Measuring Women's Agency with Unsupervised Machine Learning Approaches

Marina Rizzi, University of Turin

Losira Nasirumbi Sanya, Makerere University

Clare Clingain, IRC

Aletheia Donald World Bank

Maria Hernandez-de-Benito CUNEF Universidad

Better Data for Better Jobs and Lives: Innovations in Survey Measurement in the Age of AI Conference

December 8–9, 2025,

World Bank, Washington, D.C.

Background and Motivation

- The concept of agency , the capacity of individuals to act independently and make their own free choices, has long been central to gender and development research

  ❖ ***However, its empirical measurement remains challenging.***

- Conventional tools rely heavily on structured survey questions or composite indices that may fail to capture the contextual, relational, and linguistic nuances of empowerment, particularly in settings where social norms and intra-household dynamics shape how women describe their decision-making processes.

- This research addresses that gap by exploring ***unsupervised, text-based machine learning approaches*** to measure women's agency from open-ended qualitative survey responses.

- Using natural language features extracted from interview transcripts, the study investigates which linguistic markers most effectively capture variation in women's agency—operationalized through an independently derived ***benchmark score***.

- The overarching goal is to assess whether computational linguistics can provide scalable, culturally sensitive tools for quantifying agency in qualitative data.

  ➢ ***"Can unsupervised text-based linguistic features derived from open-ended responses reliably predict women's agency as measured by expert-coded qualitative benchmarks?"***

Contribution to Related Literature

Existing Literature	Major Authors	Our Contribution
Gender and empowerment measurement: Capture women's agency, autonomy and intra-household bargaining power using improved indicators	Donald et al., 2020; Calvi et al., 2022; Costa et al., 2023; Anderson et al., 2021; Miedema et al., 2020; Kabeer, 1999; Klugman et al., 2014	*We develop several language-based measures and we evaluate how effectively they capture latent constructs of agency.*
Survey methodology and text-derived measurement	Schober et al., 2015; Olson et al., 2020; Conrad & Schober, 2008; Maynard et al., 2002; Dykema et al., 2022; Schmieder et al., 2023; Tumasjan et al., 2021; Mullainathan & Obermeyer, 2017	*Evidence on how open-ended questions can be used to measure complex social and gender-related constructs that are difficult to capture with standard close-ended modules.*
Text-as-data: Exploits text analysis methods to gain information from high-dimensional datasets	Ash and Hansen, 2023; Gentkow et al., 2019 for general reviews; Ashwin et al., 2022; Jayachandran et al., 2023; Franzosi et al., 2012; Franzosi et al., 2013; for agency metrics	*We construct several metrics of agency from open-ended questions, and we compare the effectiveness of each measure*

Data Collected

02. “What purchase was it?”

03. “Tell me about how the decision to make the household purchase was made?”

04. “Was this decision making on the purchase the same or or different from other times?”

Major Household Purchase

07. "What was this HH expenditure on?"

08. “Tell me about how the decision to spend HH income was made?”

09. “Was this decision making on spending HH income the same or different from other times?”

Control over HH Income

12. "What was the recent decision made on spending personal income?"

13. “Tell me about how the decision to spend personal income was made?”

14. “Was this decision making on spending personal income the same or different from other times?”

Control over Own Income

- Data collected as part of the *Measures for Advancing Gender Equality* (MAGNET) survey testing women’s agency measurement tools
- Conducted in Kayunga & Mukono districts of Uganda represent diverse agricultural and business-oriented livelihoods
- Recordings of 303 respondent answers to open-ended qualitative questions used as the *benchmark measure* of women’s agency

Data

- The recorded answers during the qualitative interviews were automatically transcribed using Otter.ai.

- Transcripts reviewed and scored by two independent coders using a standardized framework reflecting three dimensions of agency:

  a) ***Decision-making power*** – *the extent to which the respondent reports initiating or influencing household or personal decisions;*

  b) ***Perceived control*** – *expressions of self-efficacy and ability to shape outcomes;*

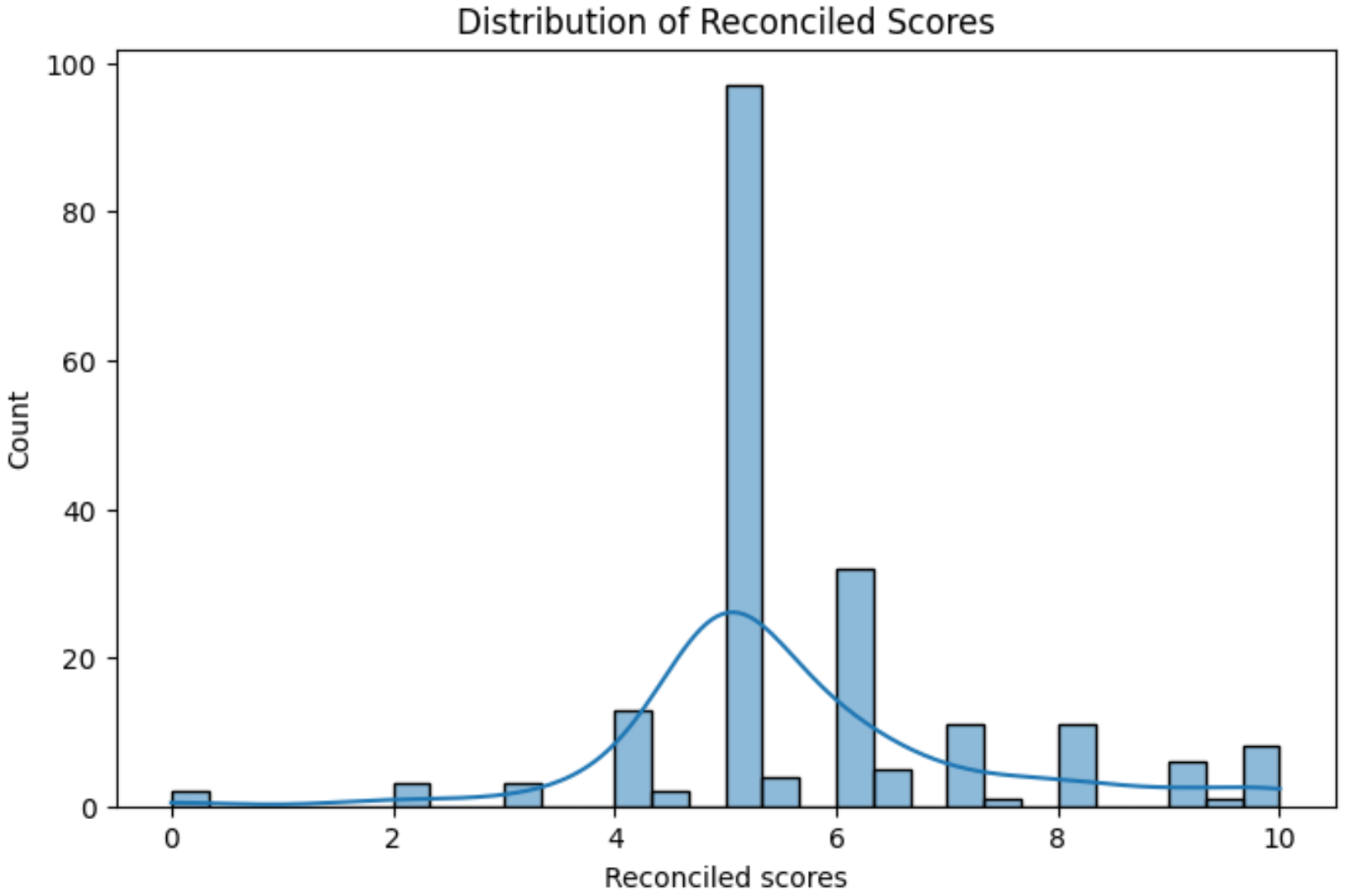
  c) ***Voice and negotiation*** – *references to communication, persuasion, or resistance in interactions*

- A rating scale developed guided by existing frameworks in literature and each dimension in the transcript assigned a score which later was aggregated to constitute a continuous *agency score* based on these qualitative assessments.

- This expert-coded score served as the external benchmark for evaluating linguistic features derived from the same text.

Distribution and Descriptive Statistics of Agency Score

Mean	5.65
Std	1.63
Min	0
25%	5
50%	5
75%	6
Max	10



Methodology

- We rely on techniques taken from text analysis/natural language processing
- We focus on *unsupervised machine learning approaches* (i.e. approached that do not require any annotation/training dataset to train the algorithms)
- We both rely on methods that focus on the structure of the phrases, on the presence of some particular words, and also on more advanced techniques that rely on state-of-the-art NLP methods (i.e. word embeddings)

Methodology

Subject Frequency Analysis

- Identified most common subjects in responses: *I, we, husband, he, decision, purchase, money, it, children*.
- Computed **percentage frequency** and **absolute counts** of each subject.
- Constructed **dummy variables** to classify women as "high agency" when usage exceeds specific thresholds.

Verb Category Classification

- Grouped verbs into semantic categories: *active agency, passive agency, causative, obligation/ necessity, volition/desire, social influence, self-directed*.
- Created percentage-based metrics and dummies for each category.

Noun Frequency Metrics

- Analyzed frequency of specific nouns (e.g., purchase, decision, husband, money, household, children).
- Absolute-count dummies show occasional correlations, but **very low overall frequency** → metrics not robust.

Word Embedding

Bigram Analysis

- Identified most frequent bigrams/trigrams after stemming (e.g., "to buy", "my husband", "we had", "want to").
- Build metrics based on the percentage mention of these bigrams.

Nominalization Detection

- Counted **nominalized words** (verbs turned into nouns).
- Calculated both **percentage of nominalizations** and **absolute number**.
- Constructed threshold-based metrics to classify high agency.

Active vs. Passive Verb Detection

- Computed share of **active verbs** and **passive verbs** for each respondent.
- Built threshold-based dummies for "high agency" based on verb type proportions.

Application of Word Embedding in defining Agency

How we applied Embeddings

- Each word or sentence was converted into a **dense vector** in a semantic space.
- Words used in similar contexts have **similar vectors**.
- Entire responses can be represented as averaged or model-generated embeddings.

Building an “Agency Dimension”

- Select seed words associated with **agency** (i.e. using the words from Ash et al. 2023) and calculate the vector representing "agency"
- Compute the average vector for each answer of each woman
- Compute the similarity of the average vector for each woman to the agency vector

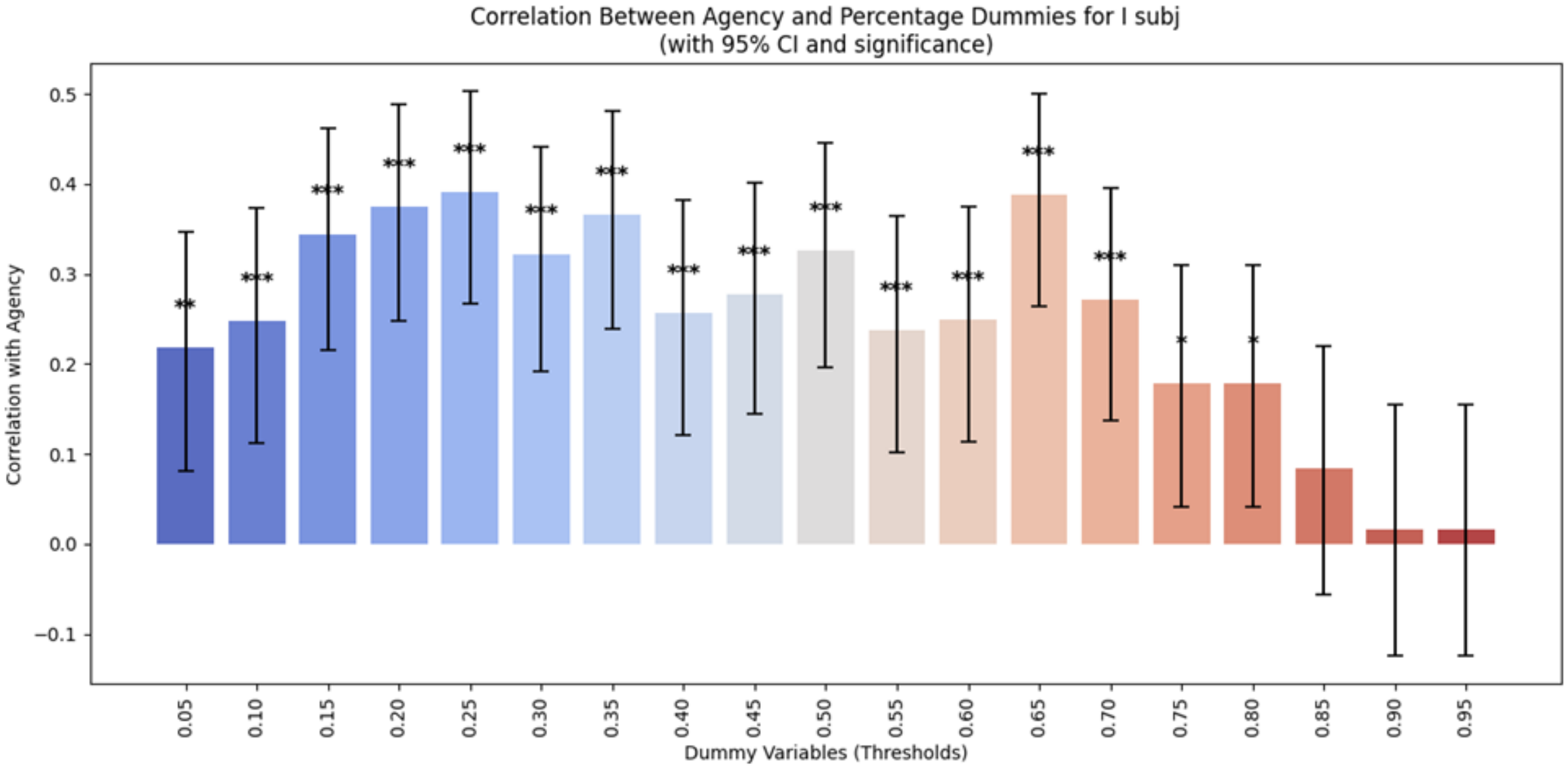
Preliminary findings: Mentioning "I" in %

- We constructed several dummy variables based on the % mentioning of "I"

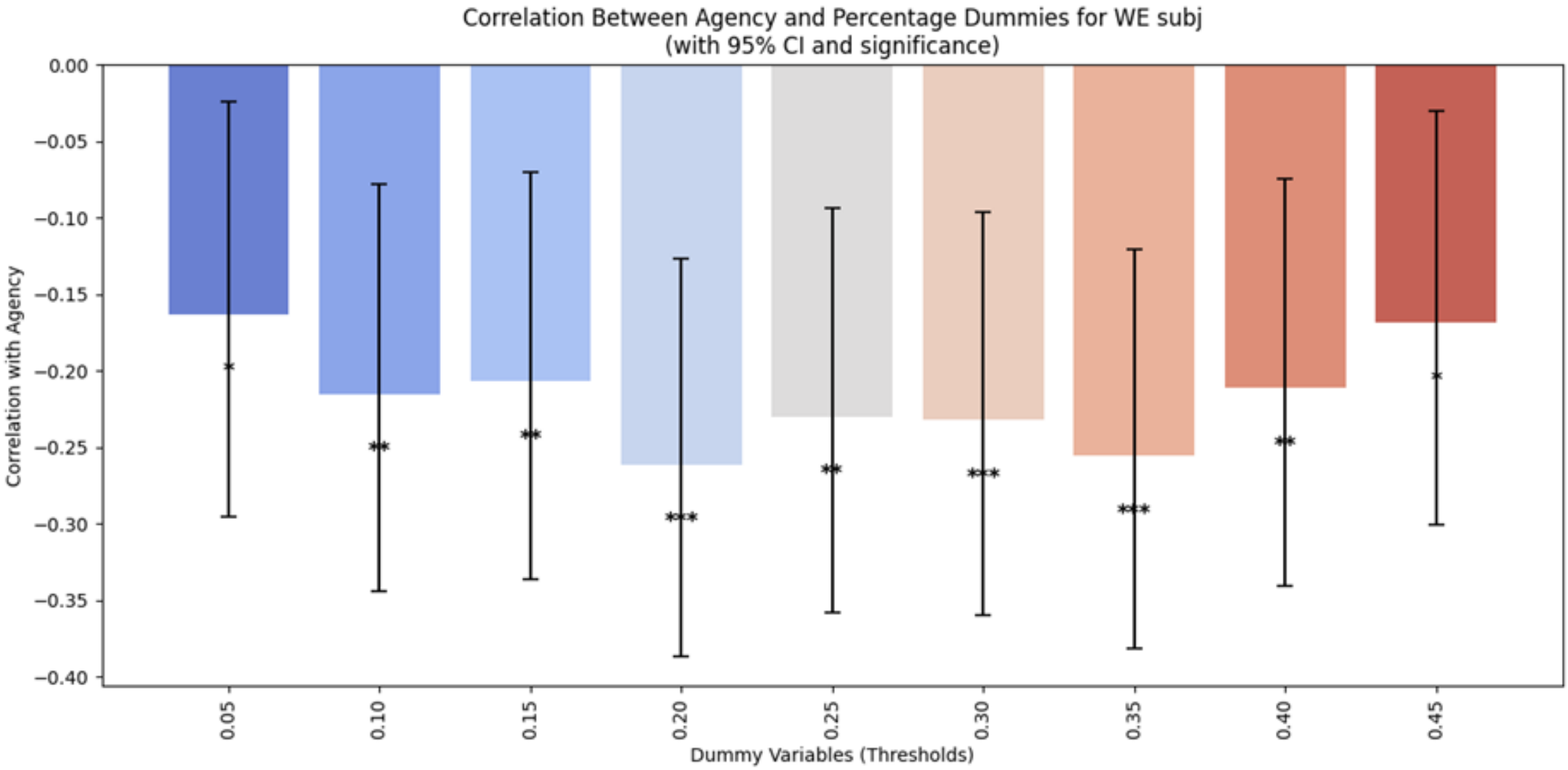
Specifically:

- We calculate the % usage of "I" with respect the total amount of subjects used by the respondent
- We construced several dummies that identify "high agency women" (dummy = 1) if the % usage of "I" exceed a certain threshold (i.e. $dummy_{50}$ = 1 if the woman used "I" more than 50% of the times in her subjects
- We correlate each dummy with the threshold agency score
- Same procedure for metrics constructed based on the % mentioning of "we"

Preliminary findings: Mentioning "I" in %



Preliminary findings: Mentioning "we" in %

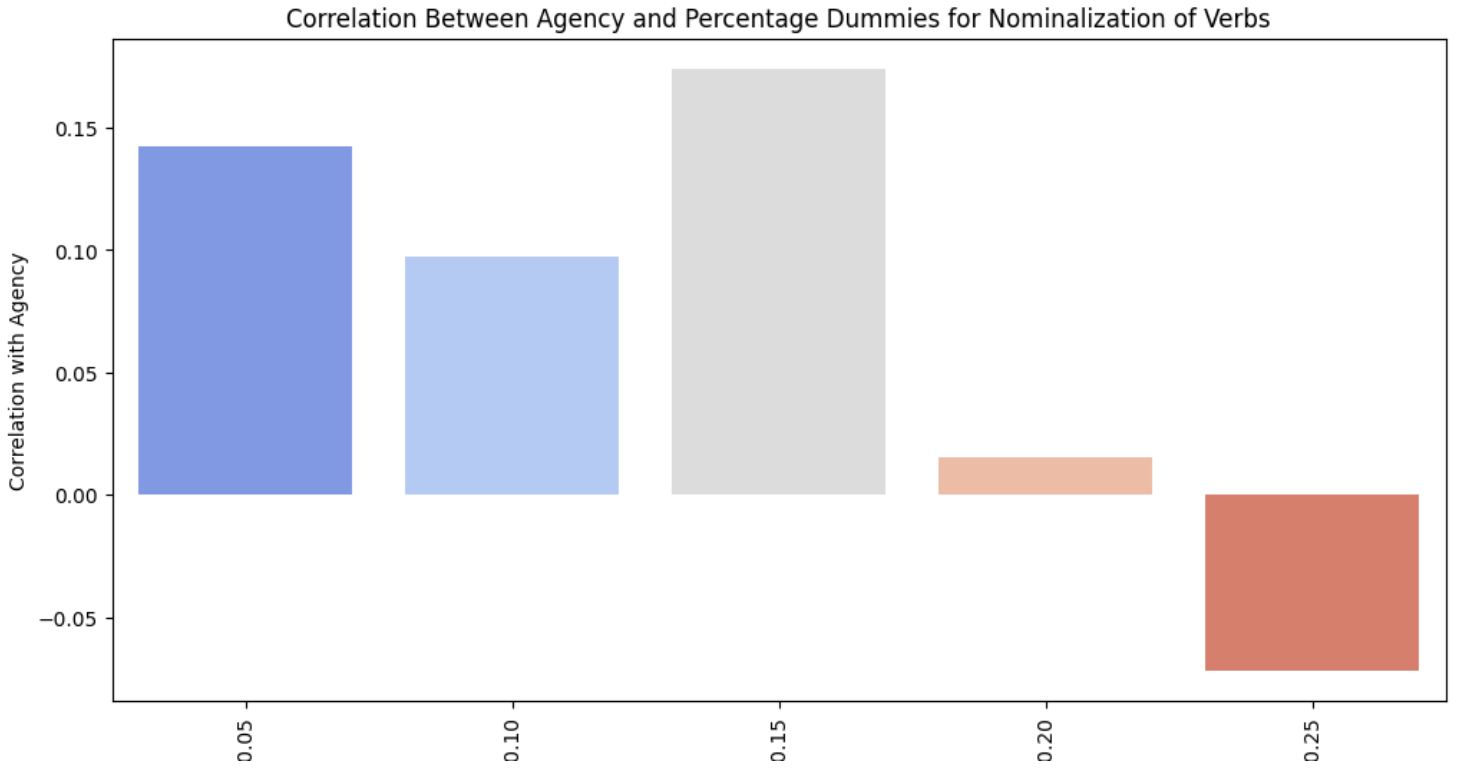


Preliminary findings: Nominalization

- The answers displayed variation in the % usage of nominalization

- The metrics constructed based on nominalization are not good predictor of agency (the highest correlation is 0.17 for a threshold of 15%, and the sign is reverted after a threshold of 20%)

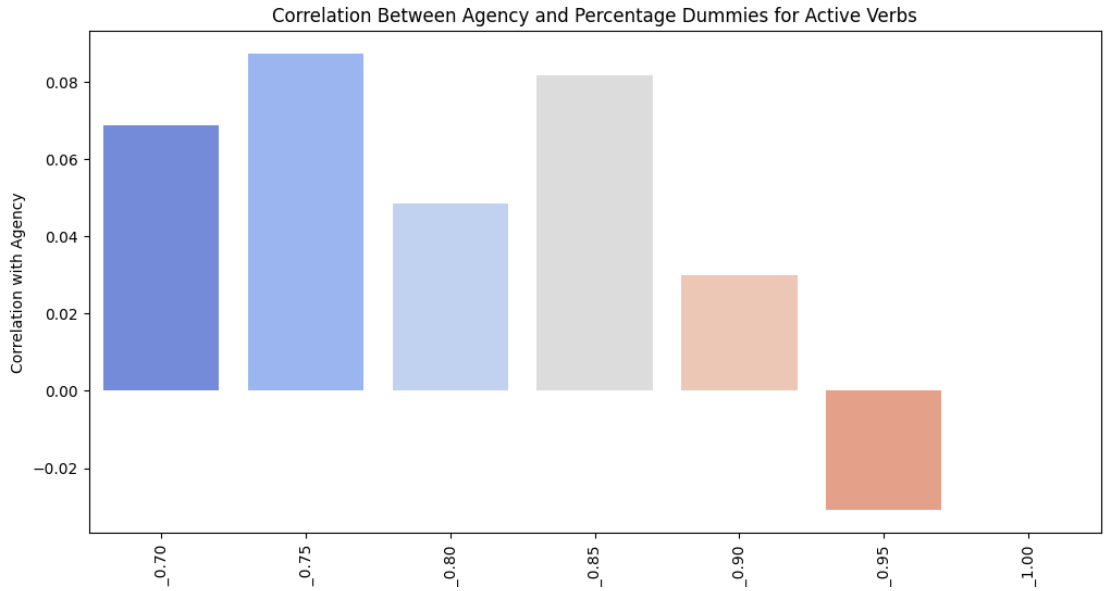
Mean	**0.06**
Std	0.05
Min	0
25%	0.02
50%	0.05
75%	0.09
Max	0.31



Preliminary findings: Passive vs Active verbs

- Not a lot of variation in the usage of Active and Passive verbs (in % to the total verbs used)
- Correlation between dummies constructed on the % usage of Active Verbs and our reference agency score measure do not exceed 0.10 (similar results for passive verbs)
- **For active verbs (in % to total verbs)**

Mean	**0.97**
std	0.06
min	0.5
25%	0.95
50%	1
75%	1
max	1



Summary of the findings and discussion

- **Metrics built based on the % frequency of "I" and "we" perform comparatively well**, displaying meaningful correlations with the main agency metric.
- **Nominalization-based metrics show weak performance**, offering limited explanatory power for identifying women's agency.
- **Passive vs. active verb measures do not reliably capture agency**, with correlations that are small and unstable across thresholds.
- **Verb-type classifications (e.g., causative, volitional, active-agency verbs)** also perform poorly, largely due to limited variation and inconsistent correlation patterns.
    - *Small correlation and flip of the sign in the metrics constructed on the % mention of these verbs over the total amount of verbs*

Next steps

Note:

- The agency concept we are analyzing is largely at an "individualistic" level
- These are preliminary results – we are continuing exploring several alternative measures **(any feedback is very welcomed!!)**

Next steps:

- Constructing metrics based on other sets of bigrams
- Constructing other measures with alternative "agency vector" for the word embedding

THANK YOU

https://magnet.ifpri.info/

MAGNET
MEASURES FOR ADVANCING GENDER EQUALITY

Further details about our work, Contact Marina Rizzi, rizzimrn@gmail.com

Source: https://magnet.ifpri.info/2025/09/02/magnet-pilots-short-form-versions-of-tools/

References

- **Anderson, S., Reynolds, T., & Biscaye, P. (2021).** Measuring women's empowerment: A review of the literature. *World Development, 146*, 105522.
- Ash, E., Stammbach, D., & Tobia, K. (2023). What is (and was) a person? Evidence on historical mind perceptions from natural language. *Cognition*, *239*, 105501.
- Ashwin, J., Rao, V., Biradavolu, M., Haque, A., Khan, A., Krishnan, N., & Peer, N. (2022). Qualitative analysis at scale
- Calvi, R., Penglase, J., & Tommasi, D. (2022). Measuring women's empowerment in collective households. *Journal of Development Economics, 158*, 102927.
- **Costa, J., Botea, M., & Oberlack, C. (2023).** Measures of women's empowerment based on individual-level data: A systematic review of methodological approaches. *World Development, 169*, 106236.
- **Conrad, F. G., & Schober, M. F. (2008).** New frontiers in standardized survey interviewing. In *F. G. Conrad & M. F. Schober (Eds.), Envisioning the survey interview of the future* (pp. 1–18). Wiley.

References

- **Dykema, J., Jones, N. A., Piché, T., & Stevenson, J. (2022).** Open-ended questions in surveys: Improving the quality of responses. *Field Methods, 34*(3), 235–252.
- Donald, A., Koolwal, G., Annan, J., & Goldstein, M. (2020). *Measuring women's agency.* World Bank Policy Research Working Paper No. 9150.
- Franzosi, R., De Fazio, G., & Vicari, S. (2012). Ways of measuring agency: an application of quantitative narrative analysis to lynchings in georgia (1875–1930). Sociological Methodology, 42(1), 1–42.
- Franzosi, R., Doyle, S., McClelland, L. E., Putnam Rankin, C., & Vicari, S. (2013). Quantitative narrative analysis software options compared: Pc- ace and caqdas (atlas. ti, maxqda, and nvivo). Quality & Quantity, 47(6), 3219–3247.
- Jayachandran, S., Biradavolu, M., & Cooper, J. (2023). Using machine learn- ing and qualitative interviews to design a five-question survey module for women's agency. World Development, 161, 106076.
- **Kabeer, N. (1999).** Resources, agency, achievements: Reflections on the measurement of women's empowerment. *Development and Change, 30*(3), 435–464.
- **Klugman, J., Hanmer, L., Twigg, S., Hasan, T., McCleary-Sills, J., & Santamaria, J. (2014).** *Voice and agency: Empowering women and girls for shared prosperity.* World Bank.

References

- **Maynard, D. W., Houtkoop-Steenstra, H., Schaeffer, N. C., & van der Zouwen, J. (2002).** *Standardization and tacit knowledge: Interaction and practice in the survey interview.* Wiley.

- **Miedema, S. S., Haardörfer, R., & Yount, K. M. (2020).** A systematic review of women's empowerment and health outcomes. *Health Psychology Review, 14*(1), 1–34.

- **Mullainathan, S., & Obermeyer, Z. (2017).** Detecting psychological patterns from narrative survey responses. *American Economic Review Papers and Proceedings, 107*(5), 168–172.

- **Olson, K., Smyth, J. D., & Wood, H. (2020).** Open-ended questions in online surveys: *Measurement, interpretation, and analysis.* Cambridge University Press.

- **Schmieder, J. F., Skopek, J., & Trappmann, M. (2023).** Machine learning and text responses in surveys: Opportunities and challenges. *Journal of Survey Statistics and Methodology, 11*(3), 567–590.

- **Schober, M. F., Conrad, F. G., & Antoun, C. (2015).** Survey interviews and new modes of data collection. In *J. D. Wright (Ed.), International encyclopedia of the social & behavioral sciences* (2nd ed., pp. 126–132). Elsevier.

- **Tumasjan, A., Kunz, R. E., & Braun, R. (2021).** Using text in surveys: A review of text-based response features for measurement. *Public Opinion Quarterly, 85*(S1), 410–445.