

Table 4. Taxonomy of Text Analysis Tools.

Approach	Common Tools	Research Questions	Benefits	Limitations and Complexities	Marketing Examples
Entity (word) extraction: Extracting and identifying a single word/n-gram	<ul style="list-style-type: none"> Named entity extraction (NER) tools (e.g., Stanford NER) Dictionaries and lexicons (e.g., LIWC, EL 2.0, SentiStrength, VADER) Rule-based classification Linguistic-based NLP tools Machine learning classification tools (conditional random fields, hidden Markov models, deep learning) 	<ul style="list-style-type: none"> Brand buzz monitoring Predictive models where text is an input Extracting psychological states and traits Sentiment analysis Consumer and market trends Product recommendations 	<ul style="list-style-type: none"> Can extract a large number of entities Can uncover known entities (people, brands, locations) Can be combined with dictionaries to extract sentiment or linguistic styles Relatively simple to use 	<ul style="list-style-type: none"> Can be unwieldy due to the large number of entities extracted Some entities have multiple meanings that are difficult to extract (e.g., the laundry detergent brand "All") Slang and abbreviations make entity extraction more difficult in social media Machine learning tools may require large human-coded training data Can be limited for sentiment analysis 	<ul style="list-style-type: none"> Lee and Bradlow (2011) Berger and Milkman (2012) Ghose et al. (2012)^a Tirunillai and Tellis (2012) Humphreys and Thompson (2014)^a Berger, Moë, and Schweidel (2019) Packard, Moore, and McFerran (2018)
Topic extraction: Extracting the topic discussed in the text	<ul style="list-style-type: none"> LSA LDA PF LDA2vec word embedding 	<ul style="list-style-type: none"> Summarizing the discussion Identifying consumer and market trends Identifying customer needs 	<ul style="list-style-type: none"> Topics often provide useful summarization of the data Data reduction permits the use of traditional statistical methods in subsequent analysis Easy-to-assess dynamics 	<ul style="list-style-type: none"> The interpretation of the topics can be challenging No clear guidance on the selection of the number of topics Can be difficult with short text (e.g., tweets) 	<ul style="list-style-type: none"> Tirunillai and Tellis (2014) Büschen and Allenby (2016) Puranam, Narayan, and Kadiyali (2017) Berger and Packard (2018) Liu and Toubia (2018) Toubia et al. (2019) Zhong and Schweidel (2019) Ansari, Li, and Yang (2018)^a Timoshenko and Hauser (2019) Liu, Singh, and Srinivasan (2016)^a Liu, Lee, and Srinivasan (2019)^a
Relation extraction: Extracting and identifying relationships among words	<ul style="list-style-type: none"> Co-occurrence of entities Handwritten rule Supervised machine learning Deep learning Word2vec word embedding Stanford Sentence and Grammatical Dependency Parser 	<ul style="list-style-type: none"> Market mapping Identifying problems mentioned with specific product features Identifying sentiment for a focal entity Identifying which product attributes are mentioned positively/negatively Identifying events and consequences (e.g., crisis) from consumer- or firm-generated text Managing service relationships 	<ul style="list-style-type: none"> Relaxes the bag-of-words assumption of most text-mining methods Relates the text to a particular focal entity Advances in text-mining methods will offer new opportunities in marketing 	<ul style="list-style-type: none"> Accuracy of current approaches is limited Complex relationships may be difficult to extract It is advised to develop domain-specific sentiment tools as sentiment signals can vary from one domain to another 	<ul style="list-style-type: none"> Netzer et al. (2012) Toubia and Netzer (2017) Boghrati and Berger (2019)

^aReference appears in the Web Appendix.

Table 4. Taxonomy of Text Analysis Tools.