Sentiment Analysis: Theoretical Foundations and Conceptual Overview

Atul Kumar Pal* and Dr. Ashok Kumar**

**Research Scholar, faculty of Library and information science, SOSS, IGNOU, New Delhi Email: atul5311@gmail.com*
*** Assistant Professor, faculty of Library and information science, SOSS, IGNOU, New Delhi Email: ashokkr@ignou.ac.in*

Abstract

Sentiment analysis or opinion mining is a computational and interpretive method of detecting, isolating, and quantifying emotional expressions, subjective views, and expression of affection in text or other multimodal data. This conceptual paper gives a theoretical analysis of sentiment analysis, its interdisciplinary origins in the fields of linguistics, psychology, computer science, and social sciences, without referring to particular empirical applications. The discussion describes three basic levels of sentiment analysis: document-level, sentence-level and aspect-based methods. Sentiment is considered at the document level as a holistic property and sentiment-level analysis at the sentence level in the analysis of mixed polarity in a single text; and sentiment analysis at the aspect level, is focused on opinion detail at the attribute / entity level. Sentiment is theorized in the dimensions of polarity (positive, negative, and neutral), intensity, and types of emotion (e.g. happiness, anger, sadness) based on psychological models, including Valence-Arousal-Dominance theory and emotion appraisal theories.

This paper describes the parallels between lexicon-based models, based on a set of predefined sentiment dictionaries and valence scores, and machine learning models, such as supervised, unsupervised, and deep learning models, which use contextual embeddings and transformer models. Some of the difficulties encountered comprise, sarcasm detecting, dealing with negation, domain dependency, cross cultural differences in expression of emotion, and the subjectivity nature of the sentiment tags. They point out ethical implications, including the amplification of bias in training data and the privacy aspects of a large-scale opinion mining as the problems of serious theoretical concern. Conclusively, this theoretical inquiry places the sentiment analysis as a mediating point between the emotional experience of humans and computational modelling, and it also highlights the importance of the rule-based linguistic knowledge combined with data-driven methods. This paper establishes the basis of building more context-sensitive, interpretable, and theoretically rigorous sentiment analysis paradigm in future studies by demystifying important theoretical constructs and methodological trade-offs.

Keywords: Sentiment analysis, opinion mining, natural language processing, emotion theory, sarcasm detection, lexicon-based methods

1. INTRODUCTION

1.1 Sentiment Analysis in the Form of Definition and Scope.

Sentiment analysis is a scientific study and computational analysis of opinions, emotions, and attitudes embedded in communication channels. It is currently spreading across multiple modes, such as pictures, video, and audio. It tries to measure the qualitative, and human speech is translated into measurable facts that can be used to make decisions across different areas. It has a broad scope today and includes not only the explicit manifestations of positivity or negativity but also subtler elements, such as irony, ambivalence, and context-dependent nuances. When examining this field, it becomes evident that it functions as more than just a channel for information; it operates entirely within the realm of emotions. These emotions arise from the intent to express oneself through the use of words, phrases, and sentences.

To define sentiment analysis accurately, it is essential to distinguish it from related fields such as affective computing and stance detection. Affective computing aims to detect emotions based on physiological signals, whereas stance detection identifies an individual's perspective or position regarding a controversial issue. Sentiment analysis focuses on understanding the emotional direction (polarity) and underlying tone within linguistic data, viewing emotions not as discrete categories but as a continuum. Its multimodal extensions involve the use of semiotics, where visual or sound technology supplements textual gestures, such as facial expressions in video, and adds a layer of sarcasm to the text. This vast range indicates the need for theoretically flexible, integrative models that can process diverse types of data with fidelity to the complexity of human emotions.

1.2 Historical Evolution

The conceptual basis of sentiment analysis can be traced back to ancient rhetorical literature, where Aristotle, in his treatise Rhetoric, discussed pathos, or the appeal to emotions, as a form of persuasion. The field then developed in the middle of the 20th century with linguistic studies of evaluative language, including Osgood's semantic differential scales in the 1950s, which were used to quantify connotations along the dimensions of evaluative, potency, and activity. It started in the 1990s, with initial NLP work, in part due to the AI boom, during which rule-based systems attempted to analyse subjective texts.

In the 2000s, there was a boom with the emergence of Web 2.0 and user-generated content, which led to theoretical shifts towards data-driven approaches—such as the 2008 survey by Pang and Lee, which conceptualised a classification problem informed by machine learning theories. The 2010s were marked by the integration of emotion models and the deep learning revolution of the late 2010s, which advanced contextual theories through embeddings. Sentiment analysis has been theoretically explored to be explainable and ethical today, as is generally true of AI. Compared to simple polarity detection, this shift points towards a framework that is more advanced and context-aware. In this exemplar, it adorns the interaction between technology and theoretical development.

1.3 Relevance in Modern Society

Sentiment analysis has deep meaning in modern society, and the theory is relevant because it can connect the personal with the common understanding. It facilitates the use of theoretical models of consumer behaviour in marketing, in which the aggregate sentiments determine market trends and brand loyalty. Politically, it educates in the theories of public opinion, enabling one to analyse discourse in the democratic process, including during elections, where societal momentum can be observed. In this study, we used psychological theories to identify emotional distress on social media, which may inform interventions during the early stages.

Socially, it combats misinformation by theorising the role of sentiment in the spread of virality, with negative emotions increasing it. From an economic perspective, stock market forecasts are based on news sentiment, which is grounded in behavioural finance theory. Nonetheless, this has its own riders: an excess reliance on sentiment may oversimplify complex human interactions, and a theoretical protective measure against misinterpretation is needed. All in all, the role of sentiment analysis in society highlights its potential as an empathy tool in the digital era, which improves human-computer interaction and requires ethical vigilance.

This introductory section serves to lay the foundation for a comprehensive theoretical exploration, covering topics such as interdisciplinary origins, levels of analysis, concepts, approaches, problems, ethical considerations, and future dimensions. This paper will bring long-term value to both the researcher and the practitioner by not concentrating on the implementations, but on the concepts.

2. RELEVANCE IN MODERN SOCIETY

2.1 Linguistic Foundations

Sentiment analysis is based on linguistics and its focus on how language reflects assessment and emotion. Theorising language as a social semiotic system, where interpersonal meanings are achieved in the form of attitude subsystems of affect (emotions), judgment (ethics), and appreciation (aesthetics), Halliday formulated (and Martin and White expanded) a theory of language called Systemic Functional Linguistics (SFL) (see in their Appraisal Framework, 2005). In sentiment analysis, this would mean finding attitudinal lexis; words such as brilliant (positive appreciation), and deceitful (negative judgment) and their grammatical instantiations, such as nominalisations or modal auxiliaries hardening or strengthening sentiment.

Pragmatic theories such as Grice's (1975) Cooperative Principle emphasise implicature, which conveys more than it states, such as sarcasm breaching the principle of quality. Semantic theories such as the Natural Semantic Metalanguage of Wierzbicka (1996) assume that emotions can be linguistically decomposed into primitives to support cross-linguistic sentiment modelling. Discourse analysis provides layers; sentiment is perceived through narrative structures, and the impact of coherence and cohesion on the general tone is applied. These language origins ensure that the sentiment analysis will be aware of the systematicity of language and will not be an immature word-counting tool that disregards syntactic and contextual dependencies.

2.2 Psychological Underpinnings

Psychology applies sentiment analysis to theories of emotion, offering a humanistic perspective. Simple emotion theories, such as Ekman's six universal emotions (1992), provide categorical theories of labelling feelings, theorising that there is a cross-cultural correlation between facial and linguistic manifestations of emotion. Dimensional theories such as Russell's (1980) Circumplex Model and the Valence-Arousal-Dominance (VAD) conceptualisation of affect (Mehrabian, 1996) conceptualise affect using continuous spaces: valence to polarity, arousal to intensity, and dominance to control, making sentiment a plot of vectors.

Theories of appraisal (Scherer, 2001; Lazarus, 1991) assume that an emotion is a product of cognitive processes that evaluate events in terms of novelty, pleasantness, goal relevance, coping potential, and norm compatibility. In sentiment analysis, this means emotion is inferred from context (e.g., event descriptions) rather than from lexical cues. Social psychology has also contributed to theories of attitude formation, such as the Theory of Planned Behaviour developed by Ajzen (1991), which posits that attitudes are based on beliefs, norms, and intentions. These psychological principles underpin the sentiment model, grounded in empirical human behaviour, with a warning against excessive generalisation and in favour of multidimensional forms.

2.3 Computer Science Perspectives

Sentiment is operationalized in computer science in terms of algorithmic theories. Parsing to extract sentiments is based on formal language theories as a by-product of the hierarchy of Chomsky (1956) and text is viewed as generative grammars. Modeling sentiment as uncertain cues Probabilistic models inspired by the information theory of Shannon (1948) polarity classification by Bayesian inference.

Theories of machine learning are split between supervised (learning with labeled data, made by statistical learning, by Vapnik, 1995) and unsupervised (clustering, algorithms such as k-means, based on pattern recognition). Deep learning is inspired by connectionism (Rumelhart et al., 1986), neural networks are based on cognitive hierarchies. The attention is presented by transformer theories (Vaswani et al., 2017, conceptually), which are theoretically interested in the relevant settings such as human cognition. Such views lay stress on scalability and flexibility, but bring out trade-offs of interpretability and performance.

2.4 The contributions of social sciences

Social sciences put sentiment into a social construct. Opinions are regarded by sociological theories such as the Habermasian public sphere (1989), as deliberative discourse, in which sentiment entails the formation of collective will, by way of aggregation. Universality is confronted through anthropological perspectives such as the cultural construction of emotion by Lutz (1988) in which he observes that cultures are involved in the expression of affect, e.g., an amae in Japanese with no direct English counterparts.

Political science combines sentiment and agenda-setting theory (McCombs and Shaw, 1972), in which the media sentiments shape the priorities of the masses. There is a theory of communication like the spiral of silence (Noelle-Neumann, 1974) and it describes the way negative feelings discourage the minority opinions. These roots highlight the importance of sentiment in the power relations, calling on models to take bias and diversity into consideration.

3. COMPARATIVE ANALYSIS AND INTEGRATION

The three main types of sentiment analysis document-level, sentence-level, and aspect-based sentiment analysis have various theoretical benefits, which serve the purposes of the three kinds of analysis, but also have predetermined trade-offs, which have to be carefully taken into consideration when designing a model.

Document-level analysis is especially good at offering macro-level information, capturing the overall affective tone of a whole piece of text, and thus is easily used on a large scale, such as tracking brand attitudes within large collections of consumer feedback, or a summary of a social media campaign, or the overall attitude to a brand in aggregates of consumer feedback. Its simplicity lies in considering sentiment as a unique, gestal-like feature, which is consistent with cognitive theories of holistic perception when the global impression prevails over the small details. Nonetheless, this methodology intrinsically loses granularity, and may not capture internal contradictions or changes of heart in the same document, and can cause aggregative misleadingness in multi-faceted and multi-layered texts.

Sentence-level analysis, on the other hand, finds a golden mean ground that allows the accommodation of ambivalent polarities and emotional swings that are common in natural discourse. Breaking texts down into manageable syntactic units, it is consistent with sequential processing theories in cognitive psychology and discourse coherence theories, enabling theorists to describe the development of sentiments on a sequential basis such as an original positive judgment being replaced by criticism in a review. It is a more accurate level that is not computationally intensive, which is why it is suitable when medium-level detail is needed, such as to trace sentiment arc of stories or dialogues. But it still needs post-processing aggregation strategies to assemble document-level meaning, which presents the possibility of ambiguity in theoretical terms in weighting mechanisms (e.g. giving more emphasis to concluding sentences as more reflective of final judgment).

Aspect-based sentiment analysis (ABSA) is known to be much more nuanced, breaking down sentiments about particular entities, attributes, or features in a text, e.g. the difference between an opinion on the battery life and an opinion on the design aesthetics in a smartphone review. Based on theories of entity-relation semantics and selective attention, ABSA is a theoretically accurate reflection of human evaluative behaviour, which is directed at discrete components instead of indiscriminate wholes, and, therefore, offers insights with fine grain that are important to such applications as product improvement advice or customized marketing. However, it has a depth that comes at the cost of greater complexity, and the aspect recognition, opinion association and disambiguation of implicit references are all challenging.

A comparison of these levels can see a clear sequential increase in priority of the breadth and efficiency in high level overviews, and in the balance between breadth and detail in sentence-level, and in the fine-tuning of targeted understanding in ABSA. Theoretical integration approaches do emphasize an increasing trend of hierarchical approaches starting from the contextual base-line setting on a document level, the sentence level parsing, and culminating in the aspect level extraction. These hierarchical designs, which are inspired by recent conceptual developments in layered neural representations and tree-structured dependencies, are more flexible, allowing for dynamically traversing across granularities, and can be better leveraged to more effectively probe and analyse a richer and richer text, as well as an adaptable analysis pipeline. This integrative resolution has the additional advantage because it not only resolves weaknesses of the individual level, but also supports the synergy with multidimensional sentiment structures, which are able to be propagated hierarchically to provide richer and contextually informed representations of the categories of polarity, intensity and emotion. Finally, these integration theories endorse scalable, flexible frameworks that are more likely to approximate human interpretive richness in the computational sentiment modelling.

4. COGNITIVE FUNDAMENTALS OF AFFECTION

Sentiment is a highly multidimensional phenomenon, which can best be conceptualized as a multidimensional construct using multiple interlocking dimensions that can describe both the qualitative and quantitative aspects of the phenomenon, and does not reduce it to the simplistic binary notions of good and bad.

4.1. Polarity Dimension

The basic dimension is polarity, which describes sentiment on a continuum of positive/negative/neutral orientations, and which is deeply rooted in traditional valence theories that represent evaluations in terms of approach motivation (positive) or avoidance motivation (negative), and neutrality as the affective equilibrium or valence-free position. This dimension is based on semantic differential paradigms, in which evaluative meaning is mapped onto bipolar scales, which allows the computation of the mapping of lexical and phrasal cues to directional sentiment. In theory, positive polarity activates motivational inclinations towards contact or attraction, negative polarity activates withdrawal or repulsion whereas neutral positions accept factual utterances, ambivalence or moderate views. The issue is the ambiguity that exists in it, such as the varying meaning it has according to context (e.g., the derogatory and the admiring use of sick as an expression of slang), which the traditional discrete categories of the language cannot capture. Theoretical solutions have gone to a higher level which includes the use of fuzzy set theory, probabilistic or graded assignment, showing partial membership in classes of polarities, and in this sense the approach to vagueness is more realistic than the crisp boundaries.

4.2. Intensity Dimension

Circumplex models of affect (e.g., on to polarity) include intensity dimensions as the experiential strength or magnitude of sentiment, as well as the dimension of arousal, which reflects the activation level of the emotion (but not its valence). Linguistic devices such as intensifiers or repetition, exclamation or capitalization are scalar modifiers, and can be theoretically viewed as multiplicative or additive functions which modify baseline valence scores. Psychological models go further to posit theorize intensity as being dependent on the perceived event importance, goal relevance or novelty, which means that contextual information such as personal stakes or cues of urgency should be included in computational scaling. The dimension can be used to differentiate between mild displeasure and strong anger, or a low level of satisfaction and high level of joy, and may be useful for applications requiring emotional modulation, such as identifying a crisis or monitoring a therapy session.

4.3. Emotion Categories

In addition to valence and arousal, the categories of emotion lead to either discrete or intermediary affective states, such as joy, anger, sadness, fear, surprise, and disgust that are drawn from the psych evolutionary wheels and promote primary emotions as primitively combinable. These types enable more detailed categorization that enables models to deduce particular psychological reactions out of textual patterns through appraisal mappings that correlate descriptions material with cognitive judgments of events. However, the cultural differences present a number of theoretical problems, including the need for adaptive taxonomies to account for the demonstrative differences between emotion perception in different sociocultural settings, which would involve removing ethnocentric bias from the relativist views and multilingual emotion ontologies.

4.4. Multidimensional Integration

Theoretical power of the latter comes from these dimensions when combined with each other in vectors space representations, with sentiment having points or trajectories in a continuous affective space that, in principle, are semantic vector models relaxed to emotional space. This integration enables subtle and holistic characterizations, such as, high-arousal negative dominance (which denotes rage) and low-arousal (which denotes melancholy) and also similarity computation and trajectory analysis of texts. The problems have remained, especially relating to assuring orthogonality between dimensions to avoid collinearity or redundancy and calibration of scales in different datasets. Recent theoretical developments focus on regularization methods, dimensionality analysis and psychologically motivated priors in order to produce more interpretable and generalizable multidimensional models spanning between categorical and continuous emotion paradigms.

5. METHODS OF SENTIMENT ANALYSIS

5.1 Lexicon-Based Approaches

The lexicon-based approaches are based on the limited set of curated dictionaries, which provide the pre-determined value scales, sentiment polarities, or emotion labels on individual words, phrases, or n-grams, which provides a theoretically transparent, rule-based underpinning based on lexical semantics. Such methods consider sentiment as compositionally derived out of atomic units, including rules to negation flip, scale intensifiers, emojis and punctuation emphasis to optimize aggregate scores. Construction is usually performed by an expert, through crowdsourcing, or by bootstrapping based on seed terms using distributional semantics, and wherever feasible, making adaptations to domain. Although theoretically static, domain-independent on the lexical level, their explain ability is bright in explainable outputs e.g. highlighting contributing terms but fails in deep contextual captivity or the changing language use.

5.2 Machine Learning Paradigms

Sentiment Within supervised paradigms, sentiment is a classification or regression problem, and the use of labeled data to train feature-sentiment maps using algorithms based on the statistical learning theory, including support vector machines that optimize the separation of margins in high dimensional spaces. Unsupervised approaches take advantage of clustering algorithms to reveal latent affective images without explicit supervision according to pattern recognition logic to cluster similar expressions according to distributional similarity. Deep learning builds upon this with hierarchical representations: word representations encode semantic similarities, with transformer networks using the self-attention architecture to approximate long-range contextual dependencies, which theoretically can be thought of as human-like attention to salient objects.

Table 1

Approach	Strengths	Weaknesses	Theoretical Basis
Lexicon	High interpretability, minimal data requirements, domain-agnostic baselines	Ignores deep context, sarcasm, evolving slang	Lexical semantics, compositional rules
Supervised ML	Strong accuracy with sufficient labeled data, feature engineering flexibility	Heavy reliance on quality/quantity of labels, overfitting risks	Statistical learning theory, generalization bounds
Unsupervised ML	No label dependency, discovers emergent patterns	Lower precision, difficulty in validation	Pattern discovery, clustering geometry
Deep Learning	Superior contextual understanding, handles complexity via layers	Model opacity ("black box"), high computational demands	Connectionism, hierarchical representation learning

5.3. Hybrid Approaches

Theoretically, hybrid approaches combine the determinism of lexicon knowledge, with the inductive ability of data-based learning techniques, such as adding lexical features to neural networks, or using inferred lexicon to weakly supervise machine learning forecasts, and/or using the lexicon output as ground truth to restrict the machine learning output as a form of ensemble fusion. A balance of explicit learning of language and statistical adaptation loosens the constraints of the paradigms of the individuals, resulting in more robust, explainable and transferable paradigms that are highly consistent with the human reasoning system of rules and experience.

6. SENTIMENT ANALYSIS IS DIFFICULT DUE TO THE FOLLOWING CHALLENGES

6.1. Sarcasm Detection

One of the modes of pragmatic inversion is sarcasm, which refers to a superficial opposition between literal expectations and the context that they are set against. Theoretical detection is based on the frames of humour and irony, which emphasize the difference between the context, and the addition of pragmatic clues such as the user's history, discourse markers or world. Multi-modal signals or advanced contextual embeddings are solutions, however, the solutions are not so easy to detect because it is subtle and embedded within the culture.

6.2. Handling Negation

The negation radically changes the scope of polarity, which is regulated by linguistic theories, defining the syntactic coverage of negation (e.g. over predicates or modifiers). Dependency parsing is a theoretical tracing of inversion paths although ambiguities with nested or implicit negations still remain and will require complex integration of logical semantics.

6.3. Domain Dependence

Terminology is particularly different in area - medical positive and consumer positive, thus questioning transferability. The theories of transfer learning suggest the domain adaptation through shared representations or adversarial alignment of distributional gaps.

6.4. Cultural Variability

Emotional expression is culturally made, where relativist frameworks emphasize divergent metaphors, norms and display rules. Multilingual theories also promote inclusive embeddings and cross-cultural datasets in order to promote fair models.

7. SUBJECTIVITY OF LABELS

Subjectivity in the annotation process can be due to personal biases and inter-annotator variance, which are theoretically resolved by building consensus, by fuzzy labeling, or by inter-subjectivity models which focus on negotiated meaning. Theoretical solutions all serve the purpose of robustness through interdisciplinary synthesis, adaptive designs and constant improvement in order to address these undying issues.

Present the analysis of collected data systematically using tables, graphs, and statistical tools. Explain how the data support or reject the hypotheses or meet the research objectives. Interpret results logically and relate them to existing theories or prior findings.

8. ETHICAL CONSIDERATIONS

8.1. Bias Amplification

The biases of the society are usually ingrained in training data and modeled, continuing to propagate gender, racial, or cultural inequalities. Fairness theories support various sourcing, debiasing algorithms, and equity auditing in order to offset the disproportionate effects.

8.2. Privacy Implications

Opinion mining at scale poses a threat to unauthorized surveillance or profiling, which is handled by ethics of care that put more emphasis on informed consent, anonymization, and differential privacy and data minimization to protect the individual autonomy.

8.3. Misuse Potential

Sentiment tools can be used to manipulate, influence or exploit sentiment in order to propagate information or cause emotional manipulation. Regulatory theories demand regulation, transparency requirements and accountability provisions to rein in dangerous usage. Responsible theorizing requires application of integrated principles of justice, beneficence, and non-maleficence which ensures that sentence analysis does not harm the society by worsening the ills.

9. FUTURE DIRECTIONS

The sentiment analysis horizon is directed at more advanced hybrid multi-modal models that combine text with visual, auditory and gestural representations to holistically represent affectivity, the fusion theories are applied to address cross-modal congruencies and incompatibilities. Explainable AI proves to be paramount with methods such as attention visualization, feature attribution and counterfactual reasoning becoming more interpretable and enhancing trust in opaque deep architectures. The cultural inclusivity requires a priority with multilingual datasets, the training which is biased and the emotion ontologies which are localized to overcome Western-centric assumptions. New interactions with quantum computing are expected to provide exponential improvements in solving complex contextual spaces, and can transform the optimization and representation learning of NLP. These paths are all imaginative of more compassionate, fair, and sentient-friendly models of sentiment that are advanced concurrently with the computational progressions.

10. CONCLUSION

Sentiment analysis is the concept of blending together a wide range of disciplines, including linguistics, psychology, computer science and social theory, which forge significant and important links between the human fluxus and the computational model. Through its strict approach in filling gaps of theory in the field of granularity, contextuality, bias, and ethics, and the uptake of integrative and hybrid developments, the area is propelled towards more reflexive, context-driven, culturally attuned, and interpretable models. The development of analytical skills is not only important, it could be useful to improve empathy and human understanding in H-AI relationships in all areas of society.

REFERENCES

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179-211. https://doi.org/10.1016/0749-5978(91)90020-T

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. G. A. Bower (Ed.), *Recent Advances in Learning and Motivation* (Vol. 8, pp. 47-89). Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1

Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and Machine Learning.* https://mitpress.mit.edu/9780262048613/fairness-and-machine-learning/

Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory, 2*(3), 113-124. https://doi.org/10.1109/TIT.1956.1056813

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273-297. https://doi.org/10.1007/BF00994018

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion, 6*(3-4), 169-200. https://doi.org/10.1080/02699939208411068

Frijda, N. H. (1986). The Emotions. *Cambridge University Press*. ISBN: 978-0521316002 https://www.cambridge.org/core/books/emotions/

Gilligan, C. (1982). In a Different Voice. *Harvard University Press.* ISBN: 978-0674445444 https://www.hup.harvard.edu/books/9780674445444

Grosz, B. J., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics, 12*(3), 175-204. https://aclanthology.org/J86-3001

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, *5*(2), 199-220. https://doi.org/10.1006/kacq.1993.1008

Gunning, D. (2017). Explainable artificial intelligence (XAI). *Defence Advanced Research Projects Agency*. https://www.darpa.mil/program/explainable-artificial-intelligence

Habermas, J. (1989). The Structural Transformation of the Public Sphere. *MIT Press*. ISBN: 978-0262581080 https://mitpress.mit.edu/9780262581080/the-structural-transformation-of-the-public-sphere/

Horn, L. R. (1989). A Natural History of Negation. *University of Chicago Press*. ISBN: 978-0226353531.

Kövecses, Z. (2000). Metaphor and Emotion. *Cambridge University Press*. https://doi.org/10.1017/CBO9780511620904

Lazarus, R. S. (1991). Emotion and Adaptation. *Oxford University Press*. ISBN: 978-0195069945 https://global.oup.com/academic/product/emotion-and-adaptation-9780195092660

Lutz, C. A. (1988). Unnatural Emotions. *University of Chicago Press*. ISBN: 978-0226497228 https://press.uchicago.edu/ucp/books/book/chicago/U/bo3624041.html

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, *1*, 281-297. https://projecteuclid.org/euclid.bsmsp/1200512992

Martin, J. R., & White, P. R. (2005). The Language of Evaluation. *Palgrave Macmillan*. https://doi.org/10.1057/9780230511910

McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly*, *36*(2), 176-187. https://doi.org/10.1086/267990

Mehrabian, A. (1996). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, *14*(4), 261-292. https://doi.org/10.1007/BF02686918

Mikolov, T., et al. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*. https://arxiv.org/abs/1301.3781

Noelle-Neumann, E. (1974). The spiral of silence: A theory of public opinion. *Journal of Communication*, *24*(2), 43-51. https://doi.org/10.1111/j.1460-2466.1974.tb00367.x

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345-1359. https://doi.org/10.1109/TKDE.2009.191

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*(1-2), 1-135. https://doi.org/10.1561/1500000011

Raskin, V. (1985). Semantic Mechanisms of Humor. D. Reidel. https://doi.org/10.1007/978-94-009-6477-8

Rumelhart, D. E., McClelland, J. L., & PDP Research Group. (1986). *Parallel Distributed Processing. MIT Press*. ISBN: 978-0262181204 (Vol. 1) https://mitpress.mit.edu/9780262181204/parallel-distributed-processing/

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161-1178. https://doi.org/10.1037/h0077714

Scherer, K. R. (2001). Appraisal processes in emotion. *Handbook of Affective Sciences* (pp. 572-595). ISBN: 978-0195155563

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379-423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Treisman, A. M. (1964). Selective attention in man. *British Medical Bulletin*, *20*(1), 12-16. https://doi.org/10.1093/oxfordjournals.bmb.a072291

Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, *37*, 141-188. https://doi.org/10.1613/jair.2934

Vaswani, A., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*. https://arxiv.org/abs/1706.03762

Vapnik, V. (1995). The Nature of Statistical Learning Theory. *Springer*. https://doi.org/10.1007/978-1-4757-2440-0

Wierzbicka, A. (1996). Semantics: Primes and Universals. *Oxford University Press.* ISBN: 978-0198700135 https://global.oup.com/academic/product/semantics-9780198700135

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338-353. https://doi.org/10.1016/S0019-9958(65)90241-X