

If You Say Something Is “Likely,” How Likely Do People Think It Is?



Likely

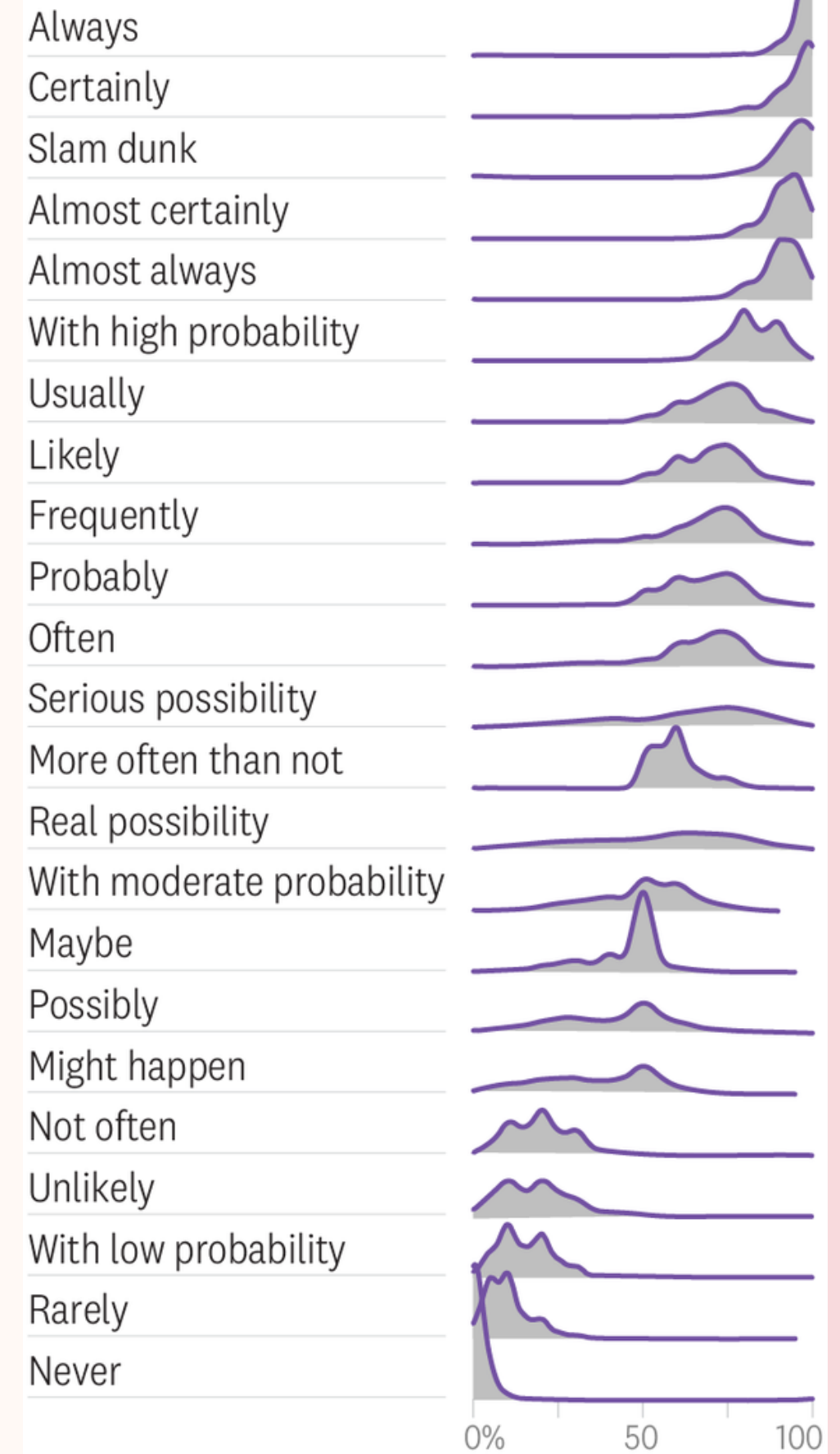
Maybe

Real Possibility

Key Words

Distribution of responses according to respondents' estimate of likelihood

Word or phrase





AMBAMBIGUOUS WORDING LEADS TO
MISUNDERSTANDING

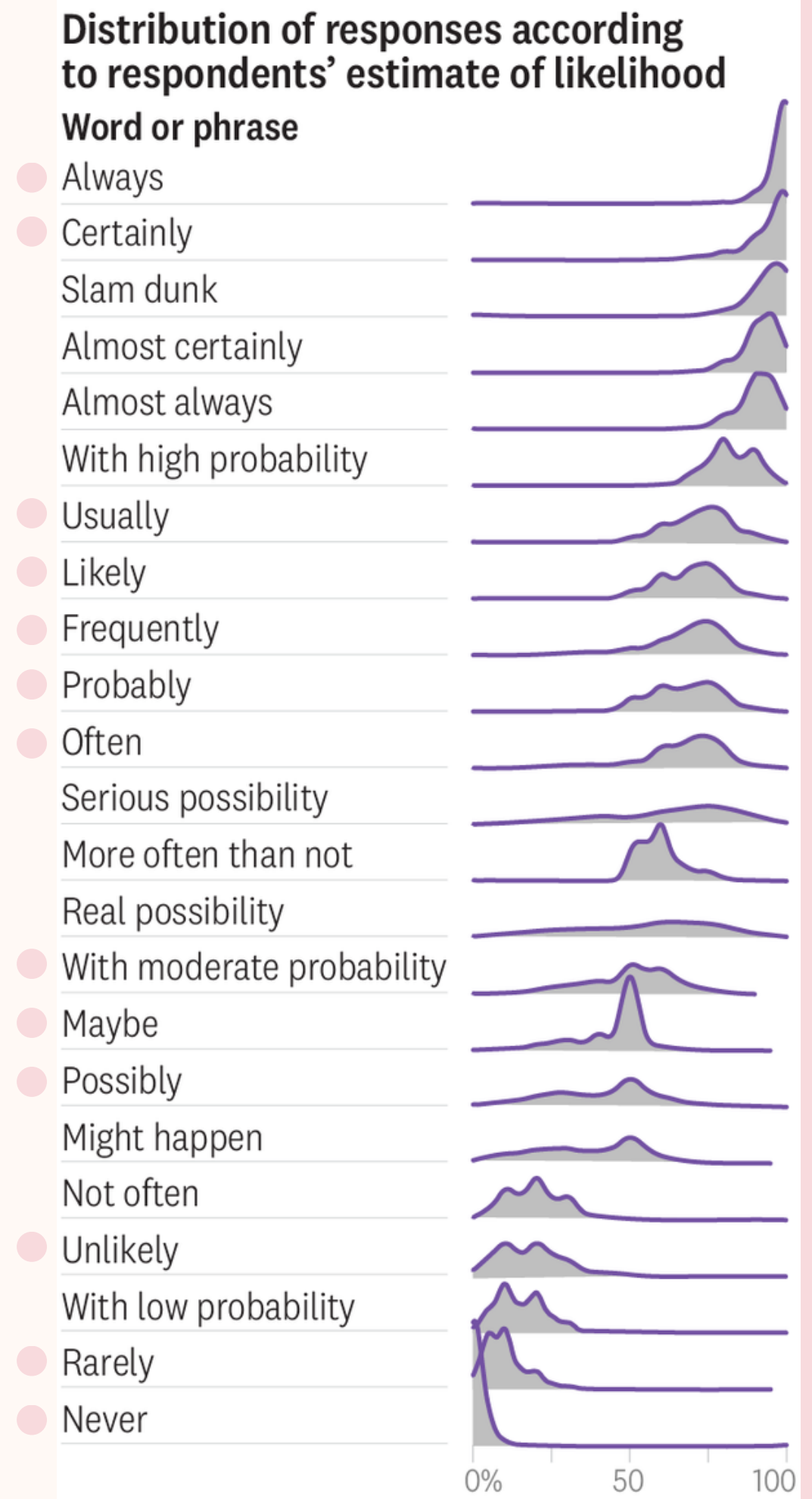


Probabilistic Words in Academic Text

Megan Church

Key Words

Always
Certainly
Usually
Likely
Frequently
Probably
Often
Probability
Maybe
Possibly
Unlikely
Rarely
Never





EDUCATIONAL PSYCHOLOGY



Scraped PubMed Articles

<https://www.ncbi.nlm.nih.gov/pmc/>



EDUCATIONAL PSYCHOLOGY



1

How to learn effectively in medical school: test yourself, learn actively, and repeat in intervals.

Cite

Augustin M.

Yale J Biol Med. 2014 Jun 6;87(2):207-12. eCollection 2014 Jun.

Share

PMID: 24910566

Free PMC article.

Review.



1

How to learn effectively in medical school: test yourself, learn actively, and repeat in intervals.

Cite

Augustin M.

Yale J Biol Med. 2014 Jun 6;87(2):207-12. eCollection 2014 Jun.

Share

PMID: 24910566

Free PMC article.

Review.

24910566

30879464

35645653

24910566
30879464
35645653

24910566

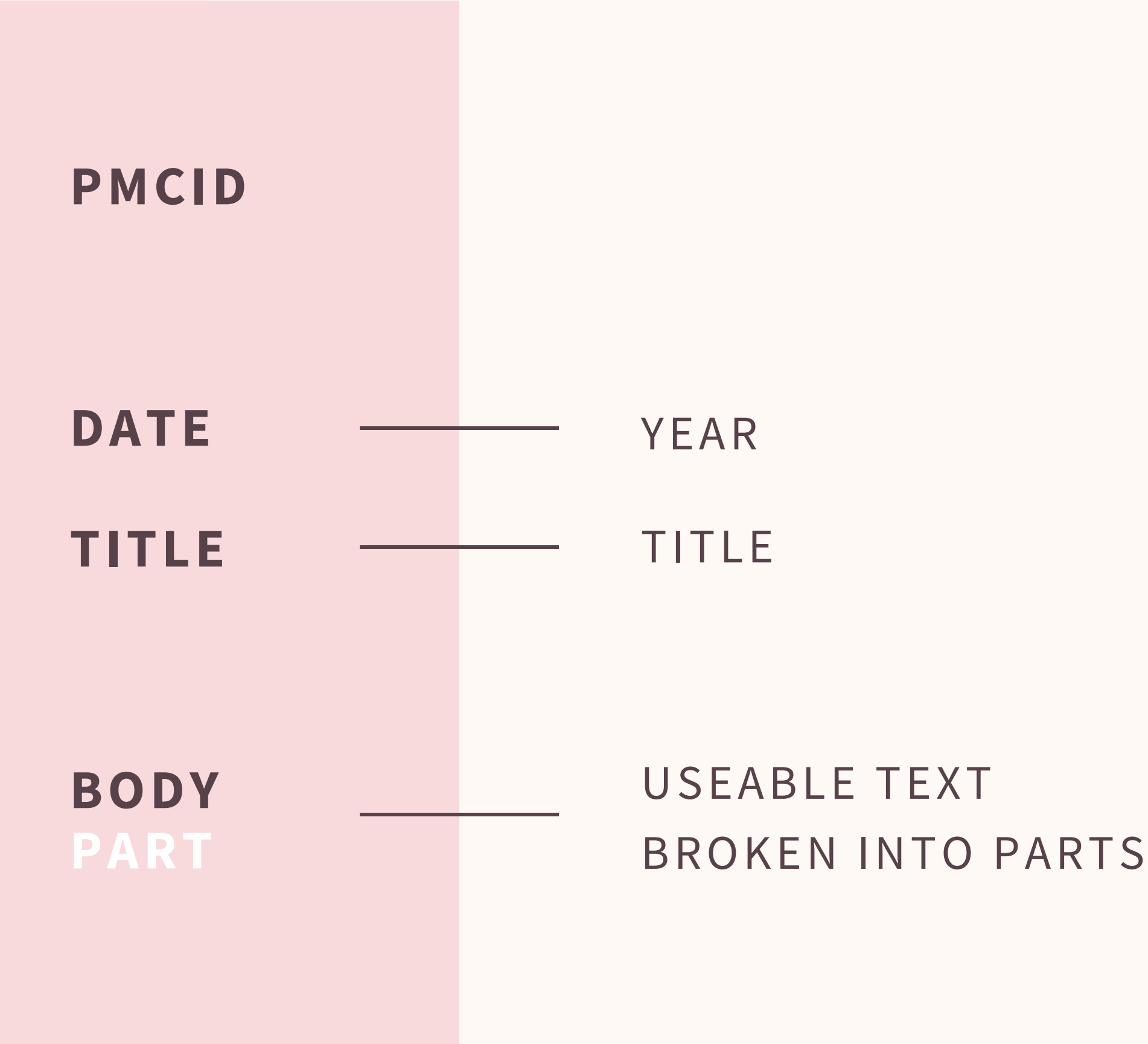


<https://www.ncbi.nlm.nih.gov/pmc/articles/{>

}



The Raw Data



PMCID

DATE

TITLE

BODY
PART

YEAR

TITLE

USEABLE TEXT
BROKEN INTO PARTS

['Published online 2018 Feb 5.']



.str.extract(r'(\d{4})')



2018

PMCID

DATE

TITLE

**BODY
PART**

YEAR

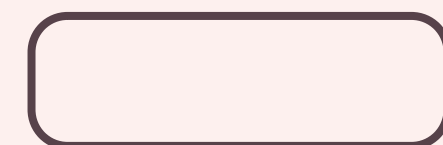
TITLE

USEABLE TEXT
BROKEN INTO PARTS

['Taking an educational psychology
course improves neuroscience literacy
but does not reduce belief in
neuromyths']



dropped



PMCID

DATE

TITLE

**BODY
PART**

YEAR

TITLE

USEABLE TEXT
BROKEN INTO PARTS

'[\Go to:AbstractEducators are
increasingly interested...neuroscience
literacy.\', \Go to:IntroductionThere
is a growing interest...\']



three rounds of regex



clean text

**BODY
PART**

USEABLE TEXT
BROKEN INTO PARTS

1. Extract each part of the body text

```
re.findall( (Go to:.*) (?=Go to:) )
```

```
'[\Go to:AbstractEducators are increasingly interested...neuroscience  
literacy.\', \'Go to:IntroductionThere is a growing interest...\']
```

**BODY
PART**

USEABLE TEXT
BROKEN INTO PARTS

2. Isolate the title of each body part

```
re.search( .(\u201cGo to.*?[a-z](?=[A-Z]))\u201c. )
```

Go to:AbstractEducators are increasingly interested ... neuroscience literacy.\', \'

Go to:AbstractBackgroundThe public Educational Psychology ... Mental health\', "

**BODY
PART**

USEABLE TEXT
BROKEN INTO PARTS

3. Split the body part text from the title

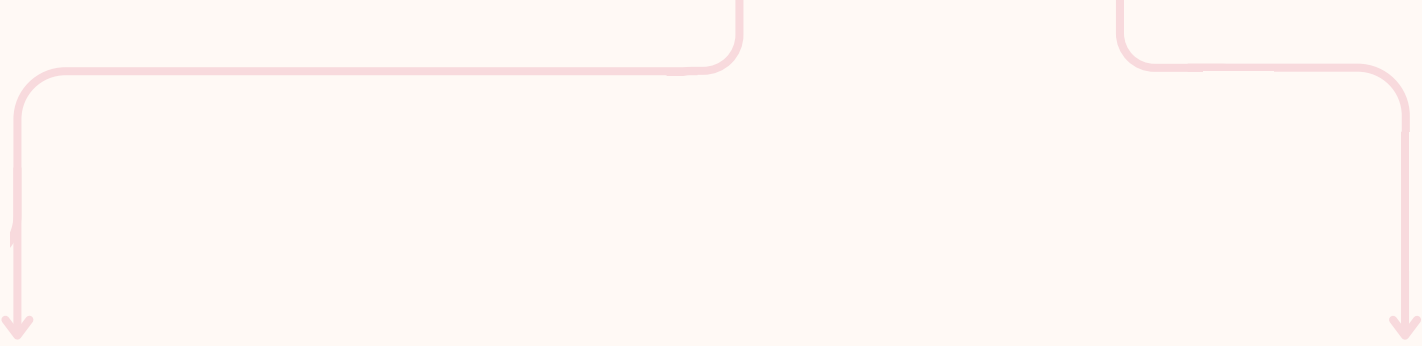
```
re.sub( .(Go to.*?[a-z](?=[A-Z])). , '' )
```

Go to:Abstract**Educators are increasingly interested ... neuroscience literacy.**\', \'

Go to:AbstractBackground**The public Educational Psychology ... Mental health**\', "

pmcid	date	title	body
PMC5798820	['Published online 2018 Feb 5.']	['Taking an educational psychology course impr...	['Go to: <u>Abstract</u> Educators are increasingly int...

pmcid	year	part	text
PMC5798820	2018	Abstract	Educators are increasingly interested in apply...
PMC5798820	2018	Introduction	There is a growing interest in applying neuros...
PMC5798820	2018	Materials and methods	ParticipantsParticipants were recruited from a...



Tokens

Lemmas

Stopwords

Punctuation

SPACY

ALWAYS

OFTEN

NEVER

STRING

Tokenized Data

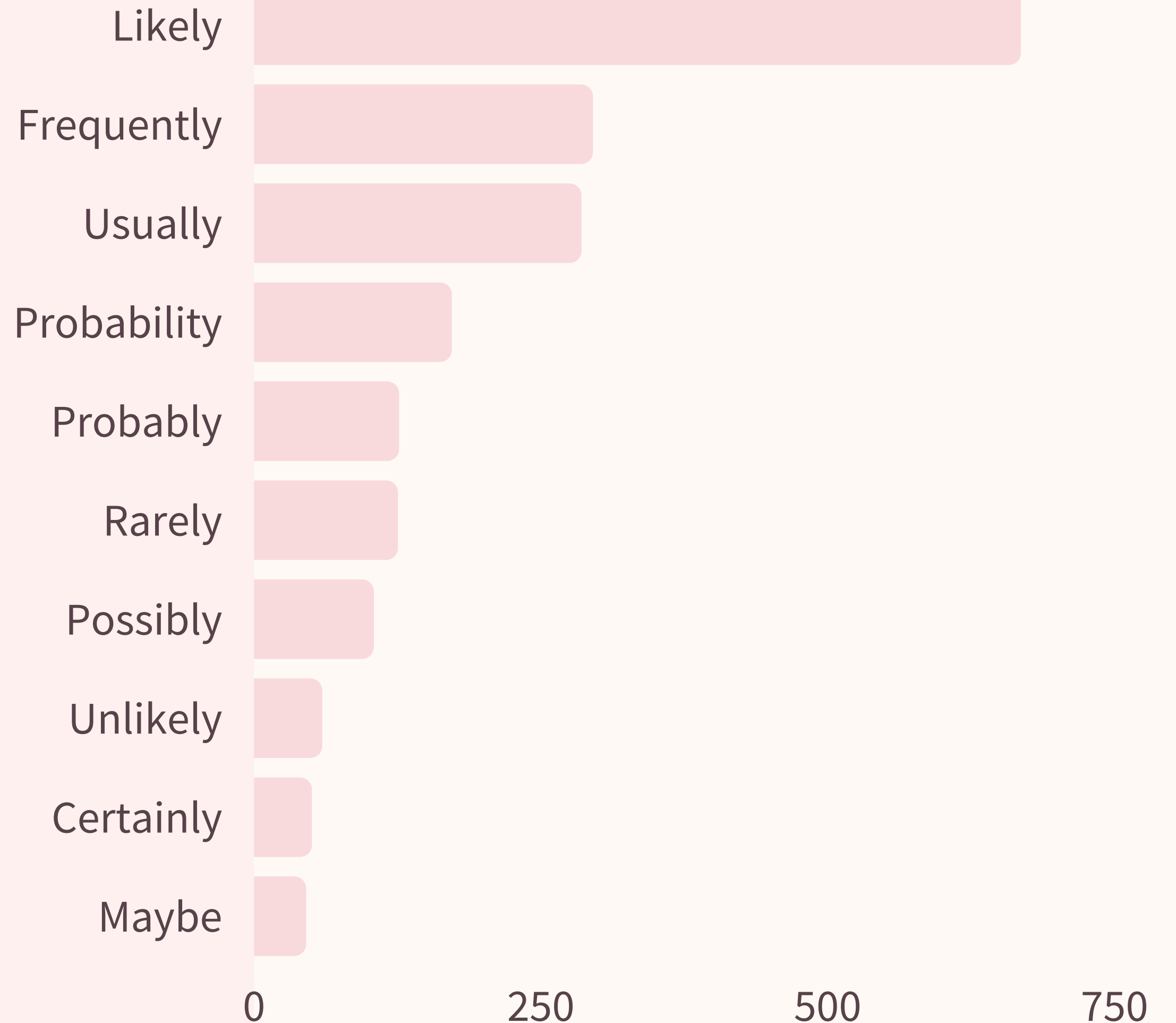
pmcid	year	part	text	tokens
PMC5798820	2018	Abstract	Educators are increasingly interested in apply...	[educator, increasingly, interested, apply, ne...
PMC5798820	2018	Introduction	There is a growing interest in applying neuros...	[grow, interest, apply, neuroscience, finding,...
PMC5798820	2018	Materials and methods	ParticipantsParticipants were recruited from a...	[participantsparticipant, recruit, public, uni...
PMC5798820	2018	Results	Neuroscience literacyThe first research questi...	[neuroscience, literacythe, research, question...
PMC5798820	2018	Discussion	It is important for educators and the general ...	[important, educator, general, public, intelli...
PMC5798820	2018	Conclusion	Prior researchers have proposed bridging betwe...	[prior, researcher, propose, bridge, neuroscie...



Explode
Filter
Frequency

Always
Certainly
Usually
Likely
Frequently
Probably
Often
Probability
Maybe
Possibly
Unlikely
Rarely
Never

Probabilistic Term Frequency



Next Steps

FREQUENCIES
CONTEXT
VARIATION

