

From Theoretical Construct to Codeable Dataset

Building a content-analysis strategy for **verbal government retaliation** against firms

JOANNA J. REDDICK, Ph.D. Candidate
UNIVERSITY OF GEORGIA

DEFINE: *Theory → Construct definition*
BUILD: *Corpus construction → Statement extraction*
TRUST: *Phenomenon validation → Human validation → Pilot coding*

Inspiration and Research Questions

The New York Times

Delta and Coca-Cola face backlash from Republicans after opposing Georgia voting law.



By David Gelles

April 1, 2021

For two weeks, [Delta Air Lines](#) and [Coca-Cola](#) had been under pressure from [activists](#) and [Black executives](#) who wanted the companies to publicly oppose a new law in Georgia that makes it harder for people to vote. On Wednesday, six days after the law was passed, both companies stated their “crystal clear” opposition to it.

Now Republicans are mad at the companies for speaking out. Hours after the companies made their statements, Gov. Brian Kemp, a Republican, took aim at Ed Bastian, the chief executive of Delta, [accusing him](#) of spreading “the same false attacks being repeated by partisan activists.” And Republicans in the Georgia state legislature [floated the idea](#) of increasing taxes on Delta as retribution.

Why do government actors retaliate against some firms that engage in CSA?

When do government retaliatory responses become more intense?

Limited examples of construct operationalization

PROBLEM #1: HOW DO I...

Define the construct, unit of analysis, & text scope?

What is the definition of the construct?

What is the context?

THEORY FIRST APPROACH

Verbal retaliation

- **What is it?** Deliberate, negative statements to firm behavior after perceived violation of expectations (Aquino et al., 2001; Skarlicki & Folger, 1997)

Corporate sociopolitical activism (CSA)

- **What is the context?** Public actions and stances that firms take on issues considered contentious yet salient (Bhagwat et al., 2020; Hambrick & Wowak, 2021; Nalick et al., 2016)

RESEARCH STRATEGY

TRANSLATE THE DECISIONS INTO RESEARCH DESIGN

- Theory first
- Search content broadly
- Refine what is collected
- Determine coding requirements
- Validation plan

Unclear how to capture the evidence

PROBLEM #2: HOW DO I...

Build the corpus?

Where is the evidence likely to occur?

How broad should the population be?

What does prior literature tell us?

SEARCH CONTENT BROADLY

Evidence-based observation search

- **Population?** S&P 1500
- **Context or Speaker?** Common denominator = SPEAKER
- **Government Speaker Identification?** Title, not name
- **Source:** Factiva

RESEARCH STRATEGY

TRANSLATE THE DECISIONS INTO RESEARCH DESIGN

- Theory first
- Search content broadly
- Refine what is collected
- Determine coding requirements
- Validation plan

Unclear how to capture the evidence

PROBLEM #3: HOW DO I...

Transform articles into codeable observations?

What is my unit of analysis?

How do I attribute statements to the correct speaker?

What tools can accomplish this reliably?

REFINE WHAT IS COLLECTED

1. Corpus construction

- Parse articles
- Clean & de-duplicate
- Crosswalk with Compustat

2. Statement extraction

- Natural Language Processing via spaCy (Grimmer & Stewart, 2013; Iqbal & Pfarrer, 2025; Schmitt, Kubler, Papadakis, & LaTraon, 2019)
- Sentence segmentation, government actor identification, quote attribution

RESEARCH STRATEGY

TRANSLATE THE DECISIONS INTO RESEARCH DESIGN

- Theory first
- Search content broadly
- Refine what is collected
- Determine coding requirements
- Validation plan

From Factiva search to a structured corpus to extracted statements

- 1 Parse**
Split each Factiva download into articles.
Extract information to create a dataset.

- 2 Clean & de-duplicate**
Standardize firm names on GVKEY
Resolve duplicate cases, multi-firm articles, same-firm articles, & fuzzy match

- 3 Crosswalk**
Merge on GVKEY to attach official S&P 1500 identifiers.
Build a firm-coverage crosswalk to combine Factiva output with Compustat.

BUILD: RESULTS

CLEANED ARTICLES

77,117
one row per article

EXTRACTED STATEMENTS

96,106
one row per statement

DISTINCT GOV ACTORS

9,508
609 firms covered

Evidence alone does not establish the construct...yet

PROBLEM #4: HOW DO I...

Measure the construct consistently?

What counts as verbal retaliation?

What information do coders need to evaluate?

How do I ensure coders apply the construct consistently?

DETERMINE CODING REQUIREMENTS

Create the replicable structure for coding

- Develop theory-driven codebook
- Define coding criteria, establish decision rules, give examples and make rules for edge cases

Retaliation Presence: is it retaliation language?

Retaliation Intensity: how intense is the retaliatory language?

RESEARCH STRATEGY

TRANSLATE THE DECISIONS INTO RESEARCH DESIGN

- Theory first
- Search content broadly
- Refine what is collected
- Determine coding requirements
- Validation plan

Evidence alone does not establish the construct...yet

Conditional on retaliation presence, intensity is analyzed as a continuous measure.

Retaliation Presence: blame attribution & consequence signaling

Retaliation Intensity: moral condemnation, & authority reaffirmation

Sources: Aquino et al., 2001; Branscombe et al., 1999; Gollwitzer & Sjöström, 2017; McCullough et al., 2012; Schelling, 1960; Skarlicki & Folger, 1997; Turner, 2017

CODEBOOK OUTCOME

Retaliation Present = 1

- ✓ Blame attribution
- ✓ Consequence signaling

INTENSITY COMPOSITE SCORE

Entire statement

LIWC dictionaries:

- ✓ Authority reaffirmation
- ✓ Moral condemnation

+

CoreNLP:

- ✓ Overall sentiment score

House Speaker David Ralston acknowledged early Thursday morning that the proposal was a response to Delta's statement.

"They like our public policy when we're doing things that benefit them, and they reap the rewards of those benefits and then turn around and do this," the Blue Ridge Republican told reporters. "As all of you know, I can't resist a country boy line or two, you don't feed a dog that bites your hand. You've got to keep that in mind."

Blame Attribution

**Authority
Reaffirmation**

**Consequence
Signaling**

**Moral
Condemnation**

Evidence alone does not establish the construct...yet

PROBLEM #5: HOW DO I...

Know my measure is valid?

Did I recover the phenomenon I intended to study?

Do the extracted statements accurately represent the speakers?

Can trained coders identify retaliation consistently?

VALIDATION PLAN

Confirm phenomenon

- **Phenomenon validation:** LLM-assisted audit = use known retaliation cases to confirm they are reflected in the dataset
- **Human Validation:** Verify extracted speakers and statements
- **Pilot Coding:** Refine the codebook and assess coding consistency

RESEARCH STRATEGY

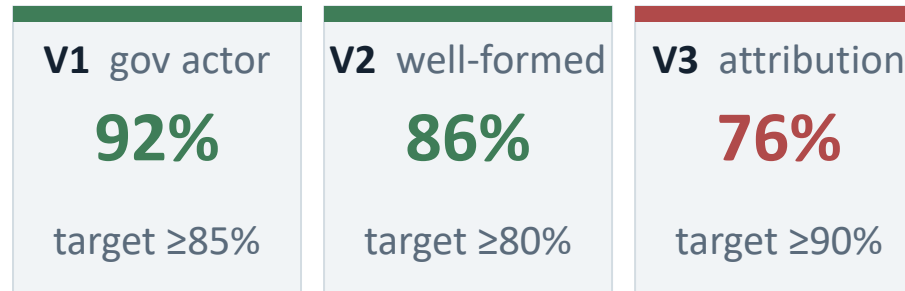
TRANSLATE THE DECISIONS INTO RESEARCH DESIGN

- Theory first
- Search content broadly
- Refine what is collected
- Determine coding requirements
- Validation plan

Four extraction validation checks, in progress:

So far...

Coding a **200-article / 553-statement pilot** to benchmark vs. research targets.

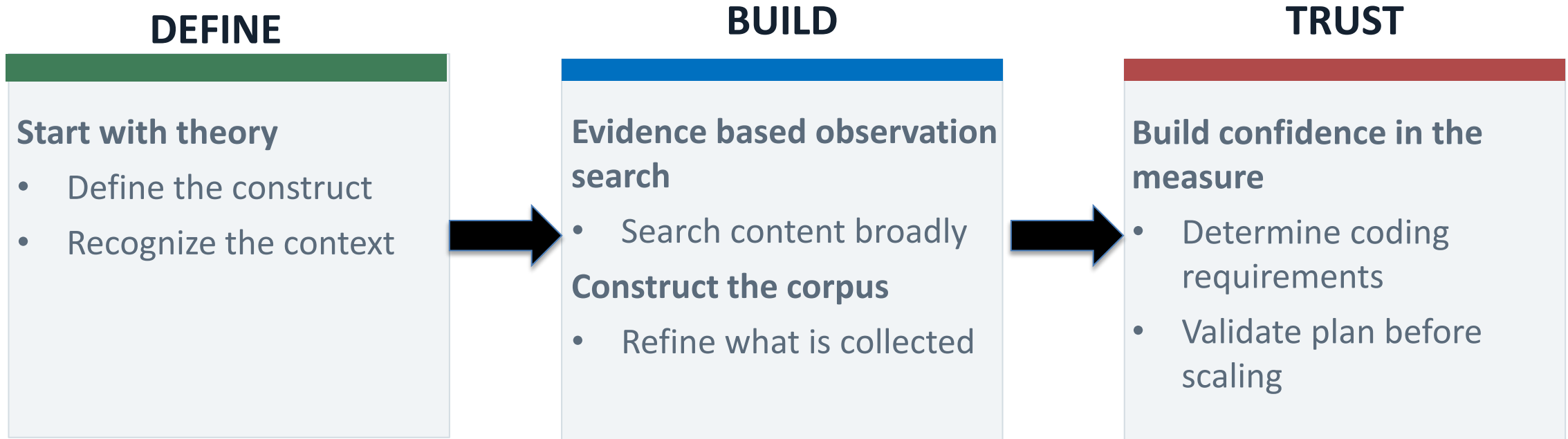


Conducted before large-scale coding.

Use real-time updates to check progress.

Pinpoints where the extraction process needs refinement.

Theoretical construct operationalization = sequential research design



Content analysis tools support theory-driven measurement.

Thank you!

Questions?

Email: joanna.reddick@uga.edu



APPENDIX

Defining the unit: the government-actor statement

How spaCy works.

One attributed statement by a government actor = one codeable observation.

Who counts as a government actor?

Two-tier title lexicon

- Strong titles (Senator, Governor, Atty General)
- Context titles need a nearby government cue.

Disambiguation

- e.g., Chairman: “Fed Chairman” vs. “3M chairman”
- Govt entities: e.g. SEC, White House, Congress

What counts as a statement?

Dependency-parsed attribution

- Attribution verbs: subject = government actor
- Direct-quote detection.

Lightweight coreference

- “he/she said” resolves to the nearest prior speaker, carried at lower confidence.

One pipeline, five transformations



Each transformation is a deliberate, documented decision — and each one was measured before it was trusted. **The talk walks the pipeline left to right.**

CLEANED ARTICLES

77,117

one row per article

EXTRACTED STATEMENTS

96,106

one row per statement

DISTINCT GOV ACTORS

9,508

609 firms covered