1 Sentiment Analysis: Background

1.1 Definition and Description of Sentiment Analysis

According to comprehensive reviews of its development and application (e.g., Feldman, 2013; Mäntylä, Graziotin & Kuutila, 2018; Zunic, Corcoran & Spasic, 2020), sentiment analysis is the process of using algorithms and computer technologies to systematically detect, extract, and classify the subjective information and affective states expressed in a text, such as opinions, attitudes, and emotions regarding a service, product, person, or topic. Subjective in nature, sentiments often appear in polarity terms (i.e., in terms of two polar opposites), such as favourable/unfavourable, good/bad, happy/unhappy, positive/negative, and pro/con, although neutral sentiment is a possibility. Given this fact, sentiment analyses, in essence, detect and extract subjective polarity in language to identify the sentiments and their strengths in words, sentences, and texts (Taboada et al., 2011, p. 268). More specifically, a given sentiment analysis identifies the subjectivity, polarity, and semantic orientation of the language regarding the thing, organization, or person that is being evaluated (D'Andrea et al., 2015; Feldman, 2013; Liu and Lei, 2018; Mäntylä et al., 2018; Zunic et al., 2020). It is necessary to note that while sentiment analysis often includes emotion analysis, the latter is a more specialized subcategory of sentiment analysis. As noted, sentiment analysis is an evaluation mainly in positive vs. negative polarity terms; in comparison, emotion analysis involves more in-depth examinations of various specific emotions, such as "anger," "anxiety," "disgust," "fear," "joy," and "sadness" (Giuntini et al. 2020; Ren & Quan, 2012). Emotion analysis is highly valuable in consumer business and healthcare.

Although sentiment analysis as a term defined here was reportedly first used by Nasukawa and Yi (2003), studies about sentiments and opinions began much earlier (D'Andrea et al., 2015; Mäntylä et al., 2018). According to Mäntylä et al.'s (2018) thorough review of the evolution of sentiment analysis, the origins of sentiment analysis were (1) public opinion studies in the early 1940s during WWII and (2) the analysis of subjectivity in a text using computational linguistic approaches in the 1990s. However, sentiment analysis as we know it today did not blossom until 2004, for, as Mäntylä et al.'s (2018, p. 16) review results show, "99% of the papers [on sentiment analysis] have been published after 2004." In other words, since the early 2000s, sentiment analysis has become a very popular research area and has been used in many different domains. This is because results from sentiment analyses may offer highly useful information for businesses, consumers, educational and healthcare institutions, government agencies, and political organizations concerning their products, services, patients' feelings and emotions, policies, and/or opinions regarding politicians and political parties respectively (Feldman, 2013; Mäntylä et al., 2018; Rambocas & Pacheco, 2018; Zhang, Gan & Jiang, 2014; Zunic et al., 2020). Another reason for the rapid growth of work in sentiment analysis is the public's increased access to the Internet and their growing use of social media (e.g., Facebook and Twitter) and other online business and social communication platforms (Rambocas & Pacheco, 2018; Pagolu et al., 2016; Zunic et al., 2020).

1.2 Sentiment Analysis vs. Appraisal, Stance, and Semantic Prosody

Based on the aforementioned definition, sentiment in sentiment analysis is quite similar in meaning to several known concepts in corpus linguistics that deal with evaluative language, such as appraisal (Martin & White, 2005), stance (Biber, 2006; Conrad & Biber, 2000), and semantic (or discourse) prosody (Sinclair, 1991, 2004). However, although these concepts are all concerned with evaluative language, their research foci, scopes, and/or analysis approaches differ from one another to various extents, thanks perhaps largely to what Hunston (2011, p. 10) calls "a variance in what kind of phenomenon 'evaluation' is taken to be."

Appraisal analysis, which originated in systemic functional linguistics, treats evaluation as the enactment of a system of meanings by speakers/writers through the use of various linguistic and discoursal resources to convey their approval or disapproval of ideas, persons, or things (Martin & White, 2005). As a result, appraisal analysis is quite broad in scope and involves intensive perusal of text by the researcher although some appraisal studies also make use of some simple corpus query and analysis tools, such as concordancing. In other words, the research method of appraisal studies is largely qualitative in nature. In comparison, stance analysis, arising from corpus linguistic research, considers evaluation to be "the expression of personal feelings and assessments" conveyed in words, phrases, and sentence structures that are frequently used to express evaluative meanings (Conrad & Biber, 2000, p. 57). Focusing on recurring evaluative linguistic items, stance analysis thus appears to have a smaller scope than appraisal. Furthermore, steeped in corpus linguistics, stance research also makes much more use of computer technology and statistics than appraisal analysis does. Of course, stance analysis also includes some close manual reading and analysis of the identified tokens (e.g., keywords in context in the form of concordance lines) to determine and classify the types of stance being expressed (e.g., epistemic, attitudinal, and modality stances) and their semantic/discoursal functions. In this sense, stance analysis also consists of both qualitative and quantitative examinations, but with the latter being more prominent.

Regarding semantic prosody, a term not as transparent as the others in the group, a definition is first in order. Semantic prosody refers to the phenomenon that certain seemingly neutral words may develop positive or negative associations through particular frequent collocations as shown in Sinclair's (1991, pp. 74–75) example of "set in," which acquires a negative meaning through its frequent collocation with negative nouns as its subjects, such as "decay sets in," "despair sets in," and "a malaise has set in." Hence, semantic prosody is a pragmatic unit of meaning that conveys or implies either an evaluation in terms of positive/negative polarity or a subtle affective feeling, such as "reluctance, frustration, or difficulty" (Hunston, 2011, p. 56). As such, semantic prosody often functions as implicit evaluation. An example of such implicit evaluation can be found in the sentence taken from Davies's (2008–) *Corpus of Contemporary American English*: "Whether the park can endure this onslaught of modernity is a hotly debated question in local cafs" where the author's wording "the onslaught of modernity" (along with the verb "endure") implies a negative assessment of modernity. Born out of corpus linguistics like stance analysis and with its close examination of words and their co-occurring items, semantic prosody analysis seems to also view evaluation as the expression of personal emotions and assessment. Yet its focus and scope are unique in that it concentrates on unit meanings in discourse. In terms of research method, semantic prosody analysis, like stance research again, involves extensive searches and analyses of keywords in context but has a heightened focus on collocations, colligations, and other co-occurring elements that display semantic preferences.

Now let us turn to sentiment analysis. As noted, because of its origin in computer science and computational linguistics, sentiment analysis uses statistical algorithms and, more recently, machine-learning algorithms, to identify, extract, and study emotional states and subjective information in texts from various fields and professions. Also, as will be explained in Section 2, although words of sentiment polarity are the focus in sentiment analysis, broad contextual information of these words, such as their co-occurring lexical and structural items, is also considered and factored into the final sentiment score of the text being analyzed. Therefore, the scope of sentiment analysis is quite wide both in content and linguistic information covered, and its methodology is almost exclusively quantitative and computer-technology based. Examples of actual texts with sentiment analysis will be given in Sections 2, 3, 4, and 5. While sentiment analysis also uses corpus data, its methods for identifying sentiments and opinions is much more automatic and involves only limited human judgment that all occurs in the form of building a sentiment lexicon or coding a small set of data for training purposes before the actual sentiment extraction. In other words, the extraction of sentiment itself is entirely automatic and there is no human analysis involved after the sentiments of a text have been extracted.

It is important to point out that these methodological differences used between semantic analysis and appraisal/semantic prosody analyses may also represent some of the differences between computational linguistics and corpus linguistics, two closely related disciplines whose main similarities and differences may be of importance and interest to the reader of this Element. Apart from both being disciplines of applied study of language, the two also are similar, but simultaneously different, in three aspects: the role of corpus data, research purposes, and methodology. In terms of the role of corpora, while both use corpus data in their research, such data appear to be the main object of study for corpus linguistics, but, for computational linguistics, corpora serve primarily as just a resource to solve various language-related problems. Concerning research purposes, whereas both have practical language-related research goals or applications, the scope of applications for computational linguistics appears to be wider than corpus linguistics because the former began as and has remained largely an "application-oriented enterprise" (Dipper, 2008, p. 77). As an application-driven discipline, computational linguistics has focused on natural language processing, understanding, and production for the purpose of developing various language processing and production programs or tools, such as automatic speech recognition, automated phone answering service, and machine translation (Dipper, 2008; Wilks, 2010). On the other hand, corpus linguistics has concentrated mostly on how language works, especially how words and other linguistic elements are used in actual communication, so as to help ensure more accurate and adequate linguistic description of language rules and usages in language textbooks/reference books as evidenced by the many corpus-based/informed dictionaries and textbooks produced in the past few decades, including the pioneering work *Collins COBUILD English dictionary* (1987).

In terms of methodology, while both use statistical analysis and computer technology, the extent of such use and the types of tools employed differ somewhat across the two. As a branch of computer science dealing with language, computational linguistics focuses on doing formal modelling of natural language via computational algorithms and computer technology (Dipper, 2008; Wilks, 2010). In other words, the work of computational linguistics is based entirely on algorithms and technology, including the increased use of machine-learning technology. Machine-learning (which may be either supervised or unsupervised, an issue we will discuss in Section 2) refers to the practice of using algorithms to create a computational model based on sample data or training data for the purpose of making automatic inferences, predictions, or decisions (Shalev-Shwartz & Ben-David, 2014). Compared with conventional computational linguistics methods, machine-learning is more into achieving a higher level of automatic language processing, understanding, prediction, and production, and its algorithms may thus be more sophisticated. Compared with computational linguistics, corpus linguistics, while often also making use of algorithms and technologies, sometimes engages in substantial qualitative analysis with limited basic computations. However, it is important to note that the difference in methodology between computational and corpus linguistics has actually become much smaller in the past two decades because of the increasing use of computational models and tools, including those of machine-learning, by corpus linguists in their research and development of computerized language teaching and assessment programs, such as those used for automated essay scoring (e.g., ETS's *c-rater: www.ets.org/accelerate/ai-portfolio/c-rater*) and automated measuring of syntactic complexity (e.g., Lu, 2010). In short, overall, with the increased use of tools from computational linguistics by corpus linguists, there now seems to be a growing amount of overlap between the two disciplines.

1.3 Existing Work of Sentiment Analysis: Major Domains/Topics, Successes, Challenges/Questions, and Principles

This section contains three subsections. Section 1.3.1 introduces the domains where sentiment analysis has been conducted most extensively and the topics most frequently covered in each of the domains including the motivation behind them. Section 1.3.2 examines the successes of the existing work and the challenges/questions it has been facing. Section 1.3.3 discusses the key principles for conducting sentiment analysis. Some existing studies will be mentioned as examples to help illustrate the main points covered.

1.3.1 Major Domains and Topics

While sentiment analysis has been carried out in many different domains, business/finance, politics, healthcare/medicine, and entertainment (mainly movies) appear to be the four domains where it has been conducted and used most extensively (Feldman, 2013; Mäntylä et al., 2018; Rambocas & Pacheco, 2018; Zunic et al., 2020). A review of the published sentiment analysis studies in these four domains indicates that the topics or targets of sentiment analysis are domain specific with each domain having its own key topics. Table 1.1 lists the most frequently covered topics in each of the four domains plus the area of

Table 1.1 Main topics of sentiment analysis across domains

Domain	Common Topics	Major Data Sources	Amount of Existing Work
Business/ Finance	consumers/media/ businesses' opinions about the economy, financial markets, products, and services	online product/service reviews, surveys, business reports, and news	enormous
Politics	voters/public's opinions about candidates for elections, governments, legislations, policies, officials/politicians, and political parties	social media postings, news, polls, surveys, interviews, candidates' speeches and writings	enormous
Healthcare/ Medicine	patients' opinions, attitudes, and/or feelings about diseases and their diagnosis and treatments, medical services and providers, and medications	discussions on social media platforms, medical reports/other medical documents, reviews of healthcare services/medicines	large
Entertainment (movies)	reviewers' evaluations of movies including aspects of acting, cinematography, directing, music, script (plot/story), etc.	movie reviews	substantial
Academic writing/ applied linguistics	positivity/negativity in academic writing in general and across disciplines	journal articles, abstracts	limited

academic writing, a subfield of applied linguistics that has recently seen some sentiment analysis studies.[1] The latter is included our discussion because of its potential interest to the reader of this Element. Table 1.1 also presents the major data sources and the amount of existing work in each domain.

As displayed in Table 1.1, for business/finance, opinions about the economy, financial markets, products, and services constitute the key topics. The reason for the prominence of such topics in this domain is rather simple. Being entirely client dependent, companies must always know how customers feel about their products and/or services in order to maintain and increase their business. In fact, sentiment analysis results about products and services are not only important and useful for businesses but also for consumers in their purchase decision-making (Feldman, 2013; Mäntylä et al., 2018; Rambocas & Pacheco, 2018; Zhang et al., 2014). Similarly, finance firms and their clients need to understand how companies and investors feel about the economy/market and their future directions as well as about corporate financial performance so they can make informed investment decisions and be successful (Feldman, 2013; Ikoro et al., 2018; Loureiro, Bilro & Japutra, 2019; Rambocas & Pacheco, 2018). In fact, for those working in the stock market, an accurate understanding of sentiment about the market is crucial in making wise investment decisions (Garcia, 2013; Hajek, Olej & Myskova, 2014; Pagolu et al., 2016).

In the domain of politics, key topics include voter/public opinions about candidates for elections, governments, legislations, policies, officials/politicians, and political parties (Antonakaki et al., 2017; Jungherr et al., 2017; Murthy, 2015, Ramteke et al., 2016; Tumasjan et al., 2011; Unankard et al., 2014). The main reason that politics has generated a large number of sentiment analyses is that politics is public-opinion dependent and policy concerned, particularly during elections. Politicians, government agencies, and social/political organizations must constantly observe the sentiments of the public in order to win their support and/or to better serve the constituents they represent and govern. Furthermore, sentiment analysis before and during elections can provide valuable information for political parties and candidates to enhance their strategies for winning the election. The results of political sentiment analyses may also help predict election results, something that is of interest to not only the candidates and political parties involved but also the general public. The main sources of data for sentiment

[1] Sentiment analysis in academic writing here refers exclusively to those studies about the positivity/negativity in the published research articles expressed by their authors, not those studies that investigate the sentiments of the target subjects in an academic discipline for practical purposes, such as consumers' sentiments in business for the purposes of increasing sales or voters' sentiments in politics for the purpose of helping election candidates or predicting election outcomes.

analysis in politics include Twitter tweets and other social media postings as well as political candidates' interviews and speeches (Antonakaki et al., 2017; Jungherr et al., 2017; Liu & Lei, 2018).

Regarding the domain of healthcare/medicine, it is important to first note that healthcare is also a business, but a unique one because it has patients as its clients, medicines as its products, and treatments as its services provided by healthcare professionals (doctors and nurses). Hence, the main topics in this domain consist of patients' opinions and feelings about diseases and diagnoses/treatments, healthcare services/providers, and medications (Oscar et al., 2017; Seabrook et al., 2018; Wang, Liu & Zhou, 2020). The importance of sentiment analysis in this domain lies in the following facts. First, healthcare providers and drug companies need to know how patients and the public view their products and services so they can make necessary improvements. Second, understanding the emotions and feelings of patients, especially mental health patients, is extremely important for successful treatment. In short, sentiment analysis in healthcare deals largely with patients' feelings and opinions about illnesses, medications, healthcare services, and treatments. It is also important to note that in terms of data used, sentiment analysis in this domain often includes not only patients' and healthcare professionals' postings on medical discussion platforms and social media but also medical reports and other documents that are not publicly available (Denecke & Deng, 2015; Weissman et al., 2019).

As for the entertainment domain, so far most of the sentiment analyses have focused on movies and the main topics, as can be expected, are reviewers' opinions about movies, especially the acting/actors, cinematography, directing, and music involved. It is important to note that most of these topics can be considered aspects of a movie that are often included in the sentiment analysis at the aspect level, as opposed to at the document or sentence level (a discussion of the three levels of sentiment analysis will be given in Section 2). Concerning the data source for sentiment analysis in this domain, movie reviews appear to have been essentially the only data used. Regarding the importance of sentiment analysis in this domain, clearly the results of such analysis are highly valuable for the entertainment industry and movie viewers. This is because often different reviews of a movie may diverge to various extents in their evaluations and it would be particularly helpful to learn the overall opinion of the reviews (i.e., systematically generated opinion information via sentiment analysis). Such information can and has been used to predict movies' performance at the box office (Hur, Kang & Cho, 2016; Hu et al., 2018). This is important because most (i.e., 78 percent) of the movies produced each year are money losers (Davenport & Harris, 2009).

Academic writing, being an emerging area for sentiment analysis, has seen a few studies recently (Cao, Lei & Wen, 2020; Vinkers, Tijdink & Otte, 2015;

Weidmann, Otto & Kawerau, 2018). The areas of academic writing covered so far are limited to two: biomedical science (Cao et al., 2020; Vinkers et al., 2015) and political science (Weidmann et al., 2018) and the data used have been confined to journal articles and/or their abstracts. The methods employed have also been largely simple with a very small sentiment lexicon. However, these limited studies have all found a significant increase of positivity in academic writing and explored various interesting political and practical reasons for such an increase. Their results should have important practical implications for academic researchers.

1.3.2 Successes and Challenges

The tremendous efforts of researchers in the field of sentiment analysis have so far not only produced an enormous amount of work but also achieved some success in at least three areas. First, most of the existing studies have attained a sentiment identification accuracy ranging between 65 percent and 90 percent (Mukhtar, Khan & Chiragh, 2018; Rout et al., 2018; Zhang et al., 2014). This accuracy range, while clearly having room for enhancement, is fairly decent considering that the known reported accuracy or agreement of human sentiment judgment is 82 percent (Wilson, Wiebe & Hoffmann, 2005). Second, new fine-grained sentimental analysis tools and methods have been developed to help enhance the accuracy and effectiveness of sentiment identification and classification (e.g., Liang et al., 2015; Ren & Quan, 2012; Unankard et al., 2014). We will return to this point in Section 1.3.3. Third, some studies have demonstrated potentially useful practical applications of sentiment analysis, such as predicting election results, market performances, and product sales as well as identifying certain mental health conditions (e.g., Garcia, 2013; Giuntini et al., 2020, Sonnier, McAlister & Rutz, 2011; Tumasjan et al., 2011; Unankard et al. 2014).

For example, in the domain of business/finance, studies of how the sentiments in financial news (Garcia, 2013) and public opinions in tweets (Pagolu et al., 2016) forecast the movements of stock markets have demonstrated this predictive power. Similarly, both Liang et al.'s (2015) and Sonnier et al.'s (2011) studies on the relationship of customer reviews and product sales found that positive, negative, and neutral sentiments in customers' feedback all had a significant effect on sales. In the domain of politics, Tumasjan et al.'s (2011) sentiment analysis of Twitter messages concerning political parties and/or politicians during the 2009 German federal election revealed that the sentiments of voters' tweets about a political candidate were a good indicator of their political preferences and "the mere number of party mentions" accurately reflected the election result (p. 402). In another study, Unankard et al.

(2014) employed an approach that combined sentiment analysis of Twitter tweets with sub-event (i.e., an incident or crisis) identification to predict election results of the 2013 elections in Australia. They examined the effectiveness of the approach via a series of experiments and the results showed that their approach could "effectively predict the election results" (Unankard et al., 2014, p. 1).

In the domain of healthcare/medicine, Wang et al. (2020) developed a mental disorder identification model (MDI-Model) to help identify four different mental disorders, including depression and obsessive-compulsive disorder by analyzing the sequential emotion patterns of social media users over time in tweets written by disorder patients. Their results indicated high accuracy and efficiency of their MDI-Model in identifying the four types of mental disorders and the level of their severity. Seabrook et al. (2018), on the other hand, investigated how emotional states of "variability" and "instability" shown in Facebook and Twitter messages might reflect the severity of depression. Their results showed that instability in emotion was a significant indicator of more serious depressions while larger variability was a harbinger of lower depression severity.

While existing work of sentiment analysis has achieved some noticeable success as mentioned, there have also been some challenges and questions regarding its accuracy and predicting power as well as some other issues. In terms of sentiment identification accuracy, although the typical accuracy range is decent with a range of 65 percent to 90 percent as reported previously, much more work is needed to enhance this overall accuracy rate. Regarding the predictive power of sentiment analysis, despite some success as noted here, the results of a substantial number of studies (e.g., Gayo-Avello, 2012a, 2012b; Giuntini et al., 2020; Jungherr et al., 2017; Murphy, 2015; Rambocas and Pacheco, 2018; Weissman et al., 2019) have shown a lack or low level of such power, especially in the prediction of election results. Gayo-Avello (2012b), Jungherr et al. (2017), and Murphy (2015) all tried to use the results of sentiment analysis of election-related tweets to predict election outcomes, but they all failed. One reason for this failure, according to Murphy (2015, p. 816), was that the sentiments of political tweets were actually "more reactive rather than predictive." Even in the domain of business/finance, Rambocas and Pacheco's (2018) review of sentiment analysis studies in marketing published between 2008 and 2016 also revealed low validity and predictive power of such research. Similarly, in healthcare/medicine, Weissman et al.'s (2019) comparative study of six sentiment analysis methods applied to the texts of clinicians' encounter notes of patients with critical illness uncovered some serious issues with these methods, including their generally low predictive validity.