You had me at hello: How phrasing affects memorability

Cristian Danescu-Niculescu-Mizil[1] Justin Cheng[1] Jon Kleinberg[1]
Lillian Lee[1]

[1]Department of Computer Science
Cornell University

March, 2012

Hello. My name is Inigo Montoya

Motivation/Related work

- Understanding influence
  - How can you be influential if I can't remember what you say?
  - marketing, politics, entertainment, social media, etc.

Hello. My name is Inigo Montoya

Motivation/Related work

- Understanding influence
  - How can you be influential if I can't remember what you say?
  - marketing, politics, entertainment, social media, etc.
- Challenge: devising an evaluation setting that separates the phrasing of a message from its context

Hello. My name is Inigo Montoya

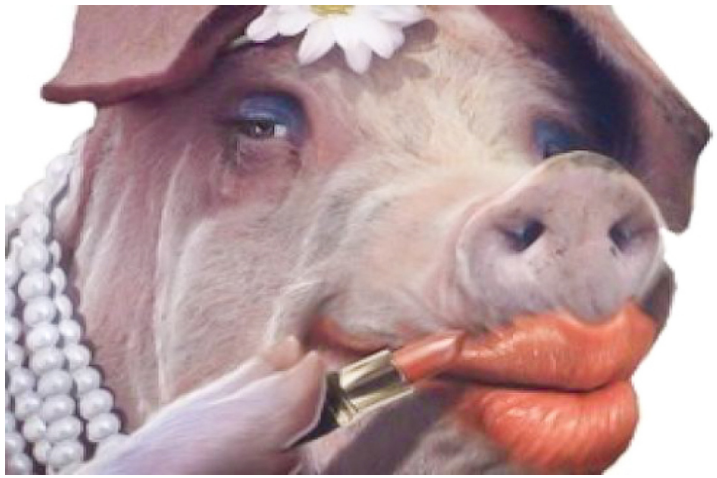
Motivation/Related work

- *You can put lipstick on a pig, but it's still a pig* - Barack Obama

Hello. My name is Inigo Montoya

Motivation/Related work

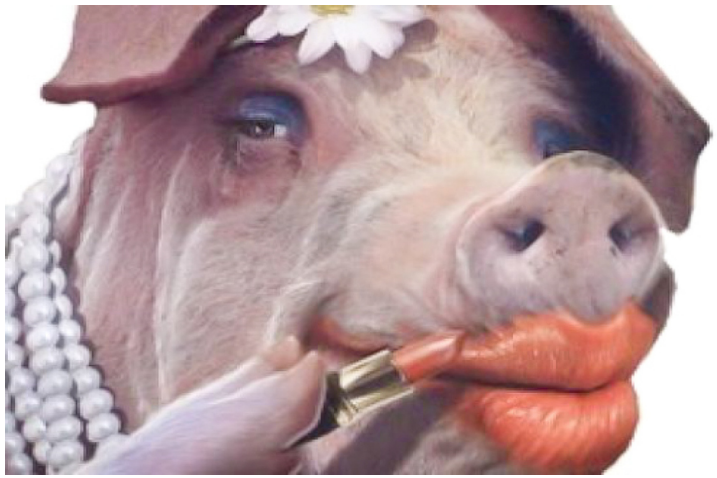
- *You can put lipstick on a pig, but it's still a pig* - Barack Obama



Hello. My name is Inigo Montoya

Motivation/Related work

- *You can put lipstick on a pig, but it's still a pig* - Barack Obama



- Actually a reference to a quote by Sarah Palin
- Did the wording "lipstick on a pig" not actually have any effects?

Hello. My name is Inigo Montoya

Motivation/Related work

- Does **phrasing** affect memorability?
  - the choice of words
  - the way it is phrased

Hello. My name is Inigo Montoya

Motivation/Related work

- Does **phrasing** affect memorability?
  - the choice of words
  - the way it is phrased
- Does the form of the language add an effect *beyond* or *independent* of context?

Hello. My name is Inigo Montoya

Motivation/Related work

- Related studies characterize memorable quotes by...
  - recognition
  - production

Hello. My name is Inigo Montoya

Motivation/Related work

- Related studies characterize memorable quotes by...
  - recognition
  - production
- Focus in domain in which...
  - rich use of language
  - already exist a large number of external human judgements
  - we can control for setting in which text was used

Hello. My name is Inigo Montoya

Motivation/Related work

- Related studies characterize memorable quotes by...
  - recognition
  - production
- Focus in domain in which...
  - rich use of language
  - already exist a large number of external human judgements
  - we can control for setting in which text was used

$\Rightarrow$ Movies!

I'm ready for my close-up

Data

- How do we select the right data so that we can show effects based in the **language** of the quotes themselves?
  - $\Rightarrow$ We need some kind of control for the speaker and context!
  - $\Rightarrow$ Movies!

I'm ready for my close-up

Data

- From Star Wars 4...

  Obi-Wan: You don't need to see his identification
  Stormtrooper: [ditto]
  Obi-Wan: These aren't the droids you're looking for
  Stormtrooper: [ditto]
  Obi-Wan: He can go about his business
  Stormtrooper: [ditto]
  Obi-Wan: Move along
  Stormtrooper: [ditto]

I'm ready for my close-up

Data

- From Star Wars 4...

  Obi-Wan: You don't need to see his identification
  Stormtrooper: [ditto]
  Obi-Wan: These aren't the droids you're looking for
  Stormtrooper: [ditto]
  Obi-Wan: He can go about his business
  Stormtrooper: [ditto]
  Obi-Wan: Move along
  Stormtrooper: [ditto]

- Only the bold-faced line went viral

  $\Rightarrow$ Apart other external factors including context, speaker, etc., what **linguistic** trait makes this phrase more memorable than others?

I'm ready for my close-up

Data

From ∼1000 movie scripts,

Pair IMDB "memorable" quotes with "non-memorable" quote such that it differs only in choice of words

i) same-scene ($\Leftrightarrow$ ∼adjacent)
ii) same-speaker
iii) same-length

$\Rightarrow$ 2200 such (Mem, Non-mem) pairs

I'm Ready for my close-up

Pilot Study: Let's try!

- Survey: 11 or 12 of these (Mem, Non-mem) pairs from the movies a participant has never watched
- Asked to predict which quote sounds more memorable out of two, so it was comparative.
- 14,000 people responded to http://memo.clr3.com/

I'm Ready for my close-up

Pilot Study: Range of possible result

50%
There exists
no language signal

100%
Language is
all that matters

I'm Ready for my close-up

Pilot Study: Turns out to be

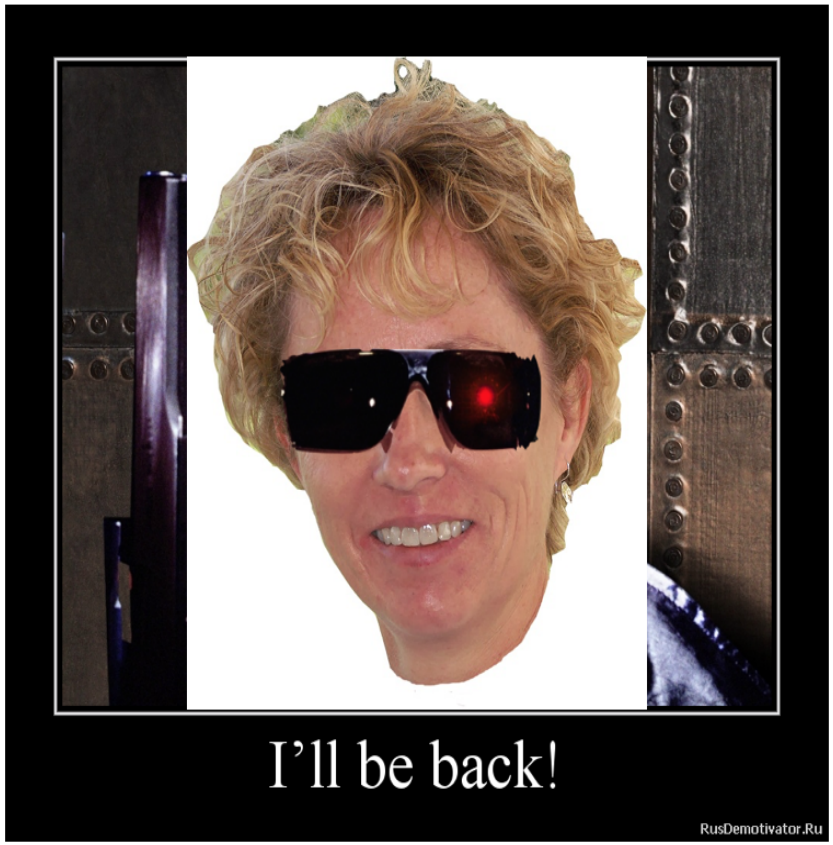
50%
There exists
no language signal

72~78%
What makes
this tasks interesting

100%
Language is
all that matters

I'm Ready for my close-up

Pilot Study: This task is not trivial for sure



I'm Ready for my close-up

Pilot Study Observations

- Subjects of pilot study suggested two basic forms textual signals could take
  - memorable quotes often involve a distinctive turn of phase
  - memorable quotes invoke general themes that aren't tied to context

I'm Ready for my close-up

Incorporating search engine counts

- Various problems to consider
- Found most effective to use search engine counts as additional filter rather than a free-standing numerical value
- +Google dataset
  - for each memorable non-memorable quote pair (M,N), only keep pairs for which M...
    - produced more than five results
    - produced at least twice as many results as N

Never send a human to do a machine's job.

Distinctiveness: How to measure distinctiveness

- Using a model of "common language" from Brown corpus, evaluate how much of lexical and syntactic **distinctiveness** these quotes have
- 1-,2-,3-gram word Language Model (lexical)
- 1-,2-,3-gram part-of-speech Language Model (syntactic)

Never send a human to do a machine's job

Distinctiveness: Result

- Lexically more distinctive
  - Obi-Wan: These aren't **the droids** you're looking for
    $\Rightarrow$ Unusual word choice is more likely to stick in head
- Syntactically less distinctive
  - "You're gonna need a **bigger** boat" vs "You're gonna need a boat **that is bigger**"
    $\Rightarrow$ Rather than complicatedly structured sentence like relative clause, simpler adjective is easier to memorize
- $\Rightarrow$ Memorable quotes consist of **unusual word sequences** built on **common syntactic scaffolding**.

Never send a human to do a machine's job

Distinctiveness: Result

"common language" model		IMDb-only	+Google
lexical	1-gram	61.13%***	59.21%***
	2-gram	59.22%***	57.03%***
	3-gram	59.81%***	58.32%***
syntactic	1-gram	43.60%***	44.77%***
	2-gram	48.31%	47.84%
	3-gram	50.91%	50.92%

Table 3: Distinctiveness: percentage of quote pairs in which the the memorable quote is more distinctive than the non-memorable one according to the respective "common language" model. Significance according to a two-tailed sign test is indicated using *-notation (***="p<.001").

Never send a human to do a machine's job

Generality

- Distinctive phrases stick better but simultaneously **generality** plays its role in memorability as well
- The more general quote is, the easier it gets for people to use the quote in their lives, outside of the specific context

Never send a human to do a machine's job

Generality: How to measure generality

- Personal Pronouns

  "he, they" vs "you, we"
- Indefinite articles

  "a, an" vs "the"
- Past tense

Never send a human to do a machine's job

Generality: Result

Generality metric	IMDb-only	+Google
fewer pers. pronoun	60.52%***	60.14%***
more indef. article	57.21%***	58.23%***
less past tense	57.91%***	59.74%***
more present tense	54.60%***	55.86%***

Table 4: Generality: percentage of quote pairs in which the memorable quote is more general than the non-memorable ones according to the respective metric. Pairs where the metric does not distinguish between the quotes are not considered.

- "**You** need **a** bigger boat" vs "**He** needs **the** bigger boat"

Never send a human to do a machine's job

Generality: Result

Generality metric	IMDb-only	+Google
fewer pers. pronoun	60.52%***	60.14%***
more indef. article	57.21%***	58.23%***
less past tense	57.91%***	59.74%***
more present tense	54.60%***	55.86%***

Table 4: Generality: percentage of quote pairs in which the memorable quote is more general than the non-memorable ones according to the respective metric. Pairs where the metric does not distinguish between the quotes are not considered.

- "**You** need **a** bigger boat" vs "**He** needs **the** bigger boat"

Never send a human to do a machine's job

Cross-Domain Application

(Non)memorable language models		Slogans	Newswire
lexical	1-gram	56.15%**	33.77%***
	2-gram	51.51%	25.15%***
	3-gram	52.44%	28.89%***
syntactic	1-gram	73.09%***	68.27%***
	2-gram	64.04%***	50.21%
	3-gram	62.88%***	55.09%***

Table 5: Cross-domain concept of "memorable" language: percentage of slogans that have higher likelihood under the memorable language model than under the non-memorable one (for each of the six language models considered). Rightmost column: for reference, the percentage of newswire sentences that have higher likelihood under the memorable language model than under the non-memorable one.

Never send a human to do a machine's job

Prediction Task