Building a Robot Judge: Data Science for the Law

11. Syntactic Dependency Parsers

Elliott Ash

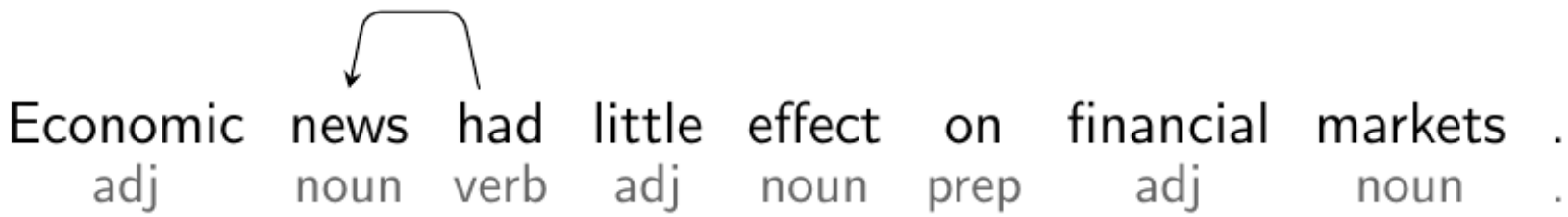
Dependency Grammar

- The basic idea:
  - **Syntactic structure** consists of **words**, linked by binary symmetric relations called **dependencies**.
  - Dependencies identify the grammatical relations between words.

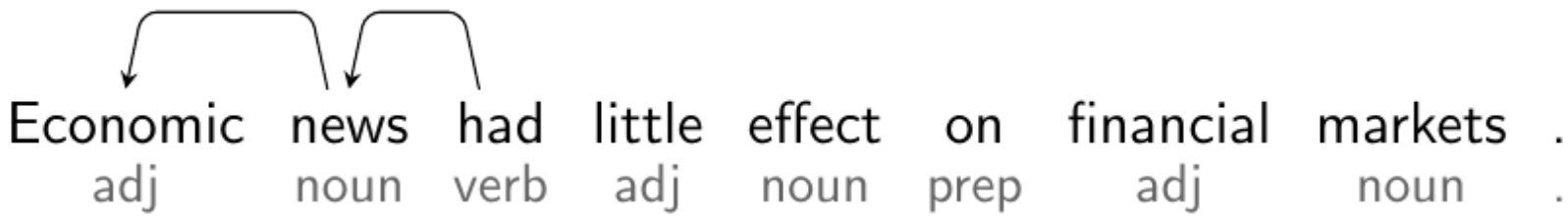
Dependency Structure

Economic	news	had	little	effect	on	financial	markets	.
adj	noun	verb	adj	noun	prep	adj	noun	.

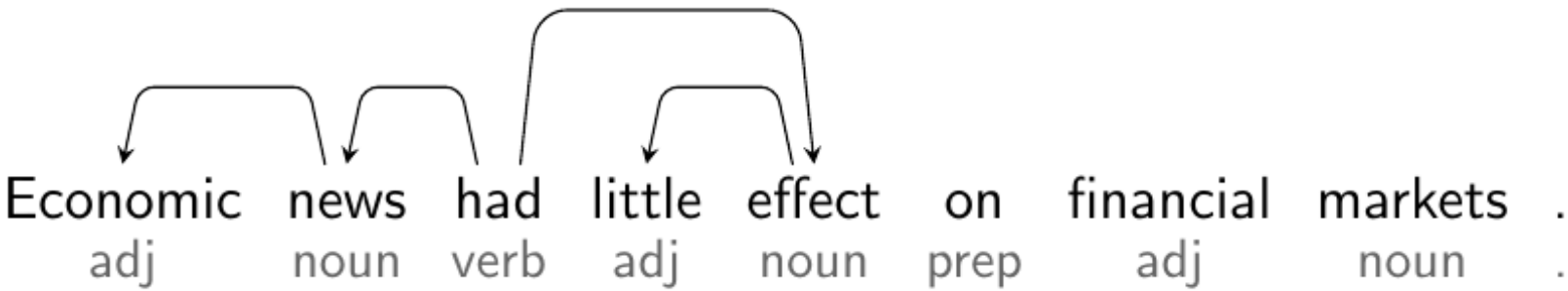
Dependency Structure



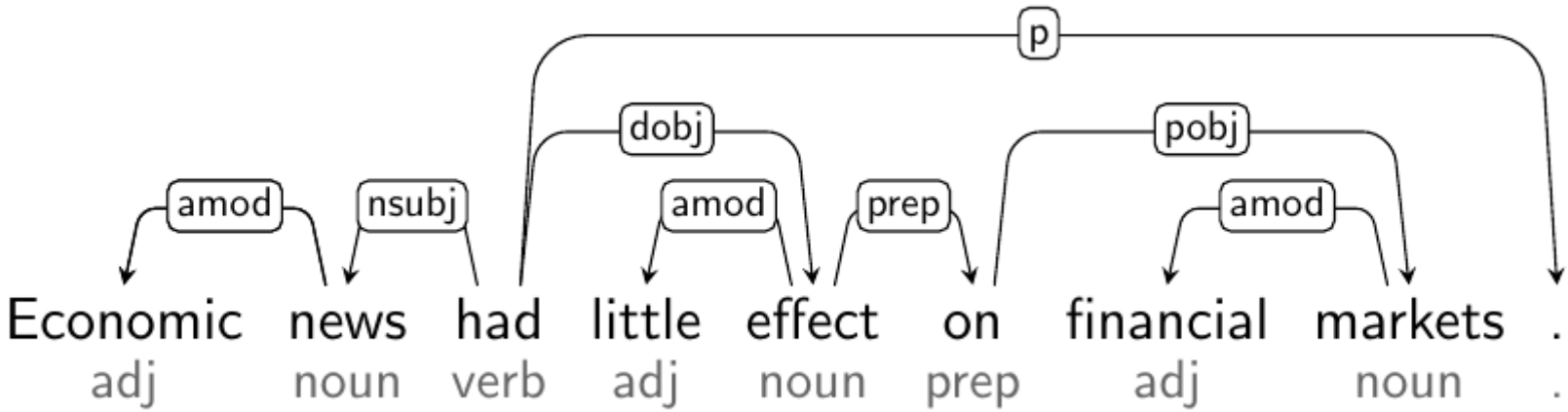
Dependency Structure



Dependency Structure



Dependency Structure

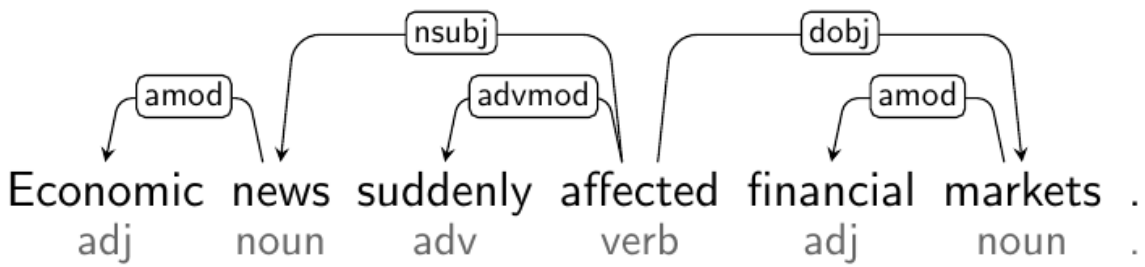


- Dependency structures represent grammatical relations between words in a sentence:
  - head-dependent relations (directed arcs)
  - functional categories (arc labels)
  - structural categories (parts-of-speech)

Heads and Dependents

- A dependency is a one-way link from a "head" token to a "dependent" token:
  - Head determines the syntactic/semantic category of the dependency.
  - Head is obligatory; Dependent may be optional.

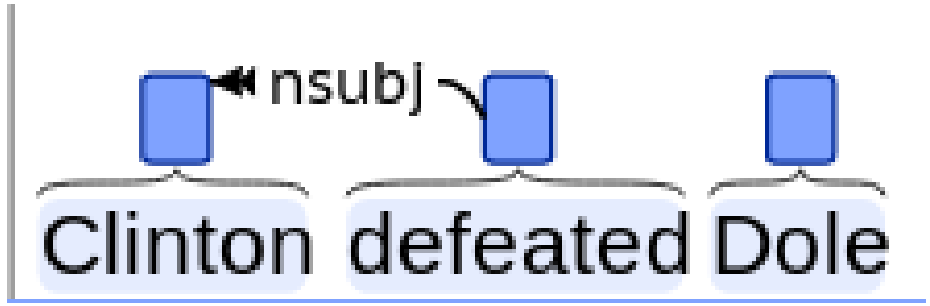
Head	Dependent
Verb	Subject (nsubj)
Verb	Object (dobj)
Verb	Adverbial (advmod)
Noun	Attribute (amod)



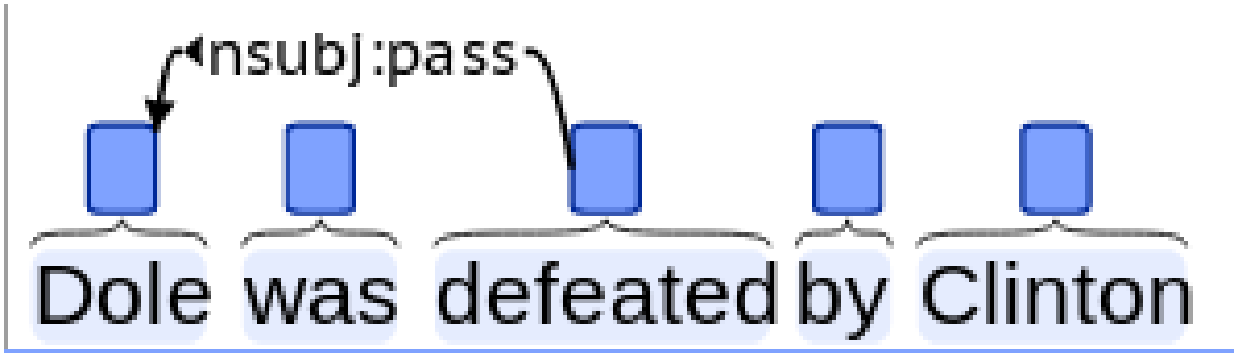
	Nominals	Clauses	Modifier words	Function Words
Core arguments	nsubj obj iobj	csubj ccomp xcomp		
Non-core dependents	obl vocative expl dislocated	advcl	advmod* discourse	aux cop mark
Nominal dependents	nmod appos nummod	acl	amod	det clf case
Coordination	**MWE**	**Loose**	**Special**	**Other**
conj cc	fixed flat compound	list parataxis	orphan goeswith reparandum	punct root dep

Subjects

- **nsubj: nominal subject**
  - non-clausal constituent in the subject position of an active verb.



- **nsubjpass: passive nominal subject**
  - non-clausal constituent in the subject position of a passive verb.



- **csubj: clausal subject**
  - clause in the subject position of an active verb.

Objects

- **dobj: direct object**
  - noun phrase, the (accusative) object of the verb.
  - "She **gave** me a **raise**": dobj(gave, raise)
- **dative: dative or indirect object**
  - noun phrase, the (dative) object of the verb.
  - "She **gave me** a raise": dative(gave, me)
- **pobj: object of a preposition**
  - noun phrase following a preposition
  - "I sat **on** the **chair**": pobj(on, chair)

Adjectives/Attributes

- **acomp: adjectival complement**
  - adjectival phrase which functions as object of verb.
  - "Bill **is honest** ": acomp(is, honest)
- **attr: attribute**
  - noun phrase that is a non-VP predicate usually following a copula verb.
  - "Bill **is** an honest **man**": attr(is, man)
- **amod: adjectival modifier**
  - modifies the meaning of the noun phrase.
  - "Sam eats red meat": amod(meat, red)
- **appositional modifier (appos)** is a noun phrase giving additional information of the preceding noun phrase.

Verb phrases

- **aux: auxiliary**
  - links between a verb and helping verb, including modals.
  - "Reagan has died": aux(died, has)
  - "He should leave": aux(leave, should)
- **auxpass: passive auxiliary**
  - links between a main verb and helping verb in passive constructions.
  - "Laws have been broken": auxpass(broken, been)
- **prt: phrasal verb particle**
  - identifies a phrasal verb: linkes verb with its particle.
  - "They shut down the station": prt(shut, down)

Etc.

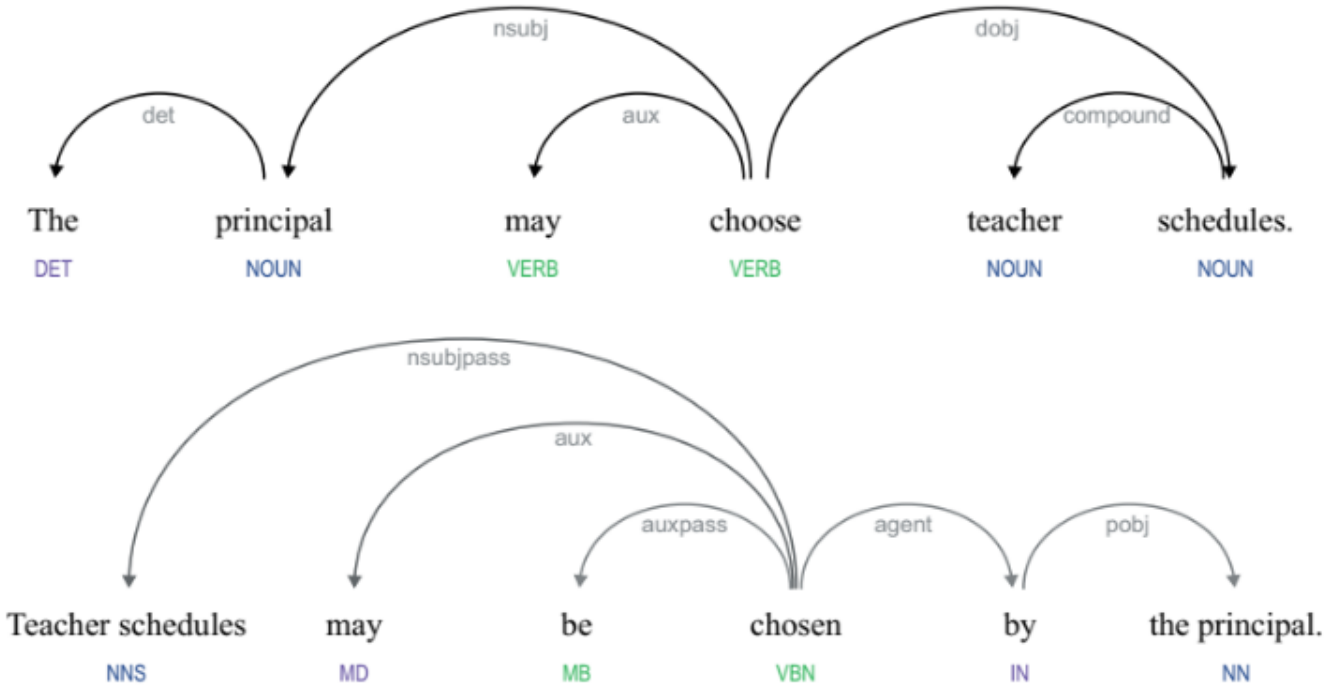
- **neg: negation modifier**
  - captures negation and the word it modifies.
  - "Bill is not a scientist": neg(scientist, not)
  - "Bill doesn't drive": neg(drive, n't)
- **poss: possession modifier**
  - holds between noun phrase and its possessive determiner, or genitive's complement
  - "Bill's clothes": poss(clothes, Bill)
- and a dozen or so more...

Dependency Parsing

- A dependency structure can be defined as a directed graph G , consisting of
  - a set V of nodes (vertices),
  - a set A of arcs (directed edges),
  - a linear precedence order $<$ on V (word order).
- Standard rules:
  - Syntactic structure is complete and hierarchical.
  - Every word has at most one syntactic head (Single-Head).

spaCy

- spaCy (spaci.io) provides an off-the-shelf state-of-the-art dependency parser.
  - by default (**spacy.load('en')**), spaCy will run the tokenizer, tagger, parser, and named entity recognizer.
- For production, use spaCy processing pipelines (`https://spacy.io/usage/processing-pipelines`)
  - customizable and parallelizable



The parser transforms sentences into parse trees, which represent the relations between words in a recursive hierarchical structure.

Noun phrase chunking

```
doc = nlp('Science cannot solve the ultimate mystery
of nature.  And that is because, in the last
analysis, we ourselves are a part of the mystery that
we are trying to solve.')
list(doc.noun_chunks)
```

[Science, the ultimate mystery, nature, the last analysis, we, ourselves, a part, the mystery, we]

- Named entity recognition:
  - New York → New_York, UBS Switzerland → UBS_Switzerland

Co-Reference and Relations

- Parsed dependencies provide important inputs to many useful tasks.

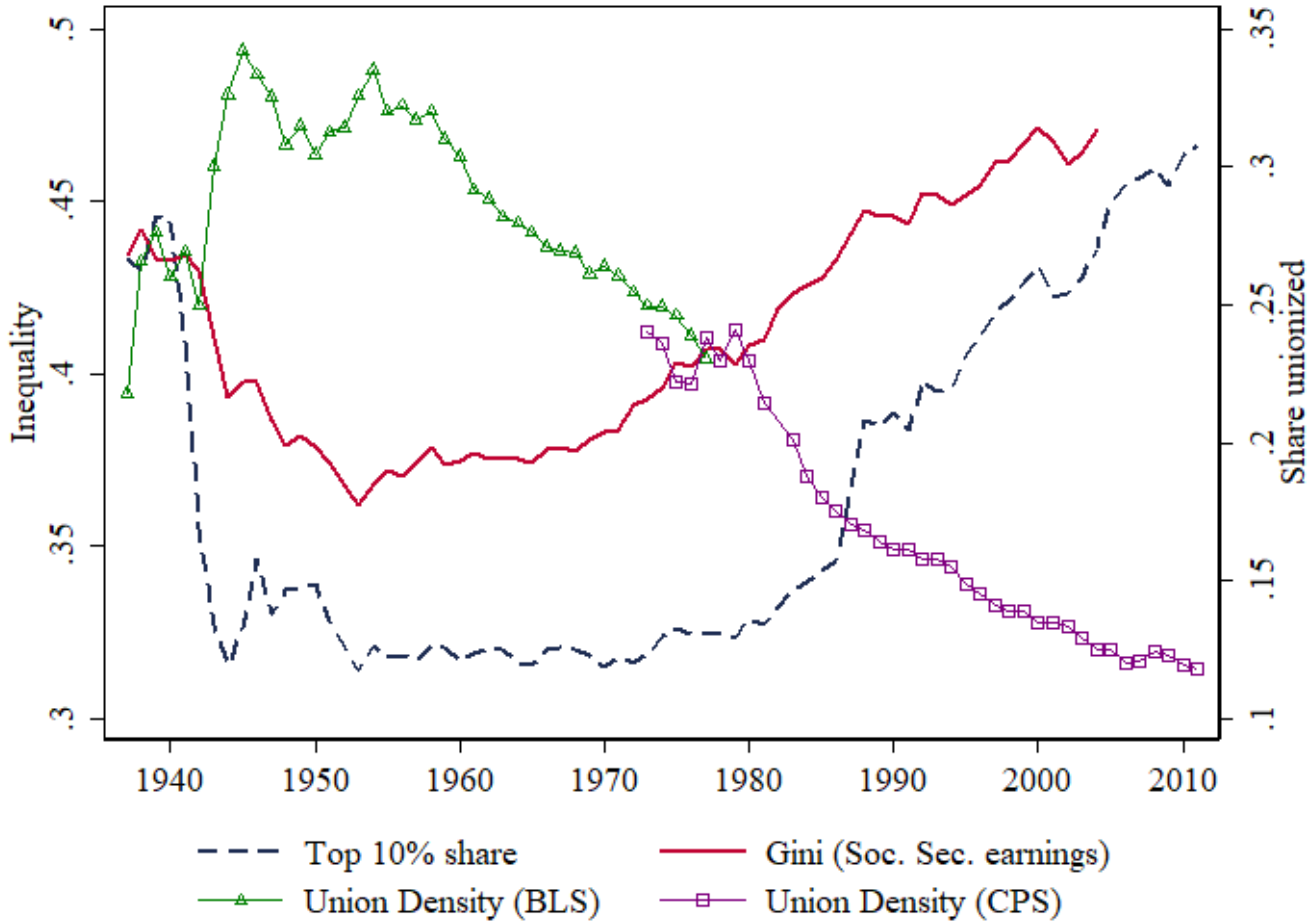
  "My sister has a dog. She loves him"

- Co-reference resolution:

  [sister] loves [dog]

- Relation extraction:
  - sister, has, dog
  - sister, loves, dog

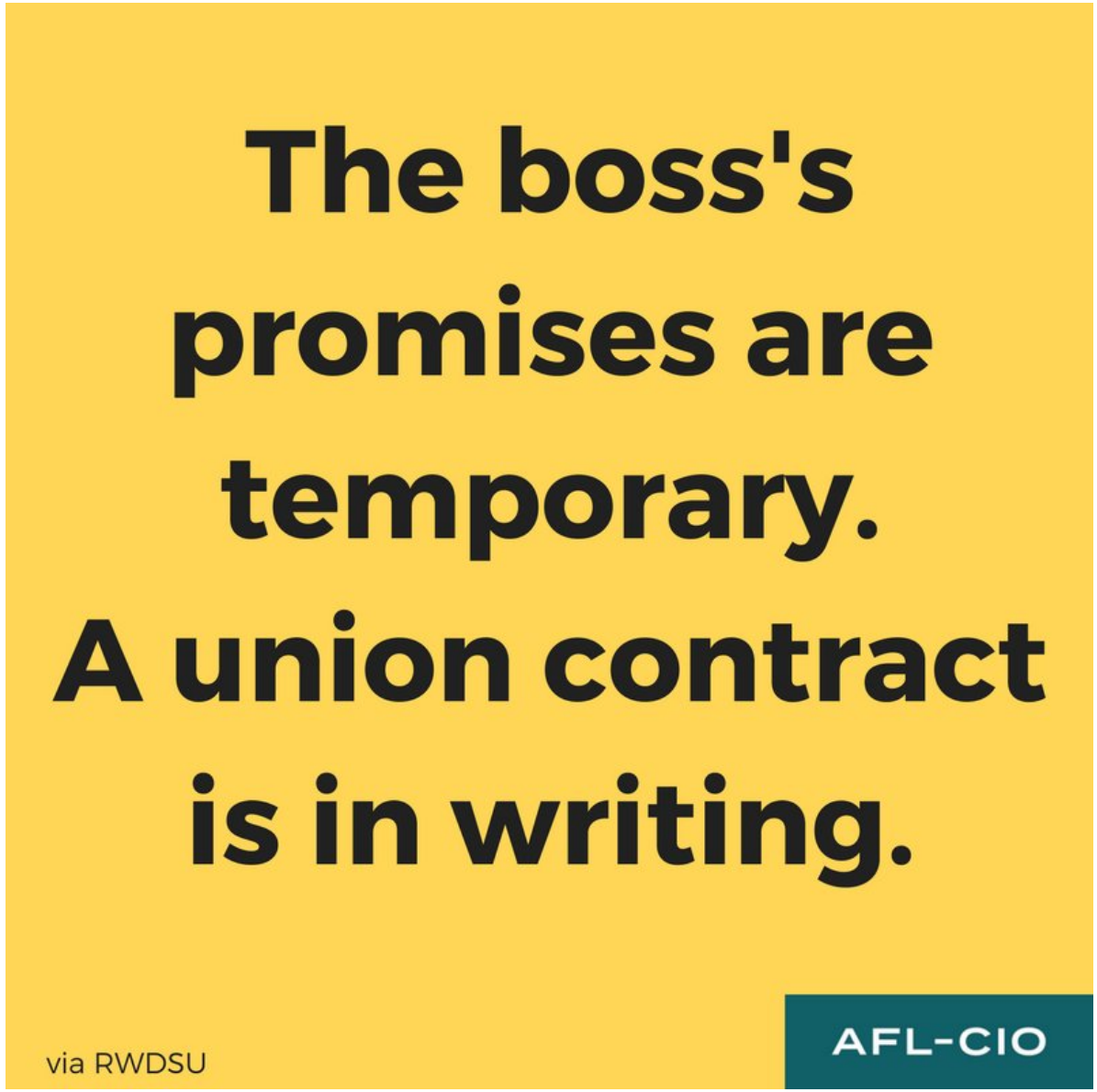
Unions and Inequality

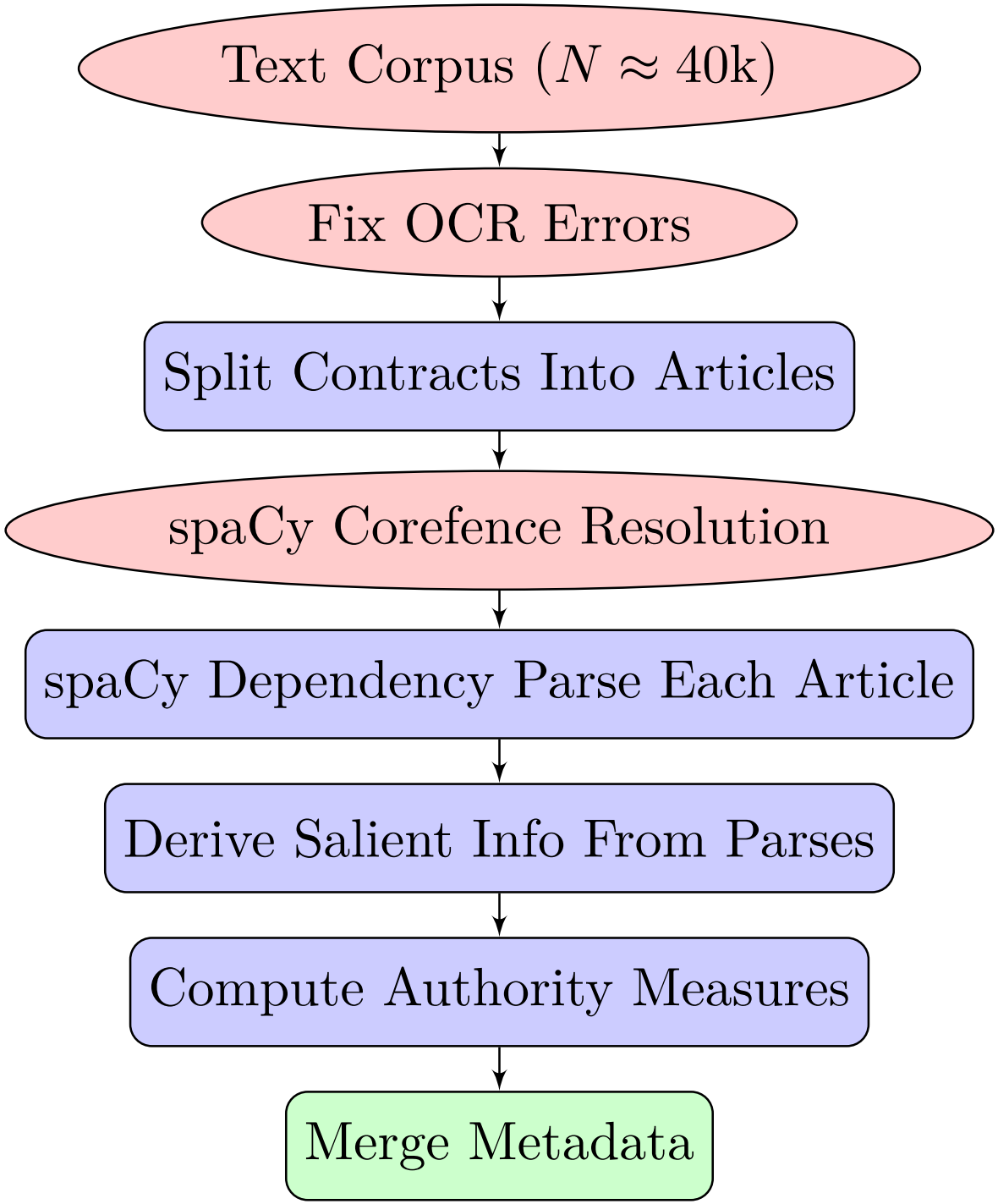


This Project

- Data:
  - new corpus of 30,000 collective bargaining agreements from Canada from 1986 through 2015
- Goal:
  - Quantify "squishy" dimensions of contracts: rigidity, non-wage amenities (e.g. health benefits) and relative worker control.
- Key ideas:
  - use tools from computational linguistics to measure economically and legally relevant contract features:
    - obligations – promises to take actions.
    - entitlements – grants of authority and amenities.
  - examine determinants and consequences of these contractual features.

What do contracts do?




Text Corpus ($N \approx 40$k)
Fix OCR Errors
Split Contracts Into Articles
spaCy Corefence Resolution
spaCy Dependency Parse Each Article
Derive Salient Info From Parses
Compute Authority Measures
Merge Metadata

Co-reference resolution and Sentence tokenization

- Within each section, we performed coreference resolution using the spaCy plugin neuralcoref.
  - convert “him” to “worker”, “it” to “company”, etc.
- Split sections into sentences using spaCy tokenizer.

Syntactic Parse for Contract Statements

- We ran each sentence through three syntactic parsers: spaCy, Stanford CoreNLP, and Google parser.
  - produce results for all three parsers, and also average them.
- Identify syntactic subjects, and form statements around each subject.
  - that is, compound sentences will contain two or more statements.

Extracting Modal Verb Structures

- Subject categories:
  - worker, union, owner, and manager.
- In law, deontic modal verb structures create legal requirements (Kratzer 1991).
  - strict (*shall*, *will*, *must*)
  - permissive (*may*, *can*)
- Statements coded as negative ("shall not" rather than "shall") and active ("shall provide") or passive ("shall be provided").

Special Verbs

- Obligation Verbs (be required, be expected, be compelled, be obliged, be obligated, have to, ought to)
- Prohibition Verbs (be prohibited, be forbidden, be banned, be barred, be restricted, be proscribed)
- Permission Verbs (be allowed, be permitted, be authorized)
- Entitlement Verbs (have, receive, retain).

Contract Statement Logic

Categorization Logic	Examples
Obligations	
Positive, Strict Modal, Active Verb	shall be, shall provide, shall include, shall notify, shall continue
Positive, Strict Modal, Obligation Verb	shall be required, shall be expected, shall be obliged
Positive, Non-Modal, Obligation Verb	is required, is expected
Prohibitions	
Negative, Any Modal, Active Verb	shall not exceed, shall not use, shall not apply, shall not discriminate
Negative, Permission Verb	shall not be allowed, is not permitted
Positive, Strict Modal, Constraint Verb	shall be prohibited, shal be restricted
Permissions	
Positive, Non-Modal, Permission Verb	is allowed, is permitted, is authorized
Positive, Strict Modal, Permission Verb	shall be allowed, shall be permitted
Positive, Permissive Modal, Active Verb	may be, may request, may use, may require, may apply
Negative, Any Modal, Constraint Verb	shall not be restricted, shall not be prohibited
Entitlements	
Strict Modal, Passive Verb	shallbe paid, shall be given, shall not be discharged
Positive, Strict Modal, Entitlement Verb	shall have, shall receive, shall retain
Negative, Any Modal, Obligation Verb	may not be required

Most Frequent Subject-Modal-Verb Tuples

Subject - Modal - Verb	Subject - Modal - Verb
agreement_shall_be	**employee**_shall_be
arbitrator_shall_have	**employee**_shall_be_allowed
board_shall_have	**employee**_shall_be_considered
case_may_be	**employee**_shall_be_entitled
committee_shall_meet	**employee**_shall_be_given
company_shall_pay	**employee**_shall_be_granted
company_shall_provide	**employee**_shall_be_laid_off
company_will_pay	**employee**_shall_be_paid
company_will_provide	**employee**_shall_be_required
decision_shall_be	**employee**_shall_continue
employee_may_request	**employee**_shall_lose

Subject - Modal - Verb
employee_shall_receive
employee_shall_retain
employee_will_be
employee_will_be_allowed
employee_will_be_entitled
employee_will_be_given
employee_will_be_granted
employee_will_be_paid
employee_will_be_required

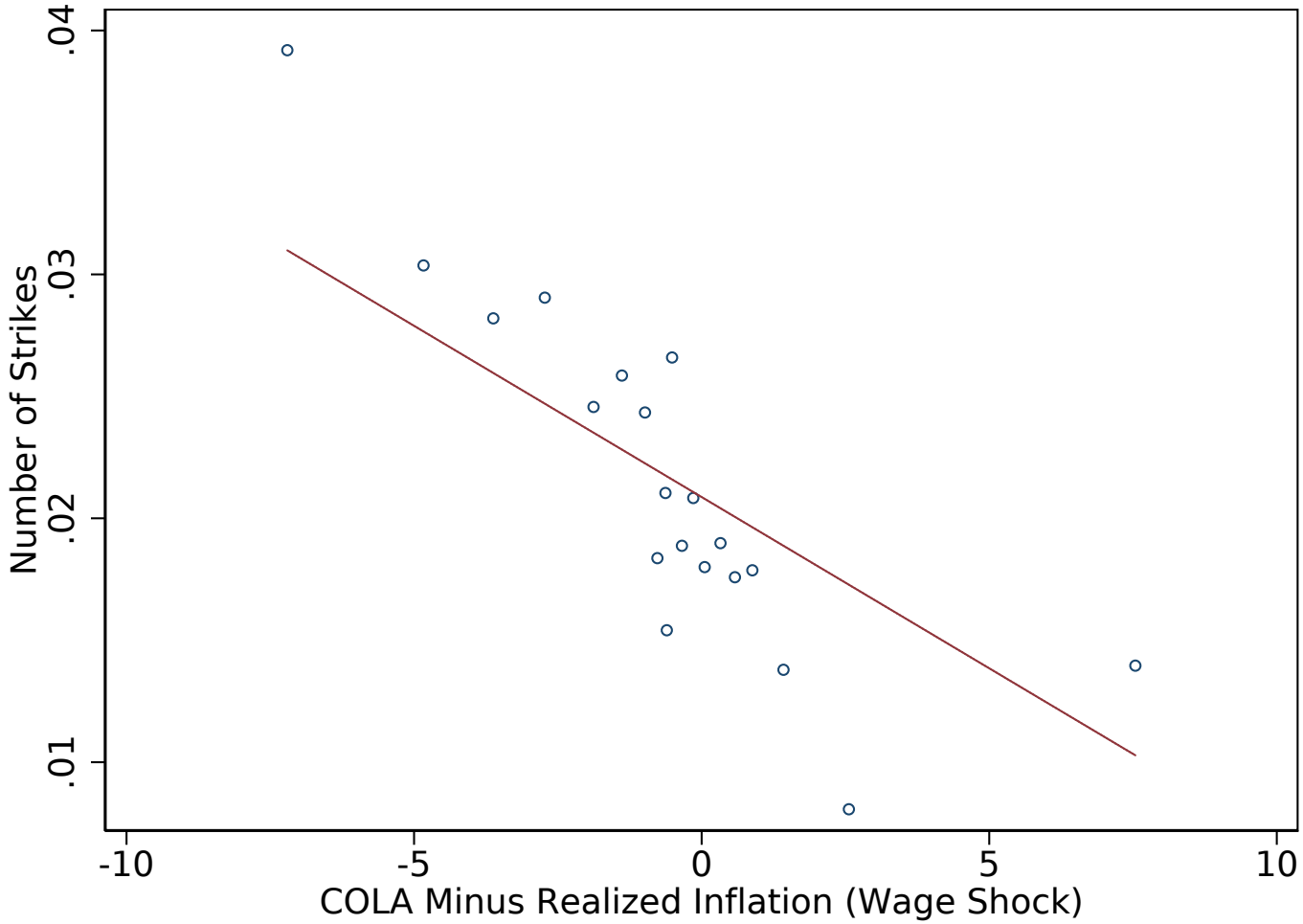
Determinants of Relative Worker Control

- Personal Income Tax (Non-Wage Compensation)↑
- Unemployment Rate (Bargaining Power) ↓
- New Democratic Party In Power (Bargaining Power) ↑
- Number of Employers (Labor Market Competition) ↑
  - All specifications use within-province, within-industry X year variation, and control for rigidity (log number of clauses).

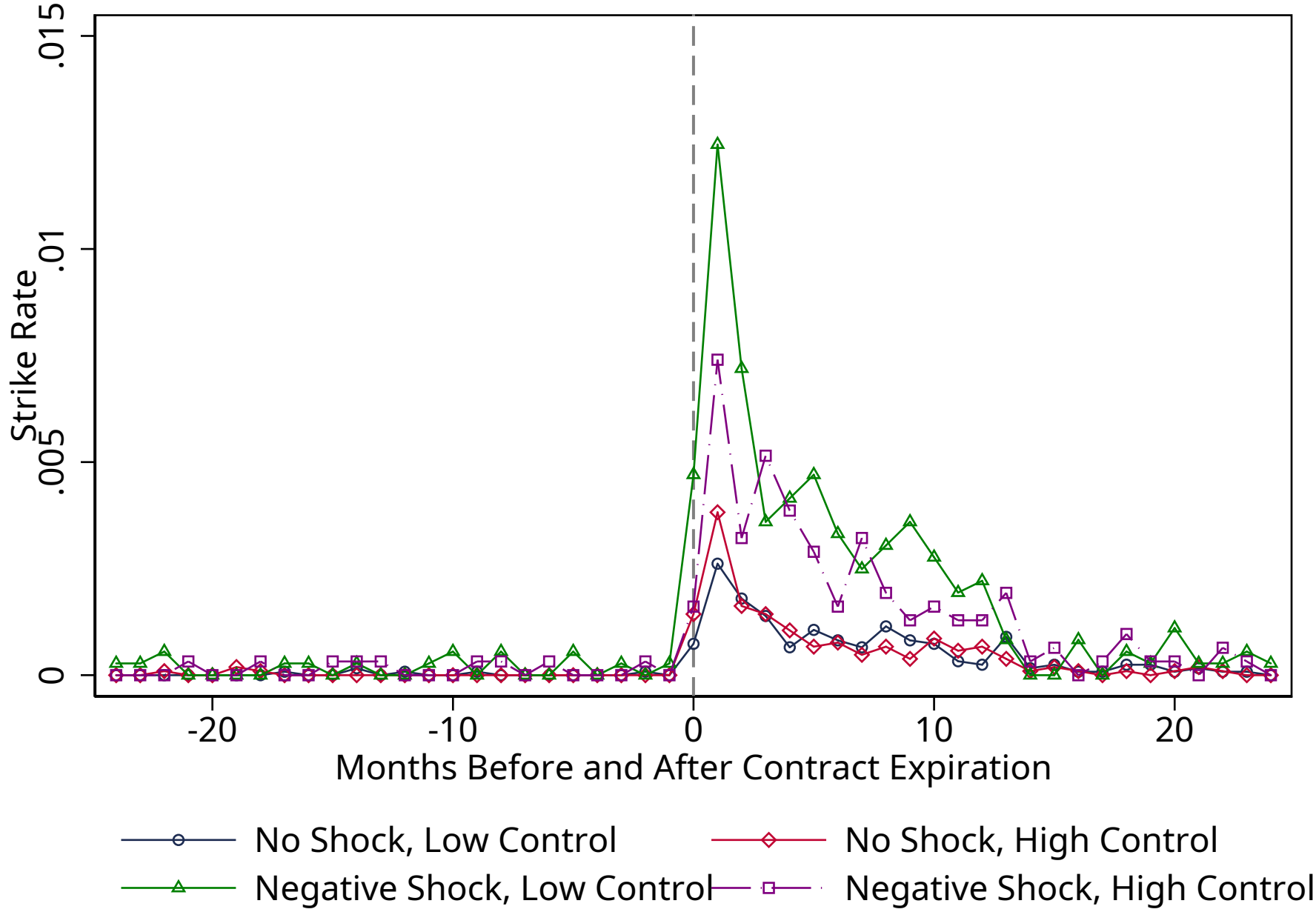
Wages, Control, and Strikes

- Union contracts often have cost-of-living adjustments (COLA), designed to keep wages on track with inflation.
  - Often (particularly in 70s and 80s) actual inflation was either below or above the COLA amount.
  - Means that real wage specified in previous contract is either too high or too low.
- If too low, stakes from improving next contract high, and union more likely to call (costly) strike.
  - If contract gives extensive control rights to workers, negative real wage shock smaller share of value of contract.
    - Hence shorter/fewer strikes with higher worker control.

Effect of COLA-Inflation Wage Shock on Strike Intensity



- Wage Adjustment − Realized Inflation = Unanticipated Wage Shock.


Strike Rate
0
.005
.01
.015
-20
-10
0
10
20
Months Before and After Contract Expiration
No Shock, Low Control
No Shock, High Control
Negative Shock, Low Control
Negative Shock, High Control