Sequencing Legal DNA

NLP for Law and Political Economy

7. Syntactic and Semantic Parsing

Beyond Word Order

- The models we have seen so far have counted words and phrases, or embedded sequences
  - the only language structure used is the ordering of words.

Beyond Word Order

- The models we have seen so far have counted words and phrases, or embedded sequences
  - the only language structure used is the ordering of words.
- How to identify whether the defendant was negligent?
  - "The negligent defendant"
  - "The defendant was negligent"
  - "The defendant, a driver, was negligent"

Beyond Word Order

- The models we have seen so far have counted words and phrases, or embedded sequences
  - the only language structure used is the ordering of words.
- How to identify whether the defendant was negligent?
  - "The negligent defendant"
  - "The defendant was negligent"
  - "The defendant, a driver, was negligent"
- Syntactic and semantic parsing will do this.

Outline

Constituency Parsing

Dependency Parsing: Linguistics

Dependency Parsing: Applications
- Hoyle et al (2019): Discovery of Gendered Language
- Vannoni, Ash, and Morelli (2020): Discretion and Delegation in Texts
- Ash, MacLeod, and Naidu (2020): The Language of Contract
- Webb (2020): AI and the Labor Market

Semantic Role Labeling

Constituency

- The idea of constituency is that groups of words behave as singular functional units in a sentence.
- Some example noun phrases:

Harry the Horse	a high-class spot such as Mindy's
the Broadway coppers	the reason he comes into the Hot Box
they	three parties from Brooklyn

  - these phrases consist of many POS's but function as nouns

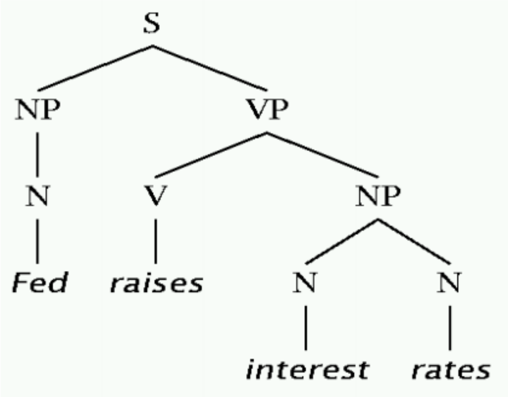
Constituency

- The idea of constituency is that groups of words behave as singular functional units in a sentence.
- Some example noun phrases:

Harry the Horse	a high-class spot such as Mindy's
the Broadway coppers	the reason he comes into the Hot Box
they	three parties from Brooklyn

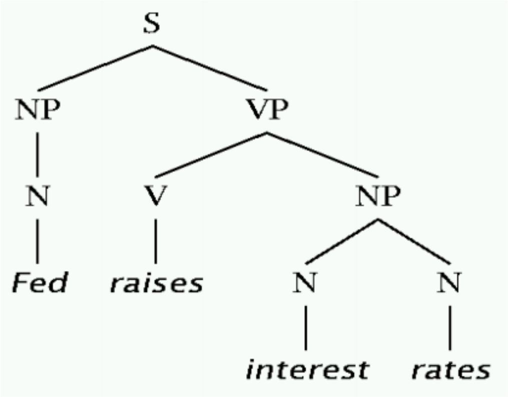
  - these phrases consist of many POS's but function as nouns
- Constituents can be moved around in a sentence (e.g. these prepositional phrases):
  - John talked [to the children] [about drugs].
  - John talked [about drugs] [to the children] .

Constituency Trees



- Phrase structure organizes words into nested constituents:
  - S = subject, NP = noun phrase, VP = verb phrase

Constituency Trees



- Phrase structure organizes words into nested constituents:
  - S = subject, NP = noun phrase, VP = verb phrase
- The "legal memes" we constructed in Week 2 were a form of constituency parsing.

Penn TreeBank Constituent Tags

Table 1.2. The Penn Treebank syntactic tagset

ADJP	Adjective phrase
ADVP	Adverb phrase
NP	Noun phrase
PP	Prepositional phrase
S	Simple declarative clause
SBAR	Subordinate clause
SBARQ	Direct question introduced by *wh*-element
SINV	Declarative sentence with subject-aux inversion
SQ	Yes/no questions and subconstituent of SBARQ excluding *wh*-element
VP	Verb phrase
WHADVP	Wh-adverb phrase
WHNP	Wh-noun phrase
WHPP	Wh-prepositional phrase
X	Constituent of unknown or uncertain category
*	"Understood" subject of infinitive or imperative
0	Zero variant of *that* in subordinate clauses
T	Trace of wh-Constituent

Phrase extraction in spacy

- Noun phrases:

```
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]
```

Phrase extraction in spacy

- Noun phrases:

```
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]

- To get, e.g., prepositional phrases, find **`token`** with tag "ADP" and get **`token.subtree`**

Phrase extraction in spacy

- Noun phrases:

```
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]
```

- To get, e.g., prepositional phrases, find **`token`** with tag "ADP" and get **`token.subtree`**
- Related to named entity recognition (NER):
  - New York → New_York, UBS Switzerland → UBS_Switzerland

Outline

- Constituency Parsing
- Dependency Parsing: Linguistics
- Dependency Parsing: Applications
  - Hoyle et al (2019): Discovery of Gendered Language
  - Vannoni, Ash, and Morelli (2020): Discretion and Delegation in Texts
  - Ash, MacLeod, and Naidu (2020): The Language of Contract
  - Webb (2020): AI and the Labor Market
- Semantic Role Labeling

Dependency Grammar

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

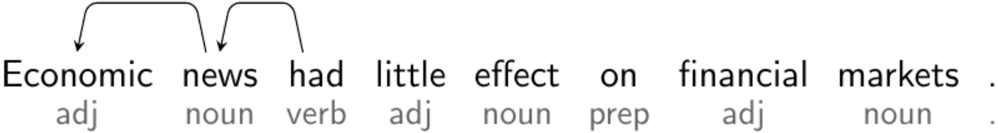
Dependency Structure

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

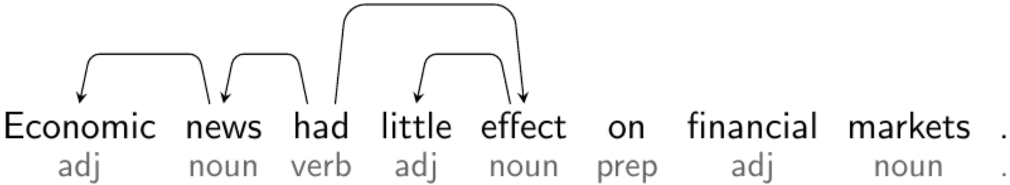
Dependency Structure

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

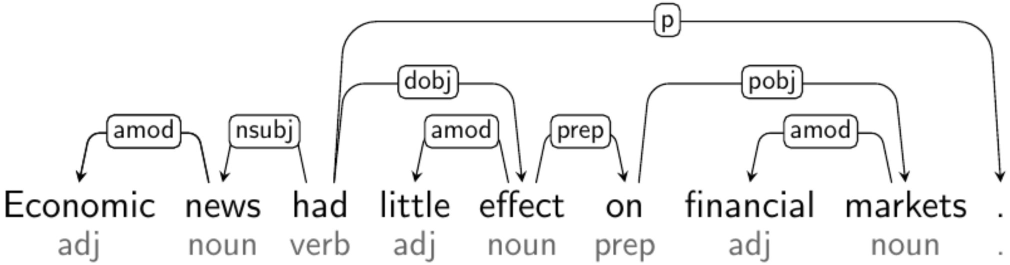
Dependency Structure



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)

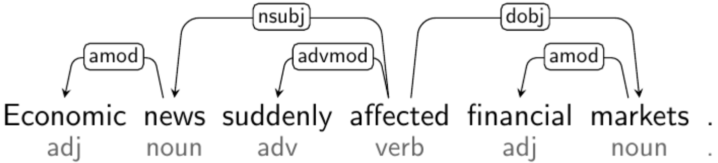
Heads and Dependents

- 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.

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.

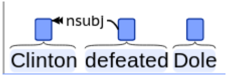
Head	**Dependent**
Verb	Subject (nsubj)
Verb	Object (dobj)
Verb	Adverbial (advmod)
Noun	Attribute (amod)



	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

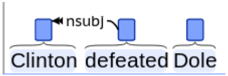
Subjects

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

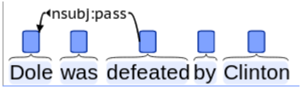


Subjects

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

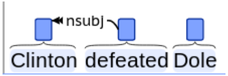


- **nsubjpass: passive nominal subject**
  - non-clausal constituent in the subject position of a passive verb.

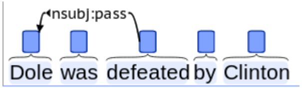


Subjects

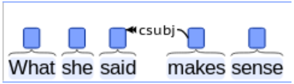
- **nsubj: nominal subject**
  - non-clausal constituent in the subject position of an active verb.



- **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.



- **clausal passive subject (csubjpass)** is a clause in the subject position of a passive verb.

Objects

- **dobj: direct object**
  - noun phrase, the (accusative) object of the verb.
  - “She **gave** me a **raise**”: dobj(gave, raise)

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)

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)

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)

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)

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)

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)

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: links 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)

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)

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)
- **agent (agent)** is the complement of a passive verb that is the surface subject of its active form.
- **expletive (expl)** is an existential there in the subject position.
- and 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.

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

Google Syntactic N-Grams

https://docs.google.com/document/d/14PWeoTkrnKk9H8_7CfVbdvuoFZ7jYivNTkBX2Hj7qLw/edit?usp=sharing

head_word<TAB>syntactic-ngram<TAB>total_count<TAB>counts_by_year

Google Syntactic N-Grams

`https://docs.google.com/document/d/14PWeoTkrnKk9H8_7CfVbdvuoFZ7jYivNTkBX2Hj7qLw/edit?usp=sharing`

head_word<TAB>syntactic-ngram<TAB>total_count<TAB>counts_by_year

- **syntactic-ngram** format: space-separated list of tokens, with format "**word/pos-tag/dep-label/head-index**".
  - word: any non-whitespace character.
  - pos-tag (POS tag)
  - dep-label: stanford-basic-dependencies label.
  - head-index: integer pointing to head of the current token.
  - Example of a syntactic-ngram:

    cease/VB/ccomp/0 **for**/IN/prep/1 **some**/DT/det/4 **time**/NN/pobj/2

Google Syntactic N-Grams

`https://docs.google.com/document/d/14PWeoTkrnKk9H8_7CfVbdvuoFZ7jYivNTkBX2Hj7qLw/edit?usp=sharing`

head_word<TAB>syntactic-ngram<TAB>total_count<TAB>counts_by_year

- **syntactic-ngram** format: space-separated list of tokens, with format "**word/pos-tag/dep-label/head-index**".
  - word: any non-whitespace character.
  - pos-tag (POS tag)
  - dep-label: stanford-basic-dependencies label.
  - head-index: integer pointing to head of the current token.
  - Example of a syntactic-ngram:

    cease/VB/ccomp/0 **for**/IN/prep/1 **some**/DT/det/4 **time**/NN/pobj/2
- **counts_by_year**: year<comma>count items when syngram appeared.
- Example of a complete line:

```
cease cease/VB/ccomp/0 for/IN/prep/1 an/DT/det/4 instant/NN/pobj/2 56 1834,2 1835,1
1856,1 1863,1 1871,1 1872,1 1874,1 1875,3 1880,2 1883,2 1889,1 1904,7 1905,2 1915,5
1918,1 1961,1 1963,5 1973,2 1975,1 1977,1 1981,2 1987,2 1988,1 1989,1 1991,1 1996,5
2000,1 2008,2
```

Outline

Outline

Unsupervised Discovery of Gendered Language

- This paper builds on the “gender bias” NLP papers by adding in syntactic information:

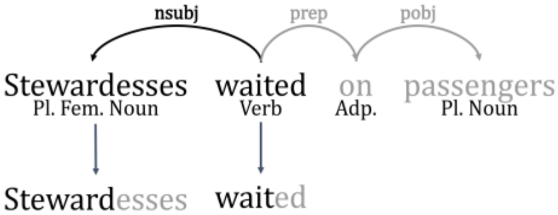


Figure 2: An example sentence with its labeled dependency parse (top) and lemmatized words (bottom).

Unsupervised Discovery of Gendered Language

- This paper builds on the “gender bias” NLP papers by adding in syntactic information:

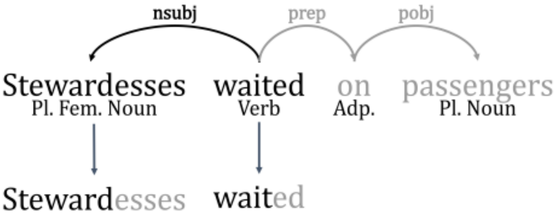


Figure 2: An example sentence with its labeled dependency parse (top) and lemmatized words (bottom).

- Corpus: dependency parse of 3.5 million books from Goldberg and Orwant (2013).
  - 37 million noun-adjective pairs
  - 41-million subject-verb pairs
  - 14 million verb-object pairs

Extracting gendered language

- Hoyle et al (2019) extract the set of adjectives and verbs attached to nouns that are predictive of the gender of the noun.
  - they use a regularized latent variable model
  - the resulting metric is (almost) proportional to PMI.

Extracting gendered language

- Hoyle et al (2019) extract the set of adjectives and verbs attached to nouns that are predictive of the gender of the noun.
  - they use a regularized latent variable model
  - the resulting metric is (almost) proportional to PMI.
- Interpreting the dimensions:
  - categorize adjectives/verbs by sentiment (positive, negative, entural)

Extracting gendered language

- Hoyle et al (2019) extract the set of adjectives and verbs attached to nouns that are predictive of the gender of the noun.
  - they use a regularized latent variable model
  - the resulting metric is (almost) proportional to PMI.
- Interpreting the dimensions:
  - categorize adjectives/verbs by sentiment (positive, negative, entural)
  - categorize adjectives/verbs as related to the body and emotions.

Gendered Adjectives

$\tau_{\text{MASC-POS}}$		$\tau_{\text{MASC-NEG}}$		$\tau_{\text{MASC-NEU}}$		$\tau_{\text{FEM-POS}}$		$\tau_{\text{FEM-NEG}}$		$\tau_{\text{FEM-NEU}}$	
Adj.	Value	Adj.	Value	Adj.	Value	Adj.	Value	Adj.	Value	Adj.	Value
faithful	2.3	unjust	2.4	german	1.9	pretty	3.3	horrible	1.8	virgin	2.8
responsible	2.2	dumb	2.3	teutonic	0.8	fair	3.3	destructive	0.8	alleged	2.0
adventurous	1.9	violent	1.8	financial	2.6	beautiful	3.4	notorious	2.6	maiden	2.8
grand	2.6	weak	2.0	feudal	2.2	lovely	3.4	dreary	0.8	russian	1.9
worthy	2.2	evil	1.9	later	1.6	charming	3.1	ugly	3.2	fair	2.6
brave	2.1	stupid	1.6	austrian	1.2	sweet	2.7	weird	3.0	widowed	2.4
good	2.3	petty	2.4	feudatory	1.8	grand	2.6	harried	2.4	grand	2.1
normal	1.9	brutal	2.4	maternal	1.6	stately	3.8	diabetic	1.2	byzantine	2.6
ambitious	1.6	wicked	2.1	bavarian	1.5	attractive	3.3	discontented	0.5	fashionable	2.5
gallant	2.8	rebellious	2.1	negro	1.5	chaste	3.3	infected	2.8	aged	1.8
mighty	2.4	bad	1.9	paternal	1.4	virtuous	2.7	unmarried	2.8	topless	3.9
loyal	2.1	worthless	1.6	frankish	1.8	fertile	3.2	unequal	2.4	withered	2.9
valiant	2.8	hostile	1.9	welsh	1.7	delightful	2.9	widowed	2.4	colonial	2.8
courteous	2.6	careless	1.6	ecclesiastical	1.6	gentle	2.6	unhappy	2.4	diabetic	0.7
powerful	2.3	unsung	2.4	rural	1.4	privileged	1.4	horrid	2.2	burlesque	2.9
rational	2.1	abusive	1.5	persian	1.4	romantic	3.1	pitiful	0.8	blonde	2.9
supreme	1.9	financial	3.6	belted	1.4	enchanted	3.0	frightful	0.5	parisian	2.7
meritorious	1.5	feudal	2.5	swiss	1.3	kindly	3.2	artificial	3.2	clad	2.5
serene	1.4	false	2.3	finnish	1.1	elegant	2.8	sullen	3.1	female	2.3
godlike	2.3	feeble	1.9	national	2.2	dear	2.2	hysterical	2.8	oriental	2.2
noble	2.3	impotent	1.7	priestly	1.8	devoted	2.0	awful	2.6	ancient	1.7
rightful	1.9	dishonest	1.6	merovingian	1.6	beauteous	3.9	haughty	2.6	feminist	2.9
eager	1.9	ungrateful	1.5	capetian	1.4	sprightly	3.2	terrible	2.4	matronly	2.6
financial	3.3	unfaithful	2.6	prussian	1.4	beloved	2.5	damned	2.4	pretty	2.5
chivalrous	2.6	incompetent	1.7	racial	0.9	pleasant	1.8	topless	3.5	asiatic	2.0

Gendered Verbs (as agent)

$\tau_{\text{MASC-POS}}$		$\tau_{\text{MASC-NEG}}$		$\tau_{\text{MASC-NEU}}$		$\tau_{\text{FEM-POS}}$		$\tau_{\text{FEM-NEG}}$		$\tau_{\text{FEM-NEU}}$	
Verb	Value	Verb	Value	Verb	Value	Verb	Value	Verb	Value	Verb	Value
succeed	1.6	fight	1.2	extend	0.7	celebrate	2.4	persecute	2.1	faint	0.7
protect	1.4	fail	1.0	found	0.8	fascinate	0.8	faint	1.0	be	1.1
favor	1.3	fear	1.0	strike	1.3	facilitate	0.7	fly	1.0	go	0.4
flourish	1.3	murder	1.5	own	1.1	marry	1.8	weep	2.3	find	0.1
prosper	1.7	shock	1.6	collect	1.1	smile	1.8	harm	2.2	fly	0.4
support	1.5	blind	1.6	set	0.8	fan	0.8	wear	2.0	fall	0.1
promise	1.5	forbid	1.5	wag	1.0	kiss	1.8	mourn	1.7	wear	0.9
welcome	1.5	kill	1.3	present	0.9	champion	2.2	gasp	1.1	leave	0.7
favour	1.2	protest	1.3	pretend	1.1	adore	2.0	fatigue	0.7	fell	0.1
clear	1.9	cheat	1.3	prostrate	1.1	dance	1.7	scold	1.8	vanish	1.3
reward	1.8	fake	0.8	want	0.9	laugh	1.6	scream	2.1	come	0.7
appeal	1.6	deprive	1.5	create	0.9	have	1.4	confess	1.7	fertilize	0.6
encourage	1.5	threaten	1.3	pay	1.1	play	1.0	get	0.5	flush	0.5
allow	1.5	frustrate	0.9	prompt	1.0	give	0.8	gossip	2.0	spin	1.6
respect	1.5	fright	0.9	brazen	1.0	like	1.8	worry	1.8	dress	1.4
comfort	1.4	temper	1.4	tarry	0.7	giggle	1.4	be	1.3	fill	0.2
treat	1.3	horrify	1.4	front	0.5	extol	0.6	fail	0.4	fee	0.2
brave	1.7	neglect	1.4	flush	0.3	compassionate	1.9	fight	0.4	extend	0.1
rescue	1.5	argue	1.3	reach	0.9	live	1.4	fake	0.3	sniff	1.6
win	1.5	denounce	1.3	escape	0.8	free	0.9	overrun	2.4	celebrate	1.1
warm	1.5	concern	1.2	gi	0.7	felicitate	0.6	hurt	1.8	clap	1.1
praise	1.4	expel	1.7	rush	0.6	mature	2.2	complain	1.7	appear	0.9
fit	1.4	dispute	1.5	duplicate	0.5	exalt	1.7	lament	1.5	gi	0.8
wish	1.4	obscure	1.4	incarnate	0.5	surpass	1.7	fertilize	0.5	have	0.5
grant	1.3	damn	1.4	freeze	0.5	meet	1.1	feign	0.5	front	0.5

Gendered Verbs (as patient)

$\tau_{\text{MASC-POS}}$		$\tau_{\text{MASC-NEG}}$		$\tau_{\text{MASC-NEU}}$		$\tau_{\text{FEM-POS}}$		$\tau_{\text{FEM-NEG}}$		$\tau_{\text{FEM-NEU}}$	
Verb	Value	Verb	Value	Verb	Value	Verb	Value	Verb	Value	Verb	Value
praise	1.7	fight	1.8	set	1.5	marry	2.3	forbid	1.3	have	1.0
thank	1.7	expel	1.8	pay	1.2	assure	3.4	shame	2.5	expose	0.8
succeed	1.7	fear	1.6	escape	0.4	escort	1.2	escort	1.3	escort	1.4
exalt	1.2	defeat	2.4	use	2.1	exclaim	1.0	exploit	0.9	pour	2.1
reward	1.8	fail	1.3	expel	0.9	play	2.7	drag	2.1	marry	1.3
commend	1.7	bribe	1.8	summon	1.7	pour	2.6	suffer	2.2	take	1.1
fit	1.4	kill	1.6	speak	1.3	create	2.0	shock	2.1	assure	1.6
glorify	2.0	deny	1.5	shop	2.6	have	1.8	fright	2.4	fertilize	1.6
honor	1.6	murder	1.7	excommunicate	1.3	fertilize	1.8	steal	2.0	ask	1.0
welcome	1.9	depose	2.3	direct	1.1	eye	0.9	insult	1.8	exclaim	0.6
gentle	1.8	summon	2.0	await	0.9	woo	3.3	fertilize	1.6	strut	2.3
inspire	1.7	order	1.9	equal	0.4	strut	3.1	violate	2.4	burn	1.7
enrich	1.7	denounce	1.7	appoint	1.7	kiss	2.6	tease	2.3	rear	1.5
uphold	1.5	deprive	1.6	animate	1.1	protect	2.1	terrify	2.1	feature	0.9
appease	1.5	mock	1.6	follow	0.7	win	2.0	persecute	2.1	visit	1.3
join	1.4	destroy	1.5	depose	1.8	excel	1.6	cry	1.8	saw	1.3
congratulate	1.3	deceive	1.7	want	1.1	treat	2.3	expose	1.3	exchange	0.8
extol	1.1	bore	1.6	reach	0.9	like	2.2	burn	2.6	shame	1.6
respect	1.7	bully	1.5	found	0.8	entertain	2.0	scare	2.0	fade	1.2
brave	1.7	enrage	1.4	exempt	0.4	espouse	1.4	frighten	1.8	signal	1.2
greet	1.6	shop	2.7	tip	1.8	feature	1.2	distract	2.3	see	1.2
restore	1.5	elect	2.2	elect	1.7	meet	2.2	weep	2.3	present	1.0
clear	1.5	compel	2.1	unmake	1.5	wish	1.9	scream	2.3	leave	0.8
excite	1.2	offend	1.5	fight	1.2	fondle	1.9	drown	2.1	espouse	1.3
flatter	0.9	scold	1.4	prevent	1.1	saw	1.8	rape	2.0	want	1.1

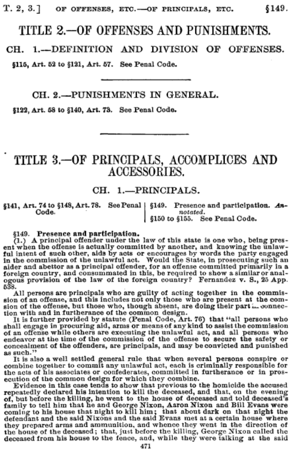
Female		Male	
Positive	Negative	Positive	Negative
beautiful	battered	just	unsuitable
lovely	untreated	sound	unreliable
chaste	barren	righteous	lawless
gorgeous	shrewish	rational	inseparable
fertile	sheltered	peaceable	brutish
beauteous	heartbroken	prodigious	idle
sexy	unmarried	brave	unarmed
classy	undernourished	paramount	wounded
exquisite	underweight	reliable	bigoted
vivacious	uncomplaining	sinless	unjust
vibrant	nagging	honorable	brutal

BODY	FEELING	MISCELLANEOUS
BEHAVIOR	SPATIAL	TEMPORAL
SUBSTANCE	QUANTITY	SOCIAL

- Female nouns were correlated with adjectives/verbs related to the body and to emotions.

Outline

Raw Text Data


T. 2, 3.] OF OFFENSES, ETC.—OF PRINCIPALS, ETC. §149.

TITLE 2.—OF OFFENSES AND PUNISHMENTS.

CH. 1.—DEFINITION AND DIVISION OF OFFENSES.

§115, Art. 52 to §121, Art. 57. See Penal Code.

CH. 2.—PUNISHMENTS IN GENERAL.

§122, Art. 58 to §140, Art. 73. See Penal Code.

TITLE 3.—OF PRINCIPALS, ACCOMPLICES AND ACCESSORIES.

CH. 1.—PRINCIPALS.

§141, Art. 74 to §148, Art. 78. See Penal Code. | §149. Presence and participation. Annotated. §150 to §155. See Penal Code.

§149. Presence and participation.

(1.) A principal offender under the law of this state is one who, being present when the offense is actually committed by another, and knowing the unlawful intent of such other, aids by acts or encourages by words the party engaged in the commission of the unlawful act. Would the State, in prosecuting such an aider and abettor as a principal offender, for an offense committed primarily in a foreign country, and consummated in this, be required to show a similar or analogous provision of the law of the foreign country? Fernandez v. S., 25 App. 538.

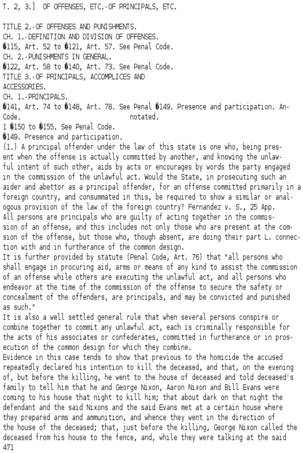
All persons are principals who are guilty of acting together in the commission of an offense, and this includes not only those who are present at the commission of the offense, but those who, though absent, are doing their part in connection with and in furtherance of the common design.

It is further provided by statute (Penal Code, Art. 76) that "all persons who shall engage in procuring aid, arms or means of any kind to assist the commission of an offense while others are executing the unlawful act, and all persons who endeavor at the time of the commission of the offense to secure the safety or concealment of the offenders, are principals, and may be convicted and punished as such."

It is also a well settled general rule that when several persons conspire or combine together to commit any unlawful act, each is criminally responsible for the acts of his associates or confederates, committed in furtherance or in prosecution of the common design for which they combine.

Evidence in this case tends to show that previous to the homicide the accused repeatedly declared his intention to kill the deceased, and that, on the evening of, but before the killing, he went to the house of deceased and told deceased's family to tell him that he and George Nixon, Aaron Nixon and Bill Evans were coming to his house that night to kill him; that about dark on that night the defendant and the said Nixons and the said Evans met at a certain house where they prepared arms and ammunition, and whence they went in the direction of the house of the deceased; that, just before the killing, George Nixon called the deceased from his house to the fence, and, while they were talking at the said

471


T. 2, 3.] OF OFFENSES, ETC.-OF PRINCIPALS, ETC.

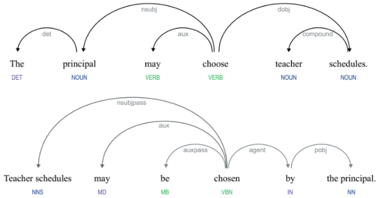
TITLE 2.-OF OFFENSES AND PUNISHMENTS.
CH. 1.-DEFINITION AND DIVISION OF OFFENSES.
§115, Art. 52 to §121, Art. 57. See Penal Code.
CH. 2.-PUNISHMENTS IN GENERAL.
§122, Art. 58 to §140, Art. 73. See Penal Code.
TITLE 3.-OF PRINCIPALS, ACCOMPLICES AND
ACCESSORIES.
CH. 1.-PRINCIPALS.
§141, Art. 74 to §148, Art. 78. See Penal §149. Presence and participation. An-
Code. notated.
1 §150 to §155. See Penal Code.
§149. Presence and participation.
(1.) A principal offender under the law of this state is one who, being pres-
ent when the offense is actually committed by another, and knowing the unlaw-
ful intent of such other, aids by acts or encourages by words the party engaged
in the commission of the unlawful act. Would the State, in prosecuting such an
aider and abettor as a principal offender, for an offense committed primarily in a
foreign country, and consummated in this, be required to show a similar or anal-
ogous provision of the law of the foreign country? Fernandez v. S., 25 App.
All persons are principals who are guilty of acting together in the commis-
sion of an offense, and this includes not only those who are present at the com-
sion of the offense, but those who, though absent, are doing their part L. connec-
tion with and in furtherance of the common design.
It is further provided by statute (Penal Code, Art. 76) that "all persons who
shall engage in procuring aid, arms or means of any kind to assist the commission
of an offense while others are executing the unlawful act, and all persons who
endeavor at the time of the commission of the offense to secure the safety or
concealment of the offenders, are principals, and may be convicted and punished
as such."
It is also a well settled general rule that when several persons conspire or
combine together to commit any unlawful act, each is criminally responsible for
the acts of his associates or confederates, committed in furtherance or in pros-
ecution of the common design for which they combine.
Evidence in this case tends to show that previous to the homicide the accused
repeatedly declared his intention to kill the deceased, and that, on the evening
of, but before the killing, he went to the house of deceased and told deceased's
family to tell him that he and George Nixon, Aaron Nixon and Bill Evans were
coming to his house that night to kill him; that about dark on that night the
defendant and the said Nixons and the said Evans met at a certain house where
they prepared arms and ammunition, and whence they went in the direction of
the house of the deceased; that, just before the killing, George Nixon called the
deceased from his house to the fence, and, while they were talking at the said
471

Full text of U.S. state session laws: all statutes enacted by state legislatures (every year or two years) from 1800 to 2012, retrieved from heinonline.com.

Extracting Legal Provisions

- Pre-processing steps:
  - segment session laws into statutes, segment statutes into sentences

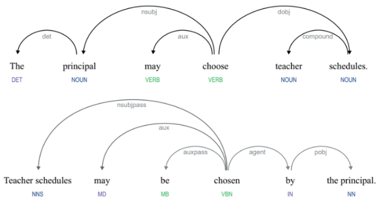
- Extract legal meaning:
  - apply syntactic dependency parser
  - extracts subjects, verbs, objects, etc.



Extracting Legal Provisions

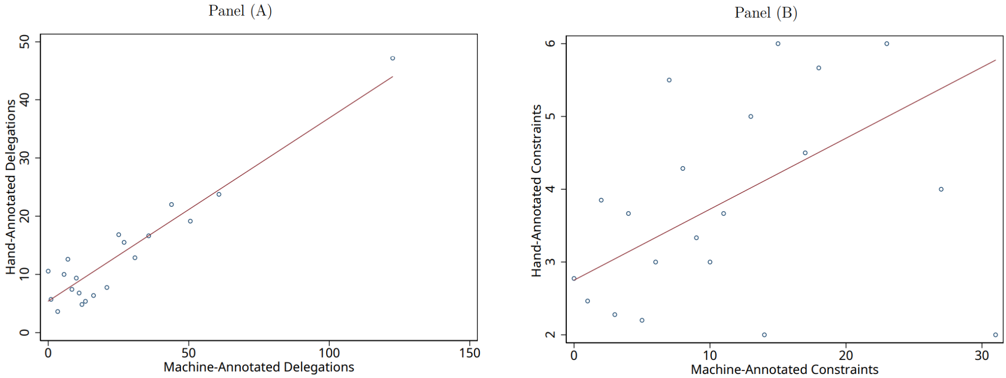
- Pre-processing steps:
  - segment session laws into statutes, segment statutes into sentences

- Extract legal meaning:
  - apply syntactic dependency parser
  - extracts subjects, verbs, objects, etc.



Lexical Units	
Strict modals	'shall','must','will'
Permissive modals	'may','can'
Delegation verbs	'require', 'expect', 'compel', 'oblige' , 'obligate', 'have to', 'ought to'
Constraint verbs	'prohibit', 'forbid', 'ban', 'bar', 'restrict', 'proscribe'
Permission verbs	'allow', 'permit', 'authorize'
Extraction Rules	
Delegation	strict modal + active verb + not negation not permissive modal + delegation verb + not negation
Constraint	modal + not delegation verb + negation strict modal + constraint verb + not negation permission verb + negation
Permission	permission verb + not negation permissive modal + not special verb + not negation constraint verb + negation
Entitlement	entitlement verb + not negation strict modal + passive + not negation delegation verb + negation

Validation Against Hand Coding



Civil Service Reform → More Legislative Detail

Figure 3: Event Study Graph

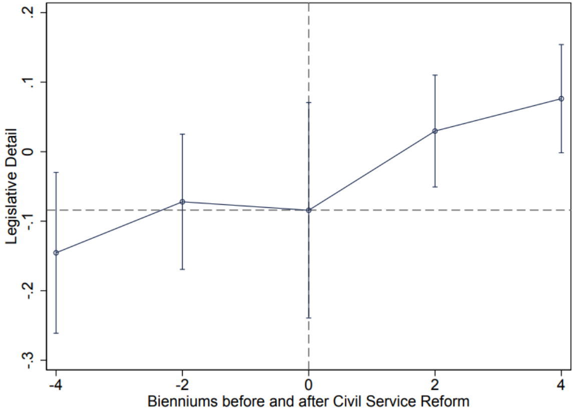


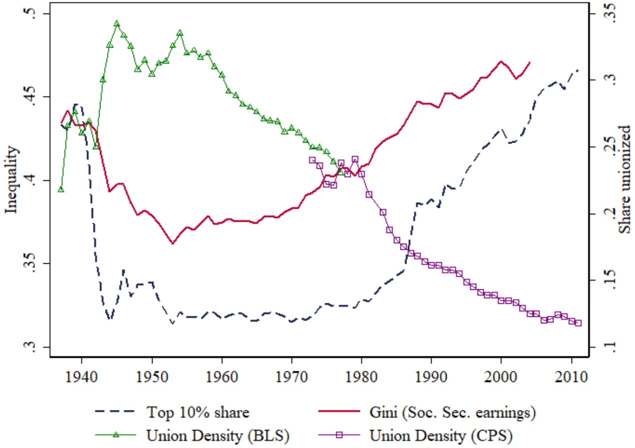
Table 3: Effect of Unified Government on Executive Delegation to the Governor

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Exec Del	Exec Del	Exec Del	Exec Del	Del Ratio	Del Ratio Gov
Unified Govt	0.0054	0.0046	0.0045	0.005	0.00678	0.008
	(0.003)	(0.0027)	(0.0025)	(0.0027)	(0.0031)	(0.004)
Observations	2,270	2,270	2,185	2,223	2,223	2,221
R-squared	0.396	0.464	0.434	0.463	0.529	0.328
State FE	X	X	X	X	X	X
Year FE	X	X	X	X	X	X
State Trends		X		X	X	X
Lagged DV			X			
Civil Service				X	X	X

Notes: Column 1 shows the results for the OLS regression model with state and biennium fixed effects. Column 2 adds state-specific time trends and Column 3 adds the lagged dependent variable. Column 4 adds the introduction of an independent civil service as control. Column 5 and Column 6 use 'Delegation Ratio' and 'Delegation Ratio Gov' as dependent variable, respectively. In all models standard errors are clustered by state.

Outline

Unions and Inequality



This Project

- Data:
  - new corpus of 30,000 collective bargaining agreements from Canada from 1986 through 2015

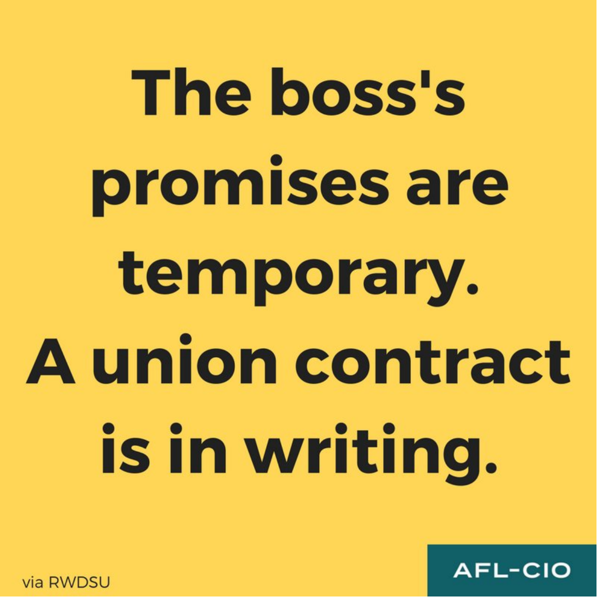
This Project

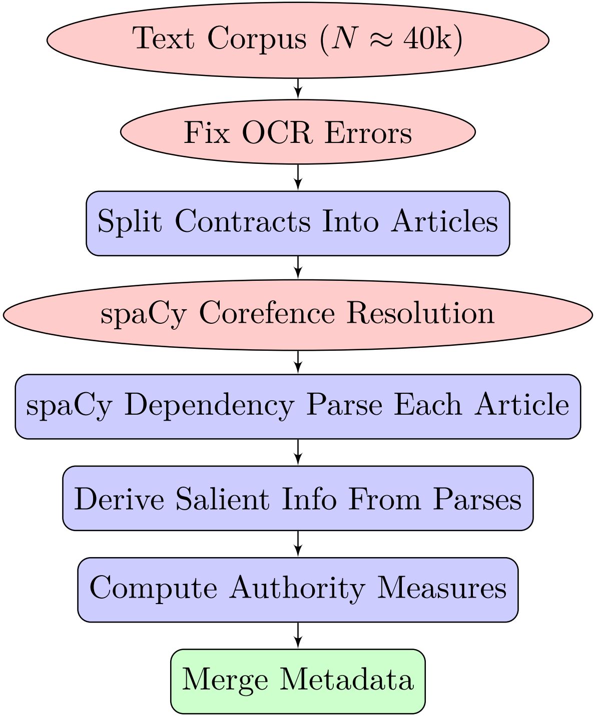
- Data:
  - new corpus of 30,000 collective bargaining agreements from Canada from 1986 through 2015
- Key ideas:
  - use tools from computational linguistics to measure economically and legally relevant contract features:
    - obligations – promises to take actions.
    - entitlements – grants of authority and amenities.

This Project

- Data:
  - new corpus of 30,000 collective bargaining agreements from Canada from 1986 through 2015
- 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.

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.

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*)

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)

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)

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)

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	Subject - Modal - Verb
agreement_shall_be	**employee**_shall_be	**employee**_shall_receive
arbitrator_shall_have	**employee**_shall_be_allowed	**employee**_shall_retain
board_shall_have	**employee**_shall_be_considered	**employee**_will_be
case_may_be	**employee**_shall_be_entitled	**employee**_will_be_allowed
committee_shall_meet	**employee**_shall_be_given	**employee**_will_be_entitled
company_shall_pay	**employee**_shall_be_granted	**employee**_will_be_given
company_shall_provide	**employee**_shall_be_laid_off	**employee**_will_be_granted
company_will_pay	**employee**_shall_be_paid	**employee**_will_be_paid
company_will_provide	**employee**_shall_be_required	**employee**_will_be_required
decision_shall_be	**employee**_shall_continue	**employee**_will_have
employee_may_request	**employee**_shall_lose	**employer**_shall_grant

Categorization of Union Contract Clauses

- Represent union contracts as a list of clauses:

$$< Agent >< Obligation/Entitlement >< Action >$$

- the "action" segment of a clause includes connected pieces of the parse tree besides the subject (agent) and modal (obligation/entitlement).
  - How to encode actions as data?

Categorization of Union Contract Clauses

- Represent union contracts as a list of clauses:

$$< Agent >< Obligation/Entitlement >< Action >$$

- the “action” segment of a clause includes connected pieces of the parse tree besides the subject (agent) and modal (obligation/entitlement).
  - How to encode actions as data?
- LDA Approach:
  - Classify each action clause by topic using Latent Dirichlet Allocation.
  - We got good results with 20 topics.

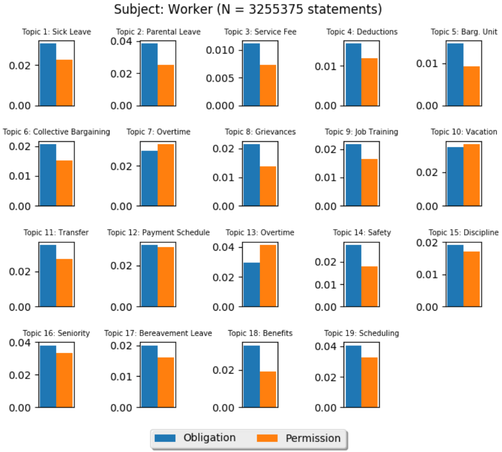
LDA Topics (1 of 2)

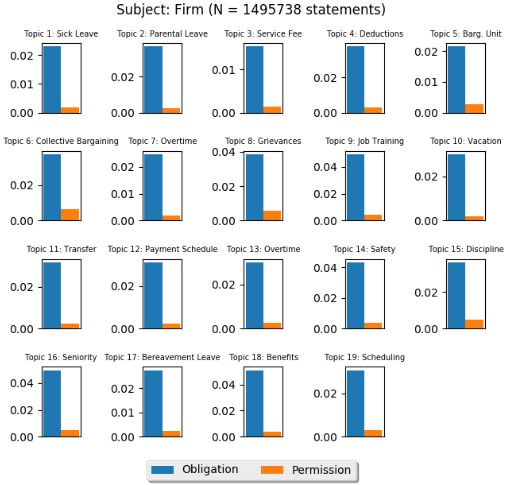
- 1 -- "**Sick Leave**" -- period month sick leave six probationary credit three complete employment twelve absent completion accumulate date exceed consecutive professional
- 2 -- "**Parental Leave**" -- leave absence pay request date grant prior week parental commencement pregnancy write maternity duty witness advance approve notice
- 4 -- "**Payroll**" -- change due result deduction amount status deduct monthly payroll reduction affect cheque technological fee employment orientation statement
- 5 -- "**Bargaining Unit**" -- unit bargaining person appointment appoint employ outside activity membership represent agent terminal sole select exercise ontario bargain behalf
- 7 -- "**Overtime**" -- hour shift work schedule overtime period call rest meal half minute start end break duty sunday weekend saturday two friday
- 8 -- "**Grievances**" -- grievance party procedure arbitration writing decision write step matter arbitrator committee complaint submit final dispute request name process
- 9 -- "**Job Training**" -- requirement operation training require equipment individual meet service responsibility provide program area manner performance" business duty operational
- 10 -- "**Vacation Leave**" -- year vacation service pay date employment week continuous effective two annual entitlement percent january salary earn termination period follow

LDA Topics (2 of 2)

- 14 – **"Medical Leave/Injuries"** medical reasonable illness reason certificate unable duty injury course require due provide information circumstance accident personal condition examination reasonably
- 15 -- **"Discipline/Firing"** -- school act safety committee health action discharge labour cause discipline disciplinary file application canada public relations suspension regulation authority accordance
- 16 -- **"Seniority"** -- seniority lay position list layoff vacancy recall transfer post temporary qualification permanent job hire fill date provide ability copy basis
- 17 -- **"Work-Related Deaths"** – article accordance law child spouse pursuant family death include immediate parent purpose require city office paragraph funeral
- 18 -- **"Insurance/Benefits"** -- benefit plan insurance payment cost premium eligible provide receive compensation disability pay coverage pension receipt term amount
- 19 -- **"Scheduling"** -- work hour day week schedule two return perform normal regular report normally excess regularly require notice eight teaching available emergency

Workers Have More Entitlements Relative To Obligations





- Workers get more authority at work than employers, consistently across work areas.

Determinants of Relative Worker Control

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

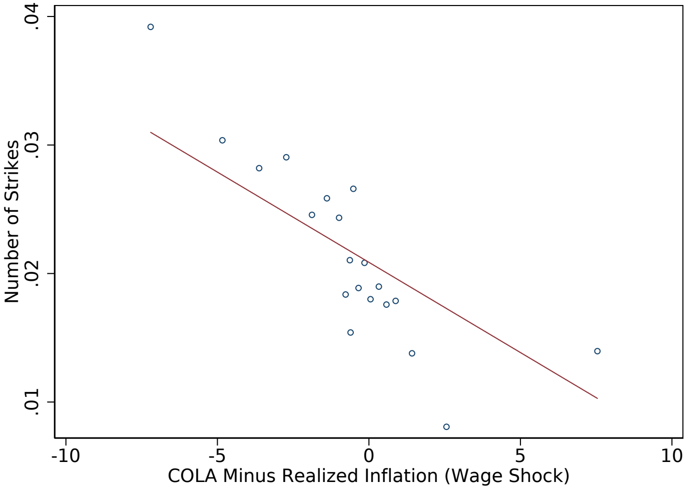
Wages, Control, and Strikes

- 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.

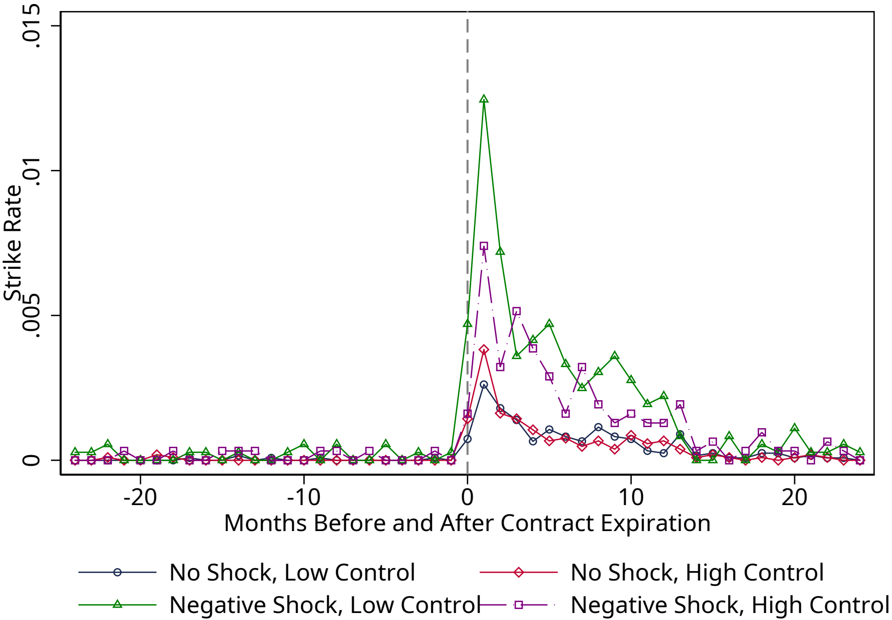
Wages, Control, and Strikes

- 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.

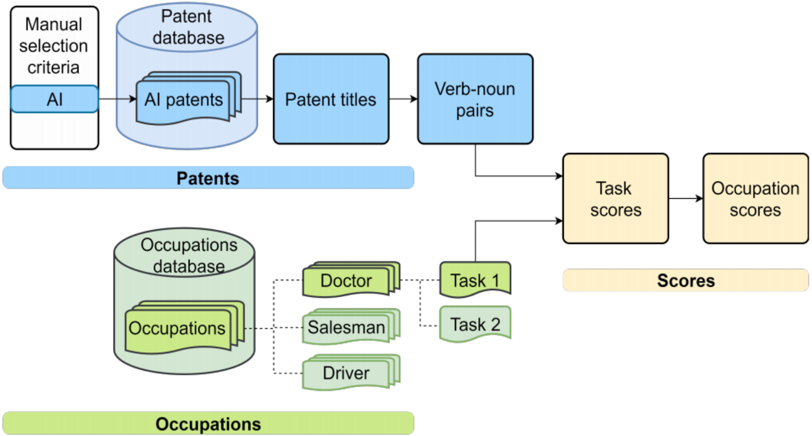


Strike Rate
.015
.01
.005
0
-20
-10
0
10
20
Months Before and After Contract Expiration
No Shock, Low Control
No Shock, High Control
Negative Shock, Low Control
Negative Shock, High Control

Outline

Webb 2020: Data

- Webb (2020) uses syntactic parsing to match the tasks in occupational descriptions to the tasks in patent texts.
- Data Pipeline:



- Job descriptions from O*NET, patents from Google Patents Public Data

Extracting verb-noun pairs: Patents

- apply spacy dependency to extract verbs and associated direct objects.

Table A1: **Extracting capabilities from patent titles.**

Text	**Extracted pairs**
Adaptive system and method for predicting response times in a service environment	(predict, time)
Method of and apparatus for determining optimum delivery route for articles	(determine, route)
Methods and apparatus for reinforcement learning	
Device for forecasting total power demand	(forecast, demand)
Method and device for classifying images on basis of convolutional neural network	(classify, image)
A method for diagnosing food allergy	(diagnose, allergy)
Neural network language model training method and device and voice recognition method	
Automatic butterfly species identification system and method, and portable terminal having automatic butterfly species identification function using the same	(have, function), (use, same)

- use WordNet to dimension-reduce nouns into higher-level categories.

Extracting verb-noun pairs: Occupational tasks

Table 1: **Tasks and exposure scores for precision agriculture technicians.**

Task	Weight in occupation	Extracted pairs	AI exposure score x100
Use geospatial technology to develop soil sampling grids or identify sampling sites for testing characteristics such as nitrogen, phosphorus, or potassium content, ph, or micronutrients.	0.050	(develop, grid)	0.050
		(identify, site)	0.234
		(test, characteristic)	0.084
Document and maintain records of precision agriculture information.	0.049	(maintain, record)	0.000
Analyze geospatial data to determine agricultural implications of factors such as soil quality, terrain, field productivity, fertilizers, or weather conditions.	0.048	(analyze, datum)	0.469
		(determine, implication)	0.837
Apply precision agriculture information to specifically reduce the negative environmental impacts of farming practices.	0.048	(apply, information)	0.000
		(reduce, impact)	0.151
Install, calibrate, or maintain sensors, mechanical controls, GPS-based vehicle guidance systems, or computer settings.	0.045	(maintain, sensor)	0.000
Identify areas in need of pesticide treatment by analyzing geospatial data to determine insect movement and damage patterns.	0.038	(identify, area)	0.234
		(analyze, datum)	0.469
		(determine, movement)	0.502

Table 2: **Top extracted verbs and characteristic nouns for robots.**

Verb	Example nouns	Verb	Example nouns
clean	surface, wafer, window, glass, floor, tool, casting, instrument	**walk**	robot, structure, base, stairs, circuit, trolley, platform, maze
control	robot, arm, motion, position, manipulator, motor, path, force	**carry**	substrate, wafer, tray, vehicle, workpiece, tool, object, pallet
weld	wire, part, tong, electrode, sensor, component, nozzle	**detect**	position, state, collision, obstacle, force, angle, leak, load, landmine
move	robot, body, object, arm, tool, part, substrate, workpiece	**drive**	unit, wheel, motor, belt, rotor, vehicle, automobile, actuator

Table 2: **Top extracted verbs and characteristic nouns for robots.**

Verb	Example nouns	Verb	Example nouns
clean	surface, wafer, window, glass, floor, tool, casting, instrument	**walk**	robot, structure, base, stairs, circuit, trolley, platform, maze
control	robot, arm, motion, position, manipulator, motor, path, force	**carry**	substrate, wafer, tray, vehicle, workpiece, tool, object, pallet
weld	wire, part, tong, electrode, sensor, component, nozzle	**detect**	position, state, collision, obstacle, force, angle, leak, load, landmine
move	robot, body, object, arm, tool, part, substrate, workpiece	**drive**	unit, wheel, motor, belt, rotor, vehicle, automobile, actuator

Table 3: **Occupations with highest and lowest exposure to robots.**

Most exposed occupations	Least exposed occupations
Forklift driver	Payroll and timekeeping clerks
Operating engineers of cranes, derricks, etc.	Art/entertainment performers
Elevator installers and repairers	Clergy
Janitors	Correspondence and order clerks
Locomotive operators: engineers and firemen	Eligibility clerks for government programs

Table 4: **Change in wages vs. exposure to robots, 1980-2010.**

	(1)	(2)	(3)	(4)	(5)
Exposure	-0.29***	-0.28***	-0.26***	-0.16***	-0.22***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Offshorability			0.84*	0.82*	-2.29***
			(0.44)	(0.44)	(0.50)
Medium education				7.84***	9.52***
				(1.75)	(1.67)
High education				10.22***	27.73***
				(1.89)	(2.01)
Wage					-0.07***
					(0.01)
Wage squared					0.00**
					(0.00)
Adjusted R^2	0.042	0.094	0.095	0.101	0.163
Industry FEs		✓	✓	✓	✓
Observations	6,708	6,708	6,708	6,708	6,708

Notes: Each observation is an occupation-industry cell. Dependent variable is 100x change in log wage between 1980 and 2010 winsorized at the top and bottom 1%. Education variables are terciles of average years of education for occupation-industry cells in 1980. Wages are cells' mean weekly wage for full-time, full-year workers in 1980. Offshorability is an occupation-level measure from Autor and Dorn (2013). Sample is restricted to industries within the manufacturing sector. Standard errors are clustered by industry. * p<0.10, ** p<0.05, *** p<0.01.

Table 5: **Change in employment vs. exposure to robots, 1980-2010.**

	(1)	(2)	(3)	(4)	(5)
Exposure	-0.37***	-0.36***	-0.35***	-0.18***	-0.16***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Offshorability			0.78	0.93*	2.02***
			(0.54)	(0.55)	(0.55)
Medium education				-0.26	-1.20
				(1.54)	(1.54)
High education				21.39***	14.42***
				(2.43)	(2.40)
Wage					0.04***
					(0.00)
Wage squared					-0.00***
					(0.00)
Adjusted R^2	0.018	0.129	0.129	0.141	0.147
Industry FEs		✓	✓	✓	✓
Observations	14,065	14,065	14,065	14,065	14,065

Notes: Each observation is an occupation-industry cell. Dependent variable is 100x DHS change of a cell's share of overall employment between 1980 and 2010, winsorized at the top and bottom 1%. Education variables are terciles of average years of education for occupation-industry cells in 1980. Wages are cells' mean weekly wage for full-time, full-year workers in 1980. Offshorability is an occupation-level measure from Autor and Dorn (2013). Observations are weighted by cell's labor supply, averaged between 1980 and 2010. Sample is restricted to industries within the manufacturing sector. Standard errors are clustered by industry. * p<0.10, ** p<0.05, *** p<0.01.

Table 6: **Top extracted verbs and characteristic nouns for software.**

Verb	Example nouns	Verb	Example nouns
record	data, position, log, location, reservation, transaction	**detect**	defect, error, malware, fault, condition, movement
store	program, data, information, image, instruction, value	**generate**	data, image, file, report, map, key, password, animation, diagram
control	access, display, unit, image, device, power, motor	**measure**	rate, performance, time, distance, thickness
reproduce	data, picture, media, file, sequence, speech, item, document, selection	**receive**	signal, data, information, message, order, request, instruction, command

Table 7: **Occupations with highest and lowest exposure to software.**

Most exposed occupations	Least exposed occupations
Broadcast equipment operators	Barbers
Water and sewage treatment plant operators	Podiatrists
Parking lot attendants	Subject instructors, college
Packers and packagers by hand	Art/entertainment performers
Locomotive operators: engineers and firemen	Mail carriers for postal service

Table 8: **Change in wages vs. exposure to software, 1980-2010.**

	(1)	(2)	(3)	(4)	(5)
Exposure	-0.13***	-0.11***	-0.09***	-0.05***	-0.04***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Offshorability			2.02***	1.42***	-0.87***
			(0.30)	(0.29)	(0.28)
Medium education				8.36***	11.80***
				(0.99)	(0.93)
High education				12.77***	32.75***
				(1.07)	(1.24)
Wage					-0.07***
					(0.00)
Wage squared					0.00***
					(0.00)
Adjusted R^2	0.008	0.064	0.067	0.079	0.168
Industry FEs		✓	✓	✓	✓
Observations	18,975	18,975	18,975	18,975	18,975

Notes: Each observation is an occupation-industry cell. Dependent variable is 100x change in log wage between 1980 and 2010 winsorized at the top and bottom 1%. Observations are weighted by cell's labor supply, averaged between 1980 and 2010. Education variables are terciles of average years of education for occupation-industry cells in 1980. Wages are cells' mean weekly wage for full-time, full-year workers in 1980. Offshorability is an occupation-level measure from Autor and Dorn (2013). Standard errors are clustered by industry. * p<0.10, ** p<0.05, *** p<0.01.

Table 9: **Change in employment vs. exposure to software, 1980-2010.**

	(1)	(2)	(3)	(4)	(5)
Exposure	-0.30***	-0.22***	-0.21***	-0.14***	-0.14***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Offshorability			2.98***	2.07***	2.66***
			(0.51)	(0.52)	(0.53)
Medium education				7.28***	6.19***
				(1.33)	(1.35)
High education				27.47***	22.26***
				(1.83)	(1.91)
Wage					0.03***
					(0.00)
Wage squared					-0.00***
					(0.00)
Adjusted R^2	0.009	0.193	0.194	0.207	0.210
Industry FEs		✓	✓	✓	✓
Observations	36,070	36,070	36,070	36,070	36,070

Notes: Each observation is an occupation-industry cell. Dependent variable is 100x DHS change of a cell's share of overall employment between 1980 and 2010, winsorized at the top and bottom 1%. Education variables are terciles of average years of education for occupation-industry cells in 1980. Wages are cells' mean weekly wage for full-time, full-year workers in 1980. Offshorability is an occupation-level measure from Autor and Dorn (2013). Observations are weighted by cell's labor supply, averaged between 1980 and 2010. Standard errors are clustered by industry. * p<0.10, ** p<0.05, *** p<0.01.

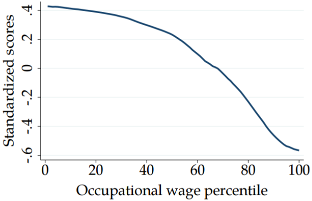
Table 10: **Top extracted verbs and characteristic nouns for AI.**

Verb	Example nouns	Verb	Example nouns
recognize	pattern, image, speech, face, voice, automobile, emotion, gesture, disease	**determine**	state, similarity, relevance, importance, characteristic, strategy, risk
predict	quality, performance, fault, behavior, traffic, prognosis	**control**	process, emission, traffic, engine, robot, turbine, plant
detect	signal, abnormality, defect, object, fraud, event, spammer, human, cancer	**generate**	image, rating, lexicon, warning, description, recommendation
identify	object, type, damage, illegality, classification, relationship, importance	**classify**	data, object, image, pattern, signal, text, electrogram, speech, motion

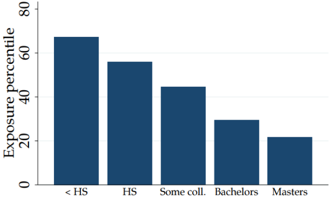
Table 11: **Occupations with highest and lowest exposure to artificial intelligence.**

Most exposed occupations	Least exposed occupations
Clinical laboratory technicians	Animal caretakers, except farm
Chemical engineers	Food preparation workers
Optometrists	Mail carriers for postal service
Power plant operators	Subject instructors, college
Dispatchers	Art/entertainment performers

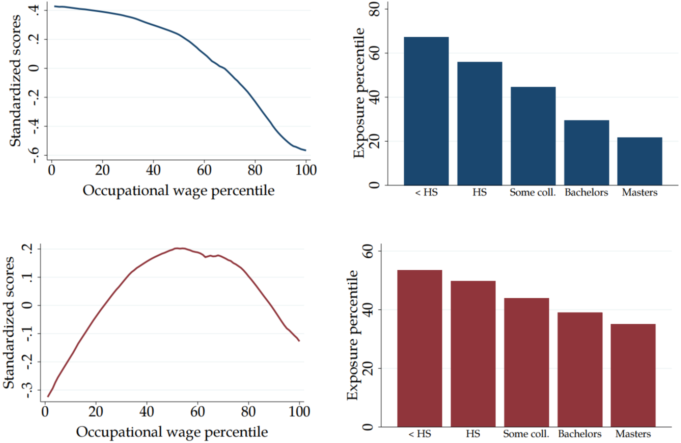
Notes: Table displays census occupation title for the five occupations with the highest exposure scores and with the lowest exposure scores above employment threshold of 150.

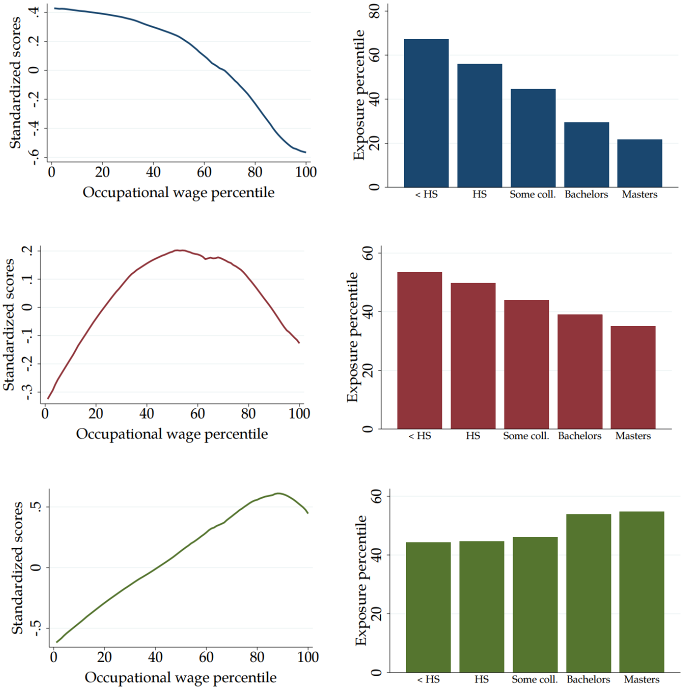


Standardized scores
.4
.2
0
-.2
-.4
-.6
0
20
40
60
80
100
Occupational wage percentile

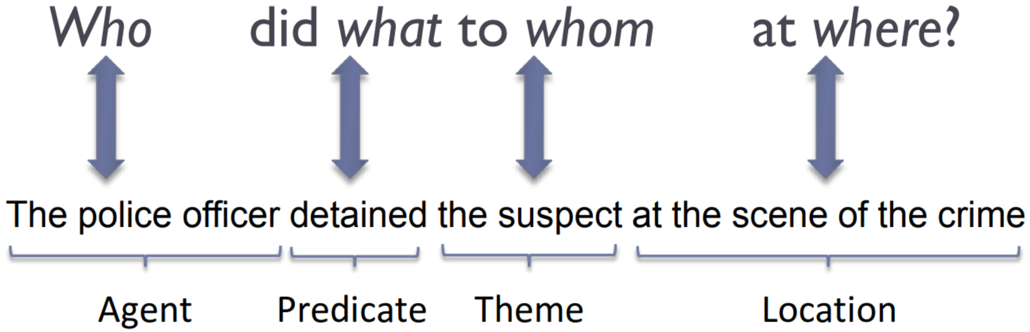


Exposure percentile
80
60
40
20
0
< HS
HS
Some coll.
Bachelors
Masters


Standardized scores
.4
.2
0
-.2
-.4
-.6
0
20
40
60
80
100
Occupational wage percentile
Exposure percentile
0
20
40
60
80
< HS
HS
Some coll.
Bachelors
Masters
Standardized scores
.2
.1
0
-.1
-.2
-.3
0
20
40
60
80
100
Occupational wage percentile
Exposure percentile
0
20
40
60
< HS
HS
Some coll.
Bachelors
Masters


Standardized scores
Occupational wage percentile
Exposure percentile
< HS
HS
Some coll.
Bachelors
Masters

Outline



Source: Jurafsky-Martin slides.

Sasha broke the window.

Pat opened the door.

- Subjects of break and open: Breaker and Opener.
- Breaker and Opener have something in common:
  - Volitional actors
  - Often animate
  - Direct causal responsibility for their events

Sasha broke the window.

Pat opened the door.

- Subjects of break and open: Breaker and Opener.
- Breaker and Opener have something in common:
  - Volitional actors
  - Often animate
  - Direct causal responsibility for their events
- Thematic roles are a way to capture this semantic commonality between Breakers and Openers:
  - They are both AGENTS.

Thematic Roles

Thematic Role	**Definition**	**Example**
AGENT	The volitional causer of an event	*The waiter* spilled the soup.
EXPERIENCER	The experiencer of an event	*John* has a headache.
FORCE	The non-volitional causer of the event	*The wind* blows debris from the mall into our yards.
THEME	The participant most directly affected by an event	Only after Benjamin Franklin broke *the ice*...
RESULT	The end product of an event	The city built a *regulation-size baseball diamond*...
CONTENT	The proposition or content of a propositional event	Mona asked *"You met Mary Ann at a supermarket?"*
INSTRUMENT	An instrument used in an event	He poached catfish, stunning them *with a shocking device*...
BENEFICIARY	The beneficiary of an event	Whenever Ann Callahan makes hotel reservations *for her boss*...
SOURCE	The origin of the object of a transfer event	I flew in *from Boston*.
GOAL	The destination of an object of a transfer event	I drove *to Portland*.

Challenges with creating standard roles

Levin and Rappaport Hovav (2015): e.g., two kinds of INSTRUMENTS:

- **intermediary instruments** that can appear as subjects:
  - The cook opened the jar with the new gadget.
  - The new gadget opened the jar.
- **enabling instruments** that cannot:
  - Shelly ate the sliced banana with a fork.
  - The fork ate the sliced banana.

Proposition Bank (PropBank)

- Agent
  - Volitional involvement in event or state
  - Sentience (and/or perception)
  - Causes an event or change of state in another participant
  - Movement (relative to position of another participant)
- Patient
  - Undergoes change of state
  - Causally affected by another participant
  - Stationary relative to movement of another participant

Labeling of Roles

- Roles are labeled verb by verb.
- Each verb sense has numbered arguments:

ARG0	agent	ARG3	starting point, benefactive, attribute
ARG1	patient	ARG4	ending point
ARG2	instrument, benefactive, attribute	ARGM	modifier

Table 1.1: List of arguments in PropBank

Labeling of Roles

- Roles are labeled verb by verb.
- Each verb sense has numbered arguments:

ARG0	agent	ARG3	starting point, benefactive, attribute
ARG1	patient	ARG4	ending point
ARG2	instrument, benefactive, attribute	ARGM	modifier

Table 1.1: List of arguments in PropBank

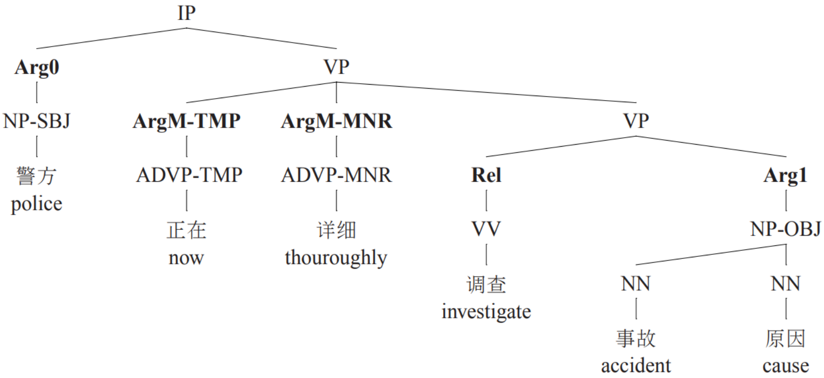
Ex1: [$_{\text{Arg0}}$ The group] *agreed* [$_{\text{Arg1}}$ it wouldn't make an offer].

Ex2: [$_{\text{ArgM-TMP}}$ Usually] [$_{\text{Arg0}}$ John] *agrees* [$_{\text{Arg2}}$ with Mary] [$_{\text{Arg1}}$ on everything].

ARG-M: Modifiers

ArgM-TMP	when?	yesterday evening, now
LOC	where?	at the museum, in San Francisco
DIR	where to/from?	down, to Bangkok
MNR	how?	clearly, with much enthusiasm
PRP/CAU	why?	because ... , in response to the ruling
REC		themselves, each other
ADV	miscellaneous	
PRD	secondary predication	...ate the meat raw

Not just English



“The police are thoroughly investigating the cause of the accident.”

AllenNLP

https://demo.allennlp.org/semantic-role-labeling

AllenNLP

`https://demo.allennlp.org/semantic-role-labeling`

- Constituency parsing:
  - Implementation of Joshi, Peters, and Hopkins (2018) using character-based ELMo embeddings.
  - F1 = 94.1 on Penn TreeBank.

AllenNLP

`https://demo.allennlp.org/semantic-role-labeling`

- Constituency parsing:
  - Implementation of Joshi, Peters, and Hopkins (2018) using character-based ELMo embeddings.
  - F1 $= 94.1$ on Penn TreeBank.
- Dependency parsing:
  - Implementation of Dozat and Manning (2017) using bidirectional LSTM with attention.
  - 92.9% accurate labeling on Penn TreeBank.

AllenNLP

`https://demo.allennlp.org/semantic-role-labeling`

- Constituency parsing:
  - Implementation of Joshi, Peters, and Hopkins (2018) using character-based ELMo embeddings.
  - F1 $=$ 94.1 on Penn TreeBank.
- Dependency parsing:
  - Implementation of Dozat and Manning (2017) using bidirectional LSTM with attention.
  - 92.9% accurate labeling on Penn TreeBank.
- Semantic role labeling
  - Implemention of Shi and Lin's (2019) BERT-based model.
  - F1 $=$ 86.49 on OntoNotes 5.0.