02.002

- Zhang, Z., Ye, Q., Zhang, Z. and Li, Y. (2011), Sentiment classification of Internet restaurant reviews written in Cantonese, *Expert Systems with Applications*, 38, 7674-7682. doi:10.1016/j.eswa.2010.12.147
- Zhou, X., Zhang, X., & Hu, X. (2008). Semantic Smoothing for Bayesian Text Classification with Small Training Data. *Proceedings of the 2008 SIAM International Conference on Data Mining*, 289-300. doi:10.1137/1.9781611972788.26
- Zipf, G.K. (1949), *Human Behavior and the Principle of Least Effort*. Oxford, England: Addison-Wesley Press.
- Zipf, G.K. (2016), *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Eastford, USA: Martino Fine Books.

Appendix A: The [Desirability] Frame (FrameNet 1.7)

[Desirability]

Definition:

This frame concerns an **Evaluee** being judged for its quality, i.e. how much it would probably be liked. In many cases, the **Evaluee** is implicitly judged good or bad relative to other instances of its type. The **Evaluee**'s desirability is determined by one or more **Parameters**, which are scalar properties of the **Evaluee**. The evaluation may also explicitly be relativized to a set of **Circumstances**, a **Comparison_set** of entities that belong to the same class as the **Evaluee**, or an **Affected_party**. The **Degree** of goodness or badness may also be expressed. Note: With some targets, desirability is conventionally aligned with quantity, i.e., GOOD is MORE.

The view was **ASTONISHING**.

On clear days, the view was **EXCELLENT**.

The book is **ASTOUNDING** in its scope.

The games have been **AWFUL** for the team.

FEs:

Core:

Evaluable [Evaluable]

The Evaluable is being judged for its quality.

He gave me another LAME excuse.

Core Unexpressed:

Parameter [Parameter]

This FE marks expressions that denote a scalar property of the Evaluable with respect to which the Evaluable is judged.

The Beaujolais Nouveau is dazzling in freshness.

Non-Core:

Affected_party [Affected_p]

The Affected_party indicates entities that benefit or suffer from the good or bad quality of the Evaluable.

The season was SPLENDID for the team.

Circumstances [Circumstances]

Circumstances marks expressions that indicate a set of conditions under which the quality of the Evaluable is being judged.

It's a GREAT game when it rains.

Comparison_set []

This FE marks expressions that indicate what class of things or events the Evaluable is being judged against.

The standard of living is POOR compared to other Western industrialized countries.

Degree [Degree]

Semantic Type: Degree

This FE identifies the Degree to which the Evaluatee is judged good or bad.

Smithers was doing an absolutely **SPLENDID** job.

Depictive [Depictive]

The Frame Element Depictive encodes a state that the Evaluatee is in and which does not represent a temporary or hypothetical set of Circumstances.

The aerial view is **SUPERB** with its view of the harbour.

FE Core set(s):

{Evaluatee, Parameter}

Frame-frame Relations:

Inherits from: [Gradable attributes](#)

Is Inherited by: [Aesthetics](#), [Desirable event](#), [Popularity](#), [Social desirability](#)

Perspective on:

Is Perspectivized in:

Uses: [Experiencer focused emotion](#)

Is Used

by: [Cause benefit or detriment](#), [Grooming](#), [Preferred alternative scenario](#), [Thriving](#)

Subframe of:

Has Subframe(s):

Precedes:

Is Preceded by:

Is Inchoative of:

Is Causative of:

See also:

Lexical Units:

amazing.a, appalling.a, astonishing.a, astounding.a, atrocious.a, average.a, awesome.a, awful.a, bad.a, best thing since sliced bread.n, cool.a, crap.n, crappy.a, decent.a, desirable.a, dreadful.a, excellence.n, excellent.a, execrable.a, extraordinary.a, fabulous.a, fair.a, fantastic.a, fine.a, first-rate.a, garbage.n, gem.n, gold.n, good.a, great.a, horrible.a, idyllic.a, incredible.a, inferior.a, junk.n, killer.a, lame.a, lamer.n, magnificent.a, marvel.n, marvellous.a, mediocre.a, miserable.a, nasty.a, okay.a, outstanding.a, pathetic.a, pitiful.a, poor.a, popular.a, rock.v, rotten.a, second-rate.a, sensational.a, shit.n, shitty.a, so-so.a, splendid.a, standout.n, stupendous.a, substandard.a, suck.v, super.a, superb.a, superlative.a, sweet.a, terrible.a, terrific.a, third-rate.a, tip-top.a, tolerable.a, top-notch.a, tremendous.a, uncool.a, unfortunate.a, well.adv, wonderful.a, wonderfully.adv, worthless.a

Created by JKR on 06/13/2002 05:24:22 PDT Thu