

Natural Language Components

Speaker and Generator

Component and level of
Representation

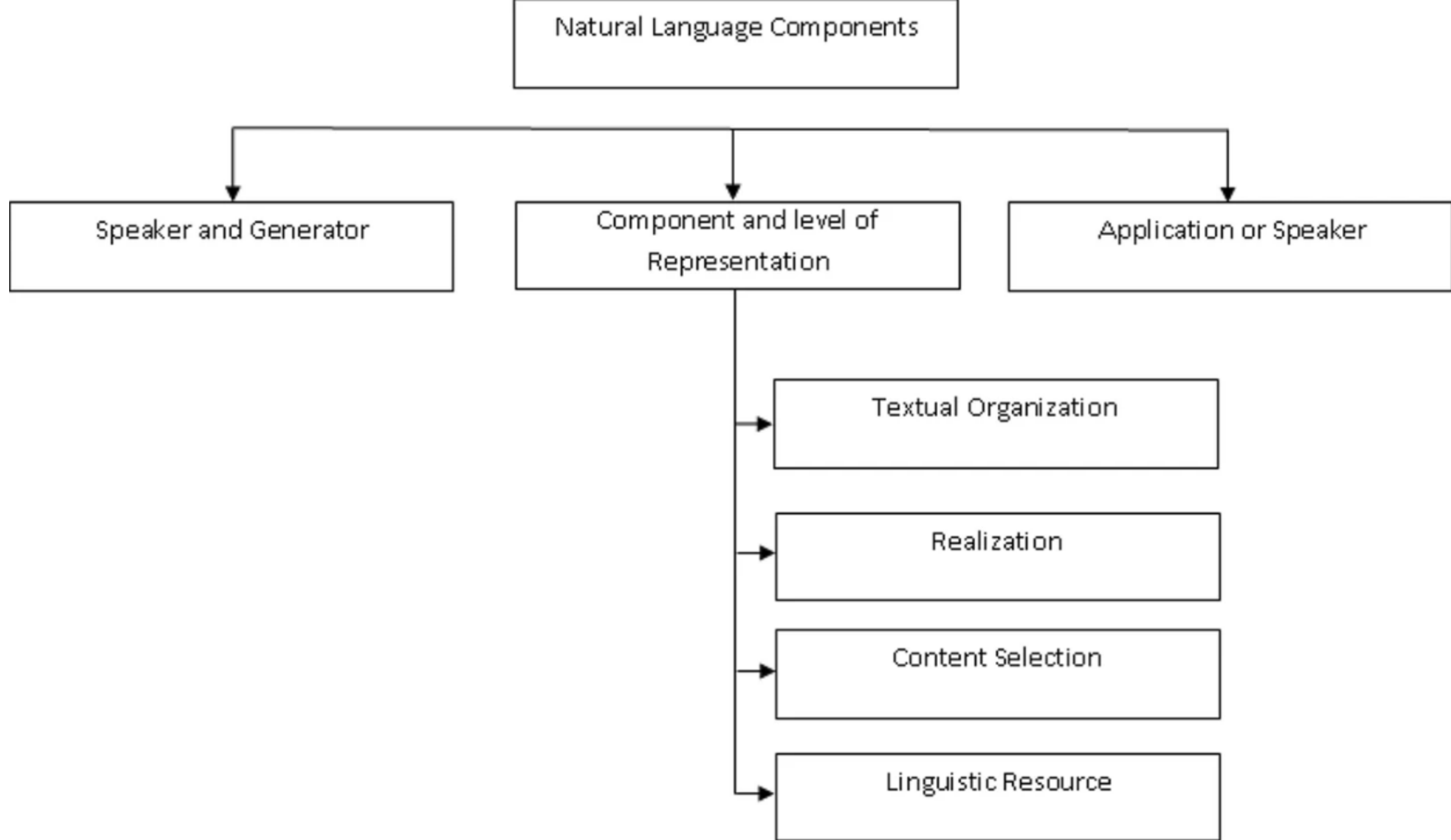
Application or Speaker

Textual Organization

Realization

Content Selection

Linguistic Resource

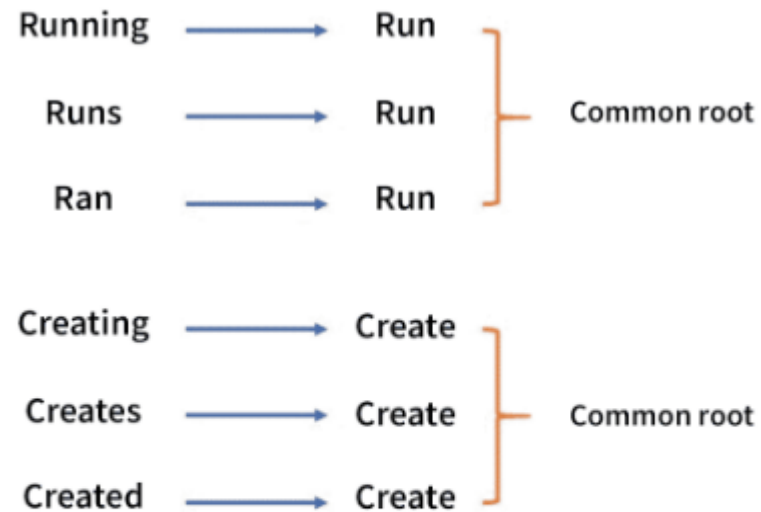


Tokenization

“What restaurants are nearby?”

‘what’ ‘restaurants’ ‘are’ ‘nearby’

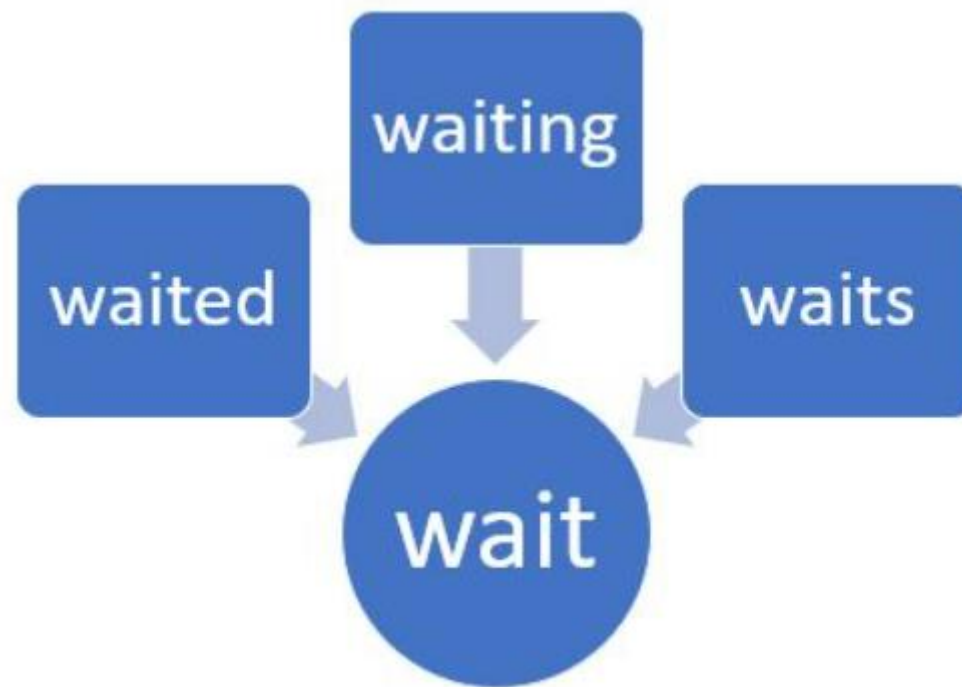
Lemmatization



The boy's cakes are of different sizes

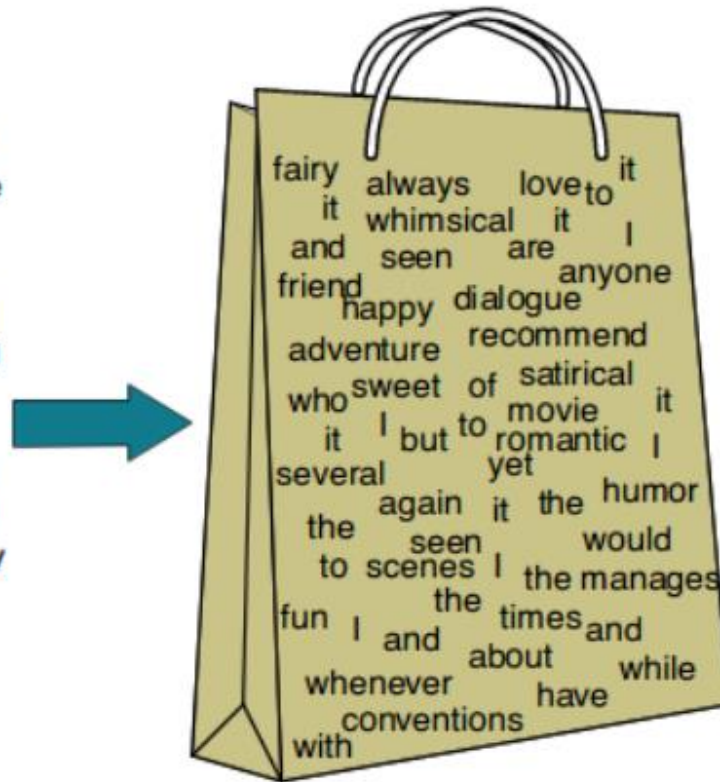
The boy cake be of differ size

Stemming



Bag of Words

I love this movie! It's sweet, but with satirical humor. The dialogue is great and the adventure scenes are fun... It manages to be whimsical and romantic while laughing at the conventions of the fairy tale genre. I would recommend it to just about anyone. I've seen it several times, and I'm always happy to see it again whenever I have a friend who hasn't seen it yet!



it	6
I	5
the	4
to	3
and	3
seen	2
yet	1
would	1
whimsical	1
times	1
sweet	1
satirical	1
adventure	1
genre	1
fairy	1
humor	1
have	1
great	1
...	...

TF-IDF (term frequency-inverse document frequency)

$$tfidf(t, d, D) = tf(t, d) \cdot idf(t, D)$$

$$tf(t, d) = \log(1 + freq(t, d))$$

$$idf(t, D) = \log\left(\frac{N}{count(d \in D : t \in d)}\right)$$