A SEMI-AUTOMATIC INTEGRATED FRAMEWORK FOR NON-ENGLISH SENTIMENT LEXICONS

MOHAMMED SALEM ABDULLAH KAITY

FACULTY OF COMPUTER SCIENCE AND
INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2020

A SEMI-AUTOMATIC INTEGRATED FRAMEWORK FOR NON-ENGLISH SENTIMENT LEXICONS

MOHAMMED SALEM ABDULLAH KAITY

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2020

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Mohammed Salem Abdullah Kaity ( )

Matric No: 17028372/1

Name of Degree: Ph.D. in Computer Science

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

A SEMI-AUTOMATIC INTEGRATED FRAMEWORK FOR NON-ENGLISH SENTIMENT LEXICONS

Field of Study: Information Science

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date: 21/4/2020

Subscribed and solemnly declared before,

Witness's Signature Date: 21/4/2020

Name:

Designation:

A SEMI-AUTOMATIC INTEGRATED FRAMEWORK FOR NON-ENGLISH SENTIMENT LEXICONS

ABSTRACT

There has been significant growth in social media networks in the last few years. Posting opinions and messages on social networking websites has become a popular activity on the Internet. The data sources are necessary for business intelligence and market analytics, as human opinions form a major indicator of human desires and behaviour. This has resulted in the development of a new study field called sentiment analysis. This includes the analysis, evaluation and interpretation of the opinions with the help of text mining and Natural Language Processing (NLP) processes, for identifying the text polarity, as positive, neutral or negative. It is important to build sentiment analysis resources before developing the sentiment analysis models. The sentiment lexicons are seen to be a major resource which includes a list of phrases and opinion words along with their sentiment orientation. Literature review revealed that though many texts are available which are written in different languages, a majority of the sentiment analysis studies have focused on those written in English. Hence, the other non-English languages noted a shortage of lexicons and resources. Also, the techniques used for building the sentiment lexicons in non-English languages display many disadvantages like their inability to handle a particular domain, informal use of language expression and vocabulary used in the social media feeds. Furthermore, a few of the non-English sentiment lexicons also have to face translation issues and are plagued by the cultural difference when they are translated from different languages. To overcome the issues which are noted while building the non-English lexicons, a language-independent integrated framework has been proposed in this work which semi-automatically builds the non-English sentiment lexicons based on the available English lexicons with an unannotated corpus from the target language. This framework includes three layers, i.e., corpus-based, lexicon-based, and human-based. The first two layers can automatically recognise and then extract the novel polarity words from the huge unannotated corpus, with the help of the initial seed lexicons. The major advantage of this framework is that it needs only an initial seed lexicon and an unannotated corpus for initiating the extraction activity. This framework is seen to be semi-supervised owing to the usage of the seed lexicons. Experiments on three languages have been carried out and the proposed framework output has shown a better performance than the existing lexicons. The F-measure values for the Arabic, French and Malay lexicons were seen to be 0.778, 0.838 and 0.686, respectively.

Keywords: Sentiment lexicon, Sentiment analysis, Text analysis, Natural language processing, Building resources.

RANGKA KERJA BERSEPADU SEMI-AUTOMATIK BAGI LEKSIKON PENDAPAT BUKAN INGGERIS

ABSTRAK

Rangkaian media sosial telah berkembang dengan pesat sejak beberapa tahun kebelakangan ini. Oleh itu, mengirim mesej dan pendapat di media sosial, menjadi suatu aktiviti yang biasa di internet. Sumber data dalam hal ini, amat berharga bagi membuat analisis pasaran dan risikan perniagaan bagi mengetahui tingkah laku dan kehendak manusia. Bahkan, ini menyebabkan adanya satu bidang kajian yang dipanggil analisis pendapat yang bertujuan menganalis, mengalih Bahasa dan menilai pendapat menggunakan Teknik Pemprosesan Bahasa Asli (NLP) dan perincian teks untuk mengenalpasti kecenderungan teks sama ada positif, negatif atau neutral. Membina sumber analisis pendapat merupakan langkah asas sebelum membangunkan apa-apa model analisis pendapat. Leksikon pendapat merupakan salah satu sumber kritikal yang boleh digambarkan sebagai senarai perkataan dan frasa pendapat yang berorientasikan pendapat mereka. Walaupun teks tersedia dalam pelbagai bahasa, tumpuan kajian analisis pendapat paling utama ialah pada bahasa Inggeris. Akibatnya, Bahasa-bahasa lain, selain bahasa Inggeris mengalami kekurangan sumber Bahasa dan masalah leksikon yang teruk. Selain itu, banyak kaedah semasa untuk membina leksikon pendapat untuk bahasa selain bahasa Inggeris, yang mempunyai kelemahan masing-masing seperti ketidakupayaan untuk menangani domain tertentu, ungkapan bahasa tidak formal dan kosa kata media sosial. Selain itu, beberapa leksikon bukan bahasa Inggeris mengalami masalah penterjemahan dan perbezaan budaya ketika ia dipindahkan dari bahasa lain. Bagi mengatasi cabaran yang dihadapi dalam membina leksikon bukan bahasa Inggeris, kami mencadangkan rangka kerja bersepadu bebas bahasa yang separa automatik bagi membina leksikon pendapat bagi pendapat bukan berasaskan leksikon Bahasa Inggeris yang sedia ada dengan korpus yang tidak diberi

notasi dari bahasa sasaran. Rangka kerja ini terdiri daripada tiga lapisan, iaitu pendapat berasaskan leksikon, berdasarkan korpus dan berasaskan manusia. Dua lapisan pertama secara automatik mengenali dan mengekstrak perkataan baru dari korpus tanpa nama yang besar menggunakan leksikon primer. Kelebihan utama rangka kerja yang dicadangkan ialah ia hanya memerlukan leksikon primer dan korpus tanpa notasi untuk memulakan proses pengekstrakdisian. Oleh itu, rangka kerja itu diawasi secara separa kerana penggunaan leksikon primer. Hasil kajian dalam tiga bahasa menunjukkan rangka kerja yang dicadangkan mengatasi leksikon sedia ada, masing-masing mencapai 0.778, 0.838 dan 0.686 F-Ukur untuk leksikon Arab, Perancis dan Melayu.

Keywords: Analisis sentimen, Pemprosesan bahasa semulajadi, Analisis teks, Leksikon sentimen, Sumber bangunan

ACKNOWLEDGEMENTS

Alhamdulillah, all praise be to Allah for granting me the opportunity, strength, knowledge and success to complete this thesis.

I owe my deepest gratitude to my supervisor Associate Prof. Dr Vimala Balakrishnan for her guidance, advice, and patience throughout my study. This research will never be accomplished without her supervision. I greatly appreciate the time she has spent on discussion and orientation sessions.

This work is dedicated to my father Salem Kaity, who passed away in 2019 while I was studying my PhD, my mother Fatima Omer, my wife Sumaya, and my sons Al-Baraa and Hamzah for their unconditional love, support and sacrifices.

Lastly, I would like to express my gratitude to those I had forgotten to mention in this section. Thanks for the assistance provided in my research.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

AISL	:	Arabic integrated sentiment lexicon
CF	:	Frequency of candidate word
CP	:	Candidate word polarity value
CS	:	Crowdsourcing
DC	:	Number of documents that contain the candidate word
DT	:	Decision tree algorithm
FISL	:	French integrated sentiment lexicon
FN	:	False negatives
FP	:	False positives
ISL	:	Integrated sentiment lexicons
L1	:	Lexicon-based layer
L2	:	Corpus-based layer
L3	:	Human-based layer
LB	:	Lexicon-based
LP	:	Label Propagation
LSA	:	Latent Semantic Analysis
MISL	:	Malay integrated sentiment lexicon
ML	:	Machine learning
MPQA	:	Multi-Perspective Question Answering subjectivity lexicon
NB	:	Naïve Bayes
ND	:	Number of documents in the corpus
NLP	:	Natural language processing
NN	:	Nearby Negative words (Negative words in the same document)
NP	:	Nearby positive words (positive words in the same document)

NW	:	Number of all words in a document
PMI	:	Pointwise mutual information
POS	:	Part of speech
SA	:	Sentiment analysis
SL	:	Sentiment lexicon
SO	:	Semantic orientation
SVM	:	Support vector machines
T	:	Threshold
TF-IDF	:	Term frequency–inverse document frequency
TN	:	True negatives
TP	:	True positives

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

This chapter presents an introduction to this study. It includes an overview of the need for building sentiment lexicons, motivation, problem statements, aims and objectives, research methodology, research scope and contributions. Moreover, the structure of this thesis is also presented.

1.1 Background and Motivation

Social media networks have grown tremendously over the prior few years. Thus, posting comments and opinions on social media channels have become one of the most common activities on the Internet today (Deng et al., 2017; Zhao et al., 2019). The huge volume of user-generated data has made social media the biggest resource of public opinions, resulting in a rapid evolvement of the Internet into a massive data warehouse consisting of user opinions and emotions (Dodds et al., 2011; Liu, 2012). Data sources are precious for market analytics and business intelligence since opinions are the key indicators of human behaviours and desires (Deng et al., 2017; Poria et al., 2016). In fact, this resulted in a field of study called sentiment analysis.

Sentiment analysis refers to analysing, interpreting and evaluating opinions using Natural Language Processing (NLP) techniques and text mining to identify text polarity, either as positive, negative or neutral (Akhtar et al., 2017; Kong et al., 2018; Yue et al., 2018). The primary aim of sentiment analysis is to extract embedded opinions and views regarding services, products, political and social events, etc. (Dey et al., 2018; Siddiqui et al., 2018). Due to the urgent need to understand user trends on a particular subject, sentiment analysis has fast become one of the most critical and value-added research areas over the past few years (Dashtipour et al., 2016; Liu, 2012). Sentiment analysis helps to analyse and interpret enormous amount of data and information thereby identifying and classifying users' opinions and emotions (Liu, 2012; Lo et al., 2016b). Sentiment classification consists of two broad categories: lexicon-based and machine learning classifications (Biltawi et al., 2016). The classifiers based on sentiment lexicons (i.e. a list of polarity words) are rule-based or lexicon-based classifiers, while machine learning classifiers depend on annotated training datasets (Wang et al., 2017).

Lexicon-based sentiment analysis for any language is dependent on sentiment lexicons that are described as a list of opinion phrases and words with their sentiment categories or orientations (Bravo-Marquez et al., 2016; Kiritchenko et al., 2014; Wu et al., 2016). In the absence of adequate training dataset, the lexicon-based approach is proven more appropriate than the machine learning approach (Deng et al., 2017). Moreover, sentiment lexicons are believed to work well in short texts (e.g. social media texts) (Bermingham & Smeaton, 2010). They are also suitable for real-time opinion classifications as their computational requirements are relatively low (Chaovalit & Zhou, 2005; Deng et al., 2017). Furthermore, these sentiment lexicons can be employed for supervised (Kiritchenko et al., 2014; Kouloumpis et al., 2011) and unsupervised classifications (Bravo-Marquez et al., 2016; Deng et al., 2017) for a given text.

Despite the enormity of texts available for multiple languages, sentiment lexicons are primarily available for the English language, while in many other languages these resources are either limited or unavailable (Kong et al., 2018; Lo et al., 2016b). Although English is recognized as the most commonly used language globally, the number of Internet users who communicate in English is less than 26%[1]. This shows that it may be necessary to create resources and tools for subjectivity purposes and sentiment analysis for non-English languages (Lo et al., 2016b; Perez-Rosas et al., 2012). Some researchers attempted to build sentiment lexicons for non-English

[1] The Stats year is 2019 on the website: http://www.internetworldstats.com/stats7.htm

languages, but they are not comparable to those in English, because they are often incomplete or developed for a specific purpose or domain (Lo et al., 2016b; Steinberger et al., 2012). An exciting and motivating factor towards creating and producing resources for non-English languages is supported by the fact that many organizations and enterprises recognize and appreciate the value and necessity to understand user feedback and associated trends, thereby gaining a competitive advantage regardless of the language or demographics (Liu, 2012; Lo et al., 2016a). Also, it is incredibly time-consuming and expensive to create sentiment lexicons for any language manually (Bravo-Marquez et al., 2016; Sun et al., 2017), therefore an automatic sentiment lexicon builder is deemed attractive.

1.2 Sentiment Analysis System Architecture

The objectives of sentiment analysis (SA) are classifying, summarising and visualising users' opinions regarding diverse entities (i.e. objectives) from internet reviews. Specifically, SA attempts to develop automated systems that can extract sentiments from a text written in natural language (Ravi & Ravi, 2015; Robaldo & Di Caro, 2013). An opinion consists of four key components as follows (Kharde & Sonawane, 2016):

- **Sentiment holder:** The holder of sentiment or opinion is a person or a group or an entity that expresses the sentiment.
- **Entity:** An object that can be a person, product, service, location, event, or text.
- **Feature:** A part (or an attribute) of the entity with respect to which evaluation is performed.

- **Sentiment orientation or polarity:** The orientation of an opinion on a feature represents whether the opinion is negative, positive or neutral. Table 1.1 shows an example comprising the four components.

Table 1.1 An example of the opinion key components

SENTENCE:	**“THE CAMERA PICTURE QUALITY IS WONDERFUL”**
SENTIMENT HOLDER	Customer
ENTITY	Camera
FEATURE	picture quality
ENTITY OPINION	Wonderful
SENTIMENT ORIENTATION	**POSITIVE**

As shown in Figure 1.1, SA systems consist of two key components: sentiment resources and sentiment classifier (Kharde & Sonawane, 2016; Zhang et al., 2016). Sentiment resources are the datasets that are used to train the classifiers, which are either annotated corpus or sentiment lexicons (SLs). The sentiment classifier differs based on the approach used for SA, namely lexicon-based (LB) approach and machine learning-based (ML) approach (Van De Kauter et al., 2015; Zhou et al., 2014). The ML approach utilises the ML algorithms and linguistic features. The LB approach relies on an SL, that is, a set of known sentiment words (Giachanou & Crestani, 2016; Medhat et al., 2014).

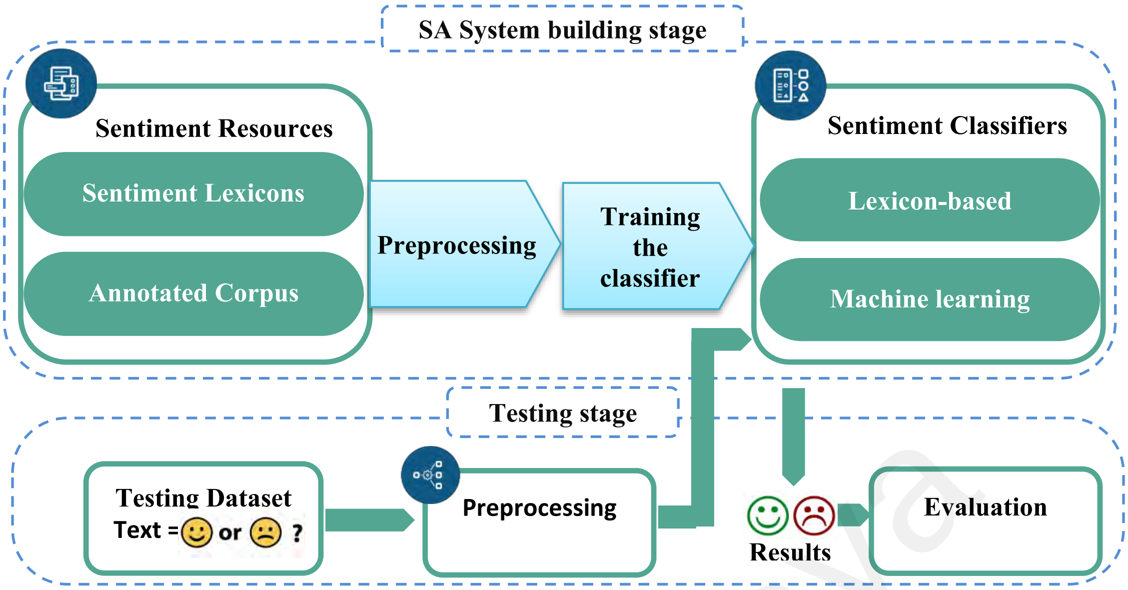


Figure 1.1 Sentiment Analysis System Architecture

The following section describes the Lexicon-based approach and sentiment lexicon concepts and components.

1.2.1 Lexicon-based Sentiment Analysis

The LB approach has received substantial attention for sentiment determination in the SA tasks (Mukhtar et al., 2018; Quan & Ren, 2014). It can be employed without any training or human labour if the lexicons exist (Mukhtar et al., 2018; Yue et al., 2018). It uses a set of sentiment words to decide whether a piece of text is positive or negative (Augustyniak et al., 2016). In the LB approach, the sentiment orientation of a given text is related to the presence of some specific words (i.e. opinion or polarity words). Consequently, text sentiment orientation (i.e. text polarity) can be decided simply by counting the number of positive and negative words, aggregating their sentiment scores and then calculating the overall text polarity (i.e. positive or negative) (Hajmohammadi et al., 2014a; Salah et al., 2013). The process has three basic steps: (i) extracting polarity words or phrases; (ii) determining the polarities of the extracted words or phrases by enquiring the SLs; and (iii) calculating the overall polarity of the given text by aggregating the polarities of that extracted words or phrases (Zhang et al., 2013). Figure 1.2 presents an example of a simple calculation of the sentiment orientation by the LB approach.

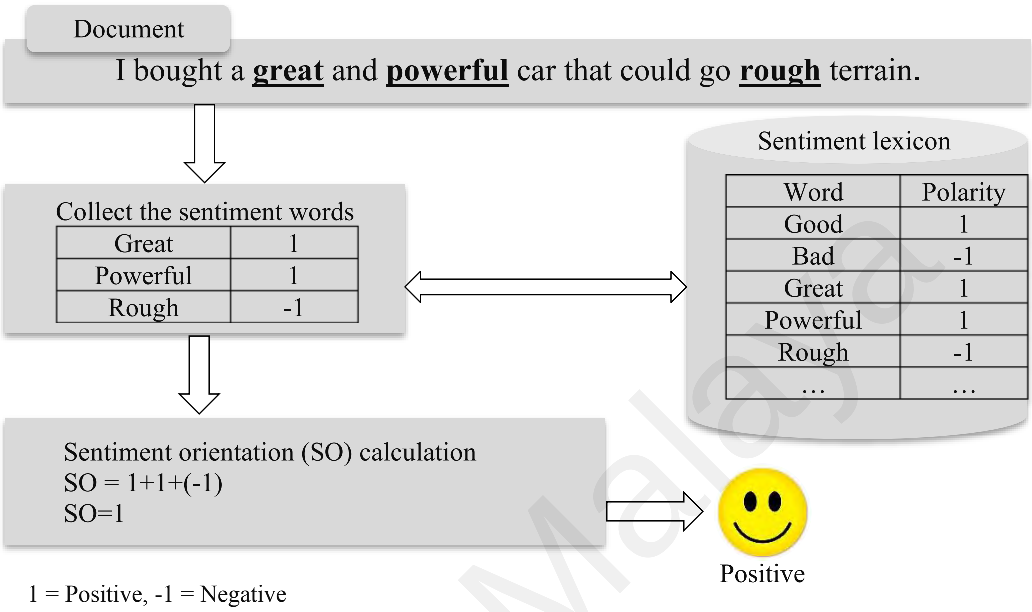


Figure 1.2 An example of a simple calculation of SO by LB approach

SLs are utilised to assign sentiment scores and orientations to the polarity words. The following section describes the SL concepts and components.

1.2.2 Sentiment Lexicons

An SL is one of the most valuable resources of SA for any language (Ahire, 2014; Cambria et al., 2013; Nusko et al., 2016; Wu et al., 2019). They are vital resources for both LB and ML approaches (Sun et al., 2017), with many researchers leveraging SLs to produce unsupervised sentiment models, or as training features to train machine learning algorithms in supervised approaches (Giachanou & Crestani, 2016). An SL is a collection of words (i.e. sentiment, polarity or opinion words) associated with their sentiment orientation, namely, positive or negative (Ahire, 2014; Medhat et al., 2014). Moreover, some SLs contain more valuable details, such as word strength (i.e. strong or weak polarity) and part of speech (POS) (e.g. adjective or noun) (Taboada et al., 2011). Details of SLs are discussed in Chapter 2.

1.3 Sentiment Lexicons for Non-English Languages

SLs are produced manually or semi-automatically (Lo et al., 2016b), and commonly stored as dictionaries or thesaurus (Mihalcea et al., 2007). Although generating SLs manually is very time consuming, it is considered to be more accurate than other methods due to the involvement of human experts (Ahire, 2014; Deng et al., 2017; Wu et al., 2019). The semi-automatic methods combine manual and automated approaches to extract the sentiment lexicon words (Biltawi et al., 2016). Creating the necessary sentiment analysis resources and making these resources available enable the construction of training data for sentiment classification tasks, or the creation of rule-based sentiment analysis. The creation of SLs, dictionaries and corpus is called resource building (Medhat et al., 2014; Montoyo et al., 2012), which is crucial to build any sentiment analysis system (Sun et al., 2017).

The majority of the studies have used translating methods to translate English lexicons to specific languages to build non-English sentiment lexicons (Abdaoui et al., 2016; Dashtipour et al., 2016; Steinberger et al., 2012). Furthermore, many of these methods will more often than not, neglect many of the important words that are used in non-English languages, especially words used in social networks (Scharl et al., 2012). Some researchers used lexical language resources containing words with synonyms and antonyms, such as translated copies of SentiWordNet (SWN) (Hassan et al., 2011; Nusko et al., 2016), whereby lexicons were built by identifying semantic relations between the words. Unfortunately, the applications using this method are limited because most languages currently lack such linguistic sources. Alternatively, other studies used annotated corpus to construct SLs using either the statistical or the semantic relations method. Statistical methods use large corpora with statistical equations to obtain polarity words, whereas semantic relations methods is dependent on the semantic relations between the words in a large corpus (Kumar & Jaiswal, 2016). However, constructing SLs by analysing the corpus requires a substantial volume of the corpus to enable an acceptable level of accuracy to be achieved (Al-Twairesh et al., 2016). Moreover, some methods depend on an annotated corpus that requires additional data annotation before analysis can commence (Pozzi et al., 2017). Chapter 2 in this study contains more details about the current methods and their limitations.

1.4 Problem statement

SLs are valuable resources for opinion-mining tasks for any language (Ahire, 2014; Cambria et al., 2013; Nusko et al., 2016). Although several researchers have studied the problem of building and expanding sentiment lexicons, there are still many unresolved limitations. For example, the majority of those studies focused on English-based sentiment lexicons (Dashtipour et al., 2016; Kong et al., 2018; Nusko et al., 2016), and hence are either limited or not available in other languages, such as Chinese (Feng et al., 2015b), French (Abdaoui et al., 2016), Swedish (Nusko et al., 2016), Arabic (El-Halees, 2011) and Polish (Haniewicz et al., 2014). Therefore, developing automatic or semi-automatic methods to build sentiment lexicons using easily available resources is a critical need as most non-English languages lack sentiment resources. Moreover, as social media languages are continually evolving, users keep innovating new expressions for their sentiments. Today, social media language varies significantly from that used in traditional media (Deng et al., 2017). Therefore, the sentiment lexicons building method will help to expand current lexicons to include new words and abbreviations that appear in the informal textual communication (Zhao et al., 2018).

Many of the current methods for building SLs for non-English languages have their respective drawbacks. The following are some challenges of building SLs for non-English languages:

- A major problem concerns dealing with informal language expressions and social media vocabularies (Rashed & Abdolvand, 2017; Zhao et al., 2018). Many of the current non-English SLs do not contain many words or shortcuts that are used on social networking sites. They cannot handle different dialects and informal or slang words, because such words do not exist in dictionaries (Wu et al., 2016).
- Some non-English SLs suffer from problems of translation and cultural differences about the sentiment orientations of words (i.e. a word may be positive in one language and negative in another, for example, the word "crazy") (Balahur & Turchi, 2012; Hajmohammadi et al., 2014a; Mihalcea et al., 2007; Perez-Rosas et al., 2012).
- Several limitations emerged when using the corpus to build SLs, such as the lack of data pre-processing tools in many languages. There is a lack of adequate corpus online too. Finally, some methods depend on an annotated corpus. This requires additional data annotation before analysis can begin (Pozzi et al., 2017).
- There is also the problem of poor lexicon coverage, where most current non-English SLs are very limited in size, which could affect the performance (Mukhtar et al., 2018; Nusko et al., 2016). Furthermore, the sentiment orientations of lexicon words are in general domain and, consequently, they may appear less accurate when used with specific domains.

- Finally, some of the production methods of SLs are time consuming, require a large number of people and are as costly as the manual methods (Mohammad & Turney, 2013a; Sun et al., 2017).

1.5 Aim of the Research

Based on the identified gaps in the previous section, this doctoral study aims to develop a language-independent framework using an unannotated corpus to build, expand and adapt SLs for non-English languages to classify sentiments in the social networks. The framework is an integration of three components or layers, namely, Lexicon-based layer, corpus-based layer and human-based layer. The first and second layers automatically recognize and extract new polarity words from a massive unannotated corpus using initial seed lexicons. A key advantage of the proposed framework is that it only needs an initial seed lexicon and unannotated corpus to start the extraction process. Therefore, the framework is considered semi-supervised and thus, it is semi-automatic, more efficient and less costly in terms of primary sources used to build the sentiment lexicon. The proposed framework will be elaborated in further details in Chapter 3.

1.6 Objectives and Research Questions

To achieve the aim, the objectives and their respective research questions are as follows:

Objective 1: To examine the current methods and languages used to build sentiment lexicons for non-English languages.

- **Research Question 1:** What are the existing methods for building sentiment lexicons for non-English languages?
- **Research Question 2:** What are the limitations of the current methods?

Objective 2: To develop a semi-automatic integrated framework to build sentiment lexicons for non-English languages.

- **Research Question 3:** What are the components of building a semi-automatic integrated framework for non-English languages?
- **Research Question 4:** How to build sentiment lexicons for non-English languages from unannotated datasets?

Objective 3: To evaluate the proposed semi-automatic integrated framework by conducting experiments and evaluations.

- **Research Question 5:** How can the proposed semi-automatic integrated framework compared with existing method(s)?
- **Research Question 6:** What metrics can be used to evaluate the proposed semi-automatic integrated framework?

1.7 Research scope

The current study addresses the problem of building sentiment lexicons for non-English languages. Specifically, the research focuses on utilizing the available resources (i.e. current sentiment lexicons, unannotated corpus and human experts) and integrating them into a single framework to address the issues identified in Section 1.4.

The proposed framework was tested using three non-English languages namely Arabic, French and Bahasa Melayu (i.e. Malay language). The dataset or corpus was collected from Facebook, specifically from various pages related to news (identities withheld due to confidentiality reason). All the datasets used in the experiments are unannotated corpora (i.e. not labelled by human experts). Finally, since this research aims to build sentiment lexicons for non-English languages and as the target non-English languages (i.e. Arabic, French and Malay) lacked the required annotated corpus, the lexicon-based approach was used in this research instead of machine learning.

1.8 Research Significance and Contributions

Developing an automatic or semi-automatic method to build SL by using available resources is a critical need, as most non-English languages lack sentiment resources. Moreover, it will save the effort of developing manual sentiment resources (i.e. annotated corpus and lexicons).

This research provides a semi-automatic framework to build, adapt and expand SLs by using easily available resources. This framework will help to expand current lexicons to add new words and shortcuts that appear in the informal texts such as blogs and social media websites. In addition, it helps to reset words' polarity in the current lexicons to the new sentiment based on the real use of those words. Furthermore, the SL-building method will be used to customise general domain lexicons to a specific domain. Finally, the proposed language-independent framework makes it easier to provide non-English SLs. The key contributions of the research are as follows:

a) The first contribution of this study is the development of a new taxonomy to classify existing studies based on the resources used to build the lexicons. This thesis provides a comprehensive review of existing studies conducted for building SLs for non-English languages. Shortcomings are highlighted, along with recommendations to improve the performance of each approach and areas for further research.

b) Another contribution is the development of an integrated framework to generate and adapt SLs for non-English languages, incorporating three available resources, namely seed lexicons, unannotated corpus and human experts.

c) This research also contributes to the development of an automated method to recognize new polarity words in the unannotated corpus, based on calculating the seed polarity values to predict the overall sentiment orientation of the candidate word. The calculation was done by formulating an equation for extracting new polarity words from a massive corpus.

d) The fourth and final contribution is the construction of three independent SLs in Arabic, French and Malay that can be made available and thus useful for future studies in similar areas of interests.

1.9 Research Methodology

Based on the research objectives, the following steps, as shown in Figure 1.3, are devised in the study:

1. **Problem Formulation**

   A literature review is conducted to identify the data resources and methods used for building SLs for non-English languages. Moreover, the limitations of the current methods are determined in order to identify the problem statements.

2. **Research Design**

   The research aim, objectives, questions, methodology and the research schedule are set, and a research framework is identified and designed.

3. **Data Preparation**

   The work domain and languages are identified. After that, a dataset based on the selected domain from social media websites is collected. Finally, the dataset is pre-processed through processes such as filtering, data cleaning-up and stop-word removing.

4. Implementation

A system is developed based on the proposed framework to build SLs for non-English languages. Then, experiments are conducted by using the collected dataset.

5. Evaluation

The output of the proposed system is evaluated with the output of the current systems in terms of accuracy, precision, recall and F-measure. Finally, the results are collected and discussed.

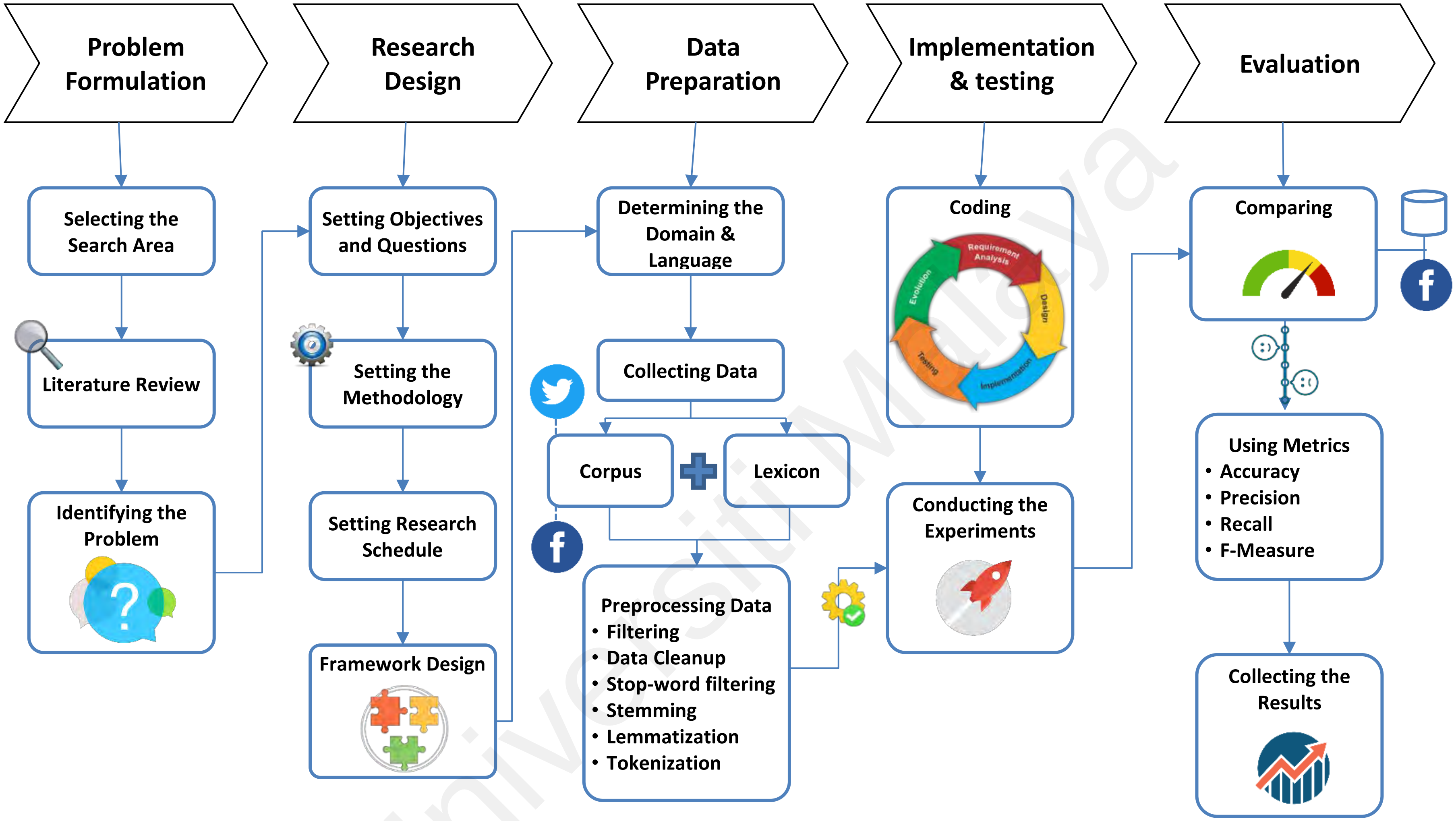


Figure 1.3 Research Methodology

1.10 Research Glossary

In this section, the main terms used in this research were defined. Table 1.2 shows these terms with their meaning.

Table 1.2 Research Glossary

Term	Meaning
Sentiment analysis	Sentiment analysis refers to analysing, interpreting and evaluating opinions using Natural Language Processing (NLP) techniques and text mining to identify text polarity, either as positive, negative or neutral (Akhtar et al., 2017; Kong et al., 2018; Yue et al., 2018).
Sentiment Lexicons	Sentiment Lexicon is a collection of words (i.e. sentiment, polarity or opinion words) associated with their sentiment orientation, namely, positive or negative (Bravo-Marquez et al., 2016; Kiritchenko et al., 2014; Wu et al., 2016).
Building sentiment resources	Building sentiment resources aims at producing lexicons and corpora in which sentiment expressions are annotated based on their polarity (i.e. positive, negative, neutral) (Medhat et al., 2014).
Sentiment holder	The holder of sentiment or opinion is a person or a group or an entity that expresses the sentiment (Kharde & Sonawane, 2016).
Entity	An object that can be a person, product, service, location, event, or text (Kharde & Sonawane, 2016).
Feature	A part (or an attribute) of the entity with respect to which evaluation is performed (Kharde & Sonawane, 2016).
Sentiment orientation or polarity	The orientation of an opinion on a feature represents whether the opinion is negative, positive or neutral (Kharde & Sonawane, 2016).
Lexicon-based Sentiment Analysis	Using a set of sentiment words to decide whether a piece of text is positive or negative (Mukhtar et al., 2018; Quan & Ren, 2014).
supervised learning method	The supervised learning method applies machine learning algorithms to learn relationships between the sentiment

	class and labelled examples in training data (i.e. corpus).
Sentiment words	Sentiment words are phrases or words that are ordinarily used to express negative or positive sentiments.
Polarity/ semantic orientation	Polarity or semantic orientation is a measure of the subjectivity of the lexicon entries that usually captures an evaluative factor (i.e. negative or positive) (Kharde & Sonawane, 2016).
Part of speech (POS)	Part of speech (POS) aims to specify parts of speech of each word in the lexicon (such as adjectives, verbs, nouns and others).
Word Strength	Word strength is the degree or power to which the word or phrase is positive or negative.
Sentiment Domain	A sentiment classifier trained in a general domain lexicon may not perform well in another domain, considering each domain has many domain-specific sentiment words.
Translation-based method	This approach relies on translating an existing sentiment lexicon into a target language. Usually, machine translation or bilingual dictionaries are used (Abdaoui et al., 2016).
Relationship-based method	This approach starts with a small group of core words (seeds) that expand by using the semantic relations between words (i.e. synonyms and antonyms) in an existing dictionary (Hassan et al., 2011).
Merge-based method	This approach uses to create large sentiment lexicons by combining predefined lexicons. It is useful in increasing the coverage and expansion of the lexicons (Badaro et al., 2014).
Frequency-based method	Statistical standards are used to calculate words frequency in a given polarity. This approach assumes that positive words appear together with positive words and vice versa (Al-Twairesh et al., 2016).
Graph-based method	This approach uses semantic relations between words in a large corpus to find new words related to predefined words (seeds) (Haniewicz et al., 2014).
Crowdsourcing method	The lexicons are built by encouraging people to answer questions or a puzzle. People select words from a text and label them with polarities using crowdsourcing and game with a purpose (Hong et al., 2013).

Manual-based method	The lexicons are created manually by researchers or linguists (Trakultaweekoon & Klaithin, 2016).

1.11 Thesis Outline

This thesis consists of six chapters, described as follows:

Chapter 1: Introduction

This chapter presents the research topic and gives an overview of the research work by describing the research problems, research questions and objectives, research methodology and the significance of the research.

Chapter 2: Literature Review

Chapter 2 provides the related work from the existing literature. Specifically, it focuses on studies addressing the issues of building SLs for the non-English languages. Moreover, the limitations and challenges of building lexicons are highlighted in this chapter.

Chapter 3: Methodology

Chapter 3 presents the research methodology employed to conduct this research. This chapter presents the proposed framework which has been developed based on the literature review to solve the research problem. It also describes the experimental dataset and setup. Furthermore, it addresses the evaluation methodology used to evaluate the proposed framework.

Chapter 4: Implementation and Evaluation

Chapter 4 presents the techniques used for the experiments including data collecting and evaluation methodology. We describe the evaluation methods and metrics that have been used to validate and evaluate the performance of the proposed framework.

Chapter 5: Results and Discussion

In this chapter, the performance of the proposed framework is compared with selected benchmark studies and lexicons. Furthermore, the chapter discusses the results of the experiments in a detailed manner.

Chapter 6: Conclusion, Limitation and Future Work

This chapter concludes this research work and presents a summary of the main contributions, and how the aim and objectives of the research are achieved. Moreover, it presents the limitations and provides relevant recommendations for future studies as well.

CHAPTER 2: LITERATURE REVIEW

This chapter provides detailed background knowledge and related work from the existing literature. Specifically, it focuses on studies addressing the issues of building sentiment lexicons for the non-English languages. In this chapter, unique perspective taxonomy of non-English sentiment lexicon building approaches is presented based on the type of resources used.

Moreover, the limitations and challenges of building lexicons are highlighted as well. In this Chapter, the methods to build non-English SLs are examined. This was supported by several studies illustrating each method and language. Comparisons were made between the works done within the span of 2006 and 2018 focusing on non-English sentiment lexicons. The studies were compared in terms of approaches, methods, languages, data sources, techniques, domains, and the number of entries.

This chapter is organised as follows: Subsection 2.1 Sentiment Classification, followed by a brief explanation of Building Sentiment Resources in Subsection 2.2. Subsection 2.3 presents the methods employed to build sentiment lexicons for non-English languages, applying three basic approaches: lexicon-based approaches; corpus-based approaches and human-based approaches. Finally, Subsection 2.4 provides challenges and open issues. Figure 2.1 shows the chapter map which displays the sequence and interconnection of the reviewed literature.

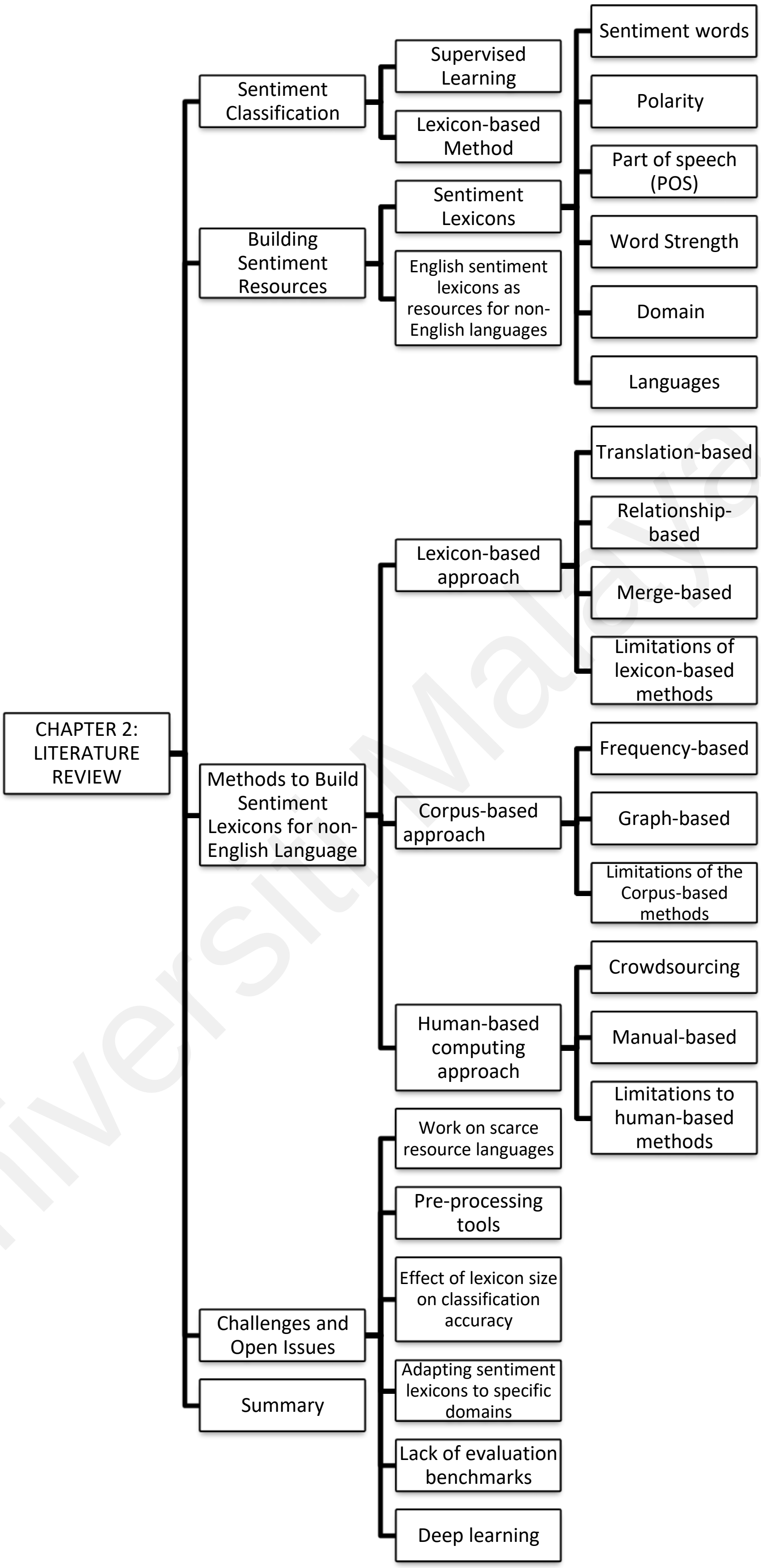


Figure 2.1 Chapter map

2.1 Sentiment Classification

2.1.1 Supervised Learning

The supervised learning method applies machine learning algorithms to learn relationships between the sentiment class and labelled examples in training data (i.e. corpus) (Araújo et al., 2020; Deng et al., 2017). Labelled training data (i.e. annotated corpus) indicate data that have been manually annotated as positive or negative classes. Supervised learning methods for sentiment analysis need a huge amount of training corpus so that the classifier can learn effectively (Lo et al., 2016b). Creating labelled training data is time-consuming and expensive. Furthermore, the requirement for social media sentiment analysis is even higher since social media texts are generally short and extremely diversified (Biagioni, 2016; Deng et al., 2017; Lo et al., 2016b).

Supervised sentiment analysis methods rely on getting a proper set of classification features. Classification features are characteristic attributes in a text and they are suitable for effectively distinguishing between positive and negative sentiment (Deng et al., 2017). Supervised sentiment analysis methods usually utilise the machine learning algorithms such as Support Vector Machine (SVM), Naive Bayes (NB), and Maximum Entropy (Lo et al., 2016b).

2.1.2 Lexicon-based Method

The lexicon-based method is the earliest method proposed for sentiment analysis (Al-Sharuee et al., 2017; Amiri et al., 2015; Osgood et al., 1957). It utilizes sentiment lexicons to find the sentiment orientation of polarity words in a given text (Wu et al., 2019). This method simply depends on the appearance of polarity words in texts where normally texts are classified by aggregating the sentiment polarity scores of the polarity words found in the text. The measure of polarity and subjectivity in the text is called semantic orientation. It is expressed as an evaluative factor (i.e. positive or negative) and strength (i.e. the positivity or negativity degree) towards a subject, idea, or product (Al-Sharuee et al., 2017; Taboada et al., 2011).

2.2 Building Sentiment Resources

Building sentiment resources (i.e., annotated corpora and sentiment lexicons) is important for sentiment analysis. It aims at producing lexicons and corpora in which sentiment expressions are annotated based on their polarity (i.e. positive, negative, neutral) (Montoyo et al., 2012; Sun et al., 2017). Building sentiment resources is not a direct task of sentiment analysis, but it is a prior task to sentiment analysis in the case of absence of the sentiment resources (Medhat et al., 2014). The performance of sentiment classification depends on the quality of the sentiment resources (Hajmohammadi et al., 2014b).

Some languages (e.g. English) have many sentiment resources for performing sentiment analysis including annotated corpus and sentiment lexicons (Lin et al., 2014). Therefore, one of the major challenges that face building sentiment resources is the multilinguality (i.e. the need to have sentiment resources for various languages) (Montoyo et al., 2012). Sentiment resources vary depending on the method used (i.e. corpus-based or lexicon-based) to analyse sentiments. The most important resources of sentiment analysis are lexicons and corpus. The following section explains the sentiment lexicon concept and components.

2.2.1 Sentiment Lexicons

Sentiment lexicons for lexicon-based methods consist of polarity words or phrases with their sentiment orientations (Ahire, 2014; Medhat et al., 2014). These words can be adjectives, nouns, verbs and adverbs (Mukhtar et al., 2018; Taboada et al., 2011). Some SLs consist of nothing more than a list of words and their associations with sentiments (i.e. negative and positive), while others include the word strength as well (Chaturvedi et al., 2018; Sun et al., 2017). On the other hand, some researchers divide the SLs into two individual files, one containing positive words and the other containing negative words (e.g. Hu & Liu’s Sentiment Lexicon (Hu & Liu, 2004)). Figure 2.2 shows the components of SLs derived from various studies (Ahire, 2014; Liu, 2012; Taboada et al., 2011; Wu et al., 2017). Those components can be defined as follows:

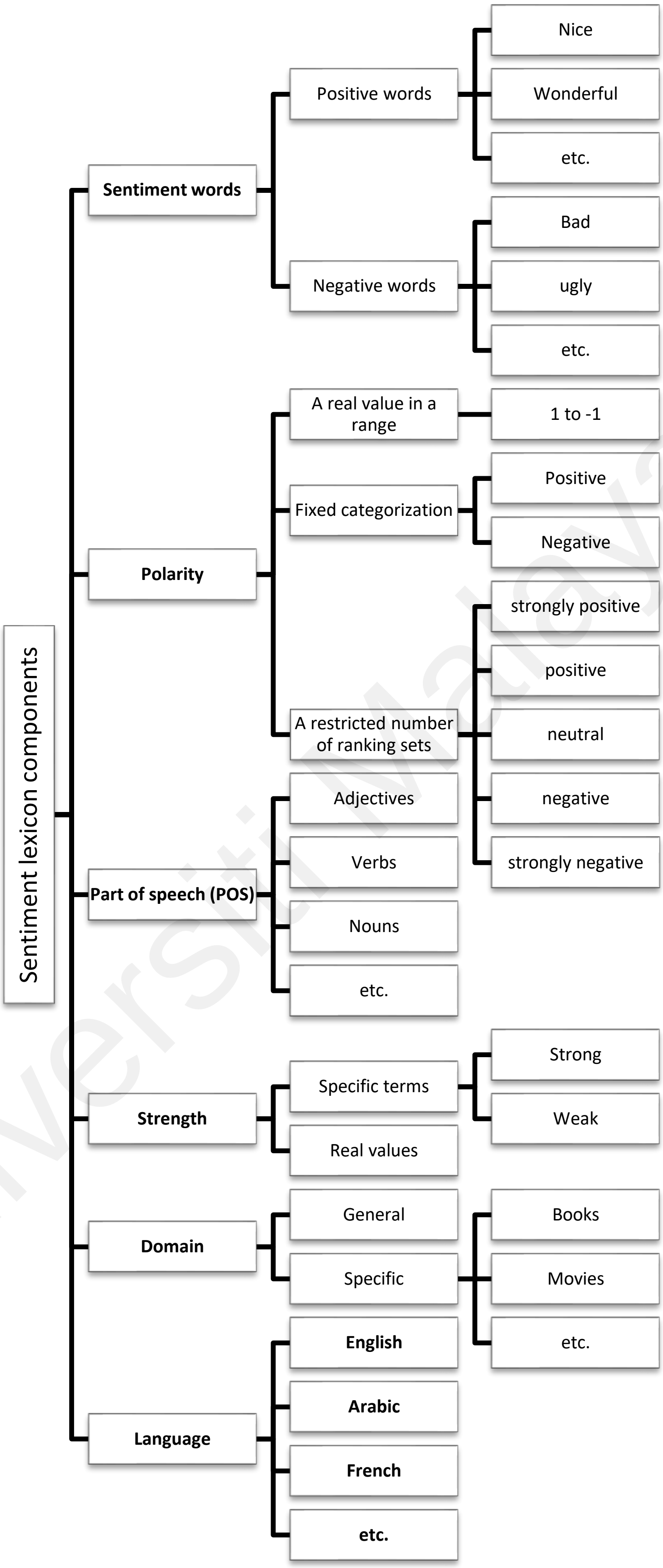


Figure 2.2 Sentiment lexicon components

2.2.1.1 Sentiment words

Sentiment words are phrases or words that are ordinarily used to express negative or positive sentiments. Examples of positive sentiment words are 'excellent', 'incredible' and 'amazing'. In contrast, examples of negative sentiment words are 'awful', 'evil' and 'wrong' (Liu, 2010; Yue et al., 2018).

2.2.1.2 Polarity

Polarity or semantic orientation is a measure of the subjectivity of the lexicon entries that usually captures an evaluative factor (i.e. negative or positive) (Taboada et al., 2011; Wu et al., 2019). The sentiment orientation or the polarity value may be represented in various forms, some of which are:

- A real value indicating sentiment strength in a range such as (−1 - +1),
- Fixed categorization into positive or negative,
- A restricted number of ranking sets such as strongly negative, negative, neutral, positive and strongly positive (Ahire, 2014).

2.2.1.3 Part of speech (POS)

Part of speech (POS) aims to specify parts of speech of each word in the lexicon (such as adjectives, verbs, nouns and others). POS presents useful information in sentiment analysis as some words are ambiguous in nature, for example, the word "novel" is a neutral noun, but a positive adjective (Taboada et al., 2011).

Most of the lexicon-based research works have concentrated on using adjectives as indicators of the sentiment orientation of the text (Hu & Liu, 2004; Taboada et al., 2011; Wu et al., 2019). However, other lexical items can carry important sentiment orientation values as well. Besides adjectives, some researchers use verbs and nouns (Kim & Hovy, 2004), adverbs (Benamara et al., 2007), adjective phrases (Whitelaw et al., 2005), verbs and adverbs (Subrahmanian & Reforgiato, 2008) as a sentiment indicator. Table 2.1 shows the types of POS commonly used in SLs with codes and examples.

Table 2.1 Types of part of speech (POS) used in sentiment lexicons

Part of speech	Examples	
	Positive	Negative
Adjectives	nice, fabulous and good	ugly, bad and harmful
Verb	love, enjoy and glorify	blister, putrefy and foul
Adverb	blessedly, okay and impressively	negatively, poorly and badly
Noun	pride, mercy and appreciation.	rubbish, junk and crap

2.2.1.4 Word Strength

Word strength is the degree or power to which the word or phrase is positive or negative. For example, the word "best" has stronger positivity than the word "good". The word strength is represented by specific terms such as "strong" and "weak" as in Multi-Perspective Question Answering subjectivity lexicon (MPQA) (Wilson et al., 2005b) or as real values such as the positive and negative score assigned to each word in SentiWordNet (Baccianella et al., 2010). Table 2.2 illustrates a fragment of the MPQA (Wilson et al., 2005b), along with the details on the word strength, length, POS, stemmed (i.e. the words are reduced to their word stems, base or root form), and polarity values for each MPQA entry. For example, the line number '3148' contains the word 'gainful' which is a strong subjective word. The polarity of this word is 'positive', and the part of speech is 'adjective'.

Table 2.2 A fragment of the MPQA subjectivity lexicon[2]

No.	Strength	Length	Word	POS	Stemmed	Polarity
1	type= weaksubj	len= 1	word1=abandoned	pos1= adj	stemmed1=n	priorpolarity=negative
…..						
3145	type=strongsubj	len= 1	word1=gaily	pos1=adverb	stemmed1=n	priorpolarity=positive
3146	type=weaksubj	len= 1	word1=gain	pos1= noun	stemmed1=n	priorpolarity=positive
3147	type=weaksubj	len= 1	word1=gain	pos1= verb	stemmed1=y	priorpolarity=positive
3148	type=strongsubj	len= 1	word1=gainful	pos1= adj	stemmed1=n	priorpolarity=positive
3149	type=strongsubj	len= 1	word1=gainfully	pos1=adverb	stemmed1=n	priorpolarity=positive
…...						
8221	type=strongsubj	len= 1	word1= zest	pos1= noun	stemmed1=n	priorpolarity=positive

2.2.1.5 Domain

A sentiment classifier trained in a general domain lexicon may not perform well in another domain, considering each domain has many domain-specific sentiment words (Wu et al., 2017). Simply said, a word may express a negative sentiment in one domain but positive sentiment in another domain. For instance, "hot" is deemed as a positive word when expressing about food (e.g., "hot pizza"). Yet, it is a negative word in the electronics domain (e.g., "my phone becomes hot"). Therefore, some researchers have built or adapted domain-specific SLs (Deng et al., 2017; Park et al., 2015).

2.2.1.6 Languages

Research work in building SLs has focused mainly on the English language, with very few dealing with non-English languages. Still, the possible market for sentiment analysis in different languages is huge. Thus, some attempts and efforts to develop methods to build sentiment resources for specific non-English languages were proposed such as: Arabic (Al-Twairesh et al., 2016), Chinese (Feng et al., 2015b), French

2 http://sentiment.christopherpotts.net/lexicons.html

(Abdaoui et al., 2016), German (Remus et al., 2010), Hindi (Bakliwal et al., 2012), Italian (Buscaldi & Hernandez-Farias, 2016), (Rouvier & Favre, 2016), Japanese (Kim et al., 2010) and Korean (Hong et al., 2013).

In this study, the proposed framework was tested using three non-English languages namely Arabic, French and Bahasa Melayu (i.e. Malay language). The selection of these three languages was made for several reasons. First, these three languages are widely used on the Internet[3] (i.e. 422 million, 281 million and 229 million for Arabic, Malay and French, respectively) and yet are considered to be limited in sentiment resources (Abdaoui et al., 2016; al Owisheq et al., 2016; Darwich et al., 2016). Furthermore, to evaluate the proposed integrated framework on more than one language group (i.e. family), the three languages were chosen from different language families where Arabic is considered as a Semitic language, French as Indo-European and Malay as Austronesian language[4] (Eifring & Theil, 2005). Table 2.3 provides a summary of some languages in which attempts were made to build its own sentiment lexicons.

Table 2.3 Distribution of studies based on the languages

No.	Languages	# References	References
1	Arabic	10	(Al-Twairesh et al., 2016), (El-Halees, 2011), (Hassan et al., 2011), (Mahyoub et al., 2014), (Badaro et al., 2014), (Abdul-Mageed & Diab, 2014), (Eskander & Rambow, 2015), (Al-Twairesh et al., 2016), (Elhawary & Elfeky, 2010), (Al-Subaihin et al., 2011)
2	Bengali	1	(Das & Bandyopadhyay, 2010)
3	Chinese	5	(Yao et al., 2006), (Feng et al., 2015b), (Zhu et al., 2009), (Kim et al., 2010) , (Yang et al., 2013)
4	French	5	(Abdaoui et al., 2016), (Lafourcade et al., 2015), (Rao & Ravichandran, 2009), (Scharl et al., 2012), (Rouvier & Favre, 2016)
5	German	6	(Remus et al., 2010), (Denecke, 2008), (Remus et al.,

[3] http://www.internetworldstats.com/stats7.htm

[4] https://www.mustgo.com/worldlanguages/language-families/

			2010), (Scharl et al., 2012), (Rouvier & Favre, 2016), (Kim & Hovy, 2006)
6	Hindi	5	(Hassan et al., 2011), (Joshi et al., 2010), (Rao & Ravichandran, 2009), (Bakliwal et al., 2012), (Jha et al., 2015)
7	Italian	4	(Basile & Nissim, 2013), (Scharl et al., 2012), (Buscaldi & Hernandez-Farias, 2016), (Rouvier & Favre, 2016), (Passaro et al., 2015)
8	Japanese	1	(Kim et al., 2010)
9	Korean	2	(Kim et al., 2010), (Hong et al., 2013)
10	Malay	1	(Darwich et al., 2016)
11	Multi-languages	1	(Steinberger et al., 2012)
12	Norwegian	1	(Hammer et al., 2014)
13	Persian	2	(Dehdarbehbahani et al., 2014), (Rashed & Abdolvand, 2017)
14	Polish	1	(Haniewicz et al., 2014)
15	Portuguese	1	(Scharl et al., 2012)
16	Romanian	3	(Mihalcea et al., 2007), (Banea et al., 2008), (Banea et al., 2013)
17	Russian	1	(Scharl et al., 2012)
18	Singlish	1	(Lo et al., 2016a)
19	Spanish	5	(Scharl et al., 2012), (Perez-Rosas et al., 2012), (Sidorov et al., 2012), (Banea et al., 2013), (Rouvier & Favre, 2016)
20	Swedish	2	(Rosell & Kann, 2010), (Nusko et al., 2016)
21	Thai	1	(Trakultaweekoon & Klaithin, 2016)

2.2.2 English sentiment lexicons as resources for non-English languages

Numerous researchers relied on SLs available in English that have been built manually and more accurately (Cho et al., 2014). These English SLs have greatly helped in saving time and effort in building new SLs for non-English languages (Abdaoui et al., 2016). Popular English SLs such as SentiWordNet (Baccianella et al., 2010; Esuli & Sebastiani, 2007), SenticNet (Cambria et al., 2010) and Opinion Lexicon (Hu & Liu, 2004) have been used in many approaches to build non-English SLs to improve the performance of sentiment classification.

For example, SentiWordNet is a publicly available lexical resource for sentiment analysis. It is built by associating each WordNet synset (i.e. sets of cognitive synonyms) to one of three categories: positive (Pos), negative (Neg) and neutral (Obj). A synset represents a set of entities (e.g. nouns, verbs, adverbs or adjectives) that share a distinct meaning or sense, and its members can be used interchangeably in the same context (Perez-Rosas et al., 2012). SentiWordNet indicates the degree of each synset with numerical scores ranging from 0.0 to 1.0 (Cho et al., 2014; Dashtipour et al., 2016; Esuli & Sebastiani, 2007). Nevertheless, like other lexicons, SentiWordNet contains some noise considering not all polarity values assigned to the terms are accurate. Moreover, some terms do not have a polarity value whereas some have conflicting values (Cambria et al., 2010; Dashtipour et al., 2016; Poria et al., 2013). For example, the term 'cruelly' has two polarity entries in SentiWordNet, that is, Pos = 0.125; Neg = 0 and Pos = 0; Neg = 0.125 (Dashtipour et al., 2016).

Likewise, SenticNet (Cambria et al., 2010) is a publicly available resource built by exploiting artificial intelligence and semantic web techniques, and based on a new dimensionality reduction approach to infer the polarity of common sense concepts. Cambria et al. (2010) used SenticNet to determine sentiments from text extracted from the Internet. The authors utilized the Hourglass model (Plutchik, 2001), in which the sentiments are organized around four independent dimensions: pleasantness, attention, sensitivity and aptitude that make up the total emotional state of the mind (Cambria et al., 2010; Poria et al., 2013). They defined Concept Polarity as an algebraic sum of the Hourglass categorization model's sentic labels, where Concept Polarity = ((Pleasantness + |Attention| - |Sensitivity| + Aptitude)/9) (Cambria et al., 2010). The following are some English lexicons widely used to produce non-English sentiment lexicons:

- **SentiWordNet**[5] (Baccianella et al., 2010; Esuli & Sebastiani, 2007) is a lexical resource publicly available for research purposes. It consists of an annotation of the WordNet indicating the degree of each term using numerical scores ranging

[5] https://sentiwordnet.isti.cnr.it/

from 0.0 to 1.0, which indicates the polarity of the word (positive, negative and neutral). Four different versions of SentiWordNet have been published: 1.0, 1.1 (Esuli & Sebastiani, 2007), 2.0 and 3.0 (Baccianella et al., 2010).

- **SenticNet**[6] (Cambria et al., 2010) is a publicly available affective lexical resource for polarity information. It consists of 14,244 polarity concepts (Cho et al., 2014) used for concept-level sentiment analysis. It provides polarity information grouped into four categories: attention, pleasantness, sensitivity, and aptitude (Poria et al., 2013).
- **Opinion Finder**[7] (Wilson et al., 2005a) includes a sentiment lexicon which is composed of manually developed resources with entries extracted from corpora. OpinionFinder consists of 6,856 unique entries associated with their polarity values (Perez-Rosas et al., 2012).
- **Bing Liu's Opinion Lexicon**[8] (Hu & Liu, 2004) is a lexicon created manually by extracting the polarity words from customer reviews. The lexicon includes 4,783 negative words and 2,006 positive words.
- **MPQA Subjectivity Lexicon**[9] (Wilson et al., 2005b) is also a manually created lexicon, consisting of 8,221 words with their subjectivities (strong or weak), polarities and POS tags.
- **Harvard General Inquirer**[10] (Stone et al., 1966) is a manually created lexicon consisting of 3,206 entries divided into 915 positive and 2,291 negative words. These marked words are divided into 182 categories such as positive, negative, strong, weak, active etc.

---

[6] http://sentic.net/downloads/
[7] http://mpqa.cs.pitt.edu/lexicons/
[8] http://www.cs.uic.edu/ ~liub/FBS/sentiment-analysis.html
[9] http://mpqa.cs.pitt.edu/lexicons/subj _ lexicon/
[10] http://www.wjh.harvard.edu/ ~inquirer/

- **AFINN**[11] (Nielsen, 2011) is a list of English polarity phrases and words rated between +5 (very positive) and −5 (very negative). The first version, AFINN-96, contained 1,468 unique words whereas the newest version has 2,477 words.

Table 2.4 provides more information about the English lexicons widely used to produce non-English SLs (Araujo et al., 2016).

Table 2.4 English lexicons used to produce non-English sentiment lexicons

Sentiment lexicon	Polarity	Entry size	License
SentiWordNet	Positive, negative, objective	117,658 Synsets	Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0) license
SenticNet	Positive, negative	14,244 Common sense concepts	MIT License
Opinion Finder	Negative, Neutral, Positive	6,856 unique entries	GNU General Public License
Bing Liu's Opinion Lexicon	-1, 0, 1	6789 Words	Free
MPQA Subjectivity Lexicon	strong or weak Positive or negative	8221	GNU General Public License
Harvard General Inquirer	Positive, negative	3206	Available for research purposes
AFINN	-1, 0, 1	2477	Open Database License (ODbL) v1.0

Table 2.5 provides a summary of the studies that used English sentiment lexicons as a base to build new non-English lexicons. The most commonly used lexicon is SentiWordNet (Baccianella et al., 2010); primarily due to its dependence on WordNet followed by General Inquirer (Stone et al., 1966) and Opinion-Finder (Wilson et al., 2005a).

Table 2.5 The English resources that used to build new lexicons

No	Sentiment Lexicons / lexical Resources	# Use	The papers that used the lexicon as a resource
1	SentiWordNet	11	(Perez-Rosas et al., 2012), (Denecke, 2008), (Basile

[11] https://github.com/fnielsen/afinn

			& Nissim, 2013), (Sidorov et al., 2012), (Badaro et al., 2014), (Joshi et al., 2010), (Abdul-Mageed & Diab, 2014), (Eskander & Rambow, 2015), (Bakliwal et al., 2012), (Das & Bandyopadhyay, 2010), (Buscaldi & Hernandez-Farias, 2016)
2	WordNet	8	(Kim & Hovy, 2006), (Hassan et al., 2011), (Joshi et al., 2010), (Rao & Ravichandran, 2009), (Dehdarbehbahani et al., 2014), (Darwich et al., 2016), (Mahyoub et al., 2014), (Perez-Rosas et al., 2012)
3	General Inquirer	4	(Remus et al., 2010) , (Hassan et al., 2011), (Abdul-Mageed & Diab, 2014), (Rouvier & Favre, 2016)
4	Opinion-Finder	4	(Perez-Rosas et al., 2012), (Mihalcea et al., 2007), (Kim et al., 2010), (Banea et al., 2013)
5	SentiStrength	1	(El-Halees, 2011)
6	NRC-EmoLex	2	(Abdaoui et al., 2016), (Rouvier & Favre, 2016)
7	AFINN	2	(Hammer et al., 2014), (Buscaldi & Hernandez-Farias, 2016)
8	Bing Liu's Opinion Lexicon	3	(Al-Twairesh et al., 2016), (Buscaldi & Hernandez-Farias, 2016), (Rouvier & Favre, 2016)
9	MPQA	2	(Rouvier & Favre, 2016), (Al-Twairesh et al., 2016)

2.3 Methods to Build Sentiment Lexicons for non-English Languages

In this section, the classifications and approaches to build non-English SLs are examined. This was supported by several studies illustrating each approach and language. Although the performance of sentiment analysis systems mainly depends on the coverage and the accuracy of the sentiment lexicon used, many languages have not received adequate attention for building lexicons (Wu et al., 2016). Thus, the current SLs available to the public have not achieved the acceptable level of precision required (Biagioni, 2016).

In this study, the approach refers to the main classification of particular methods. The approaches are described as a set of principles and correlative assumptions axiomatic in its character. The method is based upon the selected approach. The approach is a general class while the method is a procedural manner. Finally, the technique refers to a specific way to implement a method, such as an algorithm or application (Burnham, 1992; Ray, 2019). Figure 2.3 shows the relationship between approach, method and technique.

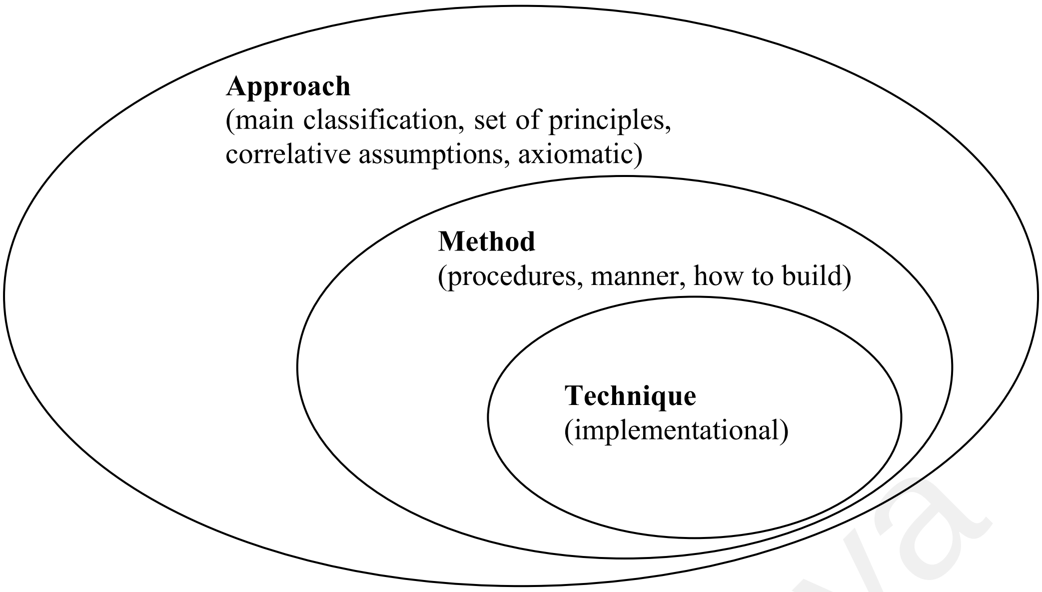


Figure 2.3 Relationship between approach, method and technique

Methods for building SLs vary from being completely manual, semi-automatic, to limited automatic approaches (Nusko et al., 2016). In this study, strategies are divided and used to construct SLs according to the type of source used. Accordingly, there are three sources employed to build SLs; pre-existing lexicons, target language corpus, and target language native speakers (i.e. humans) as shown in Table 2.6. Figure 2.4 graphically illustrates the methods used to build SLs for non-English languages.

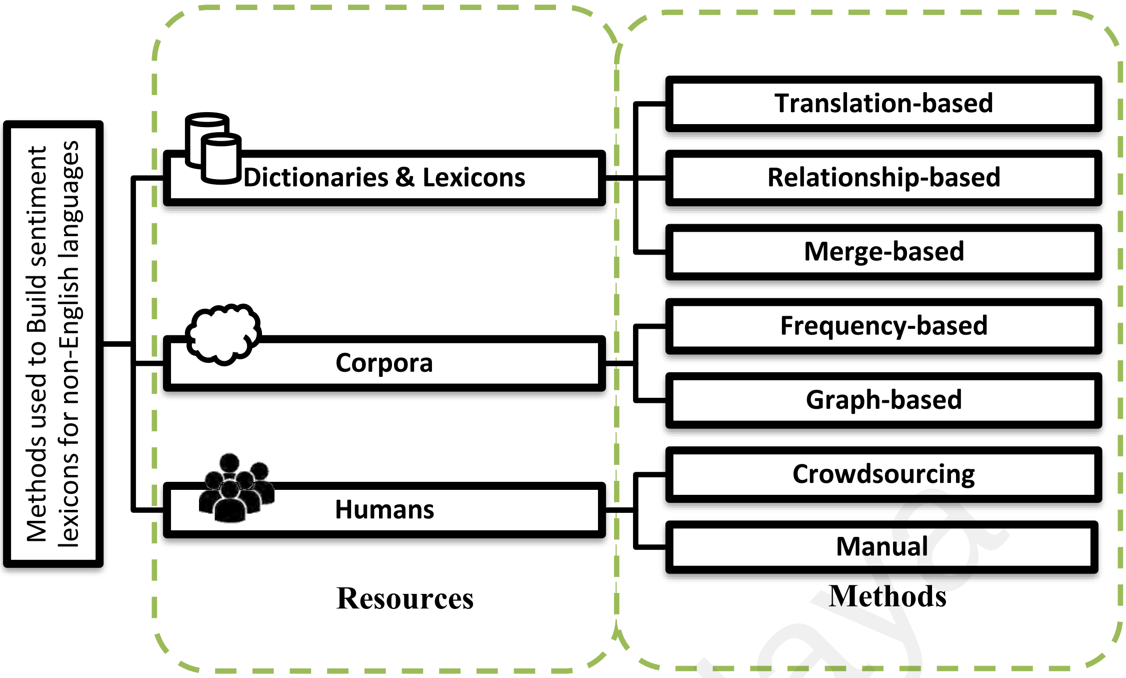


Figure 2.4 The taxonomy of the methods used to build sentiment lexicons for non-English languages

Table 2.6 Summary of the methods used to build sentiment lexicons for non-English languages

Resources	Methods	Overview	References
Lexicons	Translation-based	This approach relies on translating an existing sentiment lexicon into a target language. Usually, machine translation or bilingual dictionaries are used.	Abdaoui et al. (2016), Hammer et al. (2014), Al-Twairesh et al. (2016), Yao et al. (2006), Mihalcea et al. (2007), Steinberger et al. (2012), El-Halees (2011), Remus et al. (2010), Perez-Rosas et al. (2012), Denecke (2008), Banea et al. (2013), Kim et al. (2010), Basile and Nissim (2013), Lo et al. (2016a), Sidorov et al. (2012), Kim and Hovy (2006), Das and Bandyopadhyay (2010), Rouvier and Favre (2016)
	Relationship-based	This approach starts with a small group of core words (seeds) that expand by using the semantic relations between words (i.e. synonyms and antonyms) in an existing dictionary.	Hassan et al. (2011), Rosell and Kann (2010), Banea et al. (2008), Rao and Ravichandran (2009), Mahyoub et al. (2014), Bakliwal et al. (2012), Zhu et al. (2009), Nusko et al. (2016), Dehdarbehbahani et al. (2014), Darwich et al. (2016)
	Merge-based	This approach uses to create large sentiment lexicons by combining predefined lexicons. It is useful in increasing the coverage and expansion of the lexicons.	Badaro et al. (2014), Joshi et al. (2010), Abdul-Mageed and Diab (2014), Eskander and Rambow (2015), Buscaldi and Hernandez-Farias (2016)
Corpus	Frequency-based	Statistical standards are used to calculate words frequency in a given polarity. This approach assumes that positive words appear together with positive words and vice versa.	Al-Twairesh et al. (2016), Remus et al. (2010), Jha et al. (2015), Rashed and Abdolvand (2017), Yang et al. (2013)
	Graph-based	This approach uses semantic relations between words in a large corpus to find new words related to predefined words (seeds).	Elhawary and Elfeky (2010), Feng et al. (2015b), Haniewicz et al. (2014)
Human	Crowdsourcing	The lexicons are built by encouraging people to answer questions or a puzzle. People select words from a text and label them with polarities using crowdsourcing and game with a purpose.	Hong et al. (2013), Lafourcade et al. (2015), Al-Subaihin et al. (2011), Scharl et al. (2012),

	Manual	The lexicons are created manually by researchers or linguists.	Trakultaweekoon and Klaithin (2016), Abdul-Mageed et al. (2014)

The following presents a number of works carried out for each approach with the identification of the languages used. Comparisons were made between the works done within the span of 2006 and 2016 focusing on non-English sentiment lexicons. The studies were compared in terms of approaches, methods, languages, data sources, techniques, domains, and the number of entries.

2.3.1 Lexicon-based approach

Due to the availability of numerous sentiment lexical resources (i.e. lexicons and dictionaries) in the English language, many researchers have depended on these resources (Dashtipour et al., 2016; Pozzi et al., 2017). One of the most important methods that benefited from previous lexicons is the translation (Araújo et al., 2020; Denecke, 2008). Due to the rapid development of machine translation through sites such as Google.com, Bing.com, most researchers used different translation methods to build non-English SLs (Denecke, 2008). In order to overcome the shortcomings observed in automated translation systems, researchers have opted to use multiple translations of more than one language at the same time (Dashtipour et al., 2016; Steinberger et al., 2012). Besides the machine translation of English sentiment lexicons, other methods have been used based on existing English lexicons such as transfer learning, relationship-based and merge-based methods. Tables 2.5 to 2.9 show a survey of the studies that have built SLs for non-English using the lexicon-based approach. The following subsections describe in detail the lexicon-based approach related methods used to build SLs for non-English languages.

2.3.1.1 Translation-based

In the translation-based method, the language which has many dependable resources (i.e. lexicons) is called the source language (e.g. English), and the lacking resource language is called the target language (Hajmohammadi et al., 2014a). The target language will identify the sentiment polarities of texts using the existing resources in the source language (Dashtipour et al., 2016). Table 2.7 shows a survey of the studies that have built SLs for non-English languages using the translation-based method.

Table 2.7 Survey of the studies that have built sentiment lexicons for non-English (translation-based methods)

Approach	Method	Languages (Lexicon Name)	Data sources	Data sources ref.	Technique	Domain	Number of entries	Evaluation method	Precision	Recall	F-measure
Lexicon-based	Translation-based	French Abdaoui et al. (2016)	NRC-EmoLex	Mohammad and Turney (2013a)	Six online translators	General	14,127	SVM	74.3	74.4	72.8
		Norwegian Hammer et al. (2014)	AFINN	Nielsen (2011)	Google translate	General	2161	N/A	N/A	N/A	N/A
		Arabic (AraSenti-Trans) Al-Twairesh et al. (2016)	Opinion Lexicon MPQA	Hu and Liu (2004) Wilson et al. (2005b)	MADAMIRA tool (Pasha et al., 2014)	Twitter	N/A	N/A	78.5	78.1	76.3
		Chinese Yao et al. (2006)	10 bilingual lexicons from StarDict[12]	-	Bilingual Translator,	General	4120	SVM	83.3	91.4	N/A
								DT	84.1	92.8	N/A
		Romanian Mihalcea et al. (2007)	Opinion-Finder	Wilson et al. (2005a)	Bilingual Translator	General	4,983	LB	62.6	33.5	43.7
		Multi-languages Steinberger et al. (2012)	MicroWNOp and JRC Tonality	Cerini et al. (2007) Balahur et al. (2009)	triangulation	General	about 2000 per language	LB	N/A	N/A	N/A
		Arabic El-Halees (2011)	SentiStrength	Thelwall et al. (2010)	Manually translation	Education, Politics, Sports	8793	LB	81.3	81.7	82.7
		German (SentiWS) Remus et al. (2010)	General Inquirer	Stone et al. (1966)	Google translate	General	N/A	LB	96	74	84

12 http://goldendict.org/dictionaries.php

		Spanish Perez-Rosas et al. (2012)	Opinion-Finder SentiWordNet WordNet	Wilson et al. (2005a) Baccianella et al. (2010)	multilingual sense-level aligned WordNet structure	General	1,347	SVM	64.6	82.4	72.4
								Manual	91.8	88.2	90.0
		German Denecke (2008)	SentiWordNet	(Baccianella et al., 2010; Esuli & Sebastiani, 2007)	Translator	General	N/A	LB	66	N/A	N/A
		Romanian Spanish (Banea et al., 2013)	OpinionFinder	Wilson et al. (2005a)	Ectaco online dictionary [13]	General	1,580	SVM	67.7	38.1	48.9
							2009		66.9	50.5	57.6
		Korean Chinese Japanese Kim et al. (2010)	OpinionFinder	Wilson et al. (2005a)	multi-lingual Translator	General	3808 3980 3027	LB	59.4	71.0	64.7
									58.4	82.3	68.2
									56.9	92.4	70.4
		Italian Basile and Nissim (2013)	SentiWordNet	Baccianella et al. (2010)	Transfer learning	Twitter	N/A	N/A	55	N/A	N/A
		Singlish (Singaporean English) Lo et al. (2016a)	English-Malay lexicon Many online resources	Chen and Skiena (2014)	Matching English polarity list with Singlish list	Twitter	2666	Hybrid	N/A	N/A	77
		Spanish (SEL) Sidorov et al. (2012)	SentiWordNet	Baccianella et al. (2010)	Maria Moliner dictionary (Moliner & Moliner, 1998)	Twitter	2036	NB	78.2	N/A	N/A
								DT	83.6	N/A	N/A
								SVM	85.8	N/A	N/A
		German Kim and Hovy (2006)	WordNet	Miller (1995)	Translator	German Emails	3871	LB	N/A	N/A	N/A

13 http://www.ectaco.co.uk/free-online-dictionaries/.

		French, Italian, Spanish and German Rouvier and Favre (2016)	MPQA Opinion Lexicon General Inquirer NRC Emotion Lexicon	Wilson et al. (2005b) Hu and Liu (2004) Stone et al. (1966) Mohammad and Turney (2013b)	Transfer learning and translating	Twitter	N/A	SVM	N/A	N/A	61.7
		Bengali Das and Bandyopadhyay (2010)	SentiWordNet	Baccianella et al. (2010)	English-Bengali Dictionary	General	35805	LB	74.6	80.4	N/A

NB= Naïve Bayes, SVM= support vector machines, DT= decision tree algorithm, LP= Label Propagation test, LB= Lexicon-based, CS= crowdsourcing
Note: The columns: Precision, Recall, F-measure, are the results of the evaluation made by the researchers on their own data and not for comparing between the methods.

In general, the translated sentiment lexicon is built in three steps as shown in Figure 2.5. First, translating the source language lexicon using machine translation tools, in which simple substitution of words takes place from one language to another. This is then followed by part of speech (POS) tagging before the target language lexicon is cleaned and filtered to remove duplicates and non-translated words (Banea et al., 2013; Jha et al., 2015; Mahyoub et al., 2014). Most of the studies were found to perform the filtering manually, probably as it is easier to find non-translated or duplicate words by sorting the obtained list alphabetically (Banea et al., 2013; Jha et al., 2015; Mahyoub et al., 2014).

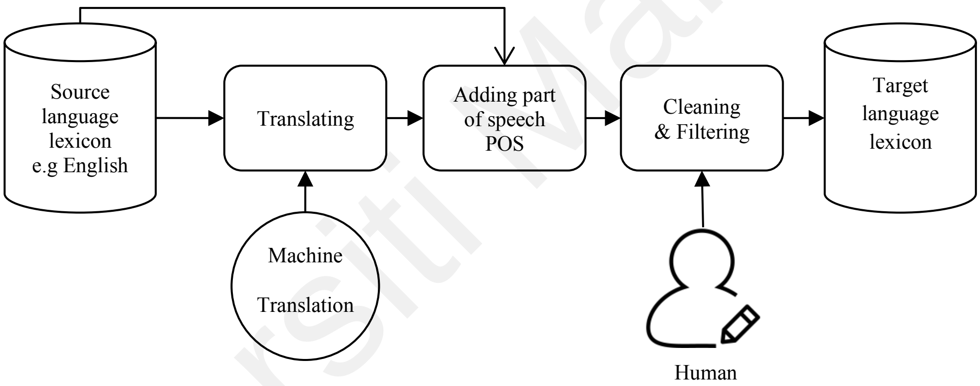


Figure 2.5 The general steps of the translating-based method

One of the first studies that used machine translation to construct a sentiment lexicon for a non-English language was conducted by Yao et al. (2006), who proposed an automatic translation method for building a Chinese sentiment lexicon. They used an electronic dictionary named StarDict[14] to translate Chinese words into English, followed by parsing to generate sequences of English words. These words were then used to determine the sentiment score for the specific words.

[14] http://stardict.sourceforge.net

Likewise, Mihalcea et al. (2007) used the same approach for the Romanian language where they used a bilingual dictionary to translate current English lexicons to the target language (i.e. Romanian) in their proposed framework. The Opinion-Finder (Wiebe & Riloff, 2005) was used as a sentiment lexicon resource. The authors used two bilingual dictionaries to perform the translation. One of them is an English-Romanian dictionary, and the other was obtained from the Universal Dictionary[15]. After translating the English resources, they built a Romanian sentiment lexicon consisting of 4,983 entries. Then, they built a rule-based subjectivity classifier using the new lexicon. The obtained precision for their rule-based classifier was 62.6%.

Steinberger et al. (2012) on the other hand, proposed a triangulation method (i.e. translating two languages to produce a new lexicon in a third language) using the semi-automatic approach. In other words, the authors used machine-translation tools to translate the English and Spanish sentiment lexicons to Arabic, Czech, French, German, Italian and Russian, followed by filtering and manually expanding the new lexicon.

Abdaoui et al. (2016) proposed a machine-translation method to build a new French sentiment and emotion lexicon, named French Expanded Emotion Lexicon (FEEL). Their method was based on the semi-automatic translation, where they used English NRC Word Emotion Association Lexicon (Mohammad & Turney, 2013a) as a resource lexicon. Their work was done in two stages: conducting an online translation on existing English lexicons to create the first version of their French lexicon, followed by validation by a human professional translator. Their final French lexicon contains 14,127 entries, whereby around 15 % were compound words and 85 % were single words. Based on their results which are shown in Table 2.7, they concluded that online

[15] http://www.dicts.info/uddl.php

translators can be used to inexpensively build sentiment resources (Abdaoui et al., 2016).

Hammer et al. (2014) created and evaluated a large set of SLs using Google translate[16] to translate the AFINN English sentiment lexicon (Nielsen, 2011) to Norwegian. Moreover, they generated one more lexicon to manually correct some errors from the machine translations. Thus, it appears that translation-based method depends on the availability of the translation engines for the required (i.e. target) languages (Balahur & Turchi, 2014).

Kim and Hovy (2006) presented a computational framework to develop a German sentiment system by translating WordNet to German to analyse emails. Similarly, Denecke (2008) used SentiWordNet to detect the polarity of a German document by translating the English lexicon into German on their multilingual framework.

Al-Twairesh et al. (2016) generated a large-scale Twitter sentiment lexicon for Arabic called AraSenti-Trans using the MADAMIRA tool (Pasha et al., 2014). After pre-processing, they used Bing Liu's Opinion Lexicon (Hu & Liu, 2004) and the MPQA lexicon (Wilson et al., 2005b) as sentiment orientation resources. The authors used the English glossary that was provided by MADAMIRA to find the word polarity by comparing with Liu's opinion and MPQA lexicons.

On the other hand, OpinionFinder (Wilson et al., 2005a) was used by Kim et al. (2010) to build multilanguage sentiment lexicons for three languages, namely Korean, Chinese and Japanese with 3,808, 3,980 and 3,027 entries respectively. Similarly, Perez-Rosas et al. (2012) presented a framework to obtain sentiment lexicons using OpinionFinder along with SentiWordNet (Baccianella et al., 2010) as English electronic

[16] translate.google.com

resources to extract Spanish sentiment lexicon. Likewise, Basile and Nissim (2013) used SentiWordNet and MultiWordNet to transfer word polarities from English to Italian. Another transfer method was used by Das and Bandyopadhyay (2010) in which an English-Bengali dictionary was used to apply a word level lexical transfer to each entry in English SentiWordNet. The result was a Bengali SentiWordNet with 35,805 Bengali entries.

Some studies have presented hybrid methods (i.e. using the translation method with other methods), but they are mainly based on the translation. A study by Sidorov et al. (2012) built a Spanish emotion lexicon called SEL containing 2,036 words. The lexicon was built in three stages: first was the automatic translation of the words from English SentiWordNet (Baccianella et al., 2010) to Spanish. Second, the Maria Moliner dictionary (Moliner & Moliner, 1998) was used to check the translated words if they had a meaning related to the basic emotions: joy, anger, fear, sadness, surprise and disgust. Finally, 19 annotators evaluated the association of the words with the emotions. The annotators put the scales such as null, low, medium and high for each entry (Sidorov et al., 2012).

The General Inquirer lexicon (Stone et al., 1966) was translated using Google translator for building a German sentiment lexicon in Remus et al. (2010). Banea et al. (2013) on the other hand, presented sentiment lexicons for Romanian and Spanish languages by automatically translating a source language lexicon into Romanian and Spanish using a multilingual dictionary. The lexicons were then expanded using the bootstrapping process (see Subsection 2.3.1.2 for further details on bootstrapping).

Instead of using machine translation on its own, the manual translation (i.e. human translators) was used as well, such as the work provided by El-Halees (2011). El-Halees (2011) manually translated an English lexicon to the Arabic language. The Arabic sentiment lexicon was built using two resources: SentiStrength (Thelwall et al., 2010) and an online dictionary. After the translation process, the initial list was manually filtered. Then, the same strength that was used in SentiStrength was used in the Arabic list. Finally, an online dictionary was used to add other common Arabic words; some of them were synonyms and the others were deemed significant. Lo et al. (2016a) also manually constructed a Singlish (Singaporean English) sentiment lexicon by combining several Internet resources such as Coxford Singlish Dictionary[17], Singlish and Singapore English[18], and Wikipedia Singlish vocabulary[19]. For English, they used many online resources such as the positive and negative lists of a Twitter sentiment analysis, and a set of positive vocabulary word lists [20]. Then, to determine the polarity of words a Malay-English sentiment lexicon (Chen & Skiena, 2014) was used. The final list contained 2,666 entries of Singlish terms.

2.3.1.2 Relationship-based

The relationship-based methods begin with a small group of core words (i.e. seed words) that are expanded using the semantic relations between the words in an existing dictionary or lexicon (Ravi & Ravi, 2015). Table 2.8 shows a survey of the studies that have built sentiment lexicons for non-English using the relationship-based methods.

Mahyoub et al. (2014) developed an algorithm that assigns sentiment scores to the entries found in the Arabic WordNet to create an Arabic sentiment lexicon. Once the seed list of positive and negative words was built, a semi-supervised learning algorithm

---

[17] http://www.talkingcock.com/html/lexec.php

[18] http://www.singlishdictionary.com/

[19] https://en.wikipedia.org/wiki/Singlish _ vocabulary

[20] https://github.com/jeffreybreen/twitter- sentiment- analysis- tutorial- 201107/ tree/master/data/opinion- lexicon- English

was used to increase the number of entries in the Arabic WordNet. Their proposed algorithm determined polarity scores to more than 600 negative, 800 positive and 6000 neutral words. Similarly, Nusko et al. (2016) presented a method to build a sentiment lexicon for the Swedish language where a small group of seed words were expanded using the semantic relations between words in SALDO (Borin et al., 2013), which is a modern Swedish lexical resource. Similarly, Rosell and Kann (2010) built a Swedish sentiment lexicon using random walk algorithms, beginning with a small group of seeds from People's Dictionary of Synonyms (Kann & Rosell, 2005). Graph method was used to calculate distances between the words, resulting in a lexicon consisting of 908 positive and 441 negative words.

Table 2.8 Survey of the studies that have built sentiment lexicons for non-English (Relationship-based methods)

Approach	Method	Languages (Lexicon Name)	Data sources	Data sources ref.	Technique	Domain	Number of entries	Evaluation method	Precision	Recall	F-measure
Lexicon-based	Relationship-based	Arabic & Hindi Hassan et al. (2011)	WordNet Arabic WordNet Hindi WordNet General Inquirer	Baccianella et al. (2010) Black et al. (2006) Narayan et al. (2002) Stone et al. (1966)	Random Walk	General	N/A	SO-PMI	83 93	N/A	N/A
		Persian Dehdarbehbahani et al. (2014)	WordNet 3.0 FarsNet 1.0	Miller (1995) Shamsfard et al. (2010)	Random Walk	General	4941	N/A	80.6	80.5	N/A
		Swedish Rosell and Kann (2010)	People's Dictionary of Synonyms	Kann and Rosell (2005)	Random Walk	General	1349	N/A	N/A	N/A	N/A
		Romanian Banea et al. (2008)	Romanian online dictionary[21]	-	Bootstrapping Method	General	4000	LB	62.8	69.9	66.2
		Hindi French Rao and Ravichandran (2009)	Hindi WordNet[22] OpenOffice thesaurus[23]	-	graph-based label propagation	General	N/A	LP	90.9	95.1	93
									73.6	93.6	82.5
		Arabic Mahyoub et al. (2014)	Arabic WordNet	Black et al. (2006)	semi-supervised learning	General	7576	NB	94	91	N/A
								SVM	73	65	N/A

21 http://www.dexonline.ro
22 http://www.cfilt.iitb.ac.in/wordnet/webhwn/
23 http://www.openoffice.org

		Malay Darwich et al. (2016)	WordNet	Baccianella et al. (2010)	graph-based label propagation	General	4206	NB	~64	N/A	N/A
		Hindi Bakliwal et al. (2012)	English-Hindi WordNet linking SentiWordNet	(Karra et al., 2009) Baccianella et al. (2010)	graph-based	General	8936	Human Judgment	~79	N/A	N/A
		Chinese Zhu et al. (2009)	HowNet semantic lexicon[24]		semantic similarity	Hotel	5573	SVM	82.1	N/A	N/A
		Multi-languages Chen and Skiena (2014)	Wiktionary[25]		knowledge graph propagation	General	Based on the language	N/A	N/A	N/A	N/A
		Swedish Nusko et al. (2016)	SALDO	Borin et al. (2013)	Semantic Relations		2127		85	N/A	N/A

NB= Naïve Bayes, SVM= support vector machines, DT= decision tree algorithm, LP= Label Propagation test, LB= Lexicon-based, CS= crowdsourcing
Note: The columns: Precision, Recall, F-measure, are the results of the evaluation made by the researchers on their own data and not for comparing between the methods.

24 http://www.keenage.com/
[25] https://www.wiktionary.org/

Hassan et al. (2011) built multilingual lexicons in two languages, Arabic and Hindi. The general goal of their work was to extract the semantic orientation of new words. They created a multilingual network of words in which the words will connect if they are semantically related. For example, the authors used Wordnet (Miller, 1995) as a source of synonyms and hypernyms for linking English words in the network (i.e. English-English). By way of an example, *colour* is a hypernym of *red* while *carmine* and *sanguine* belong to the same synset *red*. Similarly, Arabic WordNet (AWN) (Black et al., 2006; Elkateb et al., 2006) and Hindi WordNet (Narayan et al., 2002) were used for the Foreign-Foreign connections in the network whereas an English-to-foreign dictionary was used to generate the English-Foreign connections.

Banea et al. (2008) introduced a bootstrapping method to build a sentiment lexicon by generating rule-based classifiers for languages with scarce resources, beginning with manually selecting seeds. Bootstrapping was used to extract new subjective candidates. For each seed word, a query was made in an online dictionary. From the results, a list of related words was selected and added to the list of candidates. The candidate words were filtered based on their similarities with the original seed, and this is continued in the next iteration until a maximum number of iterations was reached. The new subjective words were ranked based on the Latent Semantic Analysis (LSA) similarity measure, and the top entries were used to build a sentiment classifier (Banea et al., 2008). Figure 2.6 illustrates the bootstrapping process as described in Banea et al. (2008). Although this method is useful, it requires a synonym dictionary for the target language.

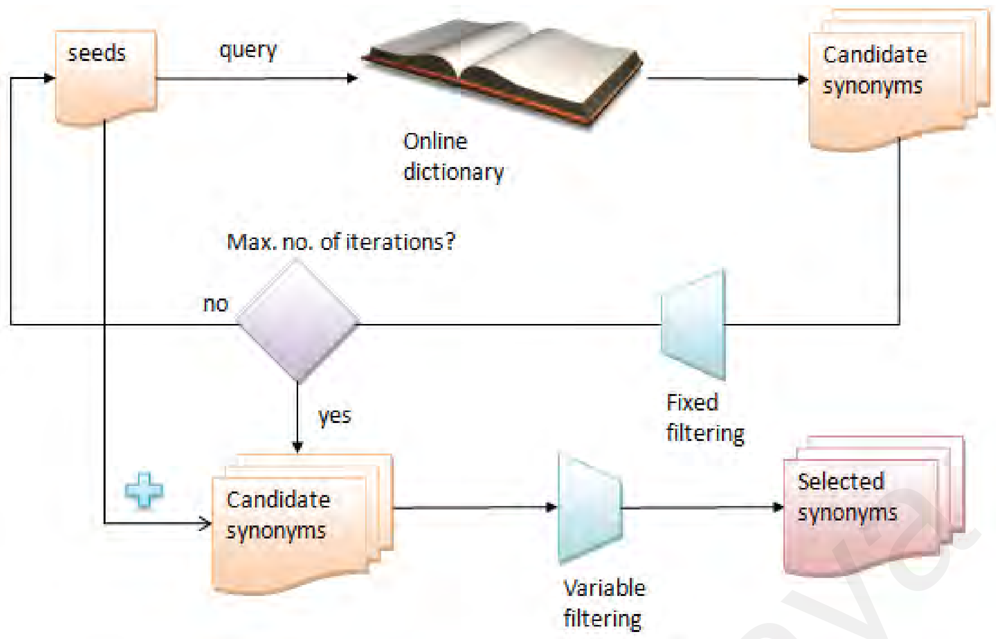


Figure 2.6 Bootstrapping process by Banea et al. (2008)

Graph-based framework was used on WordNet to construct sentiment lexicons for both Hindi and French (Rao & Ravichandran, 2009). The authors improved the label propagation results using synonymy and hypernymy relationships (i.e. semantic relation). Hindi WordNet and OpenOffice thesaurus were used for Hindi and French, respectively. Similarly, Bakliwal et al. (2012) created a Hindi lexicon by expanding its initial seed lexicon with synonym and antonym relations. Finally, Zhu et al. (2009) constructed a Chinese sentiment lexicon based on HowNet by determining the semantic similarity of Chinese words.

Additionally, many researchers implemented filtering, which is accomplished by calculating the similarity measures in the graph to remove noise from the lexicons. Common techniques used include Pointwise Mutual Information (PMI) (Hassan et al., 2011; Turney, 2001) or Latent Semantic Analysis (LSA) (Banea et al., 2008; Dumais et al., 1988). For instance, Banea et al. (2008) calculated the similarity measures between the seed and candidate words in order to choose the candidates in the next iteration, with results indicating LSA to be more efficient than PMI (i.e. faster and requires less training data).

2.3.1.3 Merge-based

The main idea behind the merge-based methods is to create large sentiment lexicons by combining predefined lexicons to increase accuracy. This is especially useful for languages in which lexical resources are lacking such as Arabic and Hindi (Joshi et al., 2010). The merging may take various forms, such as combining several lexicons in the same language, or by translating several lexicons before the merge. Table 2.9 shows a survey of the studies that have built sentiment lexicons for non-English using the merge-based methods.

Badaro et al. (2014) merged four existing sentiment lexicons to produce a new Arabic sentiment lexicon called ArSenL. The sentiment lexicons merged were Standard Arabic Morphological Analyzer (Maamouri et al., 2010), English WordNet, English SentiWordNet (Baccianella et al., 2010; Esuli & Sebastiani, 2007) and Arabic WordNet (Black et al., 2006). Likewise, Joshi et al. (2010) developed a sentiment resource for Hindi known as Hindi-SentiWordNet (H-SWN) by merging two existing resources, namely, English-Hindi WordNet (Karra et al., 2009) and English SentiWordNet (Baccianella et al., 2010). The basic premise was to keep the Hindi words unchanged, thus, if a word is found in English in SentiWordNet, the algorithm searches for the corresponding word in Hindi WordNet, and the process is repeated until the corresponding words were added to the lexicon.

Table 2.9 Survey of the studies that have built sentiment lexicons for non-English (merge-based methods)

Approach	Method	Languages (Lexicon Name)	Data sources	Data sources ref.	Technique	Domain	Number of entries	Evaluation method	Precision	Recall	F-measure
Lexicon-based	merge-based approach	Arabic (ArSenL) Badaro et al. (2014)	SAMA SentiWordNet Arabic WordNet	Maamouri et al. (2010) (Baccianella et al., 2010; Esuli & Sebastiani, 2007) Black et al. (2006)	merged existing sentiment lexicons	General	28,812	SVM	58.3	95.1	72.3
		Hindi (H-SWN) Joshi et al. (2010)	English-Hindi WordNet SentiWordNet	(Karra et al., 2009) Baccianella et al. (2010)	matching two lexical resource	General	16,253	SVM	60.3	N/A	N/A
		Arabic (SANA) Abdul-Mageed and Diab (2014)	SIFAAT and HUDA SentiWordNet General Inquirer	Abdul-Mageed and Diab (2011) Baccianella et al. (2010) Stone et al. (1966)	merged existing sentiment lexicons, translating English lexicons	General	224,564	N/A	N/A	N/A	N/A
		Italian (IRADABE2) Buscaldi and Hernandez-Farias (2016)	AFINN, Opinion Lexicon, SentiWordNet	Nielsen (2011) Hu and Liu (2004) Baccianella et al. (2010)	merged existing sentiment lexicons, translating English lexicons	General	15; 412	SVM	87	66	75
		Arabic (SLSA) Eskander and Rambow (2015)	AraMorph SentiWordNet	Buckwalter (2004) Baccianella et al. (2010)	Own linking algorithm	General	35,000	SVM	67	66.6	68.6

NB= Naïve Bayes, SVM= support vector machines, DT= decision tree algorithm, LP= Label Propagation test, LB= Lexicon-based, CS= crowdsourcing
Note: The columns: Precision, Recall, F-measure, are the results of the evaluation made by the researchers on their own data and not for comparing between the methods.

Abdul-Mageed and Diab (2014) presented an Arabic sentiment lexicon called SANA for standard Arabic and some Arabic dialects, developed both manually and automatically. The authors leveraged two existing Arabic lexicons, namely, SIFAAT and HUDA (Abdul-Mageed & Diab, 2011). SIFAAT means "adjectives" in Arabic and is composed of 3,325 Arabic adjectives whereas HUDA was extracted from an Egyptian Arabic chat data set. Pointwise mutual information (PMI) and machine translations were used to add extra words from English sentiment lexicons such as SentiWordNet (Baccianella et al., 2010) and General Inquirer (Stone et al., 1966), resulting in a total of 224,564 words. Eskander and Rambow (2015) constructed an Arabic sentiment lexicon called SLSA by linking the Arabic morphological analyser lexicon (AraMorph) (Buckwalter, 2004) with SentiWordNet (Baccianella et al., 2010). When linking the resources, the sentiment scores in SentiWordNet were applied to the entries of AraMorph to generate the new sentiment lexicon.

Finally, Buscaldi and Hernandez-Farias (2016) used a hybrid method to build a sentiment lexicon for the Italian language by creating a set of polarity words using Bing Liu's Opinion Lexicon (Hu & Liu, 2004), AFINN (Nielsen, 2011) and SentiWordNet (Baccianella et al., 2010). Their method consists of several steps beginning with the translation and merging of these three dictionaries before expanding it with the WordNet synonyms of words.

2.3.1.4 Limitations of lexicon-based methods

Several limitations were identified for the lexicon-based methods, as follows:

- The sentiment orientation of the lexicons built using lexicon-based methods are general domain lexicons, hence may appear to be less accurate when used with specific domains such as news and sports (Rao et al., 2014).

- Sentiment lexicons do not include many abbreviations or words used on social media. Therefore, they cannot handle different dialects and informal or slang words as they do not exist in the lexicons (Siddiqui et al., 2018; Wu et al., 2016).
- In machine translation, several errors may arise due to cultural and contextual differences of the sentiment orientations of words (i.e. a word may be positive in one language and negative in another, and vice versa) (Hajmohammadi et al., 2014a; Mihalcea et al., 2007; Perez-Rosas et al., 2012).
- Many words are lost when automatic translating from language to language, whereby they appear in the same translation and are considered duplicated in the new lexicon and automatically deleted (Brooke et al., 2009; Scharl et al., 2012). This happens because these automatic translators (e.g. Google translator) rely on the most common words, therefore, a number of synonyms may be translated into the same word, because the most common word is used for those synonyms. This creates a loss in the new lexicon of the target language. For example, the three words 'fabulous', 'amazing' and 'wonderful' are usually translated to one Arabic word 'رائع' while neglecting the rest of the synonyms (Lo et al., 2016b; Mihalcea et al., 2007).

2.3.2 Corpus-based approach

The corpus is a large collection of computer-readable written texts such as comments, documents, or reviews offering a rich variety of words and structures that can be relied upon to analyse the languages (Stubbs, 2001). Annotated corpora were used not only to build machine-learning based systems (Liu, 2012) but also to construct the sentiment lexicons through two types of methods: statistical and semantic relations methods. The statistical methods use large corpora with statistical equations to obtain polarity words to generate a new sentiment lexicon by calculating word frequencies in a particular class (Kumar & Jaiswal, 2016). The second method use semantic relations between words in a large corpus to produce a sentiment lexicon (Kumar & Jaiswal, 2016). Table 2.10 presents some works that were carried out for corpus-based approach. These studies were distributed in two main methods, the frequency-based method and the graph-based method, which are explained in the following.

Table 2.10 Survey of the papers that have built sentiment lexicons for non-English (Corpus-based)

Approach	Method	Languages (Lexicon Name)	Data sources	Data sources Ref.	Technique	Domain	Number of entries	Evaluation method	Precision	Recall	F-measure
Corpora-based	Frequency-based	Arabic (AraSenti-PMI) Al-Twairesh et al. (2016)	Tweets from Twitter	N/A	PMI	Twitter	N/A	N/A	90.1	89.2	89.5
		German Remus et al. (2010)	10200 product reviews	N/A	PMI	Business	N/A	N/A	96	74	84
		Persian Rashed and Abdolvand (2017)	7500 reviews	N/A	Semantic Orientation	Business	3705	N/A	80	80	N/A
		Chinese Yang et al. (2013)	Hotel review from lvping.com	N/A	Semantic Orientation-PMI	Business	N/A	LB NB	92.4	N/A	N/A
		Italian Passaro et al. (2015)	Corpus	N/A	PMI	General	N/A	CS	73	N/A	N/A
		Hindi (HMDSAD) Jha et al. (2015)	product reviews from	Amazon[26]	PMI	Business	N/A	N/A	N/A	N/A	N/A

26 http://Amazon.com

	Graph-based	Arabic Elhawary and Elfeky (2010)	large web corpus from the internet	-	similarity graph	Business	1600	N/A	75 - 88	60 - 80	N/A
		Chinese Feng et al. (2015b)	30 million Chinese microblogs	Weibo API[27]	mutual reinforcement random walk model	Microblog	12799	LB	52	74	61
		Polish Haniewicz et al. (2014)	3222 web documents	-	Random Walk	General	N/A	N/A	69.7	N/A	N/A
NB= Naïve Bayes, SVM= support vector machines, DT= decision tree algorithm, LP= Label Propagation test, LB= Lexicon-based, CS= crowdsourcing Note: The columns: Precision, Recall, F-measure, are the results of the evaluation made by the researchers on their own data and not for comparing between the methods.											

27 http://weibo.com

2.3.2.1 Frequency-based

Statistical equations and functions are used to calculate word frequency in a given polarity. This method assumes that positive words appear together with other positive words, and vice versa. Remus et al. (2010) used the co-occurrence analysis of product reviews that consisted of 5,100 positive and 5,100 negative reviews, in which users determined the rating (i.e. 1-5) of each comment. The results were lists of words that often appear together along with the polarities (positive or negative).

AraSenti-PMI is an Arabic sentiment lexicon built using pointwise mutual information (PMI) measure in a dataset of tweets (Al-Twairesh et al., 2016). PMI was used to distinguish the association between words in the corpus to be classified into positive or negative words. The PMI measure was first used in sentiment analysis by Turney (2002). The PMI between two words, *word1* and *word2*, is defined as follows.

$$PMI\ (word_1, word_2) = \log_2 \left\{ \frac{p\ (word_1\ \&\ word_2)}{p(word_1)\ p(word_2)} \right\} \qquad (1)$$

PMI was also used by Turney and Littman (2002) to find the polarity of a specific word by calculating the Sentiment Orientation SO-PMI (i.e. Sentiment Orientation-PMI) value between a word and a set of positive words (i.e. positive paradigms) minus the PMI between a word and a set of negative words (i.e. negative paradigms), as follows (Dashtipour et al., 2016; Turney & Littman, 2002):

$$\begin{aligned} SO\_PMI\ (word) \\ &= PMI\ (word, \{positive\ paradigms\}) \\ &- PMI\ (word, \{negative\ paradigms\}) \qquad (2) \end{aligned}$$

Jha et al. (2015) created a Hindi sentiment lexicon called HMDSAD based on words co-occurring frequently in a review. To calculate the relationships between the words, they used Pointwise Mutual Information (PMI). The data resource was product reviews from Amazon.com, translated into Hindi.

2.3.2.2 Graph-based

This method uses semantic relations between words in a large corpus to find new words related to some predefined words (i.e. seeds). Feng et al. (2015b) utilized a massive purified microblog dataset as training corpus to build a Chinese sentiment lexicon using emoticons to extract the polarity words. The research found that an emoticon expresses more obvious emotions if it often co-occurs with sentiment words and other important emoticons. Thus, the authors observed that the positive words frequently occur with positive emoticons, and vice versa. They integrated the emoticons and candidate sentiment words to build a graph to extract the opinion words in order to build a sentiment lexicon.

Elhawary and Elfeky (2010) on the other hand, produced a similarity graph to build an Arabic sentiment lexicon that clusters all the words/phrases of a certain language. If two words have an edge, they are similar and thus, have the same sentiment polarity, or the same meaning. A label propagation on similarity graph was performed on a seed list of 1,600 words. They built around 1,500 features such as the frequency of keywords in the document and frequency of bolded keywords to scan every document. Their lexicon consisted of two columns, one contained the word or phrase and the second represented the score of the word. For pruning purpose, filtering rules were applied to avoid both the sparseness of the data and the neglected nodes.

Haniewicz et al. (2014) attempted to build a Polish sentiment lexicon by applying the Random Walk approach on 356,275 reviews from several goods and services websites with each review having a rating score between 1 and 5. The authors designed a semantic network to store each term in the review based on the type of relation (i.e.

synonymous, hypernymous or homonymous), and their respective sentiment scores (i.e. positive, negative and neutral) in a given domain. The final Polish sentiment lexicon had about 27,000 words.

2.3.2.3 Limitations of the Corpus-based methods

The following limitations were identified for using corpus-based methods to build non-English sentiment lexicons.

- The lack of data pre-processing tools in many languages makes it difficult and complex to rely on the corpus to build lexicons (Mathur & Paul, 2012).
- There is a lack of adequate corpus online; especially for languages with fewer resources (Mukhtar et al., 2018; Quan & Ren, 2014).
- A large corpus volume is often required to construct sentiment lexicons so that an acceptable accuracy is achieved (Rashed & Abdolvand, 2017).
- The corpus-based lexicon does not contain many words and often only serves a particular domain efficiently (Tan & Wu, 2011). It can therefore not be relied upon to analyse another domain.
- Finally, some methods depend on an annotated corpus, which requires additional annotation before analysis can be performed (Pozzi et al., 2017).

2.3.3 Human-based computing approach

The human-based approach is to encourage people to answer questions or a puzzle in order to benefit from these answers in the construction of sentiment lexicons. Words are selected from a text and labelled with polarities (Hong et al., 2013). The systems are often developed using crowdsourcing platforms such as Amazon Mechanical Turk[28] (Mohammad & Turney, 2013a), building games with a purpose (Hong et al., 2013) or

[28] https://www.mturk.com/

directly by human experts (Deng et al., 2017) . Table 2.11 presents works carried out for human-based methods.

Table 2.11 Survey of the papers that have built sentiment lexicons for non-English (Human-based)

Approach	Method	Languages (Lexicon Name)	Data sources	Data sources Ref.	Technique	Domain	Number of entries	Evaluation method	Precision	Recall	F-measure
Human-Based	Crowdsourcing	Korean (Hong et al., 2013)	Players	N/A	Game (Tower of Babel)	General	N/A	N/A	N/A	N/A	N/A
		French (Lafourcade et al., 2015)	Players	N/A	Game (LikeIt)	General	385,000	N/A	N/A	N/A	N/A
		Arabic (Al-Subaihin et al., 2011)	Qaym. com Players	N/A	Game	Business	N/A	N/A	N/A	N/A	N/A
		English, Portuguese, French, Italian, Russian, German and Spanish (Scharl et al., 2012)	Players	N/A	Game and bootstrapping	Social media	N/A	NB	N/A	N/A	N/A
	Manual-based	Thai (Trakultaweekoon & Klaithin, 2016)	linguists	Pantip[29] Twitter	web-based sentiment tagging tool (SenseTag)	General Business	5120	N/A	N/A	N/A	N/A
		Arabic (Abdul-Mageed et al., 2014)	Manually	N/A	Manually	News	3982	N/A	N/A	N/A	N/A

NB= Naïve Bayes, SVM= support vector machines, DT= decision tree algorithm, LP= Label Propagation test, LB= Lexicon-based, CS= crowdsourcing
Note: The columns: Precision, Recall, F-measure, are the results of the evaluation made by the researchers on their own data and not for comparing between the methods.

29 https://pantip.com/

2.3.3.1 Crowdsourcing

Many researchers proposed crowdsourcing games to construct sentiment lexicons for resource-scarce languages (Hong et al., 2013). Lafourcade et al. (2015) developed an online game with a purpose (GWAP) that asks the players to indicate the polarity (i.e. positive, negative and neutral) of the displayed words and terms. In their extended work called *Emot*[30] (Lafourcade et al., 2016), the authors improved the game by offering the players to associate one or several emotions to a given word, either by choosing one among the displayed emotions (e.g. fear, joy, love, sadness, …), or by entering some other emotions via a text field, if none of the presented emotions were suitable. When the researchers designed their framework, they adopted the principle of simplicity in play and judgment based on the majority opinion.

Hong et al. (2013) developed a language-independent crowdsourcing game to build a Korean sentiment lexicon called Tower of Babel. Unlike previous methods, Tower of Babel required a lot of volunteers and amateurs to participate in the game to build the sentiment lexicon. Therefore, there was no need to use any previous thesaurus or provide linguistic expertise. They designed the game like Tetris where the pieces are accumulated on top of each other. The game was a collaborative game in which a pair of volunteers agree to make sentiment classifications on particular terms, and the volunteers are rewarded for making a matching classification with the partner. Another idea based on teamwork was presented by Al-Subaihin et al. (2011) who proposed a game to create Arabic sentiment lexicons by encouraging players to select words from the text and label them with polarities. The game starts with two teams of two players facing each other in three rounds. They used Qaym.com as a resource for the sentences

---

[30] http://www.jeuxdemots.org/emot.php

that were shown to every team. The winning team is the one whose members agree on the words and feelings they have chosen.

Scharl et al. (2012) presented crowdsourcing games with a purpose called Sentiment Quiz. The idea was that a number of players from different countries speaking different languages evaluate the words of their language. The results show that more than 3,500 users added approximately 325,000 evaluations in various languages such as English, Portuguese, French, Italian, Russian, German and Spanish. The next stage was to expand the sentiment lexicons by means of a bootstrapping process (see Section 2.3.1.2 for further details on bootstrapping). The results were satisfactory; however, the challenge was in convincing the required number of players or volunteers to participate in the game.

2.3.3.2 Manual-based

As the name implies, the lexicons are built manually by researchers or linguists/experts in this method. For instance, Trakultaweekoon and Klaithin (2016) developed a web-based sentiment tagging tool called SenseTag to annotate data more easily. This was accomplished by training the tool based on manual annotations provided by linguists who tag each word in randomly selected sentences (i.e. positive, negative, feature and entity). Similarly, Abdul-Mageed et al. (2014) manually created a sentiment lexicon consisting of 3,982 adjectives. The lexicon is part of the SAMAR system developed to analyse the Arabic subjectivity and sentiments in both Modern Standard Arabic and Arabic dialects. Although deemed to be time consuming and expensive, this method is still used by many researchers especially those exploring sentiment analysis in languages that lack lexical resources (Abdul-Mageed et al., 2014; Trakultaweekoon & Klaithin, 2016).

2.3.3.3 Limitations to human-based methods

Sentiment lexicons built by humans are usually more accurate than others (Hong et al., 2013), however, the production of these lexicons is time consuming, requires a large number of people and is costly (Mohammad & Turney, 2013a). To overcome these problems, several researchers built electronic games, such as Tower of Babel (Hong et al., 2013) and *Like it!* (Lafourcade et al., 2015). Figure 2.7 shows the main advantages and disadvantages of sentiment lexicons built using each approach.

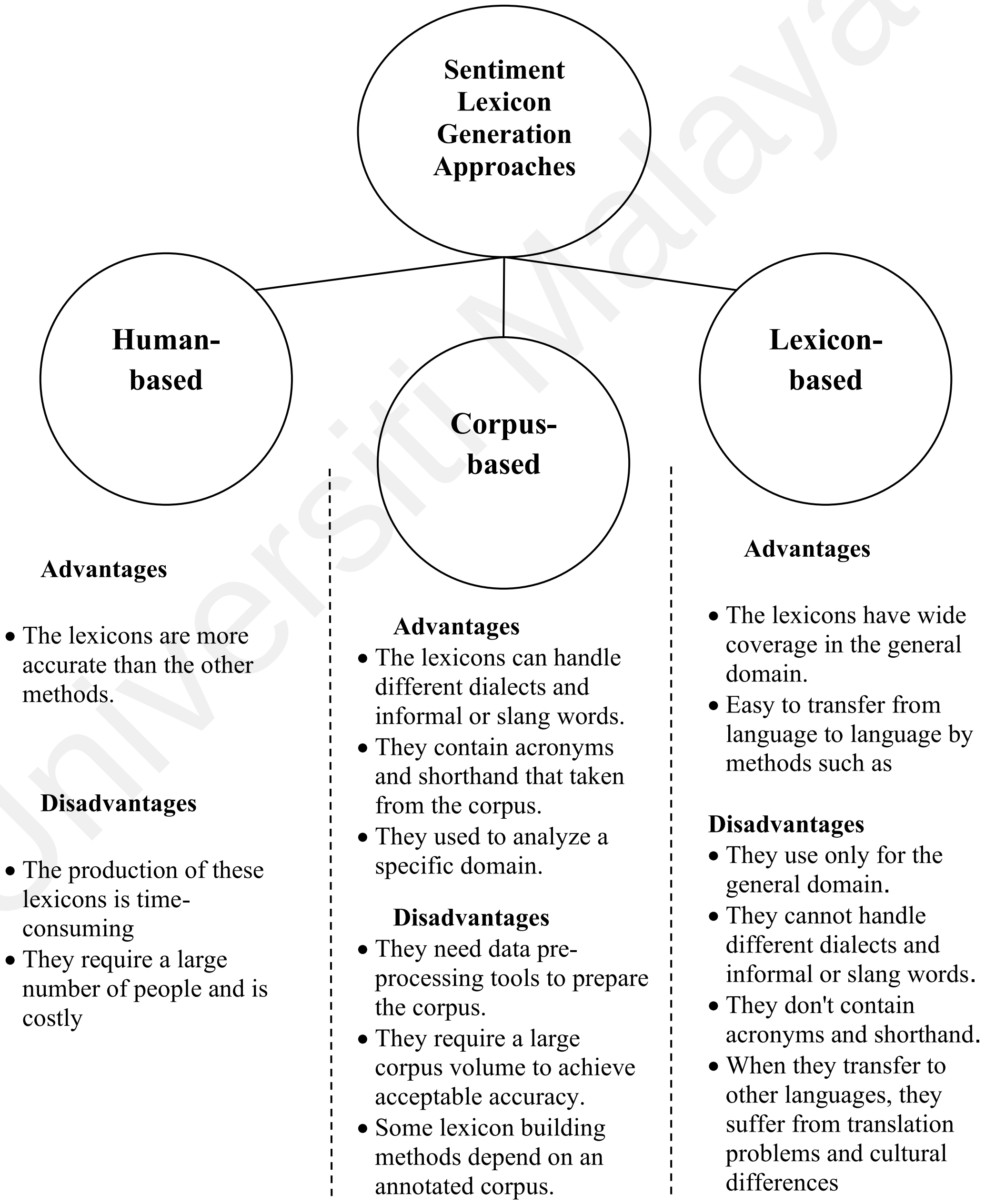


Figure 2.7 The main advantages and disadvantages of sentiment lexicons built by each method

2.4 Challenges and Open Issues

This section explains common challenges in constructing sentiment lexicons for non-English languages, including the scarcity of initial resources, the lack of pre-processing tools (Al-Twairesh et al., 2016) and translation errors (Mihalcea et al., 2007; Perez-Rosas et al., 2012). There are also open issues for research and development that include the impact of lexicon size on the accuracy of classification, adapting sentiment lexicons to a specific domain, and the use of deep learning for building sentiment lexicons. They are elaborated in the following subsections.

2.4.1 Work on scarce resource languages

Based on the literature review, it was observed that the efforts in analysing sentiment on scarce resource languages are predominantly devoted to making use of available lexicons for constructing polarity lexicons. However, many non-English languages suffer from a scarcity of primary sources and tools for the construction of sentiment lexicons (Dashtipour et al., 2016; Lo et al., 2016b). Bilingual dictionaries, annotated corpora and/or machine translation tools should be available for the construction process of any new lexicon.

2.4.2 Pre-processing tools

Sentiment analysis is a high-level NLP task that relies on pre-processing tasks, such as parsing, POS tagging, stop-word removal, stemming and word segmentation (Sun et al., 2017). Many sentiment lexicon building methods rely on the sources of the target language, and thus increases the importance of the pre-processing tools for non-English languages. Table 2.12 shows some of the tools and algorithms available for non-English languages (Sun et al., 2017). The primary pre-processing steps include:

Table 2.12 Some available NLP tools used for non-English languages

Name and Reference	Features					Languages	Developer	Programming Language	License
	Lem. /Stem.	Token/Seg	POS	NER	Others				
GATE	✓	✓	✓	✓	✓	Multi-language	University of Sheffield (1995)	JAVA	GNU Lesser General Public License
Stanford CoreNLP[31] (Manning et al., 2014)	✓	✓	✓	✓	✓	Multi-language	Stanford NLP Group	JAVA	GNU Lesser General Public License
spaCy[32] (Honnibal & Montani, 2017)	✓	✓	✓	✓	✓	Multi-language	Explosion AI, 2016	Python / Cython	MIT License
Natural Language Toolkit (NLTK)[33]	✓	✓	✓	✓	✓	Multi-language	The University of Pennsylvania, 2001	Python	Apache 2.0
FreeLing[34]	✓	✓	✓	✓	✓	Multi-language	TALP, Universitat Politècnica de Catalunya	C++	Affero GPL
FudanNLP[35] (Qiu et al., 2013)		✓	✓	✓	✓	Chinese	Fudan University	JAVA	GNU Lesser General Public License
Apache OpenNLP	✓	✓	✓	✓	✓	Multi-language	Apache Software Foundation, 2004	JAVA	Apache License, Version 2.0
FARASA[36] (Abdelali et al., 2016)	✓	✓	✓	✓	✓	Arabic	ALT Group	JAVA	Open source
MADAMIRA [37]	✓	✓	✓	✓	✓	Arabic & English	Mona Diab & et al. (Pasha	JAVA	Free for research only

[31] https://nlp.stanford.edu/software/l
[32] https://spacy.io/
[33] http://www.nltk.org/index.html
[34] http://nlp.lsi.upc.edu/freeling/node/1
[35] https://github.com/FudanNLP/fnlp
[36] http://qatsdemo.cloudapp.net/farasa/
[37] https://camel.abudhabi.nyu.edu/madamira/

							et al., 2014)		
ICTCLAS[38]		✓	✓			Chinese	Zhang Huaping & et al. (Zhang et al., 2003)	C++, Java, Python	
THULAC[39]		✓	✓			Chinese	Tsinghua University	Python	Free for research only
TextBlob	✓	✓	✓	✓	✓	Multi-language		Python	
Jieba[40]		✓				Chinese	Open source	Python	MIT License
CKIP Segmenter[41]		✓	✓		✓	Chinese	CKIP Group	Python	MIT License
Indic NLP[42]	✓	✓			✓	Indian languages	Anoop Kunchukuttan	Python	GNU General Public License
Multi-language: more than five languages. Other Features: such as parsing, n-grams chunking, coreference resolution. Lem= Lemmatization, Stem= Stemming, Token= Tokenization, POS= Part of speech tagging, NER= Named entity recognition, Seg= Segmentation.									

[38] http://ictclas.nlpir.org/
[39] http://thulac.thunlp.org/
[40] https://github.com/LiveMirror/jieba
[41] http://ckipsvr.iis.sinica.edu.tw/
[42] https://github.com/anoopkunchukuttan/indic_nlp_library

Text normalization and cleaning that include converting all letters and words to an appropriate format based on the language. Moreover, it includes converting or removing numbers, punctuations, white spaces, diacritics and stop words. Although the removal of stop words has been shown to improve sentiment analysis performance (Dashtipour et al., 2016) (Badaro et al., 2014), it has been found to be ineffective in other cases, such as in machine translation studies (Yuang et al., 2012) (Giachanou & Crestani, 2016).

-*Tokenization or segmentation* is used to split a given text into smaller pieces called tokens. The method varies from one language to another, depending on the properties of the language. A common technique used is to separate the text based on the white space, such as in English and Arabic while complex word segmentation algorithms and tools such as ICTCLAS and THULAC are used for languages with no white spaces (e.g. Chinese and Japanese). Table 2.12 shows that tokenization remains as a core process in all the tools and algorithms.

- *Stemming and lemmatization* are the processes of extracting the root of each word, in order to treat a group of words that are derived from the same root as synonyms. For example, the words "playing" and "played" will be reduced to the word "play". However, there is a danger of over-stemming and under-stemming (Al-Kabi et al., 2015). Over-stemming occurs when two different words are converted to the same stem (e.g. "universal" and "university" are converted to "universe") whereas under-stemming occurs when words of the same concept are stemmed to different roots (e.g. the words "data" and "datum" to "dat" and "datu"). Although stemming is a common step in text pre-processing, it is nevertheless language dependent (Zhang & Tsai, 2009). For example, as shown in Table 2.12, the majority of tools supporting the Chinese language do not have stemming. On the other hand, lemmatization is taking into consideration the morphological analysis of the words (Abdul-Mageed, 2017).

- *Part-of-Speech (POS) tagging* is the process of marking up each word in a given text to a particular part of speech (such as adjectives, verbs, nouns and others), based on both its definition and its context (Taboada et al., 2011).

2.4.3 Effect of lexicon size on classification accuracy

As shown in Tables 2.7 to 2.11, the size of sentiment lexicon does not have a significant impact on classification accuracy. Huge lexicons were therefore considered a challenge in sentiment analysis (Hussein, 2016). It is, therefore, necessary to study the accuracy of words in lexicons instead of the number of words. In the work presented by Badaro et al. (2014), the Arabic lexicon size was 28,812, but the total precision was only 58.3%. However, in the same language, Mahyoub et al. (2014) constructed a general-domain lexicon that consisted of 7,576 entries, with a total precision of 94%. As a result, the lexicon size is not a criterion for evaluating the lexicon accuracy. However, the lexicon should cover most of the necessary sentiment words for the classification process. Moreover, the sentiment lexicon size also differs from one language to another (Devitt & Ahmad, 2013).

2.4.4 Adapting sentiment lexicons to specific domains

Lexicons extracted from general-domain lexicons are unable to deal with sentiment information from another domain (Bravo-Marquez et al., 2016; Deng et al., 2017); the reason being general-domain lexicons include formal languages rather than informal expressions. Moreover, the sentiment orientations of some words vary from domain to domain. Therefore, resources expanded from those general-domain lexicons will exhibit limitations when used with non-English languages (Lo et al., 2016b). Based on Tables 2.5 to 2.9, 61% of the studies built general domain lexicons while only 39% built specific domain lexicons. Accordingly, the area of building specific domain lexicons for non-English languages still requires significant improvements (bin Rodzman et al., 2019).

2.4.5 Lack of evaluation benchmarks

One of the most critical challenges in sentiment lexicon evaluation is the lack of evaluation benchmarks (Giachanou & Crestani, 2016; Yue et al., 2018). As a result, the performance evaluation measures of the lexicons shown in Tables 2.7 to 2.11 vary from one research to another. Most researchers use accuracy, precision, recall and the f-measure.

Most of the researchers evaluated their proposed methods using their own data, hence a comparison of different methods with different datasets and settings is very difficult and biased. In addition, there are a number of studies that did not provide information on the evaluation of their work. In general, significant work is still required to improve precision levels in this area (Giachanou & Crestani, 2016)..

2.4.6 Deep learning

Deep learning is one of the machine learning fields applied to solve perceptual problems such as natural languages processing and speech recognition. The approach typically includes two steps for the text-related tasks: learning word embedding from the text and using them to provide the document representations (Giachanou & Crestani, 2016). Researchers such as Tang et al. (2014) presented a deep learning approach to build sentiment lexicons that could learn sentiment information based on distant supervision (Wang & Xia, 2017). However, their approach could not infer the sentiment polarity of phrases not covered in the existing vocabulary. In Kong et al. (2018), sentence-level sentiment polarity, context information and word-level sentiment polarity were combined to learn features of words in the corpus in order to construct a microblog-specific Chinese sentiment lexicon whereas Amir et al. (2015) used word embedding to define the association between positive sentiment and Twitter words. On the other hand, Dong and de Melo (2018) developed a cross-lingual propagation algorithm that generates sentiment embedding vectors for various languages, using English as the source language. Finally, although not specifically related to sentiment analysis, Bojanowski et al. (2017) published a library of generic pre-trained word vectors for efficient learning of word representations for 294 languages trained on a large-scale corpus[43]. Deep learning is promising, however, using the approach to generate new sentiment words or phrases from the corpus remains a challenge both in English and non-English texts (Giachanou & Crestani, 2016; Tang et al., 2015).

2.4.7 Language structure

The different languages have their own unique structure. For instance, in Russian language, philosophical views and thoughts are sometimes misclassified and hence some lexicon-based approaches may not be sufficient in this case (Lo et al., 2016b). Similarly, the Arabic language is a morphological language with special characteristics that create challenges for Arabic sentiment analysis. The Arabic language has a difference of challenging, complex and sophisticated grammar rules (Assiri et al., 2018). Moreover, negation rules may differ from language to language and hence may cause additional errors (Balahur & Turchi, 2013). Moreover, several dialects in some languages can be quite different in nature. For example, the Arabic language involves many different dialects have different words and rules.

[43] https://github.com/facebookresearch/fastText

2.5 Summary

This chapter provides a comprehensive review of existing research performed during the period of 2006 to 2016, on building sentiment lexicons for non-English languages. Based on the review of the literature, several gaps were identified in the field of building sentiment lexicons for non-English languages. This research is conducted with the aim to fill these gaps.

The theoretical framework guides this research and provides a novel taxonomy of the methods used to build sentiment lexicons for non-English languages. The methods employed to construct sentiment lexicons appear in three groups, each dependent on the data sources used; existing lexicons, corpus and humans. The research identifies that most researchers utilize different translation methods to build non-English lexicons. Besides the automatic translation of English sentiment lexicons, other methods include transfer learning, the graph-based approach and the merge-based approach. Conversely, a few studies were based on target language corpus using statistical methods to analyse the corpus and extract new polarity words to build new sentiment lexicons. The human-based approach adopted by several researchers collected the data directly from the individuals. Crowdsourcing and 'game with a purpose' were used to encourage people to select the correct polarity for each word. Based on the research and analysis performed in this study, multilingual sentiment analysis continues to suffer from limited resources and the recognition that this is an area (domain) of significant potential, requiring immediate attention.

CHAPTER 3: INTEGRATED FRAMEWORK FOR NON–ENGLISH SENTIMENT LEXICONS

This chapter presents the second research objective of this study, which is developing an integrated framework to build sentiment lexicons for non-English languages. The integrated framework is presented and discussed with a detailed explanation of its components in this chapter. The framework consists of three layers, namely, a corpus-based, lexicon-based and human-based layer. The first two layers automatically recognize and extract new polarity words from a huge unannotated corpus based on seed lexicons. A main strength of the proposed framework is that it only needs an initial seed lexicon and an unannotated corpus to start the extraction process. Therefore, the framework is considered to be semi-automatic due to the use of seed lexicons and human effort. In the following sections, details of the proposed framework are discussed along with figures and examples.

3.1 Overview of the Proposed Framework

Based on the literature review, the sentiment lexicons entries were extracted from three resources, namely, current lexicons, corpus and human. Each resource covers a number of lexicon entries and has some limitations, so it is beneficial to make use of all the possible entries extracted from each resource to overcome the limitations. Thus, a multi-layered framework integrating the three resources (i.e. current lexicons, corpus and human) was proposed in this study. The framework was developed based on the input and output of each layer. The first layer was the lexicon-based layer followed by the corpus-based layer, and then the human-based layer as shown in Figure 3.1. The inputs for the proposed framework are a collection of current lexicons and unannotated corpora, while the output is a set of polarity words that formed the target lexicon.

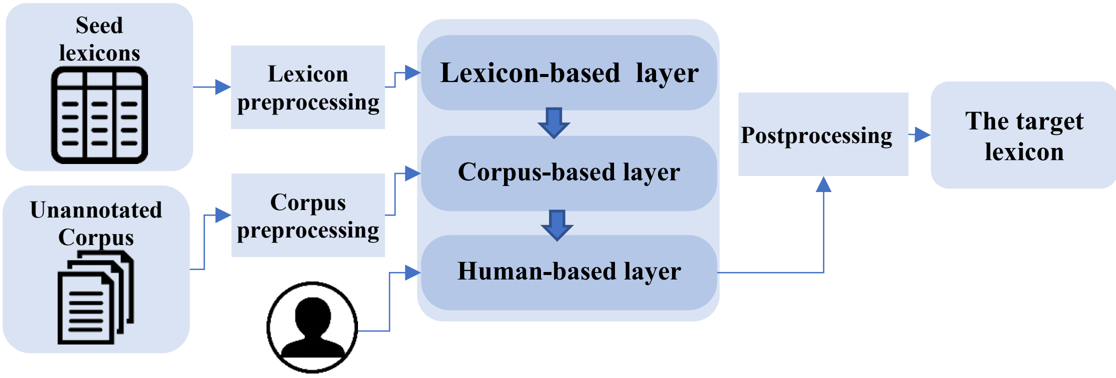


Figure 3.1 Overview of the proposed framework

The first layer (i.e. lexicon-based layer) uses machine translation tools to translate a source language lexicon in order to create an initial lexicon (i.e. seed lexicon) for the target language. In this study, languages that have plentiful and reliable sentiment resources are called source languages (e.g., English), whereas those lacking sentiment resources are called target languages (e.g., Arabic, Malay). In this layer, redundant entries were removed from the seed lexicon before proceeding to the second layer.

Then, large unannotated corpora of the target language collected to be an input of the corpus-based layer along with the seed lexicon. This step is repeated to extend the initial seed lexicon by exploring new polarity words or validating the current words in the seed lexicon. The output of the corpus-based layer, which is the new non-English sentiment lexicon, will be then evaluated by the human experts in the human-based layer. Figure 3.2 shows in detail the three layers and the flow between the layers. The following is a detailed description of these three layers.

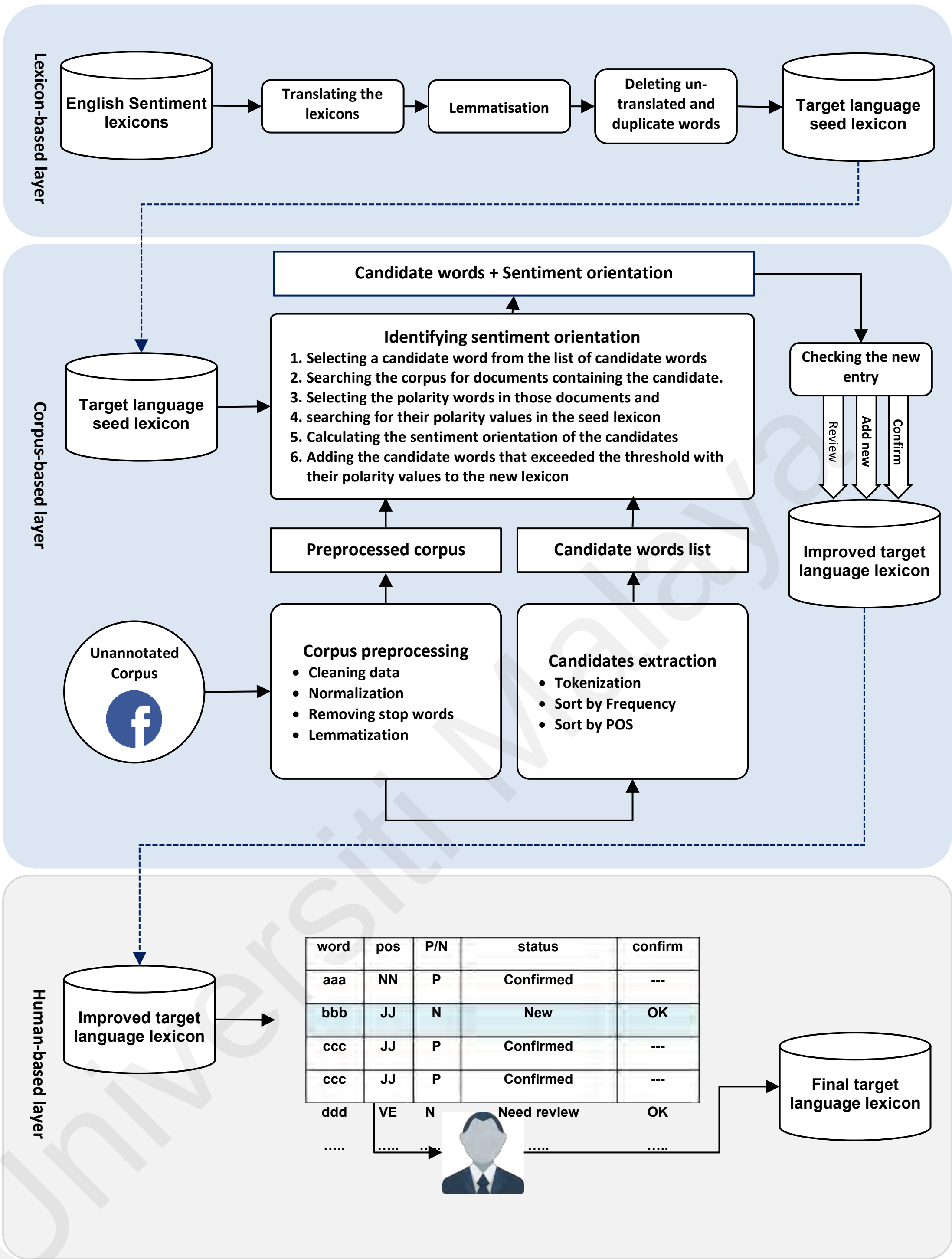


Figure 3.2 Integrated framework for non–English sentiment lexicons[1]

---

[1] The framework is registered in the Intellectual Property Corporation of Malaysia (MyIPO) under Filing No. LY2019004305

3.1.1 Lexicon-based layer

The main purpose of this layer was to generate seed sentiment lexicons, which would be used to identify the sentiment orientation of new polarity words in the corpus. One of the crucial resources used in this layer were manually built English sentiment lexicons that were translated to the target language using machine-translating tools such as Google translator[2]. To ensure that a wide variety of seed words were translated, three readily available English lexicons were used, namely Hu & Liu's Opinion Lexicon (Hu & Liu, 2004), Multi-Perspective Question Answering (MPQA) (Wilson et al., 2005b) and AFINN (Nielsen, 2011) as detailed in Table 3.1.

Table 3.1 English lexicons used to build the seed lexicon

Sentiment lexicon	Description	Polarity	Entry size	Domain	Lexicon Code*
MPQA Subjectivity Lexicon (Wilson et al., 2005b)	MPQA is an English subjective lexicon have been gathered from various resources. It presents polarity words with part-of-speech tags and strength of polarity.	strong or weak Positive or negative	8222	General Domain	MPQA
AFINN (Nielsen, 2011)	AFINN Is a strength-oriented lexicon. The words have scored based on their polarity strength from 1 to 5 in positive and from−1 to −5 in negative. In addition to ordinary words, this lexicon covers acronyms, slang, offensive words and online abbreviation.	1 to 5 -1 to -5	3382	General Domain	AFINN
Hu & Liu's Opinion Lexicon (Hu & Liu, 2004)	Opinion Lexicon is a manually generated sentiment lexicon that consists of 2,006 positive words and 4,783 negative words.	-1, 0, 1	6385	General Domain	HL
***Lexicon Code:** the lexicon abbreviation used in this thesis					

[2] translate.google.com

Lemmatisation, whereby the translated words were converted to their lemma or roots by mapping the words to their dictionary form (Abdul-Mageed, 2017; Moretti et al., 2016) was then performed. For example, the words "teachers", "were" and "different" will be converted to "teacher", "be" and "differ", respectively. Words not translated and duplicate entries were then removed from the translated lexicons. Finally, the resulting lexicon (i.e. the target language seeds lexicon) was used as the input for the next layer (i.e. corpus-based layer). Table 3.2 shows an example of the steps in the lexicon-based layer for the Arabic language.

Table 3.2 Example of lexicon-based layer steps

No	English polarity word	❶ Translation	The words after translation	❷ Lemmatisation	The words after Lemmatisation	❸ Cleaning	The words after cleaning	Status	Reason for remove
1	Nice	⇨	جميل	⇨	جميل	⇨	جميل	Keep	---
2	Bad	⇨	سيء	⇨	سيء	⇨	سيء	Keep	---
3	Terrific	⇨	رائع	⇨	رائع	⇨	رائع	Keep	---
4	Amazing	⇨	رائعة	⇨	رائع	⇨	---	Remove	duplicate
5	Righten	⇨	Righten	⇨	Righten	⇨	---	Remove	untranslated
The shaded cells indicate the cells that changed after the process									

3.1.2 Corpus-based layer

The corpus-based layer was developed to discover new polarity words based on two resources: a target language corpus and English seed sentiment lexicons. The seed sentiment lexicons were utilized to specify new sentiment words in the target language corpus depending on the relationship between the seeds and the candidate word. Figure 3.2 illustrates the corpus-based layer, which consisted of two main phases: (i) resource preparation and (ii) sentiment orientation identification. The main steps and sub-steps are elaborated in the following sub-sections.

3.1.2.1 Resource preparation

(A) Unannotated corpus

In this stage, an unannotated corpus was collected using Facebook API[3] due to the absence of an annotated corpus for many languages (Sun et al., 2017) such as Arabic and Malay. In order to improve the quality of the corpus, pre-processing techniques were applied as follows (Aldayel & Azmi, 2016; Assiri et al., 2018):

Cleaning – this involved the removal of irrelevant information such as comments containing URLs, special characters such as '@' and '#', or usernames, followed by the deletion of words, characters and punctuations unrecognized in the target language.

Normalization – this was carried out to transform the text into a consistent form (e.g. '6 cars' will be 'six cars'). In languages such as Arabic, the characters have different forms, for example, the character 'alif' is written in many forms such as 'إ','أ','آ' and 'ا', and these can be unified in a single form. Table 3.3 shows the normalization subtasks in the Arabic language (Al-Shammari, 2009; Assiri et al., 2018).

Table 3.3 Normalization subtasks in the Arabic language

No	Arabic Character(s)	Task	Output	Example	
				Before	After
1	آ,إ,ا,أ	Unifying	ا	أحسن	احسن
2	ة	Replacing	ه	جميلة	جميله
3	ى	Replacing	ي	سرى	سري
4	— (Kashida - elongation of the words)	Removing		جميــــل	جميل
5	Diacritical markings	Removing		رَائِع	رائع

[3] https://developers.facebook.com

Removing common stop words – stop words such as 'the', 'you', 'they' and 'she' do not carry any sentiment orientation (Assiri et al., 2018), therefore they were also removed from the corpus.

Lemmatization was performed to convert the words into their roots or dictionary form (Abdul-Mageed, 2017; Moretti et al., 2016). Table 3.4 presents an example of the pre-processing subtasks.

Table 3.4 Example of pre-processing subtasks

pre-processing subtasks	English example	Arabic example
Original text	This modern mobile has 10 great features #Yes http://tt.com	هذاألجوال حديث يحتوي على 10 ميزات رائعة #Yes http://tt.com
Cleaning	This modern mobile has 10 great features	هذاألجوال حديث يحتوي على 10 ميزات رائعة
Normalization	This modern mobile has ten great features	هذا الجوال حديث يحتوي على عشر ميزات رائعه
Removing stop words	modern mobile ten great features	الجوال حديث يحتوي عشر ميزات رائعه
Lemmatization	modern mobile ten great feature	جوال حديث احتوى عشر ميز رائع

(B) Candidate words extraction

In this phase, the candidate word list was extracted from the pre-processed corpus using tokenisation, a process that divides the sentences into individual words. For example, "Fish live in water" will become "Fish ", "live", "in", "water".

The number of tokens was reduced using filters. First, the term frequency was applied to group the words based on their frequencies. This step removes unusual words that were repeated fewer than 5 times in the corpus, as they were often misspelled or meaningless. The next filter was for the removal of unrecognized words in the target language, including symbols and URLs. For example, in the Arabic language, words can be sorted alphabetically to differentiate words, symbols and URLs that are not in Arabic because of the difference between the Arabic and Latin letters. Part-of-speech (POS) tagging was then performed on each candidate word in the list. Notably, adjectives and adverbs are more likely to carry sentiments compared to verbs and nouns, as indicated by majority of the previous studies (Al-Sharuee et al., 2017; Zeb et al., 2016; Zimmermann et al., 2016). Therefore, in this study, priority was given to adjectives and adverbs, followed by verbs and nouns. Table 3.5 shows an example of a candidate list pre-processing steps with the number of tokens in each step. In this example, the number of tokens starts with 161 tokens and after removing the duplicate tokens the number is reduced to 93. Then, symbols, numbers, stop words and other language words were removed. Finally, less popular words (i.e. occurrences of five and below in the corpus) were removed as they were often misspelled or meaningless.

Table 3.5 Example of candidate list pre-processing steps with the number of tokens in each step

Steps	Comments/ Tokens	Number of tokens
The Comments	Promotion should be based on merit, not race. It seems there is always this race card that's being thrown around everywhere, it seems they are individuals whose job seems to be always playing the race card. Personally I feel this could eventually lead to more racism. Let's consider merit not trying to impress a certain race. Otherwise, we are denying the world the best our societies have to offer. Why has everything got to be about race? I couldn't care if the headmaster was a blue unicorn from Jupiter. As long as they can do the job and look after school, teachers and all the #children who care what they look like. Hire based on ability, not colour. Inspire children, young adults and adults to move into a career by all means. Equality of outcome is awful and awful; we should stick to equality of opportunity but do a better job at inspiring the young to find their purpose @ALL 102	161
Removing Duplicate words	to (8) , the (7) , race (5) , a (4) , and (3) , be (3) , job (3) , not (3) , seems (3) , they (3) , adults (2) , all (2) ,	103

	always (2) , are (2) , as (2) , based (2) , card (2) , care (2) , children (2) , do (2) , equality (2) , i (2) , is (2) , it (2) , look (2) , merit (2) , of (2) , on (2) , should (2) , this (2) , we (2) , young (2) , ability (1) , about (1) , after (1) , around (1) , at (1) , awful (2) , being (1) , best (2) , better (2) , blue (1) , But (1) , by (1) , can (1) , career (1) , certain (1) , colour (1) , consider (1) , could (1) , couldn't (1) , denying (1) , eventually (1) , everything (1) , everywhere (1) , feel (1) , find (1) , from (1) , got (1) , has (1) , have (1) , @ (1) , headmaster (1) , hire (1) , if (1) , impress (1) , individuals (1) , inspire (1) , inspiring (1) , into (1) , jupiter (1) , lead (1) , let's (1) , like (1) , long (1) , means (1) , more (1) , move (1) , offer (1) , opportunity (1) , otherwise (1) , our (1) , outcome (1) , personally (1) , playing (1) , promotion (1) , purpose (1) , racism (1) , school (1) , societies (1) , stick (1) , #(1) , 102 (1)	
Removing symbols and numbers and other language words	to (8) , the (7) , race (5) , a (4) , and (3) , be (3) , job (3) , not (3) , seems (3) , they (3) , adults (2) , all (2) , always (2) , are (2) , as (2) , based (2) , card (2) , care (2) , children (2) , do (2) , equality (2) , i (2) , is (2) , it (2) , look (2) , merit (2) , of (2) , on (2) , should (2) , this (2) , we (2) , young (2) , ability (1) , about (1) , after (1) , around (1) , at (1) , awful (2) , being (1) , best (2) , better (2) , blue (1) , But (1) , by (1) , can (1) , career (1) , certain (1) , colour (1) , consider (1) , could (1) , couldn't (1) , denying (1) , eventually (1) , everything (1) , everywhere (1) , feel (1) , find (1) , from (1) , got (1) , has (1) , have (1) , headmaster (1) , hire (1) , if (1) , impress (1) , individuals (1) , inspire (1) , inspiring (1) , into (1) , jupiter (1) , lead (1) , let's (1) , like (1) , long (1) , means (1) , more (1) , move (1) , offer (1) , opportunity (1) , otherwise (1) , our (1) , outcome (1) , personally (1) , playing (1) , promotion (1) , purpose (1) , racism (1) , school (1) , societies (1) , stick (1)	154
Removing stop words	race (5) , job (3) , seems (3) , adults (2) , based (2) , card (2) , care (2) , children (2) , equality (2) , look (2) , merit (2) , young (2) , ability (1) , awful (2) , being (1) , best (2) , better (2) , blue (1) , career (1) , certain (1) , colour (1) , consider (1) , denying (1) , eventually (1) , everything (1) , everywhere (1) , feel (1) , find (1) , got (1) , headmaster (1) , hire (1) , impress (1) , individuals (1) , inspire (1) , inspiring (1) , jupiter (1) , lead (1) , let's (1) , like (1) , long (1) , means (1) , more (1) , move (1) , offer (1) , opportunity (1) , outcome (1) , personally (1) , playing (1) , promotion (1) , purpose (1) , racism (1) , school (1) , societies (1) , stick (1) , teachers (1) , thrown (1) , trying (1) , unicorn (1) , world (1)	59
Removing less popular words	race (5) , job (3) , seems (3) , adults (2) , based (2) , card (2) , care (2) , children (2) , equality (2) , look (2) , merit (2) , young (2) , ability (1) , awful (2) , best (2) , better (2)	16
The candidate list	race , job , seems , adults , based , card , care , children , equality , look , merit , young , ability , awful , best , better	16
Note: The numbers in brackets present the frequency of the tokens		

3.1.2.2 Sentiment orientation identification

In this stage, the sentiment orientation of the candidate words was identified utilising the pre-processed corpus and seed lexicon, which required the relationship between the "new" words (i.e. candidates) and the previously known polarity words (i.e. seeds) to be determined. A new candidate word was first chosen from the candidate words list and used to explore the unannotated corpus for any documents containing the word. The seed lexicon was then used to specify the polarity words in those documents. Then, the candidate sentiment orientation value (CSO) for the candidate word was specified by modifying the Semantic Orientation – CALculator (SO-CAL) (Taboada et al., 2011). SO-CAL predicts the overall sentiment orientation of a given text using a simple aggregate-and-average method, as shown in Eq. (3.1).

$$SO - CAL_{document} = \frac{\sum sentiment\ scores\ for\ each\ polarity\ word\ in\ a\ document}{\sum words\ in\ the\ document} \tag{3.1}$$

To calculate the candidate sentiment orientation (CSO) of the candidate word, the number of positive words (NP) was subtracted from the number of negative words (NN) in the documents containing the candidate word. The result was then divided by the total number of polarity words in the documents containing the candidate word. The absolute value is used for the negative words (NN) as the polarity value of the negative words in some sentiment lexicons are declared using negative values (e.g. "bad, -1"), as shown in Equations (3.2) and (3.3) below.

$$CSO_c = \frac{\sum Positive\ words - \left|\sum Negative\ words\right|}{\sum Positive\ words + \left|\sum Negative\ words\right|} \tag{3.2}$$

$$CSO_c = \frac{\sum_k NP - \left|\sum_k NN\right|}{\sum_k NP + \left|\sum_k NN\right|} \tag{3.3}$$

As the corpus still contained many unimportant repetitive words such as stop words and entity names, the sentiment orientation value of the candidate word could not be solely relied upon. Thus, the CSO was combined with the term frequency–inverse document frequency (TF-IDF) to determine the association between the candidate and seed words. The TF–IDF weight (Salton & Buckley, 1988) is a numerical statistic which reflects how important a word is to a document in a corpus, as shown in Equations (3.4) and (3.5),

$$TF_{cd} = \frac{CF_{cd}}{\sum NW_{cd}} \tag{3.4}$$

$$IDF_d = \log \frac{ND}{|DC : c \in d|} \tag{3.5}$$

where *ND* is the number of documents in the corpus, and $|DC : c \in d|$ is the number of documents containing the candidate word, CF is the frequency of candidate word c in document d, and NW is the number of all words in document d. Then, the importance of the candidate word, c to the document, d can be weighted by Eq. (3.6).

$$W_c = TF_{cd} \, . \, IDF_d \tag{3.6}$$

The combination of CSO and TF–IDF is represented by Eq. (3.7).

$$CSO - TFIDF_c = \frac{\sum_k NP - \sum_k NN}{\sum_k NP + |\sum_k NN|} \, . \left(\frac{CF_{cd}}{\sum NW_{cd}} \, . \, \log \frac{ND}{|DC : c \in d|}\right) \tag{3.7}$$

Table 3.6 The terms used to formulate the CSO-TF IDF

Term	Meaning
d	Document
c	Candidate word
k	The corpus
NN	Nearby Negative words (Negative words in the same document)

NP	Nearby positive words (positive words in the same document)
SO	Semantic orientation, is a measure of subjectivity and opinion in text.
ND	Number of documents in the corpus
NW	Number of all words in document d
CF	Frequency of candidate word c in document d
CP	Candidate word polarity value
DC	Number of documents that contain the candidate word
T	Threshold
TF-IDF	Term frequency–inverse document frequency

Table 3.6 presents the terms utilised to formulate the CSO-TFIDF equation, which is the candidate word sentiment orientation value calculated by defining the polarity values of the seed words, and multiplied by the TF-IDF value of the candidate word to ensure minimal noise due to any misspellings and stop words. The more a word is repeated in multiple documents, the more likely it is a polarity word (Taboada et al., 2011). In some cases, however, candidate words are repeated numerous times only in a single document. In other words, there is a distinction between the words being repeated five times in one document compared to it being repeated five times in five various documents.

A word can be defined as being either negative or positive based on its CSO-TFIDF value (i.e. it is a negative word if its CSO-TFIDF is negative and vice-versa). Nevertheless, this generates undesirable noise in the extracted lexicon because of the inclusion of weak polarity words. In this case, a threshold value (T) can be considered. Accordingly, if the positive value is greater or equal to the positive threshold (T+), then the word will be considered positive and vice versa, as shown in Eq. (3.8) (Wu et al., 2016) . The threshold value was selected manually by checking the polarity words from the obtained lexicon.

$$\text{Candidate word polarity} = \begin{cases} Positive \ \ if\ CSO - TFIDF(c) \geq (T+) \\ Negative \ \ if\ CSO - TFIDF(c) \leq (T-) \\ \quad\quad\quad Neutral \ \ else \end{cases} \quad (3.8)$$

3.1.3 Human-based layer

The human-based layer was included as a measure to increase the verification of the new lexicon. Reviews were performed manually for words requiring a review (e.g. new words or unrecognized words). After obtaining the lexicon extracted from the corpus, the nominated words were sorted according to their CSO-TFIDF values. The top half of the positive and the negative word lists were accepted without a manual review due to their high CSO-TFIDF values.

The remaining half of the words was tested again against the seed lexicon. If the same word was present in the resulting lexicon of the corpus-based layer and the seed lexicon, the polarities of these words were compared. If the polarities were found to be equal, it was then confirmed, else it would be placed for human review. If the particular word was absent in the seed lexicon, it would be placed for human review directly.

Figure 3.3 shows an example of the review process (i.e. human-based layer) for the improved target language lexicon (i.e. the output of the corpus-based layer). The presented example is in the English language to clarify the process. For example, the words 'bravo', 'good' and 'clean' obtained the highest values for CSO-TFIDF, so they were accepted without any human review. However, the words 'sweet', 'wow' and 'natural' obtained a lower CSO-TFIDF value, so they needed to be compared with the seed lexicon. For instance, the word 'sweet' obtained a positive polarity value (i.e. CSO-TFIDF greater than 0) in the corpus-based lexicon and it was positive in the seed lexicon as well, so it was accepted without a human review. However, the word 'wow' was added to the improved lexicon with the status 'need review' because it obtained a low CSO-TFIDF value and did not exist in the seed lexicon. The manual review process was done by distributing the generated lexicon to a group of native speakers of the target language to verify the sentiments identified in the previous two layers.

Figure 3.4 illustrates a complete example of how the integrated framework works in building non-English sentiment lexicons. The example is in Arabic with English translation. It begins from the lexicon-based layer, where the arrows illustrate the flow of the processes in the figure.

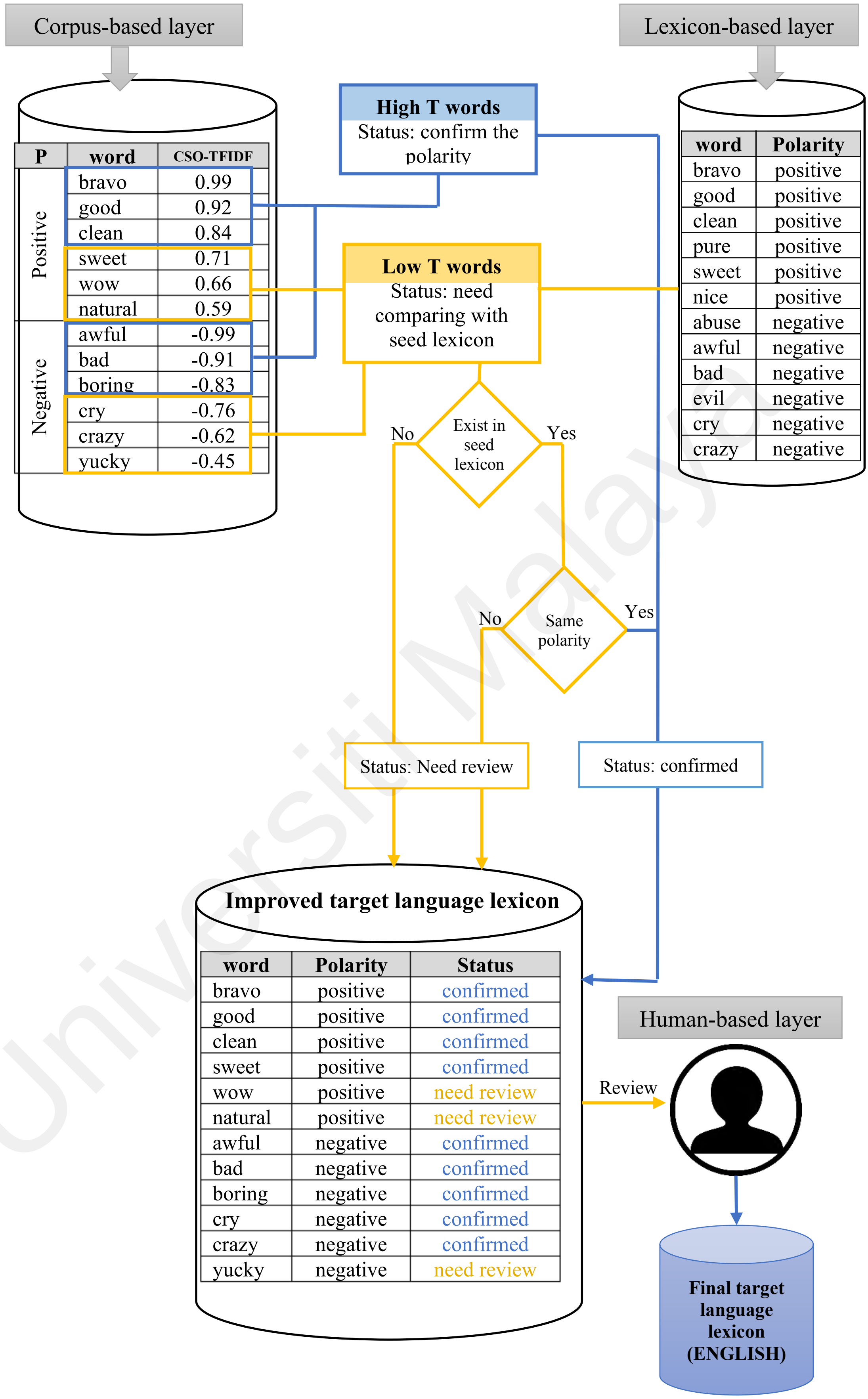


Figure 3.3 Reviewing process of the improved target language lexicon (The example in English)

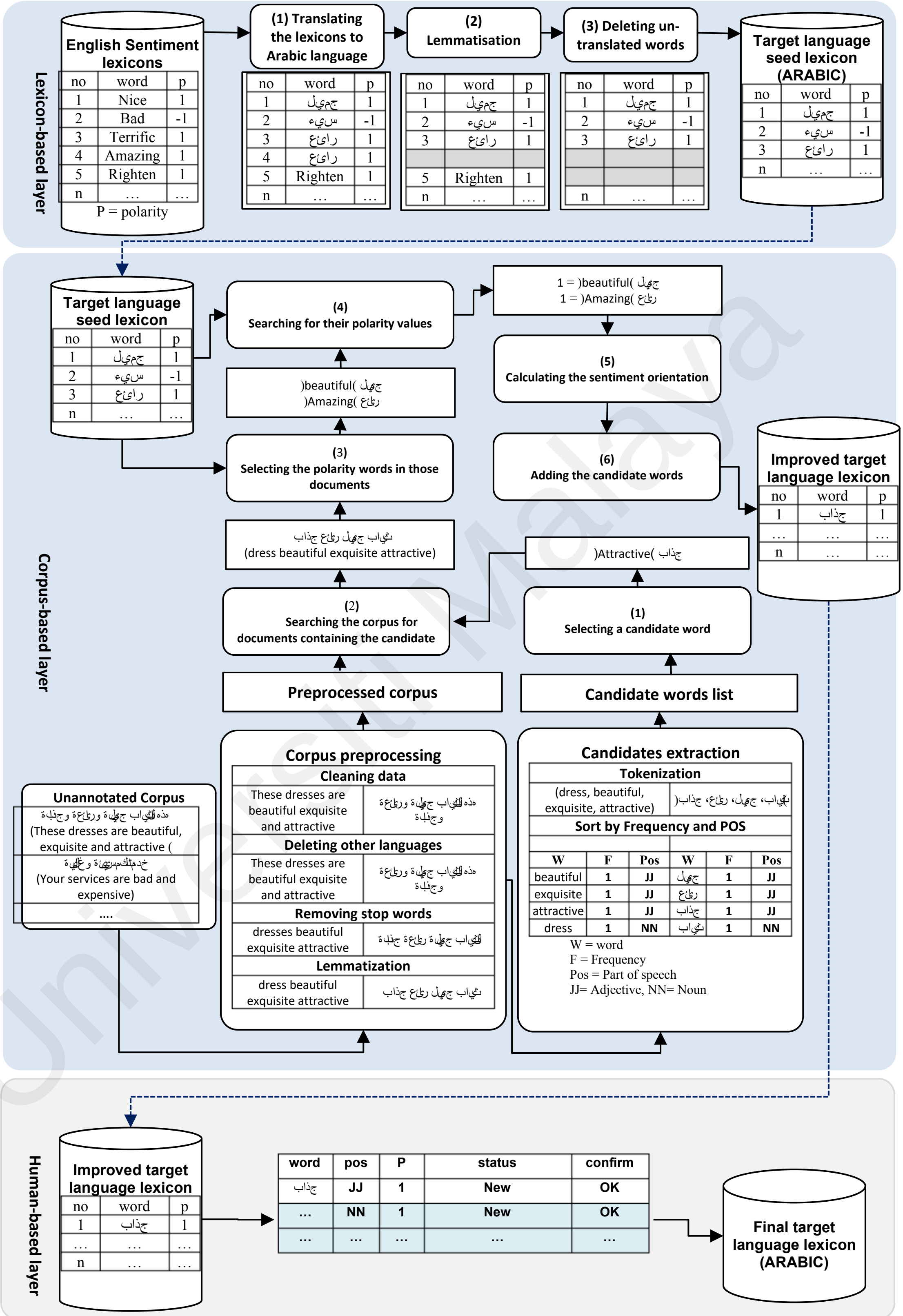


Figure 3.4 A complete example of the integrated framework (The example in Arabic with English translation)

3.2 Evaluation metrics

The lexicon scoring method was adopted for classification purposes, i.e. in any document, if the number of positive words is higher than the number of negative words, then the document is classified as positive; otherwise, it is classified as negative (Molina-González et al., 2013; Taboada et al., 2011). A confusion matrix (Powers, 2011) was used with four indices: accuracy (A), precision (P), recall (R) and F-measure (F), to measure the performance of the proposed framework as follows:

$$Accuracy\ (A) = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.9)$$

$$Precision\ (P) = \frac{TP}{TP + FP} \quad (3.10)$$

$$Recall\ (R) = \frac{TP}{TP + FN} \quad (3.11)$$

$$F - Measure\ (F) = 2.\frac{P.R}{P + R} \quad (3.12)$$

Where:

- TP (true positives) are cases assumed to be positive and appear as positive;
- TN (true negatives) are cases assumed to be negative and appear as negative;
- FP (false positives) are cases assumed to be negative but appear as positive;
- and FN (false negatives) are cases assumed to be positive but appear as negative (Giachanou & Crestani, 2016).

Table 3.7 displays an example of how the evaluation measures were calculated based on the confusion matrix. Seven cases were selected randomly from the test dataset, which was annotated manually to positive or negative (i.e. actual class). Then, these cases were tested using the lexicon-based method to predict the sentiment orientation values. The results were grouped in Table 3.8, which represents the confusion matrix for this example. Finally, Equations 3.9 to 3.12 were applied to determine the accuracy (A), recall (R) precision (P), and F-measure (F) values.

Table 3.7 Example of calculating the evaluation measures based on the confusion matrix

No	Arabic text (Translation)	Actual class	Predicted class	Result
1	شعب عظيم تحية للشعب العظيم شعب النضال، والكفاح، شعب الصمود، والتحدي تحية لمصنع الرجال ومصدرها *(A great people, greetings to the great people of struggle, struggle, people steadfastness, challenge, greetings to the factory of men and their source)*	Positive	Positive	TP
2	فلتسرحوا وافرحوا وانبسطوا وليقتل المستضعفين وليذبحوا حسبي الله ونعم الوكيل *(Be satisfied, happy, and pleased, and let the vulnerable are killed and slaughtered. Allah suffices me, for He is the best disposer of affairs).*	Negative	Positive	FP
3	في النهاية المجرمين يمزقون الاتفاقية ويهاجمون مسجد الح الدين يستهدفون النساء وا فال *(In the end, criminals tear up the agreement and attack Salahuddin mosque and target women and children).*	Negative	Negative	TN
4	يوم جميل ورهيب في حياة الشعب سيكون سبب في رفع الظلم عنه ان شاء الله *(A beautiful and awesome day in the life of the people will be a reason to lift the injustice from them, God willing)*	Positive	Positive	TP
5	أنتم دعاة فتنة لن يصدقكم الشعب الصادق النزيه *(You are Advocates of sedition, so the fair and honest people cannot believe you).*	Negative	Positive	FP
6	3 سنوات من القصف والقتل للمدنيين والكوليرا كملت عليهم حسبنا الله ونعم الوكيل *(3 years of bombing and killing of civilians and cholera completed them)*	Negative	Negative	TN
7	هذا الخبر كلام فاضي ما يحتاج تحليل و تعب خليك في حالك أحسن. *(This news is empty words, do not needs analysis and fatigue. it is better to be in your business.)*	Negative	Negative	TN
FP=false positives, TN=true negatives, TP =true positives, FN=false negatives				

Table 3.8 Confusion matrix for the example in Table 3.7

		Actual class	
		Positive	**Negative**
Predicted class	**Positive**	TP = 2	FP = 2
	Negative	FN = 0	TN = 3
TN= true negatives, TP = true positives, FN= false negatives and FP= false positives			

The result for this example will be as follows:

$$Accuracy\ (A)\ =\ \frac{2+3}{2+3+2+0} = 0.71\ ,$$

$$Precision\ (P) = \frac{2}{2+2} = 0.5,$$

$$Recall\ (R) = \frac{2}{2+0} = 1,$$

$$F - Measure\ (F) = 2.\frac{0.5 * 1.0}{0.5 + 1.0} = 0.67$$

3.3 Summary

In this chapter, the proposed integrated framework to build sentiment lexicons for non-English languages is presented and discussed. The implementation, evaluation, results and more details about the integrated framework will be introduced in chapter 4 and 5, entitled building and evaluating sentiment lexicons for Arabic, French and Malay languages. These chapters present the third objective of this study, which is evaluating the integrated framework by conducting experiments and evaluation measures.

CHAPTER 4: BUILDING AND EVALUATING SENTIMENT LEXICONS FOR ARABIC, FRENCH AND MALAY LANGUAGES

This chapter is divided into four sections. The steps for the data collection and pre-processing are explained in the first section. In the second section, the implementation of the method of extracting new polarity words from an unannotated corpus is explained. In this section, the tools and programming languages used are mentioned, and the extraction algorithm is presented. Then, in the third section, the extracted lexicons for the three evaluated languages are presented and discussed. Finally, in the fourth section, the evaluation procedures are applied to the extracted lexicons.

4.1 Data collection and pre-processing

As stated in the research scope (Section 1.7), the proposed integrated framework was evaluated using three languages, namely, Arabic, French and Malay, chosen due to their popularity based on the total number of speakers worldwide (i.e. 422 million, 281 million and 229 million for Arabic, Malay and French, respectively) (Sawe, 2019). Additionally, these languages are also considered as emerging languages in terms of sentiment resource building (Abdaoui et al., 2016; al Owisheq et al., 2016).

The application programming interface (API)[1] for Facebook (Jünger & Keyling, 2017) was used to crawl posts and comments from two different news pages for each language (identities withheld for the purpose of confidentiality). The news domain was chosen because most comments are written in a classical language with few errors. However, the collected corpora were unannotated and very noisy because they contained URLs, symbols and many typographical and spelling errors and spams. The collected corpora for the three languages were document-based, whereby each comment

[1] https://developers.facebook.com

was considered as a document. The unannotated corpora were gathered over a period of eight weeks from March 6 to May 4, 2017. Table 4.1 shows five examples of the collected unannotated comments for each language.

Table 4.1 Examples of the collected comments for each language

Doc. No.	Arabic	French	Malay
1	دولة محترم وشعب فاهم كذا يكون التطور وا الش	Sympatique intelligent constructif ##	Kat malaysia ni acc fb pun fake Mcm mana laaa kehidupan sebenar
2	رئيس فاشل وري ديمقراطية طق لي خلونا من الكلام فارغ ذه	La France il faut être clair et très claire si non?	Amalkan makn carrot sama ada mentah @ dimasak bagus unt mata
3	التعاشاقتطي شكلك اعمى ههههههههه https://tco/4p	N'importe comment vous trouverait toujours qlq chose alors quoi il a fait une bourde point	Bagusnya semua dengar arahan Senang kerja tuan dia
4	قفتونا والله من تقاي رايك الكتاف هه والكذبه والفبركه	on a jamais su ce que les noirs ont fais aux blancscest terrible	biase yg promosi mcm ni kuality x berape bagus
5	هذل كالم نثلق يم واجب العزاء لشخص ان تقتله	Mohamed aime ca Tres bien @Alic	YaAllah selamat kanlah saudara kami disana 😊

The steps for the data collection and pre-processing, which include the pre-processing of the corpus and the preparation of the candidate words list, and seed lexicons are elaborated in the following subsections.

4.1.1 Seed lexicons pre-processing

Three freely accessible English lexicons were used, namely, Hu & Liu's Opinion Lexicon (Hu & Liu, 2004), MPQA (Wilson et al., 2005b) and AFINN (Nielsen, 2011) (see Table 3.1 in Chapter 3). Google's machine translation tool[2] was used to translate

[2] translate.google.com

the three English lexicons in the three languages (i.e. Arabic, French and Malay). Table 4.2 shows the steps for the preparation of the seed lexicons and the number of words in each step.

Table 4.2 Pre-processing steps of the seed lexicons

Languages	Lexicons	Size	Un-translated words	After Translating	Duplicate words	Total	Translated Lexicon Code
Arabic	MPQA	8222	285	7937	3259	4678	Ar_ MPQA
	Hu & Liu	6385	403	5982	1426	4556	Ar_HL
	AFINN	3382	117	3265	838	2427	Ar_ AFINN
French	MPQA	8222	0	8222	2970	5252	Fr_ MPQA
	Hu & Liu	6385	0	6385	1245	5140	Fr_HL
	AFINN	3382	0	3382	575	2807	Fr_ AFINN
Malay	MPQA	8222	0	8222	4178	4044	My_ MPQA
	Hu & Liu	6385	0	6385	2390	3995	My_HL
	AFINN	3382	0	3382	1285	2097	My_ AFINN
MPQA = Multi-Perspective Question Answering							

After translating the English sentiment lexicons, any untranslated words were removed. For the Arabic language, it was easy to find words that were not written in Latin letters. However, all the untranslated words in French and Malay were adopted because there were some English words that were used in both languages. The outputs at this stage were the seed lexicons in the three languages.

4.1.2 Corpus pre-processing

The textual content of each comment (i.e. document) was pre-processed using a set of pre-processing tools for each language, including tokenization and the removal of stop words.

The Notepad++[3] free text editor was used for searching and removing symbols and URLs from the corpus. The Arabic language was the only language that needed a lot of normalization. For the lemmatisation process, FARASA[4] (Abdelali et al., 2016), which is a quick and reliable Arabic text processing toolkit, was used to convert all the Arabic words to their roots and dictionary forms. On the other hand, the Natural Language Toolkit (NLTK)[5], a platform for building Python programs to work with textual communication, was used for the French corpus. As for the Malay corpus, the pre-processing was done manually using Notepad++ text editor as no existing tool or code was found.

The initial corpus contained 446,468, 275,843 and 312,569 reviews in Arabic, French and Malay, respectively. After the pre-processing, the final corpus contained 279,107 reviews in Arabic, 245,280 reviews in French and 254,102 reviews in Malay.

4.1.3 Candidate words list pre-processing

In this step, the candidate words list was extracted from the pre-processed corpus by applying the tokenization and cleaning processes, as detailed in Section 3.1.2.1.

Table 4.3 shows the steps adopted to clean the tokens to achieve the final set of candidate words for the three languages. The social media corpora used were very dirty by containing misspelled and meaningless words. Thus, the step of removing less popular words removed unusual words and symbols that were repeated fewer than 5 times in the corpus. The remaining words after this step are the words which repeated more than 6 times in the corpus. The final numbers in Table 4.3 referred to the number

[3] https://notepad-plus-plus.org/

[4] http://qatsdemo.cloudapp.net/farasa/

[5] https://www.nltk.org/

of words without repetition. For example, in the Arabic candidate words list, the word ("مبروك " means "Congrats") repeated 4152 times, so it added to the final list as one word.

Table 4.3 Steps of pre-processing candidate list with the number of tokens in each step

No	Steps	Number of tokens		
		Arabic	**French**	**Malay**
	Number of reviews	**279,107**	**245,280**	**254,102**
1	**Total token numbers**	**38,530,015**	**18,812,284**	**16,452,796**
2	After removing Duplicate words	14,253,661	6,225,361	5,821,365
3	After removing symbols and numbers	10,427,789	5,889,132	5,004,871
4	After removing other language words	8,052,304	4,556,405	4,902,511
5	After removing stop words	5,800,417	1,282,201	1,020,197
6	After removing less popular words	10,765	8,795	7,268
7	**The candidate list**	**10,765**	**8,795**	**7,268**

4.2 Sentiment orientation identification

PHP[6] and JAVA[7] were used to develop the sentiment lexicon builder to extract new polarity words from the unannotated corpus, whereas the database was designed using MySQL[8] to store the pre-processed data. Figure 4.1 shows the algorithm for extracting new polarity words from an unannotated corpus (see Appendix A for the algorithm code). The algorithm had three main inputs, namely, the candidate words list, the unannotated corpus and the seed lexicons. The polarity threshold, T was the only parameter that had to be tuned. A high value of T would extract much fewer polarity words whereas a low value of T would add too much noise and unnecessary words. In the experiments, T was tuned to be between +5.0 and -5.0. The following are the

[6] https://www.php.net/

[7] https://www.java.com/en/

[8] https://www.mysql.com/

sentiment orientation identification steps with the algorithm line numbers shown in brackets:

1. Selecting a candidate word from the list of candidate words (1-2).
2. Searching the corpus for documents containing the candidate word (3-5).
3. Selecting the polarity words in those documents (6-11).
4. Searching for their polarity values in the seed lexicon (12-21).
5. Calculating the sentiment orientation of the candidate words (CSO) (22-24).
6. Adding the candidate words that exceeded the threshold with their polarity values to the new lexicon (25-32).

```
    Input
    Candidate word list C,
    Unannotated corpus K,
    Seed lexicons S,
    Threshold T,
    Number of documents in the corpus ND
    Output
    Non-English sentiment lexicon SL
    Begin
1:  While not end of the candidate word list C do
2:      Select a candidate word c from the candidate list C
3:      While not end of unannotated corpus K do
4:          Search the corpus K for documents d containing the candidate word
            c
5:          DC = the number of documents contain c
6:          Segment the document d into words w
7:          CF = the number of candidate word c in the document d
8:          NW = the number of words w in the document d
9:          Segment the document d into words w
```

```
                CF = the number of candidate word c in the document d
                NW = the number of words w in the document d
                For each word w
                    If c exists in S
                        Return the polarity P
                        If P=positive
                            NP=NP+1
                            Else
                            NN=NN+1
                        End If
                    Enf If
                End for
                Calculate TF-IDF = ((CF/NW) log10(ND/(DC)))) as Eq. (3.6)
            End while
            Calculate CSO-TFIDF(c) = ((∑NP-∑NN)/(∑NP+∑NN)) * TF-IDF as Eq.
            (3.7)
            Test the Threshold T using Eq. (3.8)
            If CSO-TFIDF(c) >= T+
                The polarity P(c) = Positive
            Else If CSO-TFIDF(c) <= T-
                    The polarity P(c) = Negative
            End If
            Add the candidate words c that exceeded the threshold with their polarity
            values P to the new lexicon SL
    End while
    End
```

Figure 4.1 Algorithm of extracting new polarity words from an unannotated corpus

4.3 Integrated non-English sentiment lexicons

Three new sentiment lexicons were generated for each of the three languages. These lexicons consisted of the seed words generated in the lexicon-based layer (L1) and the polarity words extracted from the unannotated corpus in the corpus-based layer (L2).

Moreover, these polarity words were reviewed by human experts in the human-based layer. These new lexicons are referred to as integrated sentiment lexicons (ISL) in this study and tagged with language to differentiate them. In other words, the lexicon is referred to as Arabic integrated sentiment lexicon (AISL), French integrated sentiment lexicon (FISL) and Malay integrated sentiment lexicon (MISL). In the following subsections, the three integrated sentiment lexicons (ISL) were described[9]. Moreover, we presented some examples of the three ISLs entries (i.e. the top positive and negative words) in order to show the effect of the cultural differences between languages on the lexicon entries even though the corpus were collected from the same domain (i.e. news).

4.3.1 Arabic integrated sentiment lexicon (AISL)

The Arabic integrated sentiment lexicon (AISL) contained 17,054 words divided into 5,287 positive and 8,630 negative words. The rest of the lexicon was comprised of neutral words. The number of seed words extracted from the lexicon-based layer was 9,672. The combination of these words and the polarity words extracted from the corpus resulted in a single lexicon, that is, AISL.

Table 4.4 shows the top 20 positive and negative words in AISL. It can be noted that the word "مبروك", which means 'congratulation' in Arabic, was repeated in 4,152 documents and had 9,522 positive seed words in the same documents. As a result, it obtained the highest CSO-TFIDF value (i.e. Eq. 3.7) of 1,250. The words from 2 to 6 are used in Arabic supplications for mercy and forgiveness, hence their appearance at the top of the list of positive words. As for the negative words, words such as "كذاب" (liar) and "عار", (shame) emerged at the top of the list. Moreover, it can also be observed that it is common to use reference to animals to indicate negative feelings. For

---

9 Available online in: https://github.com/mohkaity/ISL

example, the word "حمار", which means "donkey", appeared in third place [Item 13] in the list of negative words.

Table 4.4 Top 10 positive words and top 10 negative words in the Arabic integrated sentiment lexicon

No	Word	Translation	DC	NP	NN	N	TF-IDF (Eq 3.6)	CSO (Eq 3.3)	CSO-TFIDF (Eq 3.7)	SO
1	مبروك	Congrats	4152	9522	2338	11860	2063.85	0.605734	1250.143	P
2	فسيح	Roomy	1024	5908	1232	7140	1167.92	0.654902	764.8731	P
3	جنة	Paradise	1433	12422	3986	16408	1349.36	0.514139	693.7592	P
4	راجع	went back	2483	12925	4621	17546	1186.81	0.47327	561.6819	P
5	غفر	Forgive	876	6741	2027	8768	989.14	0.537637	531.7981	P
6	رحمة	Mercy	1504	11893	3985	15878	1052.78	0.498048	524.3346	P
7	بركة	Pond	958	6601	2160	8761	951.411	0.506906	482.2756	P
8	توفيق	Reconcile	1221	6780	2428	9208	899.189	0.472632	424.9859	P
9	شفى	Cured	1383	4417	1741	6158	932.572	0.434557	405.2554	P
10	رائع	Gorgeous	1304	5153	2008	7161	913.794	0.439184	401.3241	P
11	كذاب	Liar	1419	3014	4597	7611	1018.47	-0.20799	-211.83	N
12	عار	Shame	1931	5662	8212	13874	1132.42	-0.1838	-208.135	N
13	حمار	Donkey	1478	3408	5268	8676	834.464	-0.21438	-178.896	N
14	سفاح	Thug	670	1511	3040	4551	528.736	-0.33597	-177.639	N
15	نفاق	Hypocrisy	1814	6264	8600	14864	1065.09	-0.15716	-167.388	N
16	وسخ	Grime	840	1604	3046	4650	527.833	-0.31011	-163.685	N
17	تف	Spit	197	184	971	1250	219.165	-0.68139	-149.336	N
18	تحير	Tacky	656	1339	2464	3803	502.547	-0.29582	-148.663	N
19	غبي	Stupid	1536	4830	6760	11590	859.689	-0.16652	-143.158	N
20	خباله	Betrayal	1190	4336	6254	10590	698.186	-0.18111	-126.451	N

SO: Semantic orientation (P: positive, N: Negative), NP: Number of nearby positive words, NN: Number of nearby negative words, N: Total of nearby polarity words NP+NN, DC: Number of documents that contain the candidate word.

4.3.2 French integrated sentiment lexicon (FISL)

The French integrated sentiment lexicon (FISL) consisted of 6,572 polarity words, of which 2,220 were positive, 4,003 were negative, and the rest of the lexicon was comprised of neutral words. Table 4.5 presents the top 10 positive and negative words in FISL. The French word "bien", which means "good", achieved the highest positive CSO-TFIDF score (i.e. Eq. 3.7), and repeated in 7,248 single documents and appeared with 15,345 positive seed words. As for the negative words, "guerre" (i.e. war) had the highest negative CSO-TFIDF value of -282.21.

Table 4.5 Top 10 positive words and top 10 negative words in the French integrated sentiment lexicon

No	Word	Translation	DC	NP	NN	N	TF-IDF (Eq 3.6)	CSO (Eq 3.3)	CSO-TFIDF (Eq 3.7)	SO
1	Bien	Good	7248	15345	8142	23487	1506.47	0.30668	462.0047	P
2	merci	thanks	1356	2945	982	3927	493.585	0.499873	246.7297	P
3	Paix	Peace	2814	5778	2941	8719	750.348	0.325381	244.1492	P
4	chance	luck	1030	1878	616	2494	405.442	0.506014	205.1595	P
5	Courage	Courage	1130	2191	824	3015	441.155	0.4534	200.0195	P
6	Vrai	True	1896	4003	2220	6223	584.159	0.286518	167.3719	P
7	meilleur	better	729	1513	547	2060	275.122	0.468932	129.0135	P
8	aime	love	802	1687	694	2381	293.631	0.417052	122.4593	P
9	beau	handsome	691	1459	584	2043	265.378	0.428292	113.6592	P
10	Fort	strong	928	2005	1033	3038	296.434	0.319947	94.84327	P
11	guerre	war	3262	3349	7657	11006	720.984	-0.39142	-282.21	N
12	honte	shame	1473	1171	3146	4317	411.488	-0.45749	-188.253	N
13	triste	sad	939	708	1892	2600	355.244	-0.45538	-161.773	N
14	Mal	wrong	1861	2130	4319	6449	468.61	-0.33943	-159.061	N
15	Tue	kill	1287	1243	3214	4457	341.078	-0.44223	-150.833	N
16	mort	death	1409	1571	3499	5070	389.458	-0.38028	-148.102	N
17	terroriste	terrorist	916	897	2386	3283	285.592	-0.45355	-129.53	N
18	pauvre	poor	908	956	2230	3186	316.105	-0.39987	-126.402	N
19	Fou	crazy	784	587	1536	2123	269.92	-0.44701	-120.657	N
20	Tuer	kill	922	1057	2464	3521	259.4	-0.3996	-103.657	N

SO: Semantic orientation (P: positive, N: Negative), NP: Number of nearby positive words, NN: Number of nearby negative words, N: Total of nearby polarity words NP+NN, DC: Number of documents that contain the candidate word.

4.3.3 Malay integrated sentiment lexicon (MISL)

In the Malay integrated sentiment lexicon (MISL), the top word was "baik" which means "good" or "fine". The MISL consisted of 9,931 polarity words. As in the Arabic and French integrated lexicons, the number of negative words were more than the positive words. Specifically, there were 6,215 negative words in MISL compared to only 2,676 positive words. Table 4.6 shows the top polarity words in the MISL.

Table 4.6 Top 10 positive words and top 10 negative words in the Malay integrated sentiment lexicon

No	Word	Translation	DC	NP	NN	N	TF-IDF (Eq 3.6)	CSO (Eq 3.3)	CSO-TFIDF (Eq 3.7)	SO
1	Baik	good	3479	13790	7666	21456	1607.03	0.285421	458.6806	P
2	Sembuh	Heal	1040	3271	825	4096	605.87	0.597168	361.8062	P
3	cepat	fast	1419	5044	2110	7154	831.248	0.41012	340.9116	P
4	Semoga	hopefully	2996	7579	4195	11774	1102.79	0.287413	316.9561	P
5	betul	right	2026	6114	3682	9796	909.342	0.248265	225.7574	P
6	Rasa	feel	2396	8968	5820	14788	969.22	0.212875	206.323	P
7	Naik	go up	1817	5871	3314	9185	700.995	0.278389	195.1491	P
8	Suka	love	2070	7121	4440	11561	810.701	0.2319	188.0018	P
9	Gaji	salary	1354	4582	2073	6655	456.56	0.37701	172.1276	P
10	hidup	alive	1683	7860	4957	12817	670.791	0.226496	151.9315	P
11	hilang	gone	1784	4270	6027	10297	636.143	-0.17063	-108.546	N
12	bodoh	Stupid	1146	2070	3405	5475	418.699	-0.24384	-102.094	N
13	takut	scared	1245	2671	4169	6840	449.367	-0.21901	-98.414	N
14	buang	throw	989	1706	3035	4741	320.964	-0.28032	-89.9728	N
15	Mati	dead	1370	4195	5671	9866	524.178	-0.1496	-78.4195	N
16	Lama	old	1830	5734	6852	12586	685.365	-0.08883	-60.8802	N
17	sampah	rubbish	522	835	1673	2508	170.251	-0.33413	-56.8861	N
18	salah	false	1565	5421	6468	11889	585.875	-0.08806	-51.5948	N
19	sakit	hurts	920	3184	4269	7453	310.99	-0.14558	-45.2736	N
20	gila	crazy	575	1157	1788	2945	178.322	-0.21426	-38.2075	N

SO: Semantic orientation (P: positive, N: Negative), NP: Number of nearby positive words, NN: Number of nearby negative words, N: Total of nearby polarity words NP+NN, DC: Number of documents that contain the candidate word.

4.4 Evaluation procedure

New datasets were created by requesting three native speakers of each language to label 500 to 1000 random reviews (i.e. positive versus negative), in order to evaluate the proposed framework. Each comment was labelled by two native speakers and the third speaker's label was only used when there was a conflict, as per Deng et al. (2017). Finally, 200 positive and 200 negative reviews for each language were selected from the labelled reviews.

The baseline lexicons used in the experiments are displayed in Table 4.7. The first two letters symbolize the language whereby 'Ar' stands for Arabic, 'Fr' for French and

'My' for Malay. The suffix 'L1' refers to the output of the lexicon-based layer (see Figure 3.2 in Chapter 3). For example, 'Fr_L1' refers to the French seed lexicon, which is a combination of the seed lexicons: Fr_MPQA, Fr_HL and Fr_AFINN. The lexicons with the suffix 'L2' are the output of the second layer (i.e. corpus-based layer), which was built by extracting new sentiment words from the unannotated corpus. Several publicly available lexicons were also used to compare against the generated lexicons. They are given as follows:

- *Ar_MPQA* is the translated copy of MPQA to Arabic.
- *Ar_OL* is the translated copy of Bing Liu's sentiment lexicon to Arabic.
- *Ar_AFINN* is the translated copy of AFINN to Arabic.
- *Ar_L1* is the output of the first layer of the proposed framework described in Figure 3.1, which is a combination of the three translated lexicons, namely, Ar_MPQA, Ar_OL and Ar_AFINN in the Arabic language.
- *Ar_L2* is the output of the second layer, built by extracting new sentiment words from the Arabic unannotated corpus.

Likewise, the abbreviation of 'Fr' and 'My' were used to denote the French and Malay lexicons, respectively.

- *AISL* is the Arabic integrated sentiment lexicon, a combination of Ar_L1 and Ar_L2.
- *FISL* is the French integrated sentiment lexicon, a combination of Fr_L1 and Fr_L2.
- *MISL* is the Malay integrated sentiment lexicon, a combination of My_L1 and My_L2.

- *Chen_Ar*[10]*:* an Arabic sentiment lexicon produced via graph propagation technique to create multi-language sentiment lexicons (Chen & Skiena, 2014),
- *AraSenTi*[11] *(Arabic)*: a large-scale Arabic sentiment lexicon generated from a large dataset for social network sentiment analysis (Al-Twairesh et al., 2016).
- *NileULex*[12] includes Modern Standard Arabic and Egyptian Arabic sentiment words and their sentiment orientation (El-Beltagy, 2016),
- *Chen_Fr*[13]*:* a French sentiment lexicon produced via graph propagation technique to create multi-language sentiment lexicons (Chen & Skiena, 2014),
- *FEEL*[14]*:* a French lexicon including around 14,000 various words representing sentiments and emotions, built by automatic translation and validated by a human expert (Abdaoui et al., 2016),
- *Chen_My:* a Malay sentiment lexicon produced via graph propagation technique to create multi-language sentiment lexicons (Chen & Skiena, 2014).

Figure 4.2 shows the sentiment lexicons based on their sizes. Appendix B shows the evaluation procedure code.

Table 4.7 Lists the numbers of negative and positive entries in examined lexicons

Lexicon	Building method	Positive	Negative	Neutral	Total
Ar_MPQA	Lexicon-based (Translation)	1637 (35%)	2716 (58%)	325 (7%)	4678
Ar_HL	Lexicon-based (Translation)	1388 (30%)	3168 (70%)	0 (0%)	4556

[10] https://www.kaggle.com/rtatman/sentiment-lexicons-for-81-languages

[11] https://github.com/nora-twairesh/AraSenti

[12] https://github.com/NileTMRG/NileULex

[13] https://www.kaggle.com/rtatman/sentiment-lexicons-for-81-languages

[14] http://www.lirmm.fr/~abdaoui/FEEL

Ar_AFINN	Lexicon-based (Translation)	806 (33%)	1619 (67%)	2 (0%)	2427
Ar_L1	Ar_MPQA+Ar_HL+Ar_AFINN	3626 (37%)	6046 (62%)	4 (0%)	9676
Ar_L2	Corpus-based	2265 (24%)	3493 (38%)	3529 (38%)	9287
AISL	Ar_L1+Ar_L2	5287 (31%)	8630 (51%)	3137 (18%)	17054
AraSenTi	Corpus-based	116448 (52%)	108881 (48%)	0 (0%)	225329
Chen_Ar	Graph-based	1652 (59%)	1142 (41%)	0 (0%)	2794
NileULex	Human-based	4672 (78%)	1281 (22%)	0 (0%)	5953
Fr_MPQA	Lexicon-based (Translation)	1691 (32%)	3184 (61%)	377 (7%)	5252
Fr_HL	Lexicon-based (Translation)	1513 (29%)	3627 (71%)	0 (0%)	5140
Fr_AFINN	Lexicon-based (Translation)	999 (36%)	1806 (64%)	2 (0%)	2807
Fr_L1	Fr_MPQA+Fr_HL+Fr_AFINN	2401 (33%)	4520 (62%)	355 (5%)	7276
Fr_L2	Corpus-based	1371 (36%)	2408 (64%)	0 (0%)	3779
FISL	Fr_L1+Fr_L2	2220 (34%)	4003 (61%)	349 (5%)	6572
Chen_Fr	Graph-based	1615 (35%)	3038 (65%)	0 (0%)	4653
FEEL	Lexicon-based + Human-based	8423 (60%)	5704 (40%)	0 (0%)	14127
My_MPQA	Lexicon-based (Translation)	1399 (35%)	2426 (60%)	219 (5%)	4044
My_HL	Lexicon-based (Translation)	1131 (28%)	2863 (72%)	0 (0%)	3995
My_AFINN	Lexicon-based (Translation)	795	1302	0	2097

		(38%)	(62%)	(0%)	
My_L1	My_MPQA+My_HL+My_AFINN	1897 (34%)	3397 (62%)	224 (4%)	5518
My_L2	Corpus-based	994 (20%)	3949 (80%)	0 (0%)	4941
MISL	My_L1+My_L2	2676 (27%)	7020 (71%)	235 (2%)	9931
Chen_My	Graph-based	1150 (39%)	1784 (61%)	0 (0%)	2934

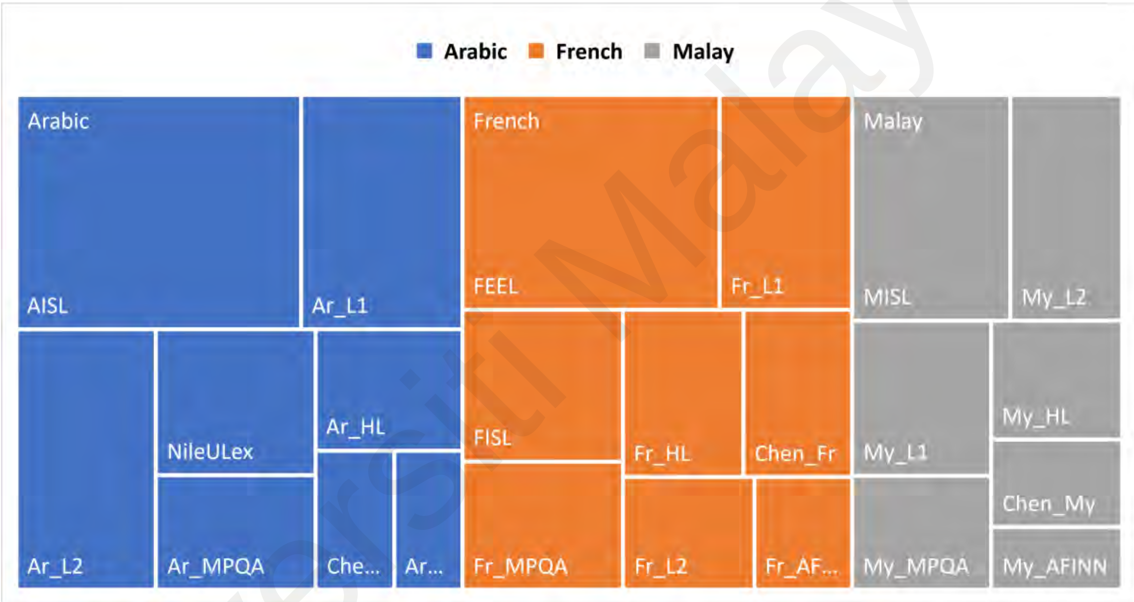


Figure 4.2 Representation of the evaluated sentiment lexicons by size (except AraSenTi)

4.5 Summary

In this chapter, the experiments, including the steps taken to build and evaluate the sentiment lexicons for Arabic, French and Malay languages, are presented along with the data collection and pre-processing stages. Finally, the evaluation procedures were applied to the extracted lexicons and compared. The evaluation results will be presented and explained in Chapter 5.

CHAPTER 5: RESULTS AND DISCUSSION

This chapter provides the results and discussion of the evaluation process of the semi-automatic integrated framework. The chapter has three sections. Section 5.1 presents the main sentiment classification results and discusses the performance of the integrated sentiment lexicons: AISL, FISL and MISL. Section 5.2 presents and discusses the classification results based on the class-level (i.e. positive and negative). Finally, Section 5.3 discusses the distribution of the building effort of lexicons.

5.1 Sentiment classifications

Table 5.1 presents the results for the performance of the integrated sentiment lexicons (i.e. AISL, FISL and MISL) which were developed using the proposed integrated framework. The results were compared with the baseline sentiment lexicons, as shown in Subsection 4.4 in Chapter 4. The results indicated that the proposed integrated sentiment lexicons showed a better performance than the other sentiment lexicons with regards to accuracy, recall, precision and F-measure values for all three languages. Results also showed that it was better to simultaneously use the three sentiment resources (i.e. seed lexicons, unannotated corpus and human) for developing high-quality sentiment lexicons. Hence, the better performances observed for these integrated lexicons was due to the coverage of polarity words extracted from the unannotated corpus.

Table 5.1 Sentiment classification results of the integrated lexicons compared to the baseline sentiment lexicons

Language		Lexicon	Accuracy	Precision	Recall	F-Measure
Arabic	**1.**	**AISL**	**0.780**	**0.786**	**0.778**	**0.778**
	2.	Ar_HL	0.487	0.488	0.487	0.481*
	3.	Ar_MPQA	0.458	0.469	0.470	0.457*
	4.	Ar_AFINN	0.600	0.601	0.594	0.591*
	5.	Ar_L1	0.634	0.640	0.627	0.622*
	6.	Ar_L2	0.707	0.706	0.708	0.706*
	7.	AraSenTi	0.707	0.726	0.706	0.700*
	8.	Chen_Ar	0.602	0.608	0.604	0.597*
	9.	NileULex	0.663	0.653	0.633	0.634*
French	**1.**	**FISL**	**0.864**	**0.876**	**0.8195**	**0.838**
	2.	Fr_HL	0.735	0.744	0.749	0.734*
	3.	Fr_MPQA	0.711	0.711	0.733	0.704*
	4.	Fr_AFINN	0.776	0.786	0.806	0.774*
	5.	Fr_L1	0.785	0.793	0.811	0.783
	6.	Fr_L2	0.825	0.816	0.758	0.777
	7.	Chen_Fr	0.736	0.711	0.732	0.715*
	8.	FEEL	0.451	0.589	0.560	0.438*
Malay	**1.**	**MISL**	**0.687**	**0.689**	**0.687**	**0.686**
	2.	My_HL	0.667	0.666	0.666	0.667*
	3.	My_MPQA	0.590	0.602	0.596	0.585*
	4.	My_AFINN	0.665	**0.689**	0.658	0.647
	5.	My_L1	0.662	0.682	0.664	0.654
	6.	My_L2	0.562	0.556	0.553	0.549*
	7.	Chen_My	0.663	0.670	0.671	0.662

- The first row of each language is the integrated sentiment lexicon.
- Highest results for each column and language are shown in bold.
- F-measure scores marked with * are statistically significantly different (with $\rho < 0.05$) from the corresponding F-measure in the row no.1 (i.e. integrated lexicon) for each language.

Results in Table 5.1 indicate that the proposed integrated sentiment lexicons showed better performances than other sentiment lexicons with regards to their accuracy. Accuracy values of 0.780, 0.864 and 0.687 were seen in the case of the Arabic, French and Malay lexicons, respectively. These were followed by the Ar_L2, Fr_L2 and My_HL lexicons, which produced accuracy values of 0.707, 0.825 and 0.667, respectively. Similarly, the AISL, FISL and MISL lexicons had better F-measure values than other sentiment lexicons, i.e., 0.778, 0.838 and 0.686, respectively.

Two-tailed t-test was conducted to test the statistical significance of the F-measure score between the integrated lexicons (i.e. AISL, FISL and MISL) and the baseline lexicons in the three languages (i.e. Arabic, French and Malay). In Table 5.1, the F-measure scores of the integrated lexicons (i.e. row no. 1 for each language) were compared with the F-measure scores of the baseline lexicons. F-measure scores marked with (*) for each baseline lexicon are statistically significantly different (with $\rho < 0.05$) from the corresponding F-measure of the integrated lexicon for each language. T-test results show that the AISL is statistically significantly different from the entire Arabic baseline lexicon in term of F-measure. In the French language, the FISL is statistically significantly different from all the French baseline lexicons except with the lexicons Fr_L1 and Fr_L2. In the Malay language, the difference between MISL and the Malay baseline is not statistically significant for some baseline lexicons, namely, My_AFINN, My_L1 and Chen_My.

The internet users make use of several dialects, colloquial and abbreviations when they interact on social media (Fersini et al., 2016; Itani et al., 2017). Hence, the lexicons that were developed by extracting the polarity words from a corpus (i.e. Ar_L2, Fr_L2 and My_L2) are expected to show a better performance compared to the seed lexicons (i.e. Ar_L1, Fr_L1 and My_L1). This is also attributed to the fact that sentiment lexicons extracted from a corpus are domain-dependent; whereas others such as AraSenTi and FEEL are general-domain lexicons (Park et al., 2015; Wu et al., 2017). Table 5.1 presented the results of the corpus-based lexicons (i.e. Ar_L2 and Fr_L2) which performed better than the seed lexicons (i.e. Ar_L1 and Fr_L1) for the Arabic and French languages. On the other hand, My_L2, a Malay corpus-based lexicon, did not show a good performance in comparison to the Malay seed lexicon (i.e. My_L1) owing to a lack of efficient pre-processing tools, as shown in Subsection 5.2.3.

In the case of the seed lexicons, it was seen that a translated version of AFINN sentiment lexicon (Nielsen, 2011) showed better results compared to other translated lexicons. This is attributed to the fact that AFINN included common terms used on different social media sites like internet slang words, acronyms like LOL (Laughing Out Loud), bad words, etc., hence, many words could be recognised. This indicated that when the accuracy of the seed lexicons was increased, better results could be seen.

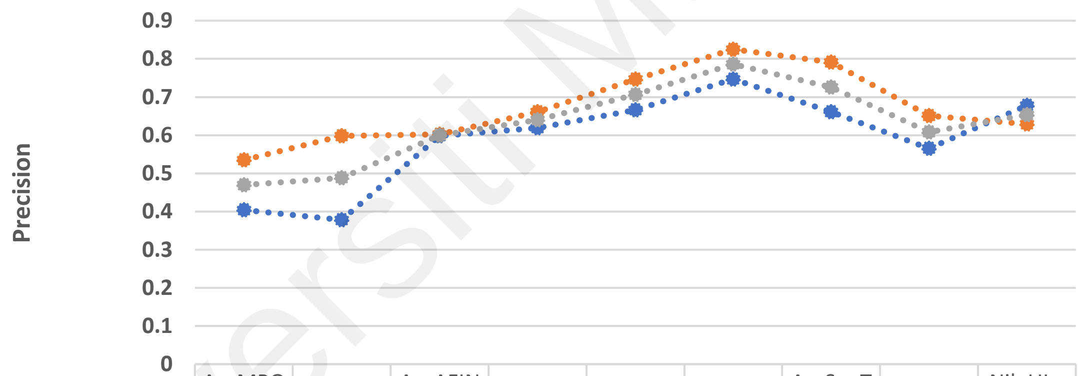
For understanding the reason behind the low-performance results of the translated lexicons, a set of words were examined. A few of the translated words were generic or were affected by cultural differences. For instance, an English word, i.e., “craftily” was regarded as negative in the Opinion Lexicon described by Hu and Liu (2004), however, when it was translated in Arabic, i.e., "للبراعة, it was seen to be a positive adjective.

5.2 Class-level sentiment classification results

This section presents and discusses the sentiment classification results based on the class-level (i.e. positive and negative). The class-level results for each evaluated language have been presented in the following subsections.

5.2.1 Evaluation results for the Arabic lexicons

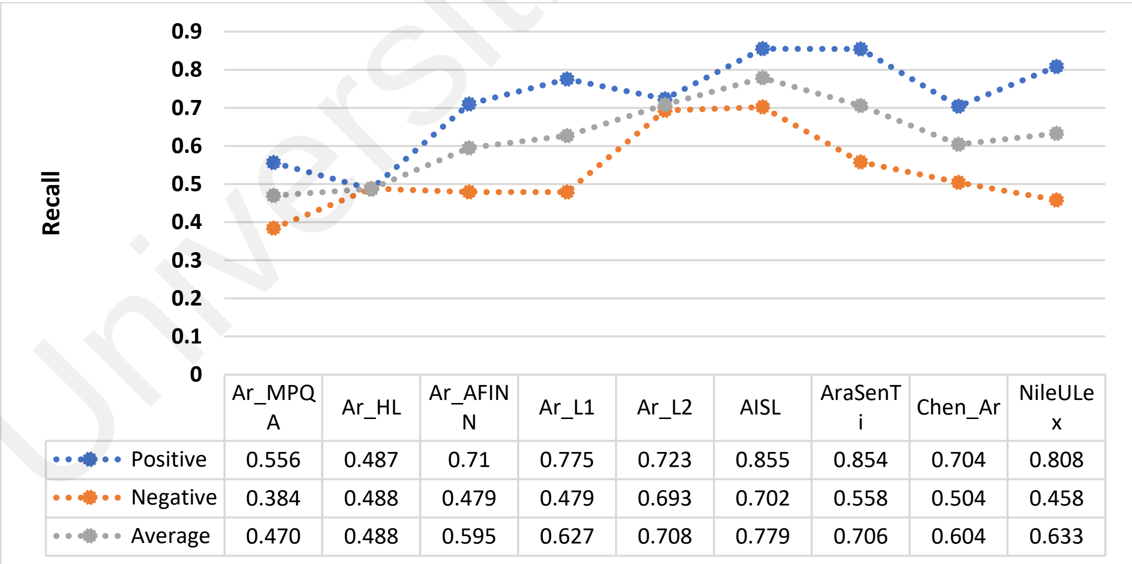
Figure 5.1 shows the values of precision in the negative class to be better than the those in the positive class across the Arabic lexicons. The precision values for the negative and positive classes in the AISL lexicon were seen to be 0.825 and 0.747, respectively. As depicted in Table 4.6 (Chapter 4), more negative words were noted in the lexicons compared to the positive words, and thus probably resulting in a higher precision, for the negative class. This result was somewhat expected due to the type of corpus studied, which was acquired from the news pages containing political disagreements and disputes (Koç et al., 2018).



	Ar_MPQA	Ar_HL	Ar_AFINN	Ar_L1	Ar_L2	AISL	AraSenTi	Chen_Ar	NileULex
Positive	0.404	0.378	0.598	0.619	0.666	0.747	0.661	0.565	0.678
Negative	0.535	0.598	0.602	0.661	0.747	0.825	0.791	0.651	0.629
Average	0.470	0.488	0.600	0.640	0.707	0.786	0.726	0.608	0.654

Figure 5.1 Precision results of the Arabic lexicons

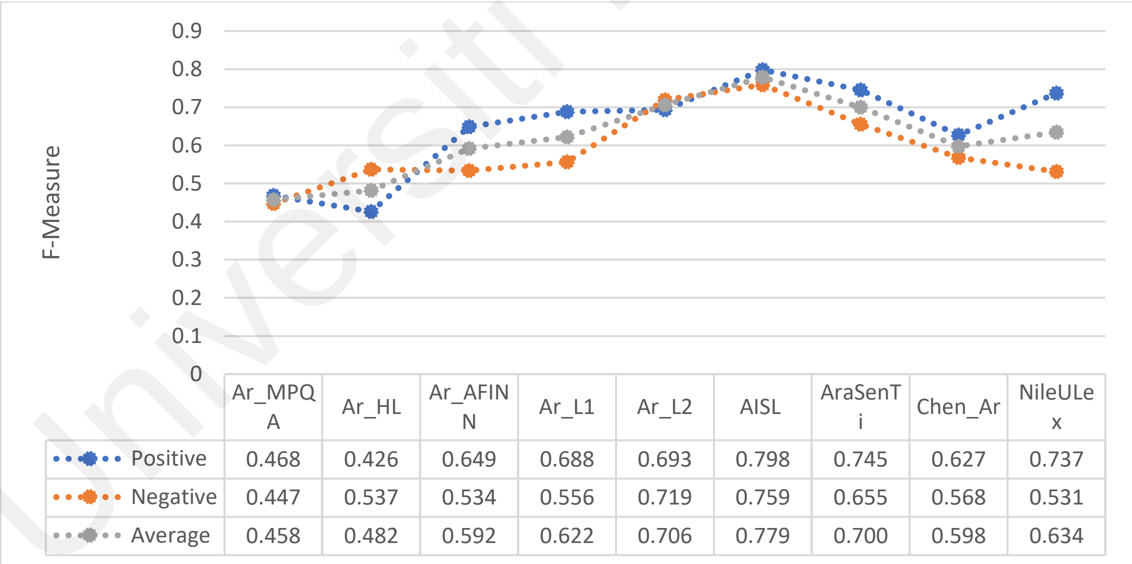
Figure 5.2 presents the recall results, where a marked difference can be noted between the positive and negative classes for all the tested Arabic lexicons, except for the Ar_HL and Ar_L2 lexicons. The seed lexicons showed a lower recall value for the negative class since they did not include polarity words used by the people on social media sites or news domains, that are often characterised by bickering or controversy. As addressed by Liu et al. (2015), the negative news articles received much more comments than positive news articles because the double controversy could be a discussion material for internet users. For instance, the Ar_L1 lexicon showed a good recall value for the positive class; however, it showed an unsatisfactory recall value for the negative class. However, the Ar_L2 lexicon showed close results for the positive and negative classes, that is, 0.723 and 0.693, respectively. Hence, when Ar_L1 and Ar_L2 lexicons were combined to generate the integrated AISL lexicon in this study, an improvement in results was observed as shown in Figure 5.2.



	Ar_MPQA	Ar_HL	Ar_AFINN	Ar_L1	Ar_L2	AISL	AraSenTi	Chen_Ar	NileULex
Positive	0.556	0.487	0.71	0.775	0.723	0.855	0.854	0.704	0.808
Negative	0.384	0.488	0.479	0.479	0.693	0.702	0.558	0.504	0.458
Average	0.470	0.488	0.595	0.627	0.708	0.779	0.706	0.604	0.633

Figure 5.2 Recall results of the Arabic lexicons

Figure 5.3 presents the values for the F-measure for the Arabic lexicons. Again, the AISL lexicon showed best performance for the positive and negative classes, followed by AraSenTi (0.745) and NileULex (0.737) in the positive class. It is interesting to note that although AraSenTi is a very large lexicon containing 225,329 polarity words, it did not always show a better performance compared to the smaller lexicons. These results are in accordance with (Al-Thubaity et al., 2018), whereby the SauDiSenti lexicon (4431 words and phrases) outperformed AraSenTi in terms of F-measure. This shows that the lexicon size is not always useful. Hussein (2016) and Feng et al. (2012) stated that a huge lexicon size could create issues such as slow processing and containing more noise. Finally, Ar_L1 lexicon which is combined three different seed lexicons (i.e. Ar_MPQA, Ar_HL and Ar_AFINN) showed a better result compared to the results of each seed lexicon, individually.

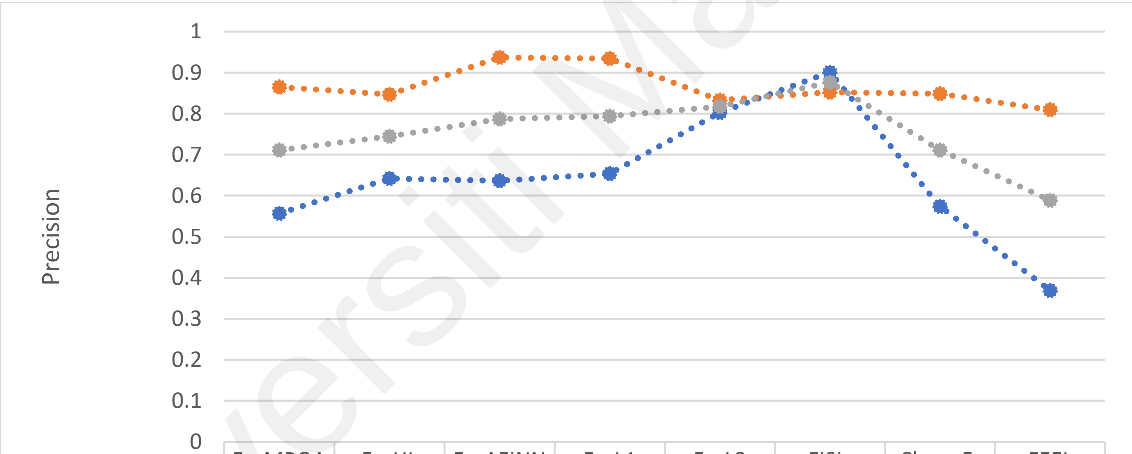


	Ar_MPQA	Ar_HL	Ar_AFINN	Ar_L1	Ar_L2	AISL	AraSenTi	Chen_Ar	NileULex
Positive	0.468	0.426	0.649	0.688	0.693	0.798	0.745	0.627	0.737
Negative	0.447	0.537	0.534	0.556	0.719	0.759	0.655	0.568	0.531
Average	0.458	0.482	0.592	0.622	0.706	0.779	0.700	0.598	0.634

Figure 5.3 F-Measure results of the Arabic lexicons

5.2.2 Evaluation results for the French lexicons

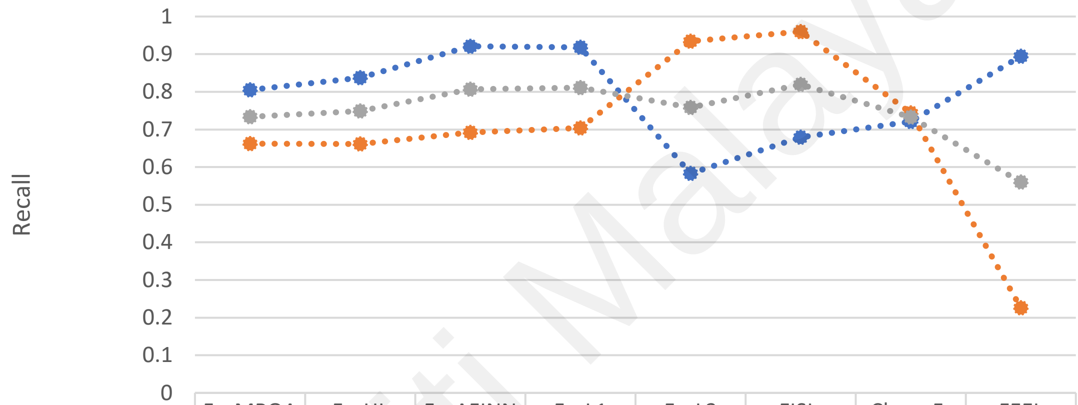
Figure 5.4 presents the precision values for the French lexicons, with the integrated. FISL showing the best results for the positive class. The high precision value in the positive class means that the classifier trained by FISL returned substantially more relevant results (i.e. true positive TP) than irrelevant ones (i.e. false positive FP). Similar to the Arabic lexicons, the French seed lexicons (i.e. Fr_MPQA, Fr_HL, Fr_AFINN and Fr_L1) also recorded a higher precision in their negative class in comparison to their positive class which recorded a lower precision compared to FISL. The lowest precision value was displayed by the FEEL French lexicon in the positive class, i.e. 0.369, however, it showed an acceptable precision value in the negative class. On average, FISL lexicon outperformed all the seed lexicons, indicating the advantage of using high-quality coverage of sentiment words from various resources (i.e. seed lexicons, corpus and human).



	Fr_MPQA	Fr_HL	Fr_AFINN	Fr_L1	Fr_L2	FISL	Chen_Fr	FEEL
Positive	0.557	0.642	0.636	0.653	0.802	0.901	0.574	0.369
Negative	0.865	0.847	0.937	0.934	0.833	0.852	0.848	0.809
Average	0.711	0.745	0.787	0.794	0.818	0.877	0.711	0.589

Figure 5.4 Precision results of the French lexicons

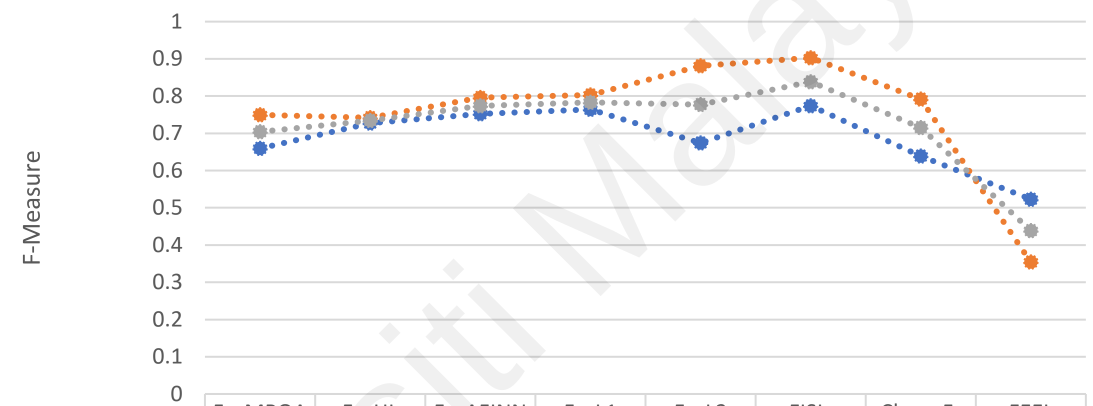
Figure 5.5 presents the recall values for the French lexicons. The seed lexicons Fr_AFINN and Fr_L1 recorded the highest recall values for the positive class while Fr_L2 and FISL produced the highest recall values for the negative class. It is to note that high recall values for the negative class were achieved by the lexicons Fr_L1 and Fr_L2 which were 0.704 and 0.934, respectively. The combination between the first- and second-layer lexicons (i.e. FISL) showed an improvement on average comparing with Fr_L1 and Fr_L2.



	Fr_MPQA	Fr_HL	Fr_AFINN	Fr_L1	Fr_L2	FISL	Chen_Fr	FEEL
Positive	0.805	0.837	0.921	0.918	0.583	0.679	0.721	0.894
Negative	0.662	0.661	0.692	0.704	0.934	0.96	0.744	0.226
Average	0.734	0.749	0.807	0.811	0.759	0.820	0.733	0.560

Figure 5.5 Recall results of the French lexicons

Figure 5.6 shows that the integrated French lexicon FISL was found to outperform the rest of the lexicons in terms of F-Measure in both classes, that is, 0.774 and 0.903 for the positive and negative class, respectively. It is clearly observed that integrating between the polarity words arising from the corpus-based layer (i.e. Fr_L2) and lexicon-based layer lexicon (i.e. Fr_L1) increased the coverage of the integrated lexicon FISL as shown in Figure 5.6. However, the FISL needs to be improved considering the class imbalance issue (2,220 positive words versus 4,003 negative words) (see Table 4.7 in Chapter 4).



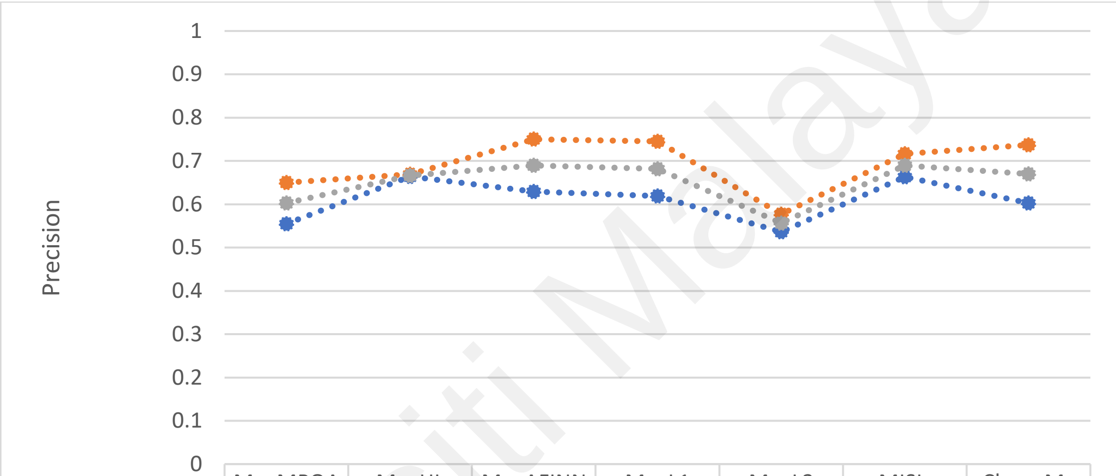
	Fr_MPQA	Fr_HL	Fr_AFINN	Fr_L1	Fr_L2	FISL	Chen_Fr	FEEL
Positive	0.659	0.727	0.752	0.764	0.674	0.774	0.639	0.523
Negative	0.75	0.742	0.796	0.803	0.881	0.903	0.792	0.354
Average	0.705	0.735	0.774	0.784	0.778	0.839	0.716	0.439

Figure 5.6 F-Measure results of the French lexicons

5.2.3 Evaluation results for the Malay lexicons

Figure 5.7 presents the precision values for the Malay lexicons. Based on the data presented in the figure, it can be noted that the Malay lexicons showed the lowest precision value in the positive and negative classes compared to the Arabic and French lexicons. This is probably attributed to a lack of pre-processing tools in the Malay language as discussed at the end of this subsection.

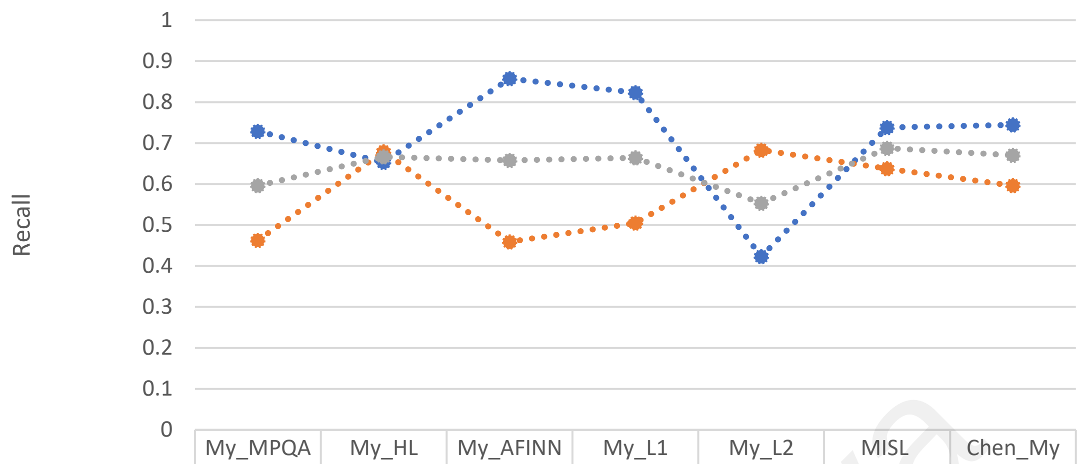
In Figure 5.7, My_HL recorded a higher precision value for the positive class (0.664) followed by MISL with a slight difference of 0.001, while My_AFINN recorded a higher precision value for the negative class (0.750) followed by My_L1 (0.745), Chen_My (0.737), then MISL (0.716). Nevertheless, on average, the precision value showed an improvement when the My_L2 lexicon was combined with the seed lexicon My_L1 to generate the integrated MISL lexicon, i.e. 0.690. This is the highest precision value recorded on average by MISL and My_AFINN lexicons. It is worth noting that the translated copies of AFINN lexicon also showed good results in some languages such as in the Norwegian language as addressed by Hammer et al. (2014).



	My_MPQA	My_HL	My_AFINN	My_L1	My_L2	MISL	Chen_My
Positive	0.555	0.664	0.629	0.619	0.536	0.663	0.603
Negative	0.65	0.669	0.75	0.745	0.577	0.716	0.737
Average	0.603	0.667	0.690	0.682	0.557	0.690	0.67

Figure 5.7 Precision results of the Malay lexicons

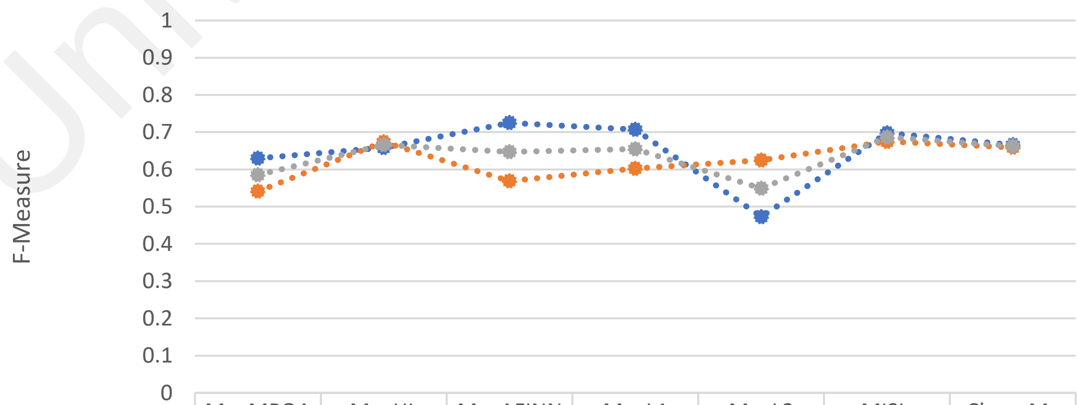
Based on the recall values shown in Figure 5.8, it can be observed that the My_AFINN and My_L1 lexicons showed better results in the positive class compared to their negative class where they achieved quite low results. On the other hand, the corpus-based lexicon My_L2 achieved 0.683 which was the highest recall values in the negative class. However, My_L2 achieved the lowest recall values, i.e. 0.423, in the positive class which means My_L2 was lacking a lot of positive words. The decrease in positive words was reinforced by increasing lexicon coverage through integration with seed lexicon My_L1 to generate MISL. Finally, on average, the integrated lexicon MISL showed the highest recall value achieved of 0.6875.



	My_MPQA	My_HL	My_AFINN	My_L1	My_L2	MISL	Chen_My
Positive	0.729	0.653	0.857	0.823	0.423	0.738	0.744
Negative	0.463	0.679	0.459	0.505	0.683	0.637	0.596
Average	0.596	0.666	0.658	0.664	0.553	0.6875	0.67

Figure 5.8 Recall results of the Malay lexicons

Finally, Figure 5.9 presents the F-measure scores for the Malay lexicons. For the negative class, My_L1 and My_L2 had F-measure values of 0.602 and 0.625, respectively, whereas MISL produced a higher value of 0.675. This is probably due to the combination of My_L1 and My_L2. Similar to Ar_AFINN and Fr_AFINN, My_AFINN achieved a higher F-measure value in the positive class compared to other seed lexicons (i.e. My_MPQA and My_HL).



	My_MPQA	My_HL	My_AFINN	My_L1	My_L2	MISL	Chen_My
Positive	0.63	0.658	0.725	0.707	0.473	0.698	0.666
Negative	0.541	0.674	0.569	0.602	0.625	0.675	0.659
Average	0.586	0.666	0.647	0.655	0.549	0.687	0.663

Figure 5.9 F-Measure results of the Malay lexicons

The results indicated that the MISL lexicon showed better performance on average compared to the other lexicons for the Malay language, however, it was not as good as the AISL and the FISL lexicons. This is probably due to the fact that the corpus that was used in the Malay lexicons required a lot of normalising and cleaning. The Malaysian users tend to use abbreviations and write in different languages beside Malay, including English, Chinese and Tamil or combination of several mainstream languages (Gill, 2013). For instance, the social media users use the number 2 for indicating plurality (Example: “buku2” means “books”) rather than repeating the same word two times as in the general Malay language (e.g., buku-buku). This fact highlights the significance of using non-English pre-processing tools for processing the corpus before extracting the polarity words. Hence, it is very important to use pre-processing tools, especially for the Malay language, since it could help in the extraction of precise and appropriate polarity words from within the corpus (Saad, 2010).

To sum up, the results indicated that the translated seed lexicons consist mostly of generally used polarity words that are applicable across domains, in line with Deng et al. (2017). It leads to the obvious need for adding special language expressions and domain-specific polarity words extracted from the target language corpus to improve the coverage of these lexicons (Bravo-Marquez et al., 2016). Moreover, the results indicate that the seed lexicon is crucial in polarity words extraction (Deng et al., 2017). Consistently with the finding of Maks and Vossen (2011), good language-dependent and domain-specific sentiment words can be extracted from the corpus by using a high-quality seed lexicon.

The results support the proposed methodology in diversifying the resources of extracting the polarity words. Thus, in this work, non-English sentiment lexicons were built from three resources (i.e. seed lexicons, unannotated corpus and human), where each resource added more valuable polarity words to the integrated lexicons. Finally, the experiments conducted on the three languages (i.e. Arabic, French and Malay) indicated that the integrated lexicons built by the proposed framework showed a better performance than the existing lexicons in terms of accuracy and the average of Precision, Recall and F-Measure.

5.3 Distribution of efforts needed for building a lexicon

The results from this study show that many polarity words can be extracted from the unannotated corpus without acquiring knowledge regarding the target language and language-specific information. This extraction process is deemed to be semi-automatic as it used human efforts only for reviewing the lexicons (i.e. Layer 3). To be specific, the semi-automatic integrated framework and the contributions based on each layer are as follows:

- ~ 50% of all polarity words in the integrated lexicons are seed lexicon entries, which were translated in the first layer (lexicon-based layer). These polarity words consisted essentially of general domain words with lack of the language-dependent and domain-based polarity words.
- ~ 40% of all polarity words were extracted from Layer 2 (i.e. a corpus-based layer), wherein the polarity words were extracted from an unannotated corpus. These polarity words included a lot of dialects, informal and slang words that use in social media.
- ~10% of the polarity words are manually reviewed in the human-based layer.

These results showed that the framework was 90% automated. It also indicated that the human effort required has decreased for building the lexicon to only 10%. Furthermore, the human efforts were limited to only reviewing the pre-defined words which needed to be reviewed, as mentioned in Subsection 3.1.3 in Chapter 3. Figure 5.10 shows the distribution of efforts needed for building a lexicon.

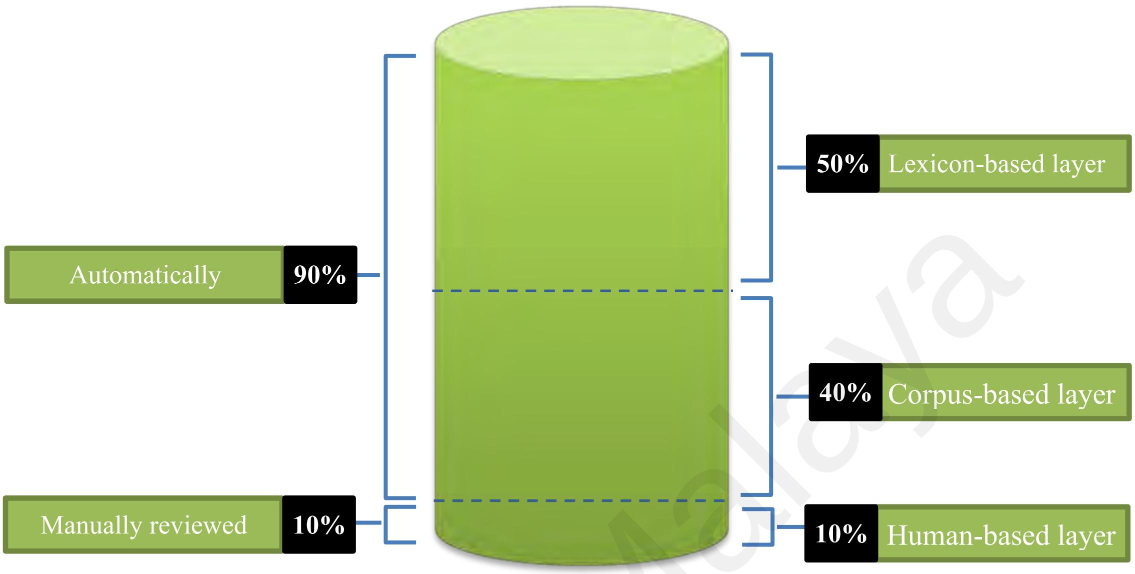


Figure 5.10 the distribution of efforts needed for building a lexicon

5.4 Summary

This chapter presents the analysis and discussion of the evaluation results for the integrated framework according to metrics identified in Chapter 3 (Subsection 3.2), namely, accuracy, precision, recall and F-measure. The experiments conducted on the three languages indicated that the integrated lexicons built using the proposed framework showed an overall F-measure value of 0.778, 0.838 and 0.686, for the Arabic, French and the Malay lexicons, respectively, outperforming the existing lexicons. The integrated lexicon consisted of ~50% seed words, ~40% novel polarity words extracted from a corpus and ~10% additional or human-reviewed words.

CHAPTER 6: CONCLUSION, LIMITATIONS AND FUTURE DIRECTION

This chapter concludes the research and discusses the research contributions, significance and limitations. Moreover, it presents possible future directions for this work.

6.1 Conclusion

A semi-automatic integrated framework was built to generate sentiment lexicons for non-English languages. Many of the shortcomings and limitations reported in previous studies were addressed in the current study through the use of available resources and minimization of human effort in data labelling (Song et al., 2019).

The integrated framework was used to build non-English sentiment lexicons based on available English lexicons with an unannotated corpus from the target language. The framework consists of three layers, namely, lexicon-based, corpus-based and human-based. The process of building the lexicon begins with the use of three current English lexicons, namely, MPQA (Wilson et al., 2005b), Opinion Lexicon (Hu & Liu, 2004), and AFINN (Nielsen, 2011) in the lexicon-based layer to obtain good coverage of the words in the target lexicon. However, the lexicon formed in this layer suffered from several limitations, namely, that it contained generic words and lacked the dialects and informal or slang words that are frequently used by the internet and social media users.

This limitation was addressed in the second layer (i.e. corpus-based layer), where an unannotated corpus was utilized to add new polarity words and to correct the polarity of the resulting words from the first layer. In the second layer, one of the most important limitations of the use of a corpus, namely, the need to annotate the corpus manually, was also addressed (Sun et al., 2017).

In this work, an unannotated corpus was used; therefore, human effort was not required in building the lexicon. Human experts only reviewed the extracted lexicon in the human-based layer. It is to note that the manual effort was greatly reduced, as it was only necessary to revise the words needed to be reviewed.

The proposed integrated framework was evaluated using posts collected from news media pages on Facebook for three languages, namely, Arabic, French and Malay. The evaluation results showed that the new lexicons that built using the framework produced better outcomes than existing sentiment lexicons such as AraSenTi (Al-Twairesh et al., 2016) in Arabic and FEEL (Abdaoui et al., 2016) in French.

Finally, this research makes several contributions. First, the development of a new taxonomy to classify existing studies based on the resources used to build the lexicons. Second, the development of an integrated framework to generate and adapt sentiment lexicons for non-English languages, incorporating three available resources namely, opinion lexicon (Hu & Liu, 2004), MPQA (Wilson et al., 2005b) and AFINN (Nielsen, 2011). Then, the development of an automated method to recognize new polarity words in the unannotated corpus. Finally, the construction of three independent sentiment lexicons in Arabic, French and Malay that can be made available and thus useful for future studies in similar areas of interests.

6.2 Research objectives and questions revisited

This section recapitulates the research objectives and questions mentioned in Section 1.6 and presents the tasks that were undertaken to achieve them.

Objective 1: To examine the current methods and languages used to build sentiment lexicons for non-English languages.

- **Research Question 1:** What are the existing methods for building sentiment lexicons for non-English languages?

  An extensive study was conducted to examine existing research on the building of sentiment lexicons, and to classify the methods with respect to non-English datasets. Additionally, the research also reviewed the tools used to build sentiment lexicons for non-English languages, ranging from those using machine translation to graph-based methods. Methods for building SLs vary from being completely manual, semi-automatic, to limited automatic approaches. The approaches are divided and used to construct SLs according to the type of source used. Accordingly, there are three sources employed to build SLs; pre-existing lexicons, target language corpus, and target language native speakers.

- **Research Question 2:** What are the limitations of the current methods?

  Shortcomings were highlighted with the methods along with recommendations to improve the performance of each method. Several limitations were identified for the lexicon-based methods, such as limited use for the general domain and inability handle different dialects and informal or slang words. Moreover, they don't contain acronyms and shorthand. On the other hand, corpus-based methods to build non-English sentiment lexicons had their limitations. For example, they need data pre-processing tools to prepare the corpus. Besides, they require a large corpus volume to achieve an acceptable accuracy and some lexicon building methods depend on an annotated corpus. Finally, sentiment lexicons built by humans are usually more accurate than others; however, the production of these lexicons is time-consuming, requires a large number of people and is costly.

Objective 2: To develop a semi-automatic integrated framework to build sentiment lexicons for non-English languages**.**

- **Research Question 3:** What are the components of building a semi-automatic integrated framework for non-English languages?

  To achieve this objective, a novel semi-automatic integrated framework was established to develop and adapt sentiment lexicons for non-English languages that incorporate three available resources (that is, seed lexicons, unannotated corpus and humans).

- **Research Question 4:** How to build sentiment lexicons for non-English languages from unannotated datasets?

  The corpus-based layer in the integrated framework was developed to automatically discover new polarity words from unlabelled datasets. The corpus-based layer relied on an unannotated corpus of the target language and seed sentiment lexicons. The seed sentiment lexicons were utilized to specify new sentiment words in the target language corpus depending on the relationship between the seeds and the candidate word, which was determined by the application of Equation 3.7 in Section 3.1.

Objective 3: To evaluate the proposed semi-automatic integrated framework by conducting experiments and evaluations.

- **Research Question 5:** How can the proposed semi-automatic integrated framework be compared with existing method(s)?

  Experiments were conducted to thoroughly evaluate the performance of the proposed semi-automatic integrated framework. In the evaluation, the performance of the lexicons built by the proposed framework was compared to that of other current lexicons.

- **Research Question 6:** What metrics can be used to evaluate the proposed semi-automatic integrated framework?

The lexicon scoring method was adopted for classification purposes, as stated in Section 3.2. Moreover, a confusion matrix was used with four indices, namely, accuracy (A), precision (P), recall (R) and F-measure (F) to assess the performance of the proposed framework. Experiments on three languages (i.e. Arabic, Malay and French) showed that the proposed framework outperformed the existing lexicons.

6.3 Research Limitations

In dealing with non-English languages, this study was further presented with a number of difficulties and limitations such as the limited size of resources or their availability to the public (Abdullah & Hadzikadic, 2017). These are specifically listed as below:

- The lack of pre-processing tools for lemmatisation and tokenisation for some languages continues to be a concern (Uysal & Gunal, 2014). Since no pre-processing tools were publicly available to process the Malay corpus, the building of Malay sentiment lexicon was negatively affected. In addition, the performance of the proposed framework may also be affected by the nature of the language as social media users frequently write in multiple dialects, incurring numerous spelling and typographical errors (Saad, 2010).
- Considering an automatic translation was used in the first layer to create the seed lexicon, some translation errors may appear as mentioned in Subsection 2.3.1.4. This shortcoming can be overcome by having high-quality seed words selected manually and then expanding them automatically (Ekinci & Omurca, 2019).

- It is necessary to have a large corpus to achieve an acceptable level of accuracy (Rashed & Abdolvand, 2017; Xing et al., 2019), while some non-English languages suffer from a lack of these resources (Abdullah & Hadzikadic, 2017; Dashtipour et al., 2016). The large corpus help in extracting a variety of high-quality polarity words by computing their occurrence inside the corpus (Feng et al., 2015a).

6.4 Future directions

This study can be extended for improving the results noted for the sentiment lexicons, using the following future directions:

- Finding and using an appropriate number of good seed words will further improve the quality of the extracted polarity words (Chao & Yang, 2018; Deng et al., 2017). Constructing a sentiment lexicon beginning from a small seed word set is time-consuming (Ekinci & Omurca, 2019). Consequently, future studies could devise appropriate method(s) to enhance the quality of the seed words before they are used to build new lexicons. For example, seed words can be manually selected and then expanded automatically by their identifying synonyms and antonyms (Huang et al., 2014).
- Many features, like negation, hashtags and emoticons can enrich the lexicon content (Mukhtar et al., 2018; Zhao et al., 2018). This needs to be investigated further, in addition to the effect of the bigrams and the trigrams (Alwakid et al., 2017; Deng et al., 2017), as the present study only focused on unigrams. For instance, the word "pain" is negative and the word "free" is a positive word, but the bigrams of the two words is "pain free" which is deemed positive.

- This study can be continued further for improving the results of sentiment analysis by making use of different approaches like Machine learning. A few studies used the lexicon-based along with Machine learning approaches for improving the accuracy of the sentiment classification process, wherein they used the sentiment lexicon words as the features in some machine learning models (Al-Moslmi et al., 2018; Dehkharghani et al., 2012; Kang et al., 2012). Hence, in future, the studies need to explore the possibility of using the machine learning classification algorithms like Random Forest and Support Vector Machine, though these process may not be straightforward for non-English languages (Al-Moslmi et al., 2018).
- Lastly, the semi-automated integrated framework, described in the study, was assessed using three different languages. i.e., Arabic, French and Malay. There are several emerging languages which need to be investigated, especially in the sentiment analysis field, like Chinese (Zhao et al., 2018) and Hindi (Jha et al., 2015). Therefore, other researchers could replicate the proposed methodology to examine the outcomes in these languages.

REFERENCES

Abdaoui, A., Azé, J., Bringay, S., & Poncelet, P. (2016). FEEL: a French Expanded Emotion Lexicon [Article in Press]. *Language Resources and Evaluation*, 1-23.

Abdelali, A., Darwish, K., Durrani, N., & Mubarak, H. (2016). Farasa: A Fast and Furious Segmenter for Arabic. HLT-NAACL Demos,

Abdul-Mageed, M. (2017). Modeling Arabic subjectivity and sentiment in lexical space. *Information Processing & Management, 56*(2), 291-307.

Abdul-Mageed, M., Diab, M., & Kübler, S. (2014, 1//). SAMAR: Subjectivity and sentiment analysis for Arabic social media. *Computer Speech & Language, 28*(1), 20-37.

Abdul-Mageed, M., & Diab, M. T. (2011). Subjectivity and sentiment annotation of modern standard arabic newswire. Proceedings of the 5th Linguistic Annotation Workshop,

Abdul-Mageed, M., & Diab, M. T. (2014). SANA: A Large Scale Multi-Genre, Multi-Dialect Lexicon for Arabic Subjectivity and Sentiment Analysis. LREC,

Abdullah, M., & Hadzikadic, M. (2017). Sentiment Analysis on Arabic Tweets: Challenges to Dissecting the Language. International Conference on Social Computing and Social Media,

Ahire, S. (2014). A Survey of Sentiment Lexicons. *Computer Science and Engineering IIT Bombay*.

Akhtar, M. S., Gupta, D., Ekbal, A., & Bhattacharyya, P. (2017, 6/1/). Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis. *Knowledge-Based Systems, 125*, 116-135.

Al-Kabi, M. N., Kazakzeh, S. A., Ata, B. M. A., Al-Rababah, S. A., & Alsmadi, I. M. (2015). A novel root based Arabic stemmer. *Journal of King Saud University-Computer and Information Sciences, 27*(2), 94-103.

Al-Moslmi, T., Albared, M., Al-Shabi, A., Omar, N., & Abdullah, S. (2018). Arabic senti-lexicon: Constructing publicly available language resources for Arabic sentiment analysis. *Journal of Information Science, 44*(3), 345-362.

Al-Shammari, E. T. (2009). A novel algorithm for normalizing noisy Arabic text. 2009 WRI World Congress on Computer Science and Information Engineering,

Al-Sharuee, M. T., Liu, F., & Pratama, M. J. a. p. a. (2017). An automatic contextual analysis and clustering classifiers ensemble approach to sentiment analysis.

Al-Subaihin, A. A., Al-Khalifa, H. S., & Al-Salman, A. S. (2011). A proposed sentiment analysis tool for modern arabic using human-based computing. Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services,

Al-Thubaity, A., Alqahtani, Q., & Aljandal, A. (2018, 2018/01/01/). Sentiment lexicon for sentiment analysis of Saudi dialect tweets. *Procedia Computer Science, 142*, 301-307.

Al-Twairesh, N., Al-Khalifa, H., & Al-Salman, A. (2016). AraSenTi: Large-Scale Twitter-Specific Arabic Sentiment Lexicons. *Association for Computational Linguistics*, pages 697–705.

al Owisheq, A., al Humoud, S., al Twairesh, N., & al Buhairi, T. (2016). *Arabic sentiment analysis resources: A survey* [Conference Paper]. 8th International Conference on Social Computing and Social Media, SCSM 2016 and 18th International Conference on Human-Computer Interaction, HCI International 2016,

Aldayel, H. K., & Azmi, A. M. (2016). Arabic tweets sentiment analysis–a hybrid scheme. *Journal of Information Science, 42*(6), 782-797.

Alwakid, G., Osman, T., & Hughes-Roberts, T. (2017). Challenges in Sentiment Analysis for Arabic Social Networks. *Procedia Computer Science, 117*, 89-100.

Amir, S., Astudillo, R., Ling, W., Martins, B., Silva, M. J., & Trancoso, I. (2015). Inesc-id: A regression model for large scale twitter sentiment lexicon induction. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015),

Amiri, F., Scerri, S., & Khodashahi, M. H. (2015). Lexicon-based sentiment analysis for Persian text. 10th International Conference on Recent Advances in Natural Language Processing, RANLP 2015,

Araújo, M., Pereira, A., & Benevenuto, F. (2020, 2020/02/01/). A comparative study of machine translation for multilingual sentence-level sentiment analysis. *Information Sciences, 512*, 1078-1102.

Araujo, M., Reis, J., Pereira, A., & Benevenuto, F. (2016). An evaluation of machine translation for multilingual sentence-level sentiment analysis. Proceedings of the 31st Annual ACM Symposium on Applied Computing,

Assiri, A., Emam, A., & Al-Dossari, H. (2018). Towards enhancement of a lexicon-based approach for Saudi dialect sentiment analysis. *Journal of Information Science, 44*(2), 184-202.

Augustyniak, L., Szymánski, P., Kajdanowicz, T., & Tuliglowicz, W. (2016). Comprehensive study on lexicon-based ensemble classification sentiment analysis [Article]. *Entropy, 18*(1), Article 4.

Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. LREC,

Badaro, G., Baly, R., Hajj, H., Habash, N., & El-Hajj, W. (2014). A large scale Arabic sentiment lexicon for Arabic opinion mining. *ANLP 2014, 165*.

Bakliwal, A., Arora, P., & Varma, V. (2012). Hindi subjective lexicon: A lexical resource for hindi polarity classification. Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC),

Balahur, A., Steinberger, R., Van Der Goot, E., Pouliquen, B., & Kabadjov, M. (2009). Opinion mining on newspaper quotations. Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT'09. IEEE/WIC/ACM International Joint Conferences on,

Balahur, A., & Turchi, M. (2012). Multilingual sentiment analysis using machine translation? Proceedings of the 3rd workshop in computational approaches to subjectivity and sentiment analysis,

Balahur, A., & Turchi, M. (2013). Improving sentiment analysis in twitter using multilingual machine translated data. Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013,

Balahur, A., & Turchi, M. (2014, 1//). Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis. *Computer Speech & Language, 28*(1), 56-75.

Banea, C., Mihalcea, R., & Wiebe, J. (2013). Porting multilingual subjectivity resources across languages. *Ieee Transactions on Affective Computing, 4*(2), 211-225.

Banea, C., Wiebe, J. M., & Mihalcea, R. (2008). A bootstrapping method for building subjectivity lexicons for languages with scarce resources.

Basile, V., & Nissim, M. (2013). Sentiment analysis on Italian tweets. Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis,

Benamara, F., Cesarano, C., Picariello, A., Recupero, D. R., & Subrahmanian, V. S. (2007). Sentiment analysis: Adjectives and adverbs are better than adjectives alone. Icwsm, Bermingham, A., & Smeaton, A. F. (2010). Classifying sentiment in microblogs: is brevity an advantage? Proceedings of the 19th ACM international conference on Information and knowledge management,

Biagioni, R. (2016). Sentiment Analysis. In *The SenticNet Sentiment Lexicon: Exploring Semantic Richness in Multi-Word Concepts* (pp. 7-16). Springer.

Biltawi, M., Etaiwi, W., Tedmori, S., Hudaib, A., & Awajan, A. (2016). Sentiment classification techniques for Arabic language: A survey. 7th International Conference on Information and Communication Systems, ICICS 2016,

bin Rodzman, S. B., Rashid, M. H., Ismail, N. K., Rahman, N. A., Aljunid, S. A., & Rahman, H. A. (2019). Experiment with Lexicon Based Techniques on Domain-Specific Malay Document Sentiment Analysis. 2019 IEEE 9th Symposium on Computer Applications & Industrial Electronics (ISCAIE),

Black, W., Elkateb, S., Rodriguez, H., Alkhalifa, M., Vossen, P., Pease, A., & Fellbaum, C. (2006). Introducing the Arabic wordnet project. Proceedings of the third international WordNet conference,

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics, 5*, 135-146.

Borin, L., Forsberg, M., & Lönngren, L. (2013). SALDO: a touch of yin to WordNet's yang. *Language Resources and Evaluation, 47*(4), 1191-1211.

Bravo-Marquez, F., Frank, E., & Pfahringer, B. (2016). Building a Twitter opinion lexicon from automatically-annotated tweets [Article]. *Knowledge-Based Systems, 108*, 65-78.

Brooke, J., Tofiloski, M., & Taboada, M. (2009). Cross-Linguistic Sentiment Analysis: From English to Spanish. RANLP, Buckwalter, T. (2004). *Buckwalter Arabic Morphological Analyzer Version 2.0. Linguistic Data Consortium, University of Pennsylvania, 2002. LDC Cat alog No.: LDC2004L02*.

Burnham, J. J. H. S. (1992). Approach-method-technique: Making distinctions and creating connections. *3*(1), 3-26.

Buscaldi, D., & Hernandez-Farias, D. I. (2016). IRADABE2: Lexicon Merging and Positional Features for Sentiment Analysis in Italian. CLiC-it/EVALITA,

Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *Ieee Intelligent Systems, 28*(2), 15-21.

Cambria, E., Speer, R., Havasi, C., & Hussain, A. (2010). SenticNet: A Publicly Available Semantic Resource for Opinion Mining. AAAI fall symposium: commonsense knowledge,

Cerini, S., Compagnoni, V., Demontis, A., Formentelli, M., & Gandini, G. (2007). Micro-WNOp: A gold standard for the evaluation of automatically compiled lexical resources for opinion mining. *Language resources and linguistic theory: Typology, second language acquisition, English linguistics*, 200-210.

Chao, A. F. Y., & Yang, H.-L. (2018, 2018/01/01/). Using Chinese radical parts for sentiment analysis and domain-dependent seed set extraction. *Computer Speech & Language, 47*, 194-213.

Chaovalit, P., & Zhou, L. (2005). Movie review mining: A comparison between supervised and unsupervised classification approaches. System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on,

Chaturvedi, I., Cambria, E., Welsch, R. E., & Herrera, F. J. I. F. (2018). Distinguishing between facts and opinions for sentiment analysis: Survey and challenges. *44*, 65-77.

Chen, Y., & Skiena, S. (2014). Building sentiment lexicons for all major languages. 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014, Baltimore, MD.

Cho, H., Kim, S., Lee, J., & Lee, J. S. (2014, Nov). Data-driven integration of multiple sentiment dictionaries for lexicon-based sentiment classification of product reviews [Article]. *Knowledge-Based Systems, 71*, 61-71.

Darwich, M., Noah, S. A. M., & Omar, N. (2016). Automatically generating a sentiment lexicon for the Malay language. *Asia-Pacific Journal of Information Technology and Multimedia, 5*(1), 49-59.

Das, A., & Bandyopadhyay, S. (2010). Sentiwordnet for bangla. *Knowledge Sharing Event-4: Task, 2*.

Dashtipour, K., Poria, S., Hussain, A., Cambria, E., Hawalah, A. Y., Gelbukh, A., & Zhou, Q. (2016). Multilingual sentiment analysis: state of the art and independent comparison of techniques. *Cognitive Computation, 8*(4), 757-771.

Dehdarbehbahani, I., Shakery, A., & Faili, H. (2014). Semi-supervised word polarity identification in resource-lean languages [Article]. *Neural Networks, 58*, 50-59.

Dehkharghani, R., Yanikoglu, B., Tapucu, D., & Saygin, Y. (2012). Adaptation and use of subjectivity lexicons for domain dependent sentiment classification. 12th IEEE International Conference on Data Mining Workshops, ICDMW 2012, Brussels.

Denecke, K. (2008). Using sentiwordnet for multilingual sentiment analysis. Data Engineering Workshop, 2008. ICDEW 2008. IEEE 24th International Conference on,

Deng, S., Sinha, A. P., & Zhao, H. (2017). Adapting sentiment lexicons to domain-specific social media texts. *Decision Support Systems, 94*, 65-76.

Devitt, A., & Ahmad, K. (2013). Is there a language of sentiment? An analysis of lexical resources for sentiment analysis [Article]. *Language Resources and Evaluation, 47*(2), 475-511.

Dey, A., Jenamani, M., & Thakkar, J. J. (2018, 2018/08/01/). Senti-N-Gram: An n-gram lexicon for sentiment analysis. *Expert Systems with Applications, 103*, 92-105.

Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., & Danforth, C. M. (2011). Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PLoS ONE, 6*(12), e26752.

Dong, X., & de Melo, G. (2018). Cross-Lingual Propagation for Deep Sentiment Analysis. Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI 2018). AAAI Press,

Dumais, S. T., Furnas, G. W., Landauer, T. K., Deerwester, S., & Harshman, R. (1988). Using latent semantic analysis to improve access to textual information. Proceedings of the SIGCHI conference on Human factors in computing systems,

[Record #2033 is using a reference type undefined in this output style.]

Ekinci, E., & Omurca, S. İ. J. I. A. J. o. I. T. (2019). A new approach for a domain-independent Turkish sentiment seed lexicon compilation. *5*, 1-11.

El-Beltagy, S. R. (2016). NileULex: A Phrase and Word Level Sentiment Lexicon for Egyptian and Modern Standard Arabic. LREC,

El-Halees, A. (2011). *Arabic opinion mining using combined classification approach* The International Arab Conference on Information Technology,

Elhawary, M., & Elfeky, M. (2010). Mining Arabic business reviews. Data Mining Workshops (ICDMW), 2010 IEEE International Conference on, Elkateb, S., Black, W., Rodríguez, H., Alkhalifa, M., Vossen, P., Pease, A., & Fellbaum, C. (2006). Building a wordnet for arabic. Proceedings of The fifth international conference on Language Resources and Evaluation (LREC 2006),

Eskander, R., & Rambow, O. (2015). SLSA: A sentiment lexicon for Standard Arabic. Conference on Empirical Methods in Natural Language Processing, EMNLP 2015,

Esuli, A., & Sebastiani, F. (2007). SENTIWORDNET: A high-coverage lexical resource for opinion mining. *Evaluation*, 1-26.

Feng, S., Song, K., Wang, D., & Yu, G. (2015a). A word-emoticon mutual reinforcement ranking model for building sentiment lexicon from massive collection of microblogs [Article]. *World Wide Web-Internet and Web Information Systems, 18*(4), 949-967.

Feng, S., Song, K. S., Wang, D. L., & Yu, G. (2015b, Jul). A word-emoticon mutual reinforcement ranking model for building sentiment lexicon from massive collection of microblogs [Article]. *World Wide Web-Internet and Web Information Systems, 18*(4), 949-967.

Feng, S., Wang, L., Xu, W., Wang, D., & Yu, G. (2012). *Unsupervised learning Chinese sentiment lexicon from massive microblog data* [Conference Paper]. 8th International Conference on Advanced Data Mining and Applications, ADMA 2012, Nanjing.

Fersini, E., Messina, E., & Pozzi, F. A. (2016, 1//). Expressive signals in social media languages to improve polarity detection. *Information Processing & Management, 52*(1), 20-35.

Giachanou, A., & Crestani, F. (2016). Like it or not: A survey of twitter sentiment analysis methods. *ACM Computing Surveys (CSUR), 49*(2), 28.

Gill, S. K. (2013). *Language policy challenges in multi-ethnic Malaysia* (Vol. 8). Springer Science & Business Media.

Hajmohammadi, M. S., Ibrahim, R., & Selamat, A. (2014a, 9//). Bi-view semi-supervised active learning for cross-lingual sentiment classification. *Information Processing & Management, 50*(5), 718-732.

Hajmohammadi, M. S., Ibrahim, R., & Selamat, A. (2014b, 11//). Cross-lingual sentiment classification using multiple source languages in multi-view semi-supervised learning. *Engineering Applications of Artificial Intelligence, 36*, 195-203.

Hammer, H., Bai, A., Yazidi, A., & Engelstad, P. (2014). Building sentiment lexicons applying graph theory on information from three norwegian thesauruses. *Norsk Informatikkonferanse (NIK)*.

Haniewicz, K., Kaczmarek, M., Adamczyk, M., & Rutkowski, W. (2014). *Polarity lexicon for the polish language: Design and extension with random walk algorithm* [Conference Paper]. International Conference on Systems Science, ICSS 2013,

Hassan, A., Abu-Jbara, A., Jha, R., & Radev, D. (2011). Identifying the semantic orientation of foreign words. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2,

Hong, Y., Kwak, H., Baek, Y., & Moon, S. (2013). Tower of babel: A crowdsourcing game building sentiment lexicons for resource-scarce languages. 22nd International Conference on World Wide Web, WWW 2013, Rio de Janeiro.

Honnibal, M., & Montani, I. (2017). spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing. *To appear*.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining,

Huang, S., Niu, Z., & Shi, C. (2014). Automatic construction of domain-specific sentiment lexicon based on constrained label propagation [Article]. *Knowledge-Based Systems, 56*, 191-200.

Hussein, D. M. E.-D. M. (2016). A survey on sentiment analysis challenges. *Journal of King Saud University - Engineering Sciences*.

Itani, M., Roast, C., & Al-Khayatt, S. (2017, 2017/01/01/). Developing Resources For Sentiment Analysis Of Informal Arabic Text In Social Media. *Procedia Computer Science, 117*, 129-136.

Jha, V., Savitha, R., Hebbar, S. S., Shenoy, P. D., & Venugopal, K. (2015). Hmdsad: Hindi multi-domain sentiment aware dictionary. Computing and Network Communications (CoCoNet), 2015 International Conference on,

Joshi, A., Balamurali, A., & Bhattacharyya, P. (2010). A fall-back strategy for sentiment analysis in hindi: a case study. *Proceedings of the 8th ICON*.

Jünger, J., & Keyling, T. (2017). *Facepager an application for generic data retrieval through APIs. Código y nuevas versiones disponibles en*. https://github. com/strohne/Facepager

Kang, H., Yoo, S. J., & Han, D. (2012, Apr). Senti-lexicon and improved Naive Bayes algorithms for sentiment analysis of restaurant reviews [Review]. *Expert Systems with Applications, 39*(5), 6000-6010.

Kann, V., & Rosell, M. (2005). Free construction of a free Swedish dictionary of synonyms. Proc. of NODALIDA 2005, Karra, A. K., Pande, P., Railkar, R., Sharma, A., & Bhattacharyya, P. (2009). *Hindi english wordnet linkage* [Dual degree thesis, IIT Bomobay]. Bomobay.

Kharde, V., & Sonawane, P. (2016). Sentiment analysis of twitter data: a survey of techniques. *arXiv preprint arXiv:1601.06971*.

Kim, J., Li, J.-J., & Lee, J.-H. (2010). Evaluating multilanguage-comparability of subjectivity analysis systems. Proceedings of the 48th annual meeting of the association for computational linguistics,

Kim, S.-M., & Hovy, E. (2004). Determining the sentiment of opinions. Proceedings of the 20th international conference on Computational Linguistics,

Kim, S.-M., & Hovy, E. (2006). Identifying and analyzing judgment opinions. Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics,

Kiritchenko, S., Zhu, X., & Mohammad, S. M. (2014). Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research, 50*, 723-762.

Koç, S. Ş., Özer, M., Toroslu, İ. H., Davulcu, H., & Jordan, J. (2018, 2018/06/15/). Triadic co-clustering of users, issues and sentiments in political tweets. *Expert Systems with Applications, 100*, 79-94.

Kong, L., Li, C., Ge, J., Yang, Y., Zhang, F., & Luo, B. (2018). Construction of Microblog-Specific Chinese Sentiment Lexicon Based on Representation Learning. Pacific Rim International Conference on Artificial Intelligence,

Kouloumpis, E., Wilson, T., & Moore, J. D. (2011). Twitter sentiment analysis: The good the bad and the omg! *Icwsm, 11*(538-541), 164.

Kumar, P., & Jaiswal, U. C. (2016). A comparative study on sentiment analysis and opinion mining [Article]. *International Journal of Engineering and Technology, 8*(2), 938-943.

Lafourcade, M., Joubert, A., & Le Brun, N. (2015). Collecting and Evaluating Lexical Polarity with A Game With a Purpose. RANLP,

Lafourcade, M., Le Brun, N., & Joubert, A. (2016). Mixing Crowdsourcing and Graph Propagation to Build a Sentiment Lexicon: Feelings Are Contagious. In E. Metais, F. Meziane, M. Saraee, V. Sugumaran, & S. Vadera (Eds.), *Natural Language Processing and Information Systems, Nldb 2016* (Vol. 9612, pp. 258-266). Springer Int Publishing Ag.

Lin, Z., Jin, X., Xu, X., Wang, Y., Wang, W., & Cheng, X. (2014). A cross-lingual joint aspect/sentiment model for sentiment analysis. 23rd ACM International Conference on Information and Knowledge Management, CIKM 2014,

Liu, B. (2010). Sentiment analysis and subjectivity. In *Handbook of Natural Language Processing, Second Edition* (pp. 627-666). Chapman and Hall/CRC.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies, 5*(1), 1-167.

Liu, Q., Zhou, M., & Zhao, X. (2015, 2015/11/01/). Understanding News 2.0: A framework for explaining the number of comments from readers on online news. *Information & Management, 52*(7), 764-776.

Lo, S. L., Cambria, E., Chiong, R., & Cornforth, D. (2016a). A multilingual semi-supervised approach in deriving Singlish sentic patterns for polarity detection [Article]. *Knowledge-Based Systems, 105*, 236-247.

Lo, S. L., Cambria, E., Chiong, R., & Cornforth, D. (2016b). Multilingual sentiment analysis: from formal to informal and scarce resource languages. *Artificial Intelligence Review*, 1-29.

Maamouri, M., Graff, D., Bouziri, B., Krouna, S., Bies, A., & Kulick, S. (2010). Standard Arabic morphological analyzer (SAMA) version 3.1. *Linguistic Data Consortium, Catalog No.: LDC2010L01*.

Mahyoub, F. H. H., Siddiqui, M. A., & Dahab, M. Y. (2014, 12//). Building an Arabic Sentiment Lexicon Using Semi-supervised Learning. *Journal of King Saud University - Computer and Information Sciences, 26*(4), 417-424.

Maks, I., & Vossen, P. (2011). Different approaches to automatic polarity annotation at synset level. Proceedings of the First International Workshop on Lexical Resources, WoLeR,

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations,

Mathur, P., & Paul, S. (2012). Integration of a noun compound translator tool with moses for english-hindi machine translation and evaluation. International Conference on Intelligent Text Processing and Computational Linguistics,

Medhat, W., Hassan, A., & Korashy, H. (2014, 12//). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal, 5*(4), 1093-1113.

Mihalcea, R., Banea, C., & Wiebe, J. M. (2007). Learning multilingual subjective language via cross-lingual projections. Proceedings of the 45th annual meeting of the association of computational linguistics,

Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM, 38*(11), 39-41.

Mohammad, S., & Turney, P. (2013a). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence, 29*(3), 436-465.

Mohammad, S., & Turney, P. (2013b). *Nrc emotion lexicon* National Research Council, Canada.

Molina-González, M. D., Martínez-Cámara, E., Martín-Valdivia, M.-T., & Perea-Ortega, J. M. (2013, 12/15/). Semantic orientation for polarity classification in Spanish reviews. *Expert Systems with Applications, 40*(18), 7250-7257.

Moliner, M., & Moliner, M. (1998). *Diccionario de uso del español* (Vol. 2). Gredos Madrid.

Montoyo, A., Martínez-Barco, P., & Balahur, A. (2012, 11//). Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments. *Decision Support Systems, 53*(4), 675-679.

Moretti, G., Sprugnoli, R., Menini, S., & Tonelli, S. (2016, 11/1/). ALCIDE: Extracting and visualising content from large document collections to support humanities studies. *Knowledge-Based Systems, 111*, 100-112.

Mukhtar, N., Khan, M. A., & Chiragh, N. (2018, 2018/12/01/). Lexicon-based approach outperforms Supervised Machine Learning approach for Urdu Sentiment Analysis in multiple domains. *Telematics and Informatics, 35*(8), 2173-2183.

Narayan, D., Chakrabarti, D., Pande, P., & Bhattacharyya, P. (2002). An experience in building the indo wordnet-a wordnet for hindi. First International Conference on Global WordNet, Mysore, India,

Nielsen, F. A. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. 1st Workshop on Making Sense of Microposts 2011: Big Things Come in Small Packages, #MSM 2011 - Co-located with the 8th Extended Semantic Web Conference, ESWC 2011, Heraklion, Crete.

Nusko, B., Tahmasebi, N., & Mogren, O. (2016). Building a Sentiment Lexicon for Swedish. Digital Humanities 2016. From Digitization to Knowledge 2016: Resources and Methods for Semantic Processing of Digital Works/Texts, Proceedings of the Workshop, July 11, 2016, Krakow, Poland, Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. University of Illinois press.

Park, S., Lee, W., & Moon, I. C. (2015, Apr). Efficient extraction of domain specific sentiment lexicon with active learning [Article]. *Pattern Recognition Letters, 56*, 38-44.

Pasha, A., Al-Badrashiny, M., Diab, M. T., El Kholy, A., Eskander, R., Habash, N., Pooleery, M., Rambow, O., & Roth, R. (2014). MADAMIRA: A Fast, Comprehensive Tool for Morphological Analysis and Disambiguation of Arabic. LREC,

Passaro, L. C., Pollacci, L., & Lenci, A. (2015). Item: A vector space model to bootstrap an italian emotive lexicon. *CLiC it, 60*(15), 215.

Perez-Rosas, V., Banea, C., & Mihalcea, R. (2012). Learning Sentiment Lexicons in Spanish [Proceedings Paper]. *Lrec 2012 - Eighth International Conference on Language Resources and Evaluation*, 3077-3081.

Plutchik, R. (2001). The Nature of Emotions Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist, 89*(4), 344-350.

Poria, S., Cambria, E., & Gelbukh, A. (2016, 9/15/). Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems, 108*, 42-49.

Poria, S., Gelbukh, A., Hussain, A., Howard, N., Das, D., & Bandyopadhyay, S. (2013). Enhanced SenticNet with affective labels for concept-based opinion mining. *Ieee Intelligent Systems, 28*(2), 31-38.

Powers, D. M. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation.

Pozzi, F. A., Fersini, E., Messina, E., & Liu, B. (2017). Chapter 1 - Challenges of Sentiment Analysis in Social Networks: An Overview. In *Sentiment Analysis in Social Networks* (pp. 1-11). Morgan Kaufmann.

Qiu, X., Zhang, Q., & Huang, X. (2013). Fudannlp: A toolkit for chinese natural language processing. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations,

Quan, C., & Ren, F. (2014). Unsupervised product feature extraction for feature-oriented opinion determination [Article]. *Information Sciences, 272*, 16-28.

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,

Rao, Y., Lei, J., Wenyin, L., Li, Q., & Chen, M. (2014). Building emotional dictionary for sentiment analysis of online news [Article]. *World Wide Web-Internet and Web Information Systems, 17*(4), 723-742.

Rashed, F. E., & Abdolvand, N. (2017). A Supervised Method for Constructing Sentiment Lexicon in Persian Language. *Journal of Computer & Robotics, 10*(1), 11-19.

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

Ray, H. M. J. L. i. I. (2019). The Nature of Approaches, Methods and Techniques in Teaching Korean: A Case Study. *19*(8).

Remus, R., Quasthoff, U., & Heyer, G. (2010). SentiWS-A Publicly Available German-language Resource for Sentiment Analysis. LREC, Robaldo, L., & Di Caro, L. (2013, 9//). OpinionMining-ML. *Computer Standards & Interfaces, 35*(5), 454-469.

Rosell, M., & Kann, V. (2010). Constructing a swedish general purpose polarity lexicon random walks in the people's dictionary of synonyms. *SLTC 2010*, 19.

Rouvier, M., & Favre, B. (2016). Building a robust sentiment lexicon with (almost) no resource. *arXiv preprint arXiv:1612.05202*.

Saad, M. K. (2010). The impact of text preprocessing and term weighting on arabic text classification. *Gaza: Computer Engineering, the Islamic University*.

Salah, Z., Coenen, F., & Grossi, D. (2013). *Generating domain-specific sentiment lexicons for opinion mining* [Conference Paper]. 9th International Conference on Advanced Data Mining and Applications, ADMA 2013, Hangzhou.

Salton, G., & Buckley, C. (1988, 1988/01/01/). Term-weighting approaches in automatic text retrieval. *Information Processing & Management, 24*(5), 513-523.

Sawe, B. E. (2019, 7/6/2019). *What Is The Most Spoken Language In The World?* worldatlas.com. Retrieved 25/7/2019 from https://www.worldatlas.com/articles/most-popular-languages-in-the-world.html

Scharl, A., Sabou, M., Gindl, S., Rafelsberger, W., & Weichselbraun, A. (2012). Leveraging the wisdom of the crowds for the acquisition of multilingual language resources.

Shamsfard, M., Hesabi, A., Fadaei, H., Mansoory, N., Famian, A., Bagherbeigi, S., Fekri, E., Monshizadeh, M., & Assi, S. M. (2010). Semi automatic development of farsnet; the persian wordnet. Proceedings of 5th Global WordNet Conference, Mumbai, India, Siddiqui, S., Monem, A. A., & Shaalan, K. (2018). Evaluation and Enrichment of Arabic Sentiment Analysis. In *Intelligent Natural Language Processing: Trends and Applications* (pp. 17-34). Springer.

Sidorov, G., Miranda-Jiménez, S., Viveros-Jiménez, F., Gelbukh, A., Castro-Sánchez, N., Velásquez, F., Díaz-Rangel, I., Suárez-Guerra, S., Treviño, A., & Gordon, J. (2012). Empirical study of machine learning based approach for opinion mining in tweets. Mexican International Conference on Artificial Intelligence,

Song, M., Park, H., & Shin, K.-s. (2019, 2019/05/01/). Attention-based long short-term memory network using sentiment lexicon embedding for aspect-level sentiment analysis in Korean. *Information Processing & Management, 56*(3), 637-653.

Steinberger, J., Ebrahim, M., Ehrmann, M., Hurriyetoglu, A., Kabadjov, M., Lenkova, P., Steinberger, R., Tanev, H., Vázquez, S., & Zavarella, V. (2012, 11//). Creating sentiment dictionaries via triangulation. *Decision Support Systems, 53*(4), 689-694.

Stone, P. J., Dunphy, D. C., & Smith, M. S. (1966). *The general inquirer: A computer approach to content analysis*. M.I.T. Press.

Stubbs, M. (2001). Computer-assisted text and corpus analysis: lexical cohesion and communicative competence. *The handbook of discourse analysis, 18*, 304.

Subrahmanian, V. S., & Reforgiato, D. J. I. I. S. (2008). AVA: Adjective-verb-adverb combinations for sentiment analysis. *23*(4), 43-50.

Sun, S., Luo, C., & Chen, J. (2017, 7//). A review of natural language processing techniques for opinion mining systems. *Information Fusion, 36*, 10-25.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011, Jun). Lexicon-Based Methods for Sentiment Analysis [Article]. *Computational Linguistics, 37*(2), 267-307.

Tan, S., & Wu, Q. (2011). A random walk algorithm for automatic construction of domain-oriented sentiment lexicon [Article]. *Expert Systems with Applications, 38*(10), 12094-12100.

Tang, D., Qin, B., & Liu, T. (2015). Deep learning for sentiment analysis: successful approaches and future challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 5*(6), 292-303.

Tang, D., Wei, F., Qin, B., Zhou, M., & Liu, T. (2014). Building Large-Scale Twitter-Specific Sentiment Lexicon: A Representation Learning Approach. COLING,

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*(12), 2544-2558.

Trakultaweekoon, K., & Klaithin, S. (2016). SenseTag: A tagging tool for constructing Thai sentiment lexicon. Computer Science and Software Engineering (JCSSE), 2016 13th International Joint Conference on,

Turney, P. D. (2001). Mining the web for synonyms: PMI-IR versus LSA on TOEFL. European Conference on Machine Learning,

Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. Proceedings of the 40th Annual Meeting on Association for Computational Linguistics,

Turney, P. D., & Littman, M. L. (2002). Unsupervised learning of semantic orientation from a hundred-billion-word corpus. *arXiv preprint cs/0212012*.

Uysal, A. K., & Gunal, S. (2014, 2014/01/01/). The impact of preprocessing on text classification. *Information Processing & Management, 50*(1), 104-112.

Van De Kauter, M., Breesch, D., & Hoste, V. (2015). Fine-grained analysis of explicit and implicit sentiment in financial news articles [Article]. *Expert Systems with Applications, 42*(11), 4999-5010.

Wang, L., & Xia, R. (2017). Sentiment Lexicon Construction with Representation Learning Based on Hierarchical Sentiment Supervision. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing,

Wang, Q., Jin, Y., Yang, T., & Cheng, S. (2017, 1/15/). An emotion-based independent cascade model for sentiment spreading. *Knowledge-Based Systems, 116*, 86-93.

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,

Wiebe, J., & Riloff, E. (2005). Creating subjective and objective sentence classifiers from unannotated texts. International Conference on Intelligent Text Processing and Computational Linguistics,

Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., Cardie, C., Riloff, E., & Patwardhan, S. (2005a). OpinionFinder: A system for subjectivity analysis. Proceedings of hlt/emnlp on interactive demonstrations,

Wilson, T., Wiebe, J., & Hoffmann, P. (2005b). Recognizing contextual polarity in phrase-level sentiment analysis. Proceedings of the conference on human language technology and empirical methods in natural language processing,

Wu, F., Huang, Y., Song, Y., & Liu, S. (2016, Jul). Towards building a high-quality microblog-specific Chinese sentiment lexicon [Article]. *Decision Support Systems, 87*, 39-49.

Wu, F., Huang, Y., & Yuan, Z. (2017). Domain-specific sentiment classification via fusing sentiment knowledge from multiple sources [Article]. *Information Fusion, 35*, 26-37.

Wu, S., Wu, F., Chang, Y., Wu, C., & Huang, Y. (2019, 2019/02/01/). Automatic construction of target-specific sentiment lexicon. *Expert Systems with Applications, 116*, 285-298.

Xing, F. Z., Pallucchini, F., & Cambria, E. (2019, 2019/05/01/). Cognitive-inspired domain adaptation of sentiment lexicons. *Information Processing & Management, 56*(3), 554-564.

Yang, A. M., Lin, J. H., Zhou, Y. M., & Chen, J. (2013). Research on Building a Chinese Sentiment Lexicon Based on SO-PMI. In J. Zhang, Z. J. Wang, S. R. Zhu, & X. M. Meng (Eds.), *Information Technology Applications in Industry, Pts 1-4* (Vol. 263-266, pp. 1688-1693). Trans Tech Publications Ltd.

Yao, J., Wu, G., Liu, J., & Zheng, Y. (2006). Using bilingual lexicon to judge sentiment orientation of Chinese words. Computer and Information Technology, 2006. CIT'06. The Sixth IEEE International Conference on,

Yuang, C. T., Banchs, R. E., & Siong, C. E. (2012). An empirical evaluation of stop word removal in statistical machine translation. Proceedings of the Joint Workshop on Exploiting Synergies between Information Retrieval and Machine Translation (ESIRMT) and Hybrid Approaches to Machine Translation (HyTra),

Yue, L., Chen, W., Li, X., Zuo, W., & Yin, M. (2018). A survey of sentiment analysis in social media. *Knowledge and Information Systems*, 1-47.

Zeb, S., Qamar, U., & Hussain, F. (2016). *Sentiment analysis on user reviews through lexicon and rule-based approach* [Conference Paper]. 18th International Conference on Web Technologies and Applications, APWeb 2016 and Workshop on 2nd International Workshop on Web Data Mining and Applications, WDMA 2016 and 1st International Workshop on Graph Analytics and Query Processing, GAP 2016 and 1st International Workshop on Spatial-temporal Data Management and Analytics, SDMA 2016,

Zhang, H.-P., Yu, H.-K., Xiong, D.-Y., & Liu, Q. (2003). HHMM-based Chinese lexical analyzer ICTCLAS. Proceedings of the second SIGHAN workshop on Chinese language processing-Volume 17, Zhang, P., Wang, S., & Li, D. (2016, 9/1/). Cross-lingual sentiment classification: Similarity discovery plus training data adjustment. *Knowledge-Based Systems, 107*, 129-141.

Zhang, Y., Dang, Y., & Chen, H. (2013, 6//). Research note: Examining gender emotional differences in Web forum communication. *Decision Support Systems, 55*(3), 851-860.

Zhang, Y., & Tsai, F. S. (2009). Chinese novelty mining. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3,

Zhao, C., Wang, S., & Li, D. (2018, 2018/11/08/). Exploiting social and local contexts propagation for inducing Chinese microblog-specific sentiment lexicons. *Computer Speech & Language*.

Zhao, C., Wang, S., & Li, D. (2019, 2019/05/01/). Exploiting social and local contexts propagation for inducing Chinese microblog-specific sentiment lexicons. *Computer Speech & Language, 55*, 57-81.

Zhou, S., Chen, Q., & Wang, X. (2014, 5/5/). Fuzzy deep belief networks for semi-supervised sentiment classification. *Neurocomputing, 131*, 312-322.

Zhu, Y., Wen, Z., Wang, P., & Peng, Z. (2009). A method of building Chinese basic semantic lexicon based on word similarity. 2009 Chinese Conference on Pattern Recognition, CCPR 2009 and the 1st CJK Joint Workshop on Pattern Recognition, CJKPR, Nanjing.

Zimmermann, M., Ntoutsi, E., & Spiliopoulou, M. (2016, 2/1/). Extracting opinionated (sub)features from a stream of product reviews using accumulated novelty and internal re-organization. *Information Sciences, 329*, 876-899.