

# RE-EVALUATING THE TOMES FOR THE TIMES

LREC-COLING 2024

---

Ryan Brate, Marieke van Erp, Antal van den Bosch

KNAW Humanities Cluster, DHLab, Amsterdam  
Utrecht University, Institute for Language Sciences, Utrecht

**Disclaimer:** This paper contains derogatory words and phrases. They are provided solely as illustrations of the research results and do not reflect the opinions of the authors or their organisations.

*But Mrs. Tome Gallien's Adventure? A woman like Mrs. Tome Gallien wouldn't stop at any- thing! It might be a pair of llamas from Peru! Or a **greasy witchy-gypsy** to tell his fortune! Or a **homeless little jet-black pickaninny with a banjo and-consumption!***

[“The Stingy Receiver”, by Eleanor Hallowell Abbott]

*stereotypical images of black people buttressed the view that black people were unfit to attend racially integrated schools, live in safe neighborhoods, work in responsible jobs, vote, and hold public office.*

[by David Pilgrim, Curator, Jim Crow Museum]

Aim: Can we *make explicit* such problematic tropes ?

RQ: In applying transparent *descriptor:context-feature* metrics, as the basis for a trope ranking system; to what extent do we observe known tropes in highly ranked regions?

adjectives:

*The ignorant native*

verbs for which the descriptor is the agent

*The gypsy stole ...*

verbs for which the descript is the patient

*The slave was sold*

- Project Gutenberg English Literature and American Literature corpora according to their LoCC tags
- 9,179 English Literature books and 10,873 American Literature books.

1. segmentation into sentence parts
2. dependency parsing of sentence parts
3. define patterns linking descriptors and adjectives/verbs according to dep-parse output and POS tags
4. extract descriptor:context feature pairs

Results in:

- approx. 42M noun–adj pairs;
- approx. 30M agent–verb pairs;
- approx. 16M patient–verb pairs;



Log likelihood ratio can measure the affinity of context features, *but it requires context to be meaningful ...*

central idea: For a descriptor – context feature pair to be indicative of a trope, they should be mutually predictive of one-another.

# MUTUAL AFFINITY

mutually highly predictive ...

Relative LLR rank of each feature wrt., *Noun i*

	Feature j-2	Feature j-1	Feature j	...
Noun i-1				
Noun i	L	M	H	M
Noun i+1				
...				

(a)

X

Relative LLR rank of each noun wrt., each feature

	Feature j-2	Feature j-1	Feature j	...
Noun i-1				
Noun i	L	L	H	L
Noun i+1				
...				

(b)

Aggregation ...

$$\text{noun-feature mutual affinity score} = \sqrt{\left(1 - \frac{(R_f - 1)}{N_c}\right) \left(1 - \frac{(R_c - 1)}{N_r}\right)}$$

Word	Known problematic connotations
Oriental	slit-eyed, inscrutable
Chicano	sombrero-wearing peon, fiesta-loving, macho bandito
savage, primitive, lazy, conniving, superstitious treacherous, wily, crafty inscrutable, docile, backwards	loaded language (typically racist)

**Table 1:** Known–problematic connotations from “Anti-Bias Curriculum: Tools for Empowering Young Children” [1]

Noun	Corresponding adjectives	Adjective rank by noun-feature mutual affinity score
Oriental	skew-eyed	2 / 65
	not slant-eyed	4 / 65
Chinese	slant-eyed	7 / 192
	slit-eyed	30 / 192
savage	naked	2/902
	african	9/902

**Table 2:** Nouns and corresponding adjectives indicative of Table 1 known problematic connotations, and their mutual affinity score rankings according to co-occurrence frequencies in the data.

Selected descriptors: *cripple, dwarf, gypsy, native, negress, negro, nigger, oriental, servant, slave, tribe, and tribesman*

Noun	Associated adjectives
cripple	helpless (1 / 189), poor (2), hopeless (3), un-presentable (5), miserable (6)
dwarf	misshapen (2 / 201) , shriveled (5), ugly (7), cunning (8), hideous (14), grotesque (15), hunchbacked (18)
gypsy	vagrant (10 / 117), witchy (13), not trustworthy (14)
native	hostile (6 / 694), ignorant, superstitious, little, full-blooded
negress	slatternly (5 / 142), fat (7), good-natured (8), stout (9), big (13)
negro	full-blooded (4 / 872), ignorant (5), giant (8), gigantic (10), burly (12)
nigger	runaway (1 / 609), lazy (6), damned (7), dead (9), onery (12), damn (13)
oriental	godless (3 / 39)
slave	fugitive (1 / 1026), runaway (2), negro (5)
tribe	savage (2 / 803), hostile (3), warlike (7), primitive (23)
tribesman	savage (5 / 114), wild (6), intransigent (7), ferocious (13), hostile (18)
	<b>Associated verbs for which the noun is the agent</b>
gypsy	steal (1 / 139)

- data-driven discovery of negative tropes
- Examination of the differences between representations and tropes between English and American cultures





L. Denman-Sparks.

***Anti-bias curriculum.***

National Association for the Education of Young Children,  
Washington, D.C., DC, 7 edition, Sept. 1993.