Sequencing Legal DNA

NLP for Law and Political Economy

7. Syntactic and Semantic Parsing

Q&A Page

`bit.ly/NLP-QA07`

Activity: Ash, Chen, and Ornaghi (2021)

see link in zoom chat

Using Grammar: 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

Using Grammar: 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] .
- The "legal memes" we constructed in Week 2 were a form of constituency 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

Dependency Parsing: Linguistics

Dependency Parsing: Applications

- Ash, Jacobs, MacLeod, Naidu, and Stammbach (2020): Unsupervised Extraction of Rights and Duties from Collective Bargaining Agreements
- Hoyle et al (2019): Discovery of Gendered Language
- 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	.

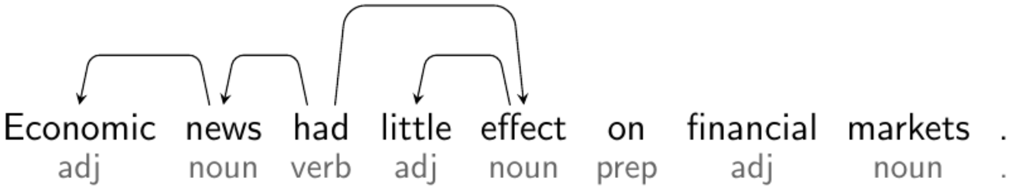
Dependency Structure

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

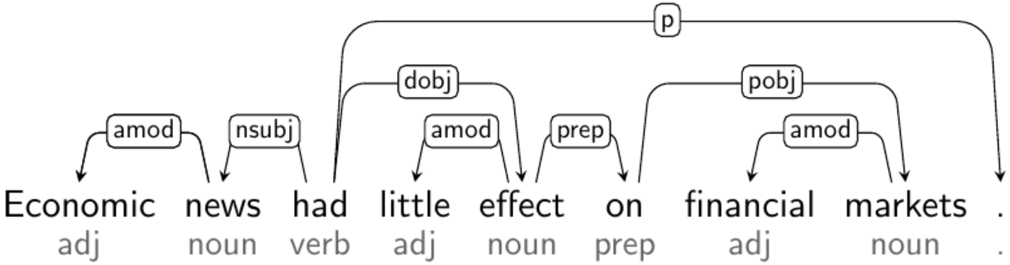
Dependency Structure

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

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)

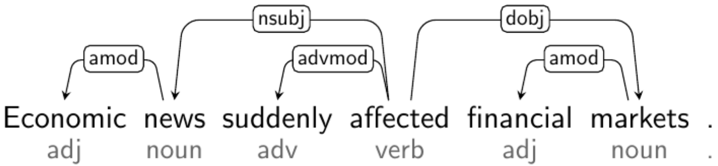
Heads and Dependents

- 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

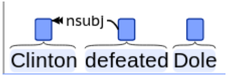
- A dependency is a one-way link from a "head" token to a "dependent" token:
  - Head determines the syntactic/semantic category of the dependency.
  - Head is obligatory; Dependent may be optional.

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



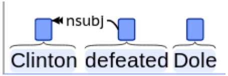
Subjects

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

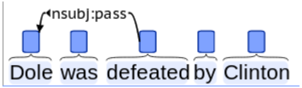


Subjects

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

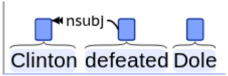


- **nsubjpass: passive nominal subject**
  - non-clausal constituent in the subject position of a passive verb.

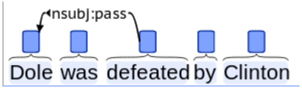


Subjects

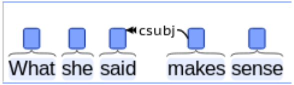
- **nsubj: nominal subject**
  - non-clausal constituent in the subject position of an active verb.



- **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...

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

Dependency Parsing: Linguistics

Dependency Parsing: Applications

- Ash, Jacobs, MacLeod, Naidu, and Stammbach (2020): Unsupervised Extraction of Rights and Duties from Collective Bargaining Agreements
- Hoyle et al (2019): Discovery of Gendered Language
- Webb (2020): AI and the Labor Market

Semantic Role Labeling

Outline

Dependency Parsing: Linguistics

Dependency Parsing: Applications

- Ash, Jacobs, MacLeod, Naidu, and Stammbach (2020): Unsupervised Extraction of Rights and Duties from Collective Bargaining Agreements
- Hoyle et al (2019): Discovery of Gendered Language
- Webb (2020): AI and the Labor Market

Semantic Role Labeling

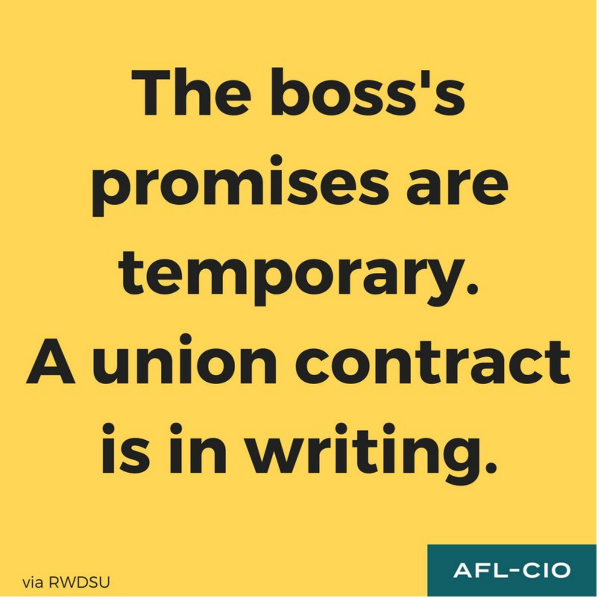
Overview

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

Overview

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

What do contracts do?



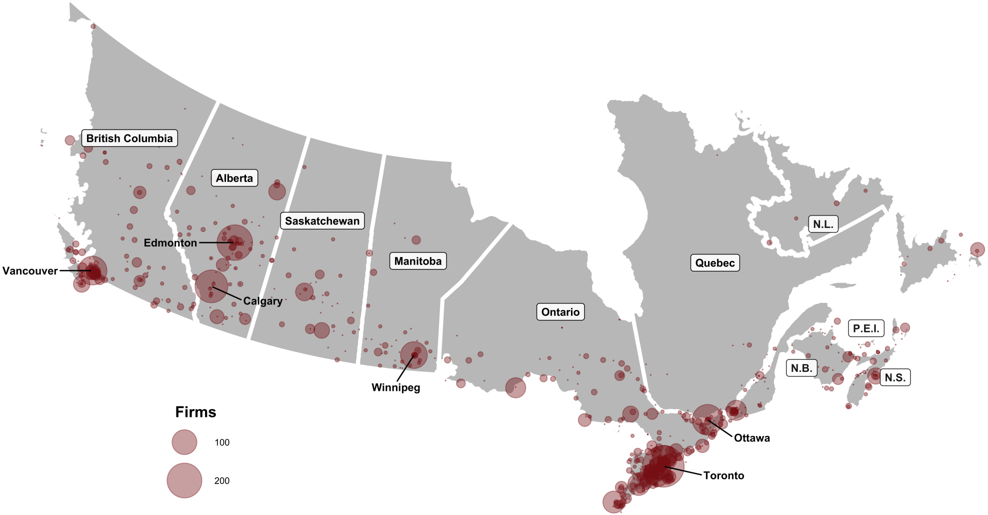

British Columbia
Alberta
Saskatchewan
Manitoba
Ontario
Quebec
N.L.
P.E.I.
N.B.
N.S.
Vancouver
Edmonton
Calgary
Winnipeg
Ottawa
Toronto
Firms
100
200

TABLE OF CONTENTS

AGREEMENT

This Agreement ratified December 16,2005 is made and entered into between ST. CLAIR TECHNOLOGIES INC., Wallaceburg, Ontario (hereinafter called “the Company”), and the International Union, united Automobile, Aerospace and Agricultural Implement Workers of America (UAW-CLC) and its Local No. 251, (hereinafter called “the Union”).

ARTICLE 1
RECOGNITION

1. The provisions of this Agreement shall apply to all employees covered by this Agreement without discrimination on account of race, creed, colour, sex, marital status, nationality, ancestry or place of origin.

2. Wherever the male noun or pronoun is used, it shall also mean the female.

3. The Company recognizes the Union as the sole bargaining agent of all its employees at Wallaceburg, Ontario, save and except supervisor, those above the rank of supervisor, office and sales staff, students for not more than twenty-four hours per week and students employed during the school vacation period (May 1st-September 15th). In case of reduction in force, students would be laid off first. Students will be paid at a rate to be determined by the Company, but will not be less than the Employment Standards Act.

4. The word “employee” or “employees” wherever used in this Agreement shall mean only the employees in the bargaining unit defined above unless the context otherwise provides.

5. The Company will negotiate with the Union for the purpose of adjusting any disputes which may arise concerning sickness and accident, wages, hours and working conditions.

ARTICLE 2
UNION SECURITY

1. All employees covered by this Agreement who are members of the Union at the signing date of this Agreement or who after become members thereof during the term of this Agreement, must retain their membership in the Union for the duration of the Agreement by paying the regular monthly dues levied against all members, as a condition of employment. All employees covered by this Agreement who are not members of the Union shall pay regular monthly dues, the same as the dues that are levied against those who are members of the union as a condition of employment.

2. All new employees, upon completion of thirty (30) days employment shall become members thereof in good standing in accordance with the constitution and bylaws of the Union for the life of this Agreement.

3. The Company will during the term of the Agreement, deduct initiation fees, monthly dues and assessments on a monthly basis from the pay cheque of all seniority employees, probationary employees and full-time students who have worked or been compensated for forty (40) hours in any one (1) month, or as required by the U.A.W. constitution, (full-time student being a student who works all or any time between May 1st and September 15th of the same year). Such deductions shall be credited to the Secretary-Treasurer of Local 251, not later than the tenth (10th) day of the calendar month next following the month in which such deductions are made. The Company and the Union will work out a mutually satisfactory arrangement by which the Company will furnish monthly records to the Financial Secretary of Local 251 of those from whom deductions were made, together with the amount of such deductions.

ARTICLE 3
MANAGEMENT RIGHTS

The Union recognizes and acknowledges that the management of the plant and direction of the working force are fixed exclusively in the Company and, without restricting the generality of the foregoing, the Union acknowledges that it is the exclusive function of the Company to:

1. Maintain order and efficiency

2. Hire, promote, demote, classify, transfer, suspend and retire employees, and to discipline or discharge for just cause, any employee provided that a claim by an employee who has acquired seniority that he has been discharged or disciplined without just cause may be the subject of a grievance and dealt with as here in before provided.

3. Make, enforce, and alter, from time to time, rules and regulations to be observed by the employees, such rules not to be inconsistent with the provisions of this Agreement. The Company agrees to give a copy of any changes in plant rules to the Union Chairperson prior to posting of same on bulletin boards.

4. Determine the nature and kind of business conducted by the Company, the kinds and locations of plants, equipment and material to be used, the control of materials and parts, the use of incentive programs, the methods and techniques of work, the content of jobs, the schedules of production, the number of employees to be employed, the extension, limitations, curtailment or cessation of operations or any part thereof, and to determine and exercise all other function and prerogatives which shall remain solely with the Company except specifically limited by the express provisions of this Agreement.

ARTICLE 4
NO STRIKES - NO LOCKOUTS

1. The Union agrees that during the term of this agreement, there shall be no strikes, sit-downs, work stoppage, slowdowns, or suspension of work, either complete or partial, for any reason, by an employee or employees. There shall be no lockout of employees by the Company, for the duration of this Agreement.

ARTICLE 5
REPRESENTATION

1. The Union shall elect or appoint, and the Company shall recognize, from those employees who have completed at least one (1) year service with the Company a plant committee of four (4) people, one of whom will be the chairperson and one of whom will be vice-chairperson. The committee people shall be employed on the day shift.

2. The Company shall also recognize a steward who will be elected or

appointed by the Union and work on the afternoon or midnight shift during such periods as the Company schedules these shifts and is equal to or greater than five (5) employees. Stewards will have preferred seniority on their shift for lay off and recall purposes only.

3. The Union will inform the Company in writing of the names of the stewards and members of the Grievance Committee and of any subsequent changes in the names of any steward or members of the Grievance Committee. The Company shall not be asked to recognize any steward or member of the Grievance Committee until such notification from the Union has been received.

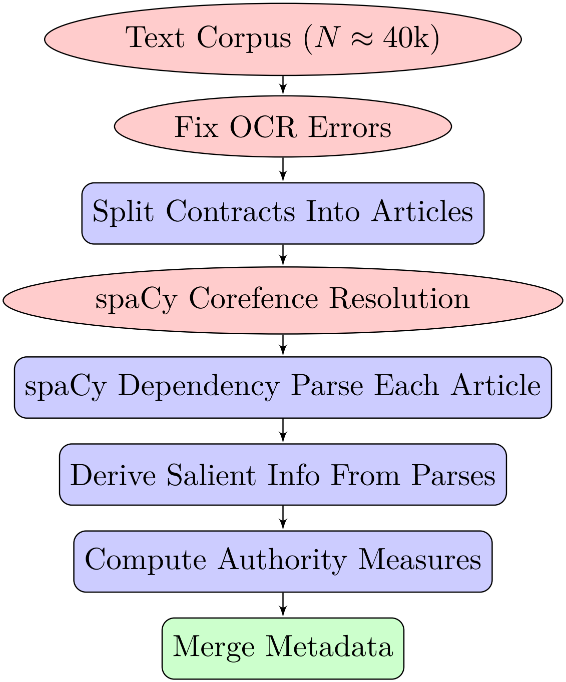
4. The Union acknowledges that committee persons and stewards have their regular duties as employees to perform and that such persons will not leave their regular duties without first obtaining permission from their supervisor. Such permission shall not be unreasonably withheld. In the application of representation language “such permission shall not be unreasonably withheld” it is understood that a supervisor will not detain a Union representative from performing legitimate representation and by the same token the Union representative will understand the occasional need to complete a job in the interest of continuing production before leaving for legitimate union business. In any event, no such Union representative shall be detained any longer than thirty (30) minutes to perform their union representation duties.

5. The Company shall schedule a meeting, date and time within the time limits prescribed for any grievance submitted to Step 2 and/or Step 3 of the grievance procedure. The grievance committee only shall be compensated at their job rate for any regular scheduled work hours lost during such meeting with Company representatives. Overtime shall be paid when the meeting has been requested by the Company or the meeting goes beyond the Union representatives scheduled shift.

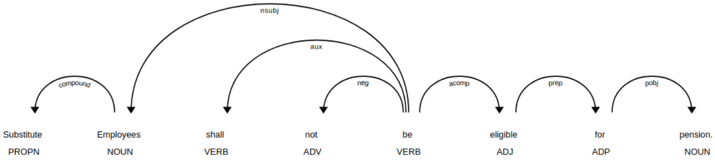
6.

The plant committee referred to in Section 1, shall head the seniority list during their term of office for layoff and recall purposes only.

A committee person will be required where more than nine (9) employees on the day shift at any one plant are required to work on Saturdays, Sundays and Statutory Holidays.


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

Syntactic Parse for Contract Statements



- We ran each sentence through spaCy'ssyntactic parser
- Identify syntactic subjects, and form statements around each subject.

Modal Verbs in Law

- Modal verbs:
  - strict (*shall*, *will*, *must*) modals express necessity.
  - permissive (*may*, *can*) modals express possibility.

Modal Verbs in Law

- Modal verbs:
  - strict (*shall*, *will*, *must*) modals express necessity.
  - permissive (*may*, *can*) modals express possibility.
- Deontic modal verbs (deontic indicating "duty") capture necessity/possibility in social freedoms to act.
  - In the law, they impose legal requirements.

Parse Information on Subjects and Verbs

- Subject categories:
  - worker, union, owner, manager.

Parse Information on Subjects and Verbs

- Subject categories:
  - worker, union, owner, manager.
- Negation (“shall **not**”)
- Active/passive (“shall provide” vs “shall be provided”).

Parse Information on Subjects and Verbs

- Subject categories:
  - worker, union, owner, manager.
- Negation (“shall **not**”)
- Active/passive (“shall provide” vs “shall be provided”).
- Special verbs:
  - *Obligation* Verbs (have to, ought to, be required, be expected, be compelled, be obliged, be obligated)
  - *Prohibition* Verbs (be prohibited, be forbidden, be banned, be barred, be restricted, be proscribed)

Parse Information on Subjects and Verbs

- Subject categories:
  - worker, union, owner, manager.
- Negation ("shall **not**")
- Active/passive ("shall provide" vs "shall be provided").
- Special verbs:
  - *Obligation* Verbs (have to, ought to, be required, be expected, be compelled, be obliged, be obligated)
  - *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).

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	shall be 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

What do union contracts consist of?

Firm obligations, worker obligations, and worker entitlements

role	obligation	prohibition	permission	entitlement	total
worker	**579K (16.2%)**	83K (2.3%)	266K (7.4%)	**1115K (31.2%)**	**2045 (57.1%)**
firm	**787K (22.0%)**	46K (1.3%)	109K (3.1%)	90K (2.5%)	**1033 (28.9%)**
union	207K (5.8%)	17K (0.5%)	62K (1.8%)	130K (3.6%)	418 (11.7%)
manager	54K (1.5%)	3K (0.1%)	11K (0.3%)	16K (0.4%)	85 (2.4%)
total	1628 (45.5%)	151 (4.2%)	451 (12.6%)	1352 (37.7%)	3582K

- clauses are mostly about workers (57.1%), followed by the firm/employer (29.9%)
- single most important clause: worker entitlements, followed by firm obligations
- worker obligations: workers also make some long-term promises that the firm can rely on

Most Frequently Occurring Subject-Verb Prefixes

Table: Most frequently occurring subject-verb prefixes

subject	obligations	prohibitions	permissions	entitlement	others
worker	employee is required (41789) employee shall be (21968) employees shall be (14350)	employee shall not lose (3578) employee shall not be (3517) employee will not be (2997)	employee may request (11120) employee may elect (9148) employee shall be allowed (7524)	employee shall be paid (61643) employee shall receive (57367) employee has (54772)	employee is (181457) employee works (42449) employees are (24868)
firm	company agrees (83488) employer agrees (76739) employer shall provide (19909)	board shall not be authorized (2397) company shall not be (1403) company will not be (1133)	employer may require (4992) employer may grant (4307) company may require (2705)	company has (9725) board shall have (7767) employer has (7506)	employer recognizes (13744) company recognizes (13531) company is (9089)
union	union agrees (46060) union shall notify (6113) member is required (3034)	union will not cause (967) union will not engage (590) representatives shall not suffer (316)	representative may be (3452) union may refer (1983) union may submit (1785)	union shall have (9463) union has (5231) member shall receive (4184)	union recognizes (15091) member is (12139) union is (10315)
manager	supervisor shall give (1278) management agrees (1272) manager shall give (1057)	supervisors shall not perform (343) supervisors will not perform (284) management will not take (139)	administrator may desire (566) director may grant (384) administrator may grant (377)	principal shall receive (973) administrator may have (808) principal shall be paid (516)	supervisor is (2126) management is fix (1375) management is vest (1334)
other	there shall be (73307) parties agree (70143) there will be (33167)	provisions shall apply (4494) leave shall not exceed (4242)	case may be (14213) which may arise (6131) which may be (6042)	who has (36467) leave shall be granted (15557) leave will be granted (10311)	who is (141114) there is (116098) it is understood (102328)

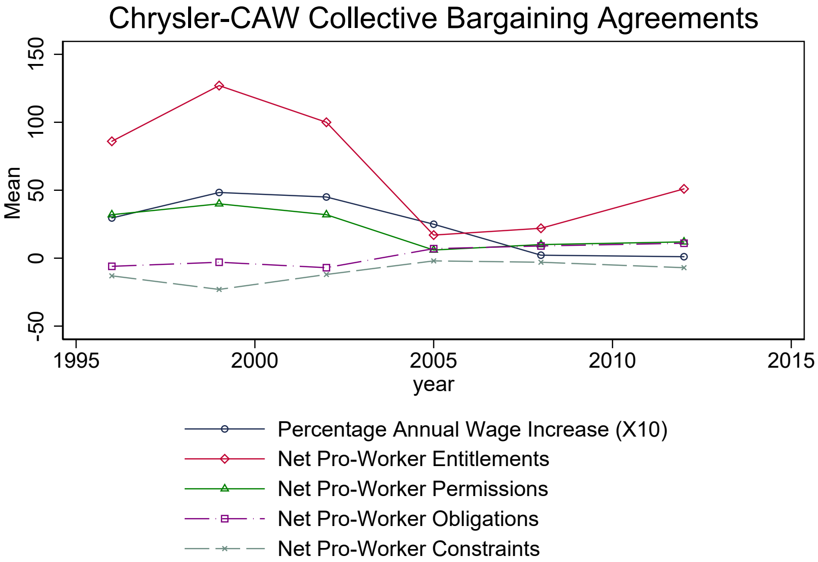
Validating Rule-Based Annotations

Method	**Precision**
weighted (freq)	0.91 – 0.99
not weighted	0.84 – 0.96

- Precision for the 317 most frequent subject-verb tuples (3.8 million statements, corresponding to 19% of the whole corpus)
- Example error: "employee shall not lose" is coded as a prohibition, but it should be an entitlement.

Case Study: Canadian Auto Workers Union Contract

Case Study: Canadian Auto Workers Union Contract



Empirical Analysis 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) $\uparrow$

Outline

Unsupervised Discovery of Gendered Language

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

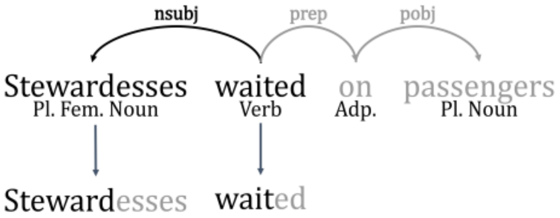


Figure 2: An example sentence with its labeled dependency parse (top) and lemmatized words (bottom).

Unsupervised Discovery of Gendered Language

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

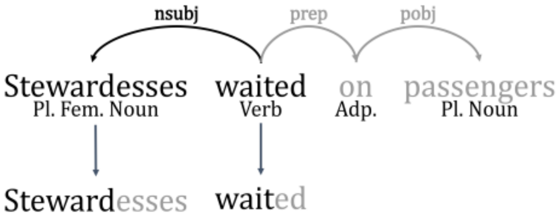


Figure 2: An example sentence with its labeled dependency parse (top) and lemmatized words (bottom).

- 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, neutral)

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, neutral)
  - 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.

Abstract

Studying the ways in which language is gendered has long been an area of interest in sociolinguistics. Studies have explored, for example, the speech of male and female characters in film and the language used to describe male and female politicians. In this paper, we aim not to merely study this phenomenon qualitatively, but instead to quantify the degree to which the language used to describe men and women is different and, moreover, different in a positive or negative way. To that end, we introduce a generative latent-variable model that jointly represents adjective (or verb) choice, with its sentiment, given the natural gender of a head (or dependent) noun. We find that there are significant differences between descriptions of male and female nouns and that these differences align with common gender stereotypes: Positive adjectives used to describe women are more often related to their bodies than adjectives used to describe men.

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

Figure 1: Adjectives, with sentiment, used to describe men and women, as represented by our model. Colors indicate the most common sense of each adjective from Tsvetkov et al. (2014); black indicates out of lexicon.

1. What is the research question?
2. Why is it important?
3. What is the problem solved?
4. What is being measured?
5. How does the measurement help answer the research question?
6. Pros/cons of this method, relative to "Gender Attitudes in Judiciary"?

Outline

Dependency Parsing: Linguistics

Dependency Parsing: Applications

- Ash, Jacobs, MacLeod, Naidu, and Stammbach (2020): Unsupervised Extraction of Rights and Duties from Collective Bargaining Agreements
- Hoyle et al (2019): Discovery of Gendered Language
- Webb (2020): AI and the Labor Market

Semantic Role Labeling

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

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.

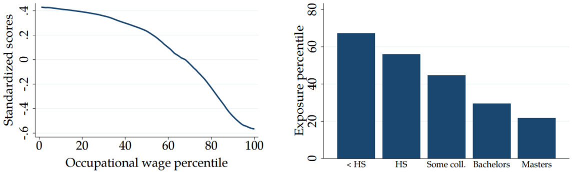
Table 10: **Top extracted verbs and characteristic nouns for AI.**

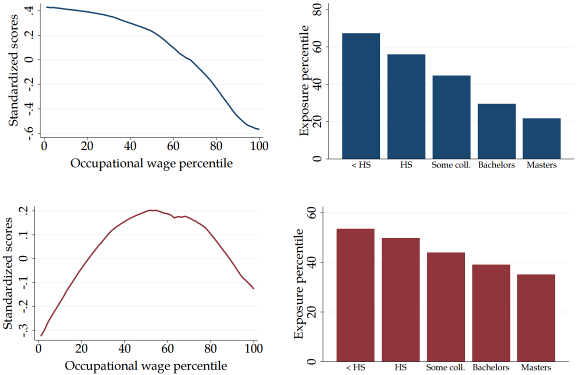
Verb	Example nouns	Verb	Example nouns
recognize	pattern, image, speech, face, voice, automobile, emotion, gesture, disease	**determine**	state, similarity, relevance, importance, characteristic, strategy, risk
predict	quality, performance, fault, behavior, traffic, prognosis	**control**	process, emission, 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

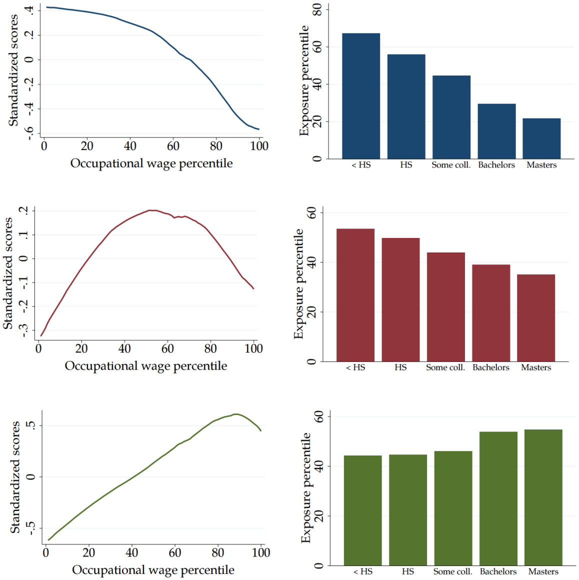
Table 11: **Occupations with highest and lowest exposure to artificial intelligence.**

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



1. What is the research question?
2. Why is it important?
3. What is the problem solved?
4. What is being measured?
5. How does the measurement help answer the research question?

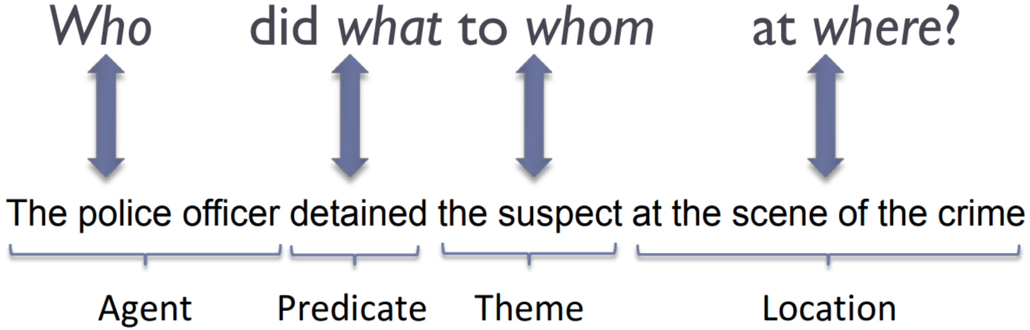
Outline

Dependency Parsing: Linguistics

Dependency Parsing: Applications

- Ash, Jacobs, MacLeod, Naidu, and Stammbach (2020): Unsupervised Extraction of Rights and Duties from Collective Bargaining Agreements
- Hoyle et al (2019): Discovery of Gendered Language
- Webb (2020): AI and the Labor Market

Semantic Role Labeling



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.

Proposition Bank (PropBank)

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

- 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:

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

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.