

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

2019 Workshop: Interviewers and Their Effects
from a Total Survey Error Perspective

Sociology, Department of

2-26-2019

Humans vs. Machines: Comparing Coding of Interviewer Question-Asking Behaviors Using Recurrent Neural Networks to Human Coders

Jerry Timbrook

University of Nebraska-Lincoln, jerry.timbrook@gmail.com

Adam Eck

Oberlin University, aeck@cse.unl.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/sociw>

Part of the [Quantitative, Qualitative, Comparative, and Historical Methodologies Commons](#)

Timbrook, Jerry and Eck, Adam, "Humans vs. Machines: Comparing Coding of Interviewer Question-Asking Behaviors Using Recurrent Neural Networks to Human Coders" (2019). *2019 Workshop: Interviewers and Their Effects from a Total Survey Error Perspective*. 30.

<https://digitalcommons.unl.edu/sociw/30>

This Article is brought to you for free and open access by the Sociology, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in 2019 Workshop: Interviewers and Their Effects from a Total Survey Error Perspective by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.



Comparing Coding of Interviewer Question-Asking Behaviors Using Recurrent Neural Networks to Human Coders

Jerry Timbrook (University of Nebraska-Lincoln) and Adam Eck (Oberlin College)

BACKGROUND

Interviewer (I'wer) Question-Asking Behaviors

- Exact Reading:** The I'wer reads the question exactly as worded in the questionnaire

What is your sex?

- Minor Change:** The I'wer omits or adds words that do not alter question meaning

And so, what is your sex?

- Major Change:** The I'wer omits or adds words that do alter question meaning

What is your sex, male or female?

Why Study Interviewer Question-Asking?

- Pre-test survey questions (Fowler & Cannell 1996)
- Explore I'wers' cognitive processing (Fowler & Cannell 1996; Schaeffer & Maynard 1996)
- Evaluate I'wers' field performance (Fowler & Mangione 1990)
- Evaluate I'wers' effect on measurement (Dykema et al. 1997)

Human Coding

- Humans use transcripts/recordings to code
- Pros: Easy visual/audio review, humans can differentiate between major and minor changes
- Cons: Costly and time-consuming

Computer Coding

- Text Alignment (i.e., String Matching)**
 - Computer programs code by comparing question-asking from transcript to questionnaire text
 - Pros: No extra per case cost once program is written
 - Cons: No automatic way to differentiate between major and minor changes
- Machine Learning**
 - Computer learns to code automatically on its own using previously coded examples
 - Pros: No per case cost once models are trained, can potentially differentiate between major and minor changes
 - Cons: Requires specialized knowledge/tech

Research Questions

- Can Text Alignment or Machine Learning approaches partially automate the coding of I'wer question-asking behaviors?
- How does the reliability of computer coding compare to human coding?

METHODS AND RESULTS

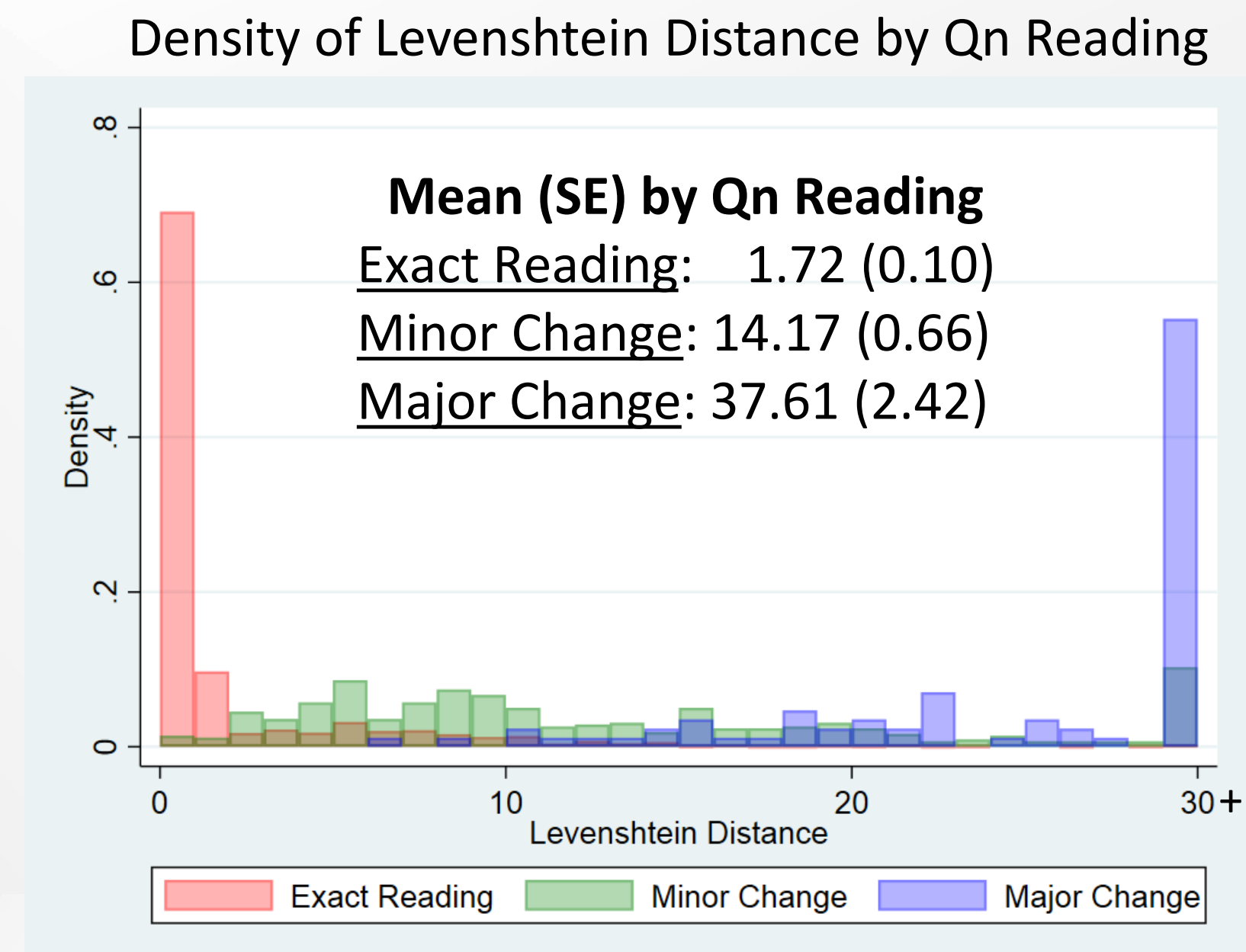
Text Alignment via Exact String Matching and Levenshtein Distance

- Exact String Match:** coded as 1) exact reading (if text of transcript = questionnaire), or 2) read with a change (if not)
- Levenshtein Distance:** coded as # of insertions, deletions, or substitutions needed to change one string into the other

Overall Reliability by Coding Method (Exact vs. Change)

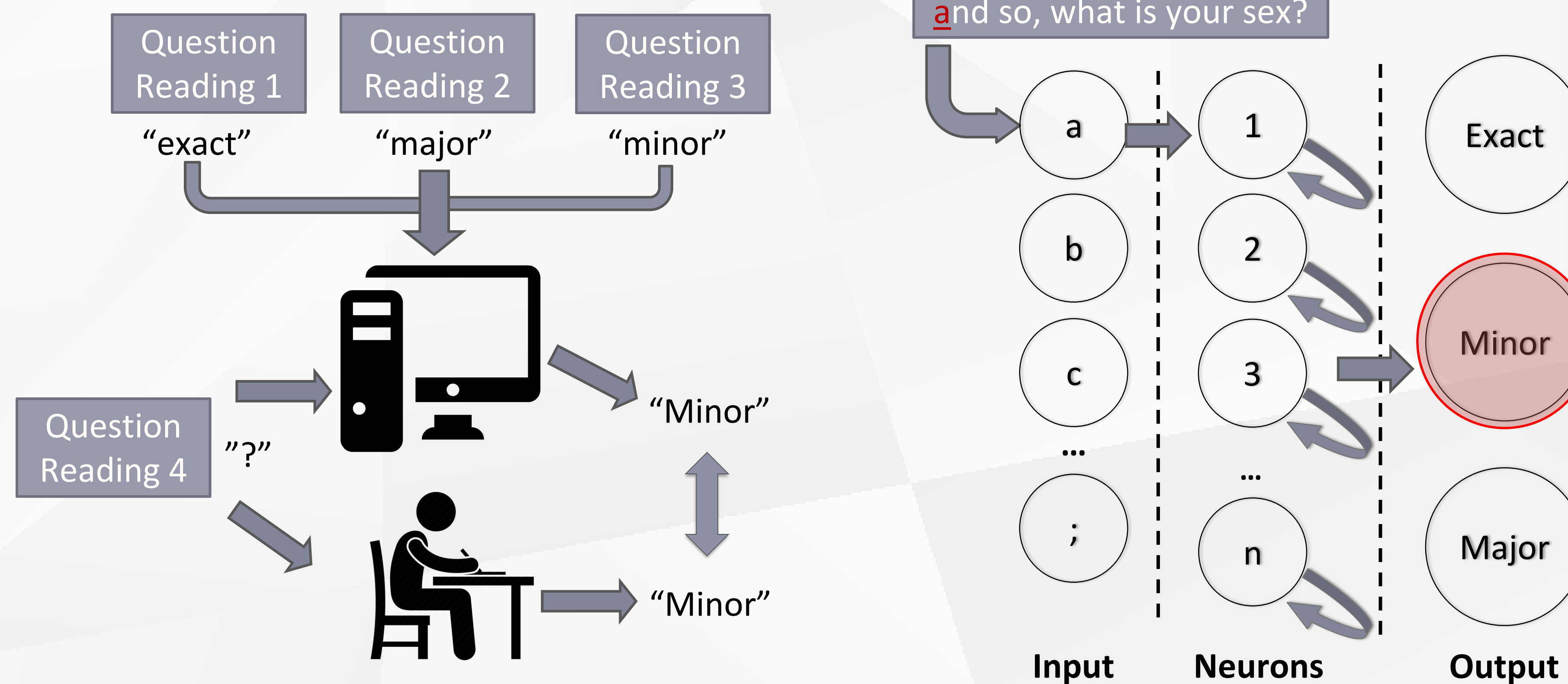
Master vs:	Kappa	CI LL	CI UL	% Agreement
Undergrad	0.57	0.53	0.61	88%
RNN (NP)	0.36	0.31	0.41	85%
String (NP)	0.09	0.08	0.11	36%
String (P)	0.47	0.44	0.51	75%

P = Text was Processed, NP = Text was Not Processed



Machine Learning via Recurrent Neural Networks (RNNs)

- Coded as 1) an exact reading, 2) a minor change, or 3) a major change
- Training=80% undergrad-coded data; Validation=20% of undergrad-coded data; Test=100% of master-coded data
- k-fold cross validation with k=5 networks per questions



High Prop. Of Changes in Training Data

Question Topic	Kappa (% Agree) w/ Master Coder		
	Undergrad	RNN	Diff
Respondent Sex	0.39 (62%)	0.47 (66%)	-0.08 (-4%)
Leisure Activities	0.60 (85%)	0.57 (83%)	0.03 (2%)
Job Equipment	0.70 (86%)	0.80 (91%)	-0.10 (-5%)

*p<.05

Low Prop. Of Changes in Training Data

Question Topic	Kappa (% Agree) w/ Master Coder		
	Undergrad	RNN	Diff
Adults in HH	0.20 (74%)	-0.02 (67%)	-0.22* (7%)
Personal Priv.	0.80 (95%)	0.55 (90%)	0.25 (5%)
Cell Ownership	0.38 (64%)	0.11 (46%)	-0.27* (18%)

*p<.05

DATA

- Survey:** Work and Leisure Today 2 (Dual-Frame CATI survey; n=902; AAPOR RR3=7.8%; 26 I'wers)
- Unit of Analysis:** I'wer question-asking turns for n=58 questions
- Human Coding:** Each question-asking coded as 1) an exact reading, 2) a minor change, or 3) a major change
 - Undergrad-Coded Data:** 16 students coded Work and Leisure Today 2's 45,078 question-asking turns using SequenceViewer (Dijkstra 1999)
 - Master-Coded Data (Ground Truth):** 10% random subsample of undergrad-coded data (94 cases; 4,688 question-asking turns)

FINDINGS

- Exact String Matches:** Without text preprocessing, exact string matches were less reliable than all other coding methods. With text preprocessing, string matches were slightly more reliable than RNNs trained using unprocessed text, but slightly less reliable than undergrad human coding.
- Levenshtein Distance:** This measure did not identify clear classifications for major or minor changes, nor does it fully classify exact readings.
- RNNs vs. Human Coding:** Coding using RNNs trained with unprocessed text is comparable to undergrad human coding when there is a high prevalence of deviations from exact reading in the training data, but worse than undergrad coding when there is a low prevalence of deviations.

CONCLUSIONS

- The exact string match method is comparable to other methods in coding deviations from exact readings, but it cannot differentiate between major and minor changes.
- RNNs are a promising method for partially automating coding of I'wer question-asking behaviors, especially when many major changes are expected.
- Future work should refine the RNNs (e.g., train RNNs using processed text, employ class imbalance techniques, change parameters)