Sentiment Analysis - 2

Text Analytics

Ramaseshan Ramachandran

Definitions

Sentiment Analysis	The process of assigning a polarity, positive, negative or neutral, to an unstructured text	Lexicon-based approach Supervised classification approach - statistical and machine learning
Semantic Orientation	Measuring subjectivity and opinion from an unstructured text	Identifying polarity and strength of words and/or phrases
Opinion Segment	Capturing individual sentiment/ segment from a compound sentence	It is a beautiful phone case but it is also hard to remove
Explicit aspect	The feature appears in the sentence	It is a very cute ***case***
Implicit aspect	The feature does not appear in the text	Arrived broken and very flimsy ***package and case?***
Subjectivity	Expressing desires, beliefs, proclamations, preferences, etc	Don't believe that these screen protectors have glue in them

Sentiment Definition

Given a text find (t_i, a_{ij}, o_{ijk}) ①

It can be extended with more parameters when the opinion was created, who created, comparing entity, etc.

t_i is the target entity or product in the text and $t_i \in T$ where
$T = \{t_1, t_2, t_3, \ldots, t_n\}$

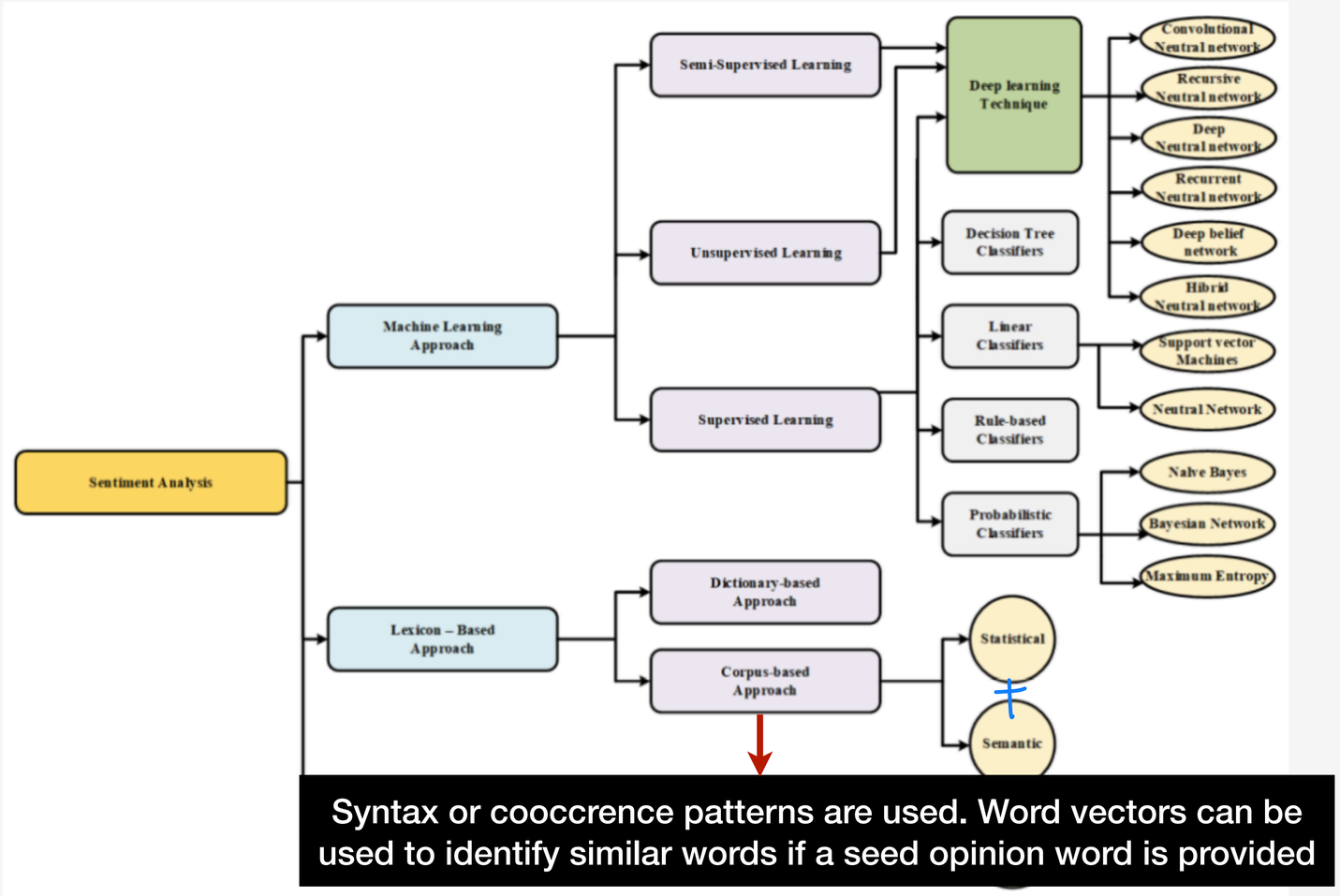
a_{ij} is the set of aspects for the product t_i

O_{ijk} is the opinion expressed on the aspect a_{ij}. The scale of O_{ijk} could be one of the values such as positive, neutral or negative of could be a value
$[a, b] = \{x \mid a \leq x \leq b\}$

We may include other parameters such as time, owner of the opinion, product version, platform where the opinion was recorded, country, etc.

Using ①, we can perform various analysis on the text

Sentiment Classification Techniques



Source: Nhan Cach Dang María N. Moreno-García and Fernando De la Prieta , "Sentiment Analysis Based on Deep Learning: A Comparative Study", https://www.mdpi.com/2079-9292/9/3/483/htm

Lexicon-based Model

- Dictionary used to classify the word based on sentiment lexicon
- Requires no training data, but is constructed from a generalizable, valence-based, human-curated gold standard[1]
- Libraries such as NLTK, Valence Aware Dictionary for sentiment Reasoning (VADER), SentiWordnet
- $$S = \frac{number\ of\ positives - number\ of\ negatives}{total\ number\ of\ words}$$

$$Sentiment = \begin{cases} positive, & if\ S = 1 \\ neutral, & if\ S = 0 \\ negative, & if\ S = -1 \end{cases}$$

[1] Hutto, C.J. & Gilbert, Eric. (2015). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014.

VADER based Sentiment Analysis

- Requires no training data, but is constructed from a generalizable, valence-based, human-curated gold standard[1]
- Every lexicon has a score depending on its intensity [-4,4] -using 5 rules
  1. Punctuation mark
  2. Capitalisation
  3. Degree of modifiers/intensifiers
  4. Contrastive conjunction,
  5. Tri-gram before sentimentally loaded phrase
- The sentiment score is calculated by summing up the sentiment scores of each VADER-dictionary-listed word in the sentence $x = \sum_i s_{w_i}$
- The ``compound`` score is computed by $\frac{x}{\sqrt{\sqrt{x^2 + \alpha}}}$

[1] Hutto, C.J. & Gilbert, Eric. (2015). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014.

Lexicon-based SA - Vader Results

It is a beautiful phone case but it is also hard to remove

compound: 0.2144, neg: 0.114, neu: 0.712, pos: 0.174,

It is a very cute case

compound: 0.5095, neg: 0.0, neu: 0.548, pos: 0.452,

Arrived broken and very flimsy

compound: -0.4767, neg: 0.437, neu: 0.563, pos: 0.0,

Don’t believe that these screen protectors have glue in them

compound: 0.0, neg: 0.0, neu: 1.0, pos: 0.0,

I chose this case because it was beautiful

compound: 0.5994, neg: 0.0, neu: 0.606, pos: 0.394,

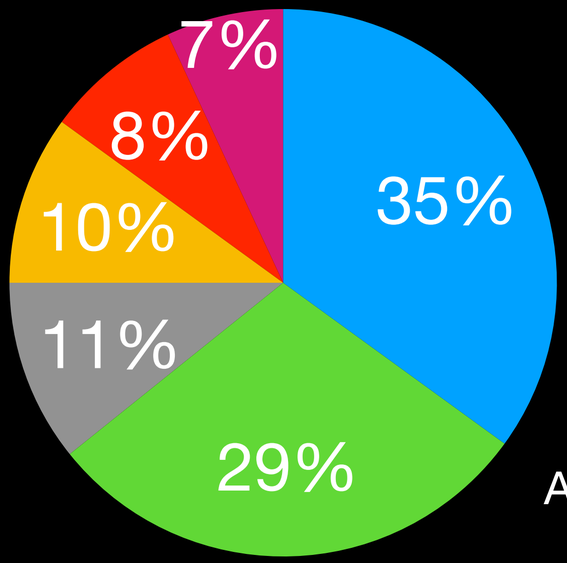
I liked it because it was cute, but the studs fall off easily and to protect a phone this would not be recommended. Buy if you just like it for looks

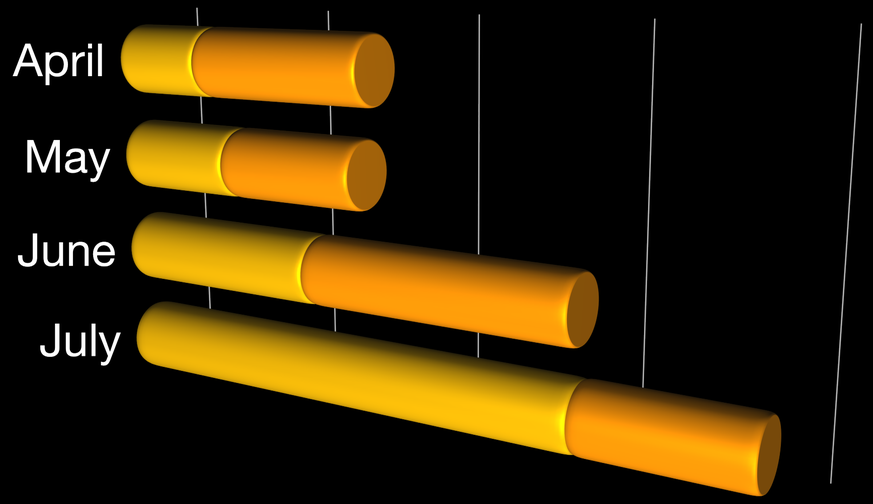
compound: 0.8948, neg: 0.049, neu: 0.597, pos: 0.354,

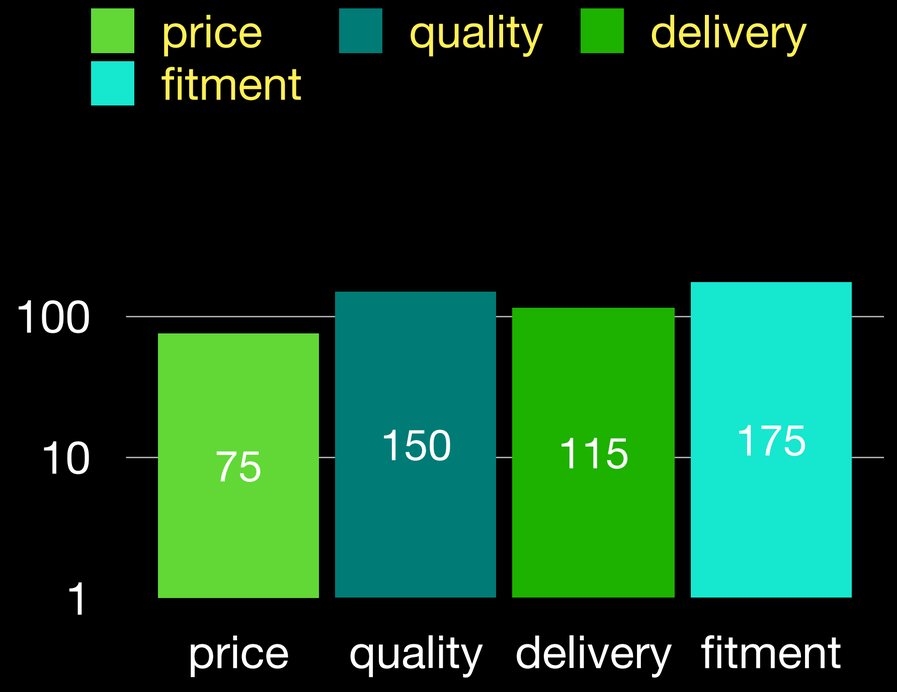
it works good but after few weeks it just stop working. the light just on but not charging

compound: -0.2144, neg: 0.135, neu: 0.771, pos: 0.094,

Visualisation of Sentiments





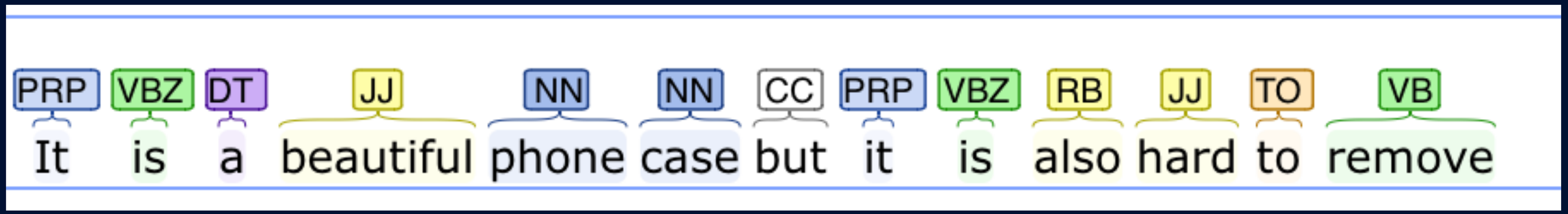


POS tags

CC	Coordinating conjunction
CD	Cardinal number
DT	Determiner
EX	Existential there
FW	Foreign word
IN	Preposition or subordinating conjunction
JJ	Adjective
JJR	Adjective, comparative
JJS	Adjective, superlative
LS	List item marker
MD	Modal
NN	Noun, singular
NNS	Noun, plural
NNP	Proper noun, singular
NNPS	Proper noun, plural's
PDT	Predeterminer
POS	Possessive ending
PRP	Personal pronoun
PRP$	Possessive pronoun
RB	Adverb
RBR	Adverb, comparative
RBS	Adverb, superlative
RP	Particle
SYM	Symbol
TO	to
UH	Interjection
VB	Verb, base form
VBD	Verb, past tense
VBG	Verb, gerund or present participle
VBN	Verb, past participle
VBP	Verb, non3rd person singular present
VBZ	Verb, 3rd person singular present
WDT	Whdeterminer
WP	Whpronoun
WP$	Possessive whpronoun
WRB	Whadverb

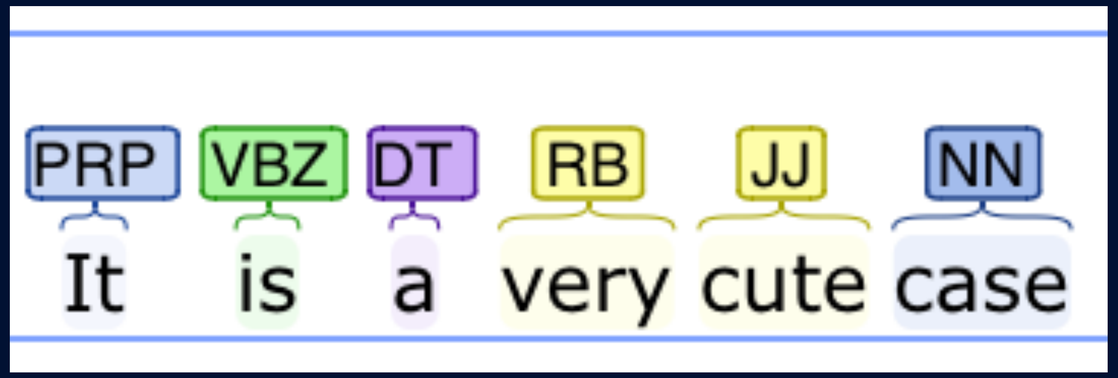
Little bit of Grammar :)

It is a beautiful phone case but it is also hard to remove



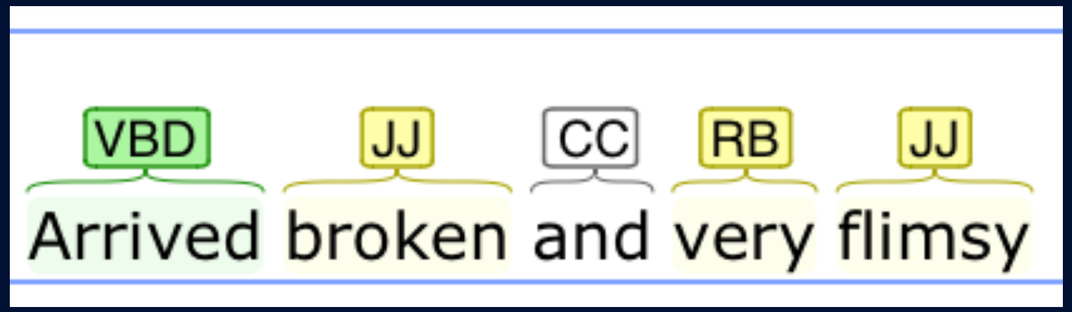
Little bit of Grammar

It is a very cute ***case***



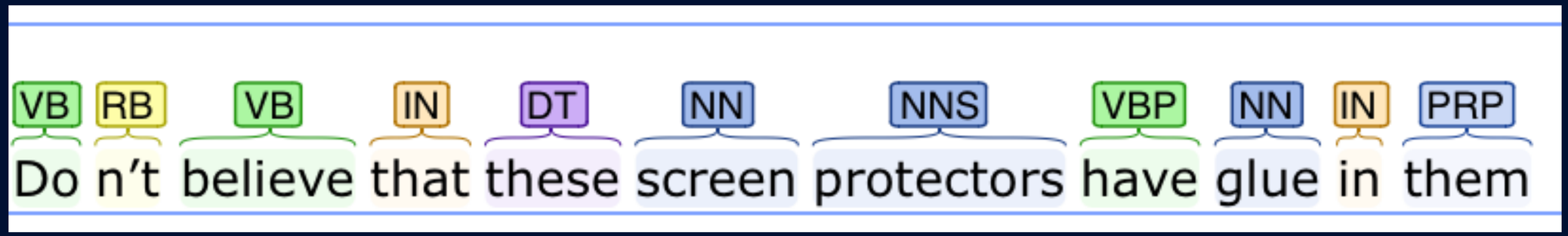
Little bit of Grammar

Arrived broken and very flimsy



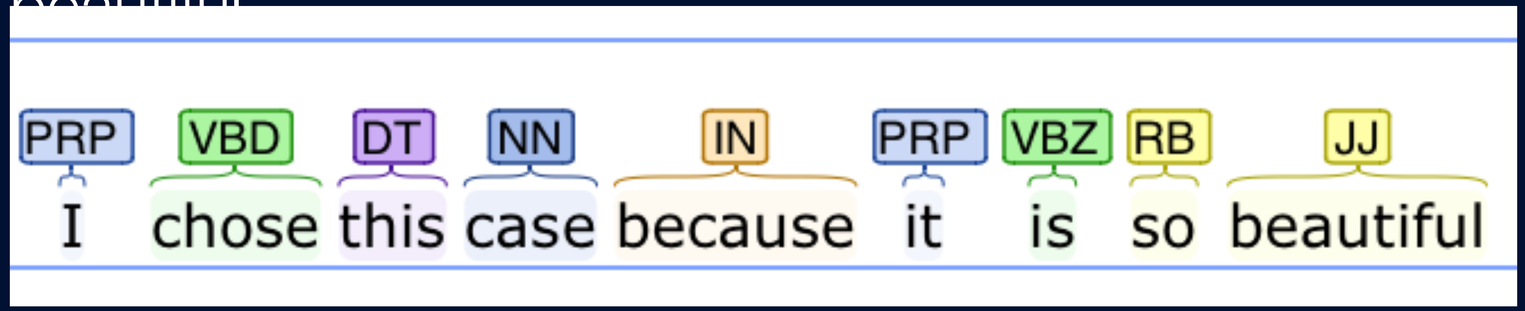
Little bit of Grammar

Don’t believe that these screen protectors have glue in them



Little bit of Grammar

I chose this case because it was beautiful



Entities/Aspects/Features

Sentiment/opinion can be expressed on each node

Explicit and Implicit Aspects

The quality of the cover is very good

I love,love,love it fits great so pretty and femine

this case is so cute it looks good on my white iphone its pretty good quality

This case may look very breakable but it is very sturdy

Another very good phone charger that does the job and has a very affordable price that anyone can afford to buy it

Another very good phone charger that does the job and has a very affordable price that anyone can afford to buy it

Was here right on time. Wall adapter what I expected and works well but dose Not Have Quick-Charge…

Was here right on time. Wall adapter what I expected and works well but dose Not Have Quick-Charge…

Without question, it was the worst handset I've ever owned

Using the sentiment words, implicit and explicit words can be detected using POS and dependency relationships.

Patterns of POS tags for Aspect Mining

syntactic templates

	First Word	Second Word	Third Word (Not Extracted)
1.	JJ	NN or NNS	anything
2.	RB, RBR, or RBS	JJ	not NN nor NNS
3.	JJ	JJ	not NN nor NNS
4.	NN or NNS	JJ	not NN nor NNS
5.	RB, RBR, or RBS	VB, VBD, VBN, or VBG	anything

Frame your own rules

	First word	Second word	Third word
	NN	VBZ	RB or/and JJ

Reference: Peter D. Turnkey, “Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews”, Proc. of the 40th Annual Meeting on the Association of Computationsl Linguistics (ACL), July 2002, https://arxiv.org/abs/cs/0212032

Explicit/Implicit Aspect Extraction

Looks even better in person. Be careful to not drop your phone so often because the rhinestones will fall off (duh). More of a decorative case than it is protective, but I will say that it fits perfectly and securely on my phone. Overall, very pleased with this purchase."

Rule for ***fits perfectly and securely?***

[('not', 'RB'), ('drop', 'VB')]
ASPECT: not drop

[('decorative', 'JJ'), ('case', 'NN')]
ASPECT: decorative case

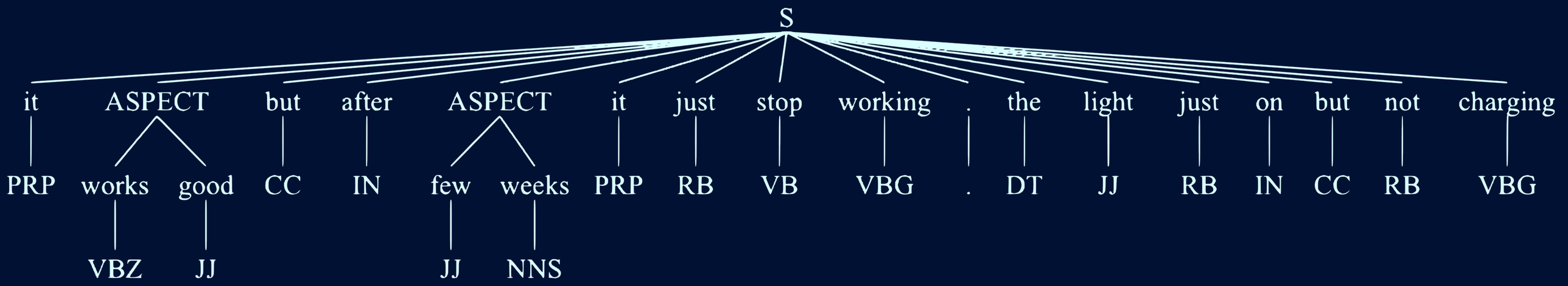
[('it', 'PRP'), ('is', 'VBZ'), ('protective', 'JJ')]

ASPECT: it is protective

Implicit or explicit

[('very', 'RB'), ('pleased', 'JJ')]
ASPECT: very pleased

Aspects using POS



Classification of Aspects - Unsupervised model

- Capture all the probable aspects using POS chunking and Chinking rules
- Use pointwise Mutual Information to determine the aspects
  - $$PMI = \log_2 \left(\frac{p(w_1, w_2)}{p(w_1)p(w_2)} \right)$$
  - Statistically determine whether two cooccurring words are independent or dependent
  - Pick up words or phrases that form the set of aspects