

Brighter than Gold:
Figurative Language in
User Generated Comparisons

Vlad Niculae

MPI-SWS

Cornell University

Cristian Danescu-Niculescu-Mizil

MPI-SWS

Cornell University

Comparisons

Assert likeness of two things;

Point out shared properties.

Aren't always meant to be taken literally!

Literal comparison ●

“Sterling is much cheaper than gold.”

“Her voice makes the song shine brighter than gold.”

Figurative comparison

(*simile*) ●

This talk:

What makes comparisons figurative?

Literal comparison ●



“Sterling is much cheaper than gold.”

“Her voice makes the song shine brighter than gold.”



Figurative comparison

(*simile*) ●

Figurative language

Metaphor, **simile**, etc.

Most research is on **metaphor**.

- Conceptual interpretation (Shutova et al., 2013)
- Identification
 - tree kernels (Hovy et al., 2013)
 - conceptual features (Tsvetkov et al., 2014)

Less (computational) research on **simile**.

- Irony, humor (Veale et al., 2012)

Role of comparisons

- Describe new things in terms of given things.

“A **unicorn** is like a **horse**, but it has a horn.”

- Rank with respect to a property.

“The cord is **more durable** than the adapter.”

Applications in **Knowledge Extraction**

(Tandon et al., 2014, Lofi et al., 2014, *inter alia*)

Assumption: comparisons are literal.

But can we tell if a comparison is literal?

our main insight:

**For figurative language,
context is important!**

social, topical...

applying our insight:

Does **topic** information differentiate literal and figurative comparisons?

Can we take advantage of the **structure** of the comparison?

How does **social context** drive figurative language use?

Data we analyze: Amazon.com product reviews

Rich in comparisons

need to describe products

Rich in **social context**

helpfulness, star ratings

Topical context comes for free
product categories

Customer Review

16 of 16 people found the following review helpful

★★★★★ **ladies sing the blues...**,
April 25, 2008

By [Matthew G. Sherwin](#)

This review is from: Great Ladies Sing the Blues (Audio CD)

"My Man" sounds better than ever when Billie sings it; she is able to use her vocal talents to make this number shine brighter than gold!

Help other customers find the most helpful reviews

Was this review helpful to you?

The screenshot shows a customer review for 'Great Ladies Sing the Blues (Audio CD)' by Matthew G. Sherwin. The review is dated April 25, 2008, and has a 5-star rating. The text of the review is: "My Man" sounds better than ever when Billie sings it; she is able to use her vocal talents to make this number shine brighter than gold! The review is marked as helpful by 16 of 16 people. At the bottom, there are buttons for 'Yes' and 'No' to indicate if the review was helpful to the user. Several green circles and ovals are drawn over the text to highlight specific elements: one circle around the helpfulness count, one oval around the product title, and two ovals around the review text.

Figurative Comparison Dataset

1400 sentences with marked comparisons.

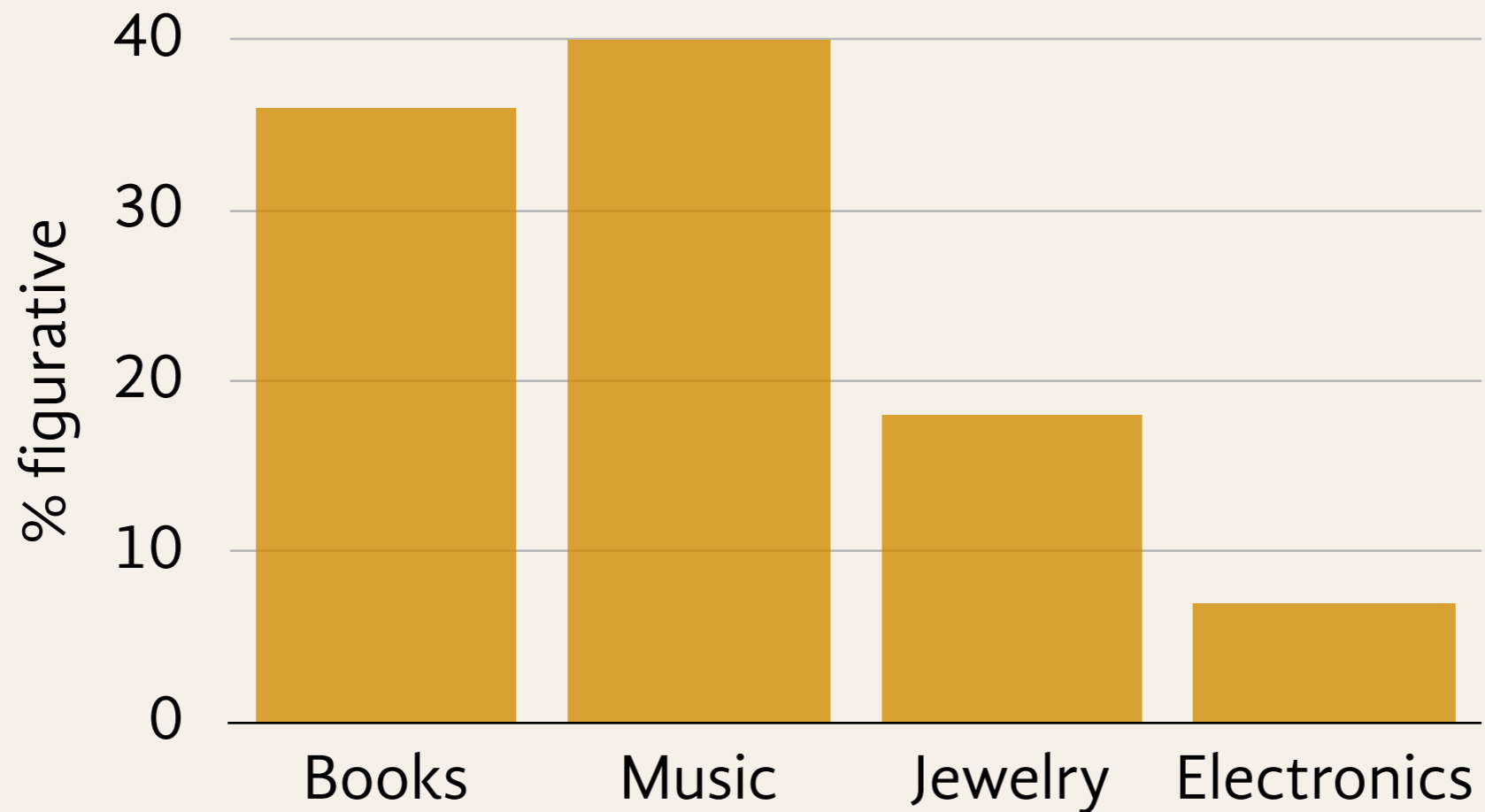
Figurative scores from three
Amazon Mechanical Turk annotators.

Freely available: <http://vene.ro/figurative-comparisons/>

Characterizing figurative comparisons

Topical context

Is figurativeness more likely in certain product categories?



Topical context

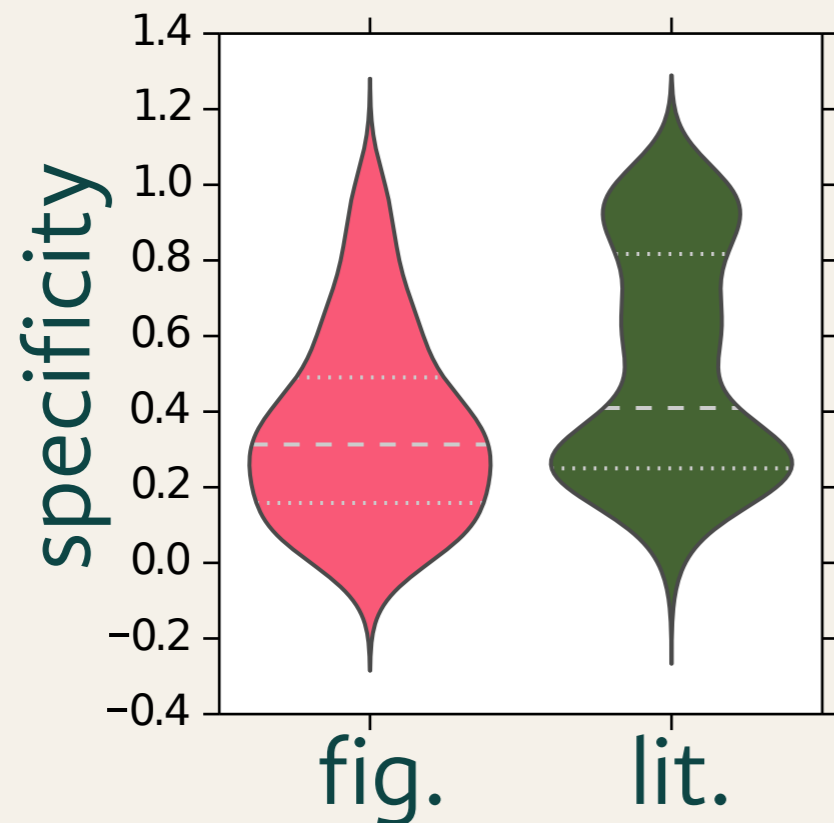
Is figurative word choice
topic-specific?

*The song feels like a boring **melody** (music)*
*Her book is like a sweet **melody** (books)*

$$\text{specificity}(\text{word}, \text{topic}) = \frac{\#(\text{word appears in topic})}{\#(\text{word appears in all topics})}$$

Topical context

Is figurative word choice
topic-specific?

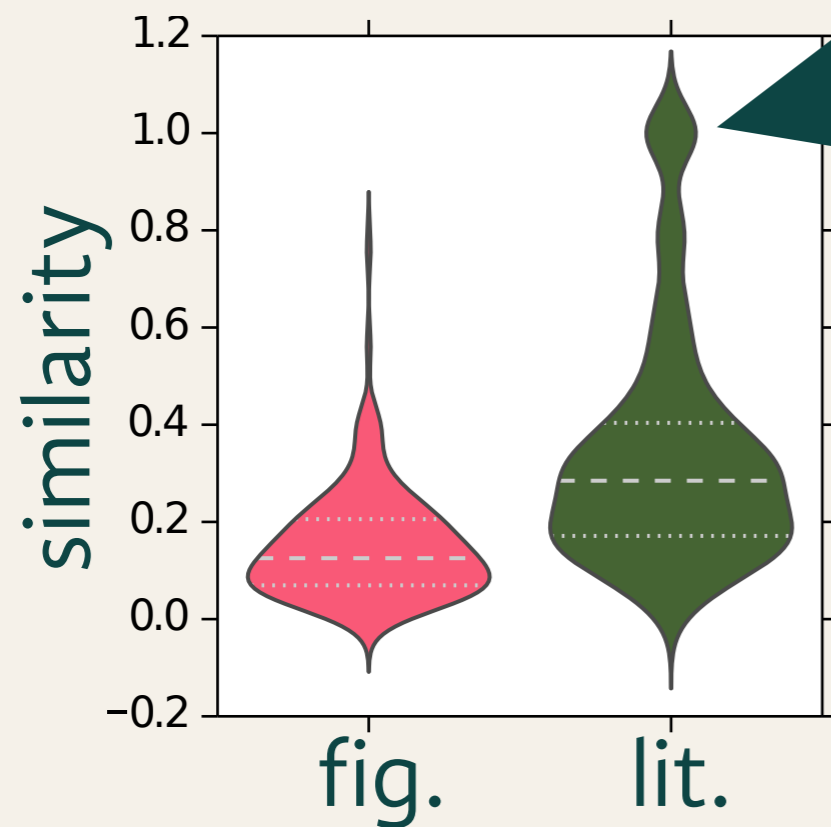


96%
song : *melody* (music)
book : *melody* (books)
3%

$$\text{specificity}(\text{word}, \text{topic}) = \frac{\#(\text{word appears in topic})}{\#(\text{word appears in all topics})}$$

Comparison structure

Similarity between
the things being compared.



*The **song** shines brighter than **gold**.*
***Sterling** is cheaper than **gold**.*

≈

≇

*cosine similarity in a vector space model

Metaphor features

Inspired by Tsvetkov et al. (2014)

- Concreteness

kiss vs. happiness

- Imageability

cinnamon vs. casque

- Supersenses: coarse categories

annotate: {communication, creation}

- Vector space embeddings

**Are these insights
predictive?**

Classification (logistic regression)

Linguistic insight features:

Topic-aware

- Review category
- Word specificity

Topic-agnostic

- Structural similarity
- Metaphor features
(concreteness, imageability, supersenses, embeddings)
- Definiteness

Classification (logistic regression)

| method | acc. | F |
|--------------|------|-----|
| bag-of-words | 80% | 75% |

Why so strong?

- “*the hero is truly larger than **life***”
- “*the headphones sounded like **crap***”

Classification (logistic regression)

| method | acc. | F |
|--------------------|------|-----|
| bag-of-words | 80% | 75% |
| linguistic insight | 86% | 79% |

Top features:

1. **Structural** similarity
*“the older **man** was wiser than the **boy**”*
2. **Topic** specificity
*“the cord’s more durable than the **adapter**”*
3. Imageability
*“the explanations are as clear as **mud**”*

Classification (logistic regression)

| method | acc. | F |
|--------------------|------------|------------|
| bag-of-words | 80% | 75% |
| linguistic insight | 86% | 79% |
| both | 88% | 82% |

The two approaches are complementary.

($p < 0.05$)

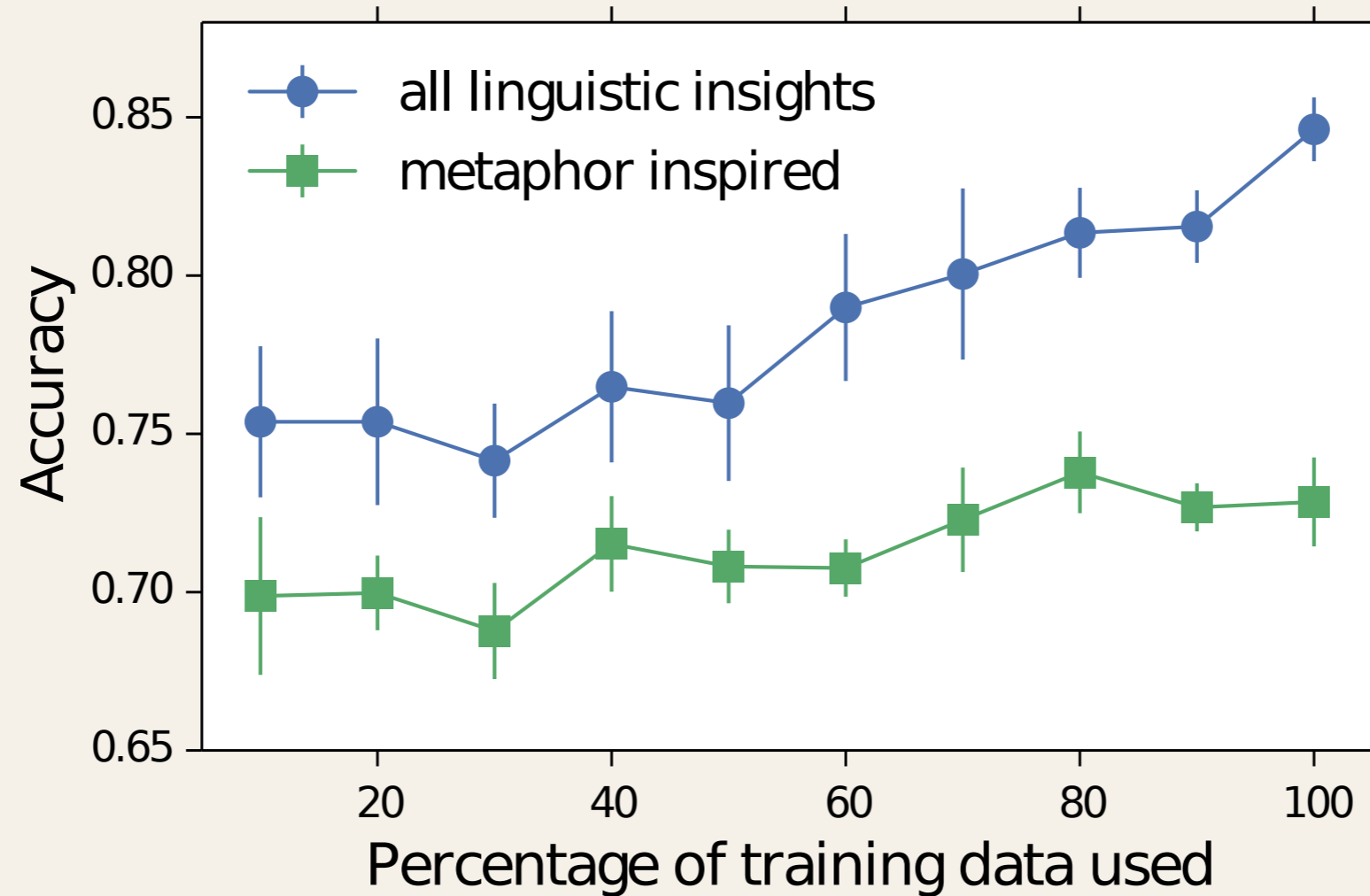
Classification (logistic regression)

| method | acc. | F |
|--------------------|------------|------------|
| bag-of-words | 80% | 75% |
| linguistic insight | 86% | 79% |
| both | 88% | 82% |
| human | 96% | 94% |

Optimistic estimation:

Majority voting among
three Mechanical Turk
workers

Why not just use a metaphor system?



($p < 0.05$)

Practical importance

Verified Purchase ([What's this?](#))

This review is from: Sony MDR-NC6 Noise Canceling Headphones (Discontinued by Sony)
Initially the headphones sounded like crap (no bass and no high end,

The image shows a screenshot of an Amazon search bar. The search bar is labeled "Search" and has a dropdown menu set to "Electronics". The search term "headphones" is entered in the search bar. Below the search bar, there are several suggestions: "headphones", "bluetooth headphones", and "crap". The word "crap" is highlighted in red. The search bar also shows "Best Sellers" and "TV & Video" as categories. The text "Results for Electronics : 'headphones'" is visible at the bottom of the search bar.

Figurative language and social context

Social context of reviews

Verified Purchase ([What's this?](#))

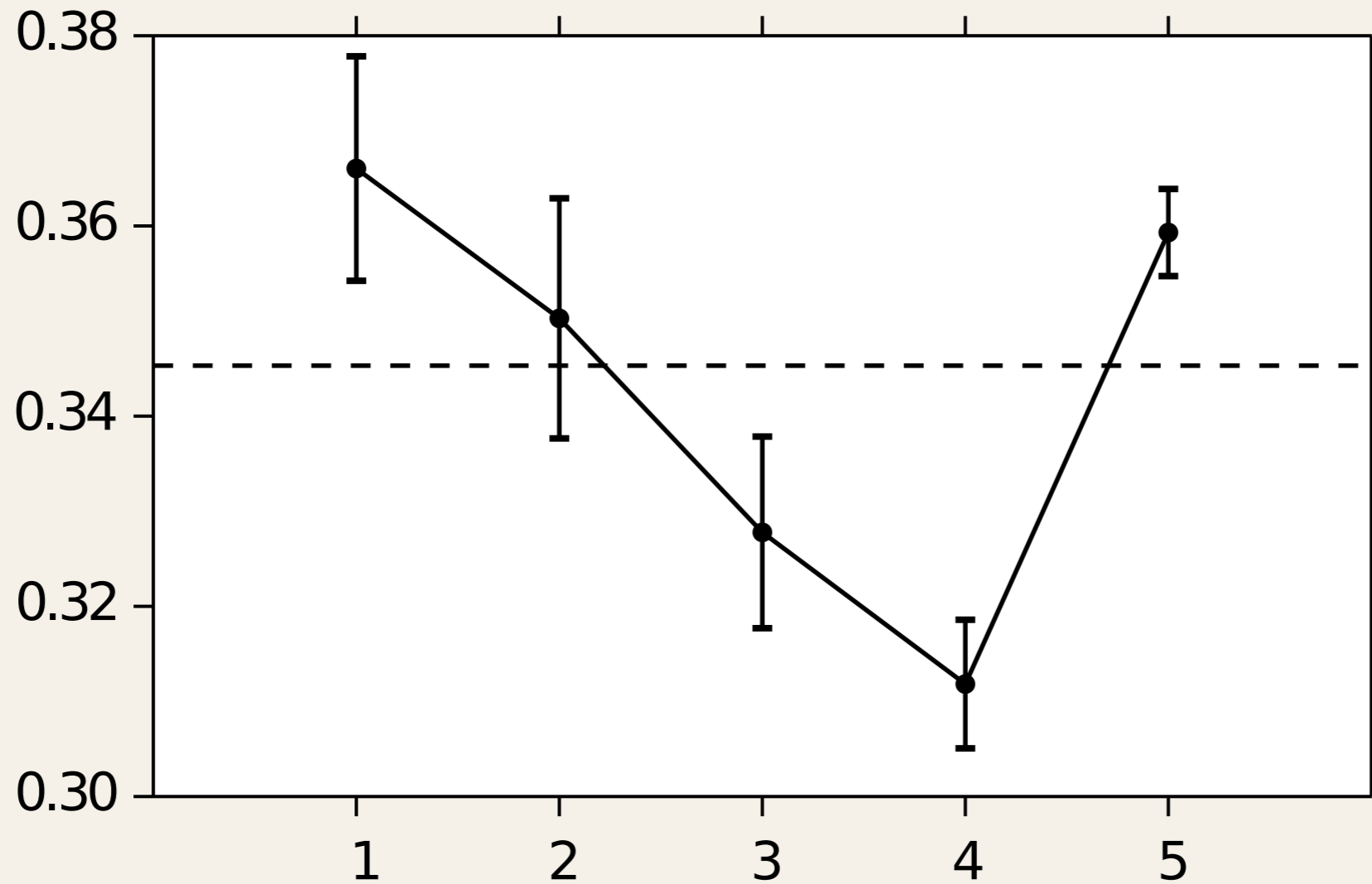
This review is from: Sony MDR-NC6 Noise Canceling Headphones (Discontinued b
Initially the headphones sounded like crap (no bass and no high end,

8 of 10 people found the following review helpful

★★★★★ Not bad for the price, December 18, 2004

Intrinsic opinion

ratio of figurative comparisons



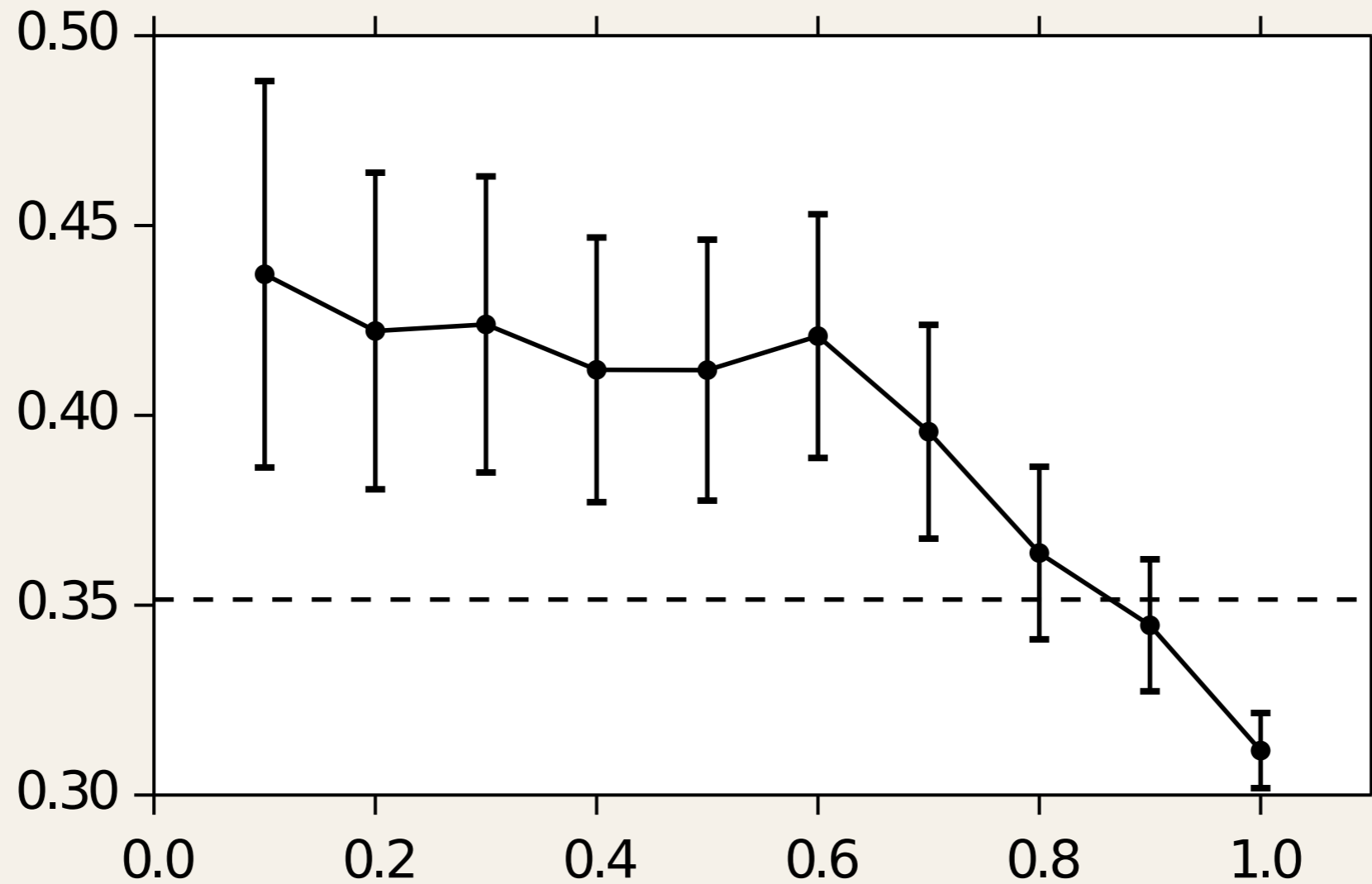
star rating

($p < 0.05$)

Extrinsic opinion

(how the review is perceived)

ratio of figurative comparisons



helpfulness

($p < 0.05$)

Conclusions

Compiled a **dataset** of figurative & literal comparisons.

Figurative language is tied to a **social** context.

Topic information is relevant for figurativeness.

Similes benefit from specific computational treatment.

“He uses statistics as a drunken man uses lamp-posts... for support rather than illumination.”

— Andrew Lang

“Getting an education was a bit like a communicable sexual disease. It made you unsuitable for a lot of jobs and then you had the urge to pass it on.”

— Terry Pratchett, Hogfather

Extra slides

| method | # feat | acc. | P | R | F | AUC |
|-------------------------|---------------|-------------|----------|----------|----------|------------|
| bag-of-words | 1840 | .80 | .64 | .90 | .75 | .89 |
| metaphor-inspired | 345 | .75 | .60 | .72 | .65 | .84 |
| domain-specific | 8 | .69 | .51 | .81 | .63 | .76 |
| all insight features | 365 | .86 | .76 | .83 | .79 | .92 |
| all of the above | 2202 | .88 | .80 | .84 | .82 | .94 |
| human | | .96 | .92 | .96 | .94 | |

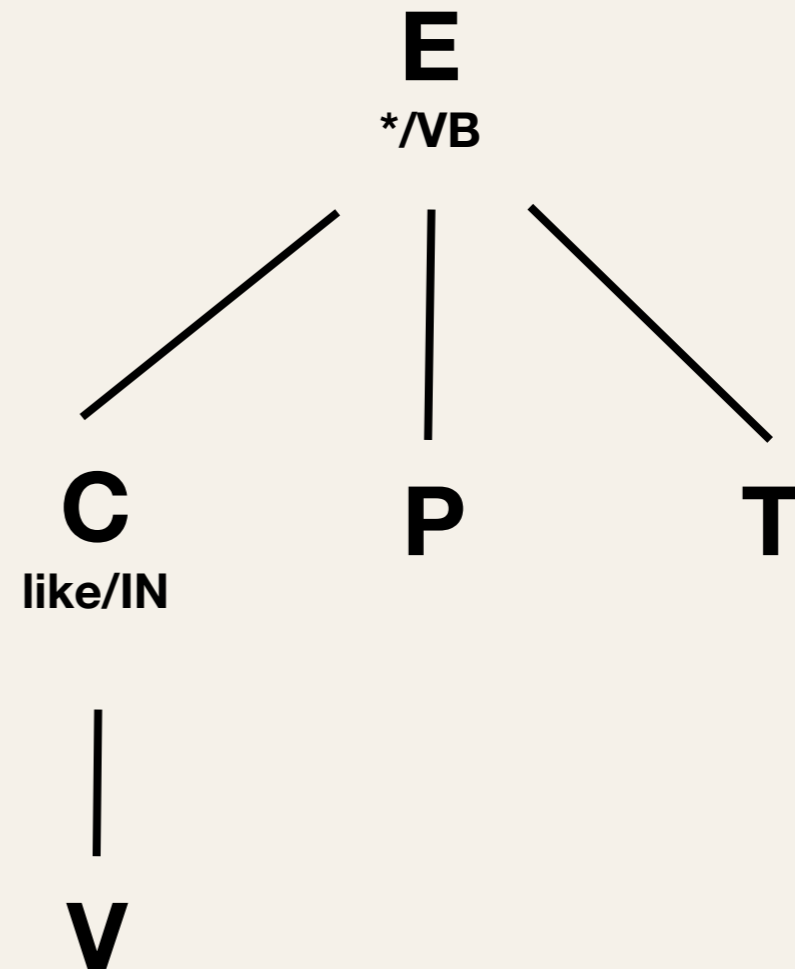
Training set size = test set size
(408 comparisons, 66% literal)

Matching comparisons

The [song^{Topic}] [shines^{Event}] [brighter^{Property}]
[than^{Comparator}] [gold^{Vehicle}]

3 types of patterns:

- T is like V
- T is as P as V
- T is P-er than V



Data collection

Amazon.com
reviews

extract
comparisons

-
-
-
-
-

Validate
(3x MTurk)

-
-
-

Annotate 1-4
(3x MTurk)

-(3, 4, 4)
-(4, 2, 1)
-(1, 2, 1)

-
-

consensus,
binarize

| | more imageable | less imageable |
|---------------|-----------------------------|-----------------------------|
| more concrete | <i>cinnamon, kiss</i> | <i>casque, pugilist</i> |
| less concrete | <i>devil, happiness</i> | <i>aspect, however</i> |

Concreteness and imageability (Tsvetkov et al., 2014 for metaphor)

Concreteness, imageability:
MRC Psycholinguistic Database

Supersenses:
WordNet for nouns, verbs
Tsvetkov et al. (2014) for adjectives

Logistic regression on word embeddings.

(Tsvetkov et al., 2014 for metaphor)

Metaphor vs. Simile

Israel et al (2004)

Metaphors that can't be rephrased as similes:

- “Her argument was somewhat murky.”
- “The house had great bones.”

Similes that can't be rephrased as metaphors:

- “The windshield wipers made a great clatter like two idiots clapping in church.”
- “Two bears sat facing each other like two matrons having tea.”