

Text Data in Business and Economics

Basel University – Autumn 2024

7. Linguistic Parsing

Beyond Word Order

- ▶ The models we have seen so far have counted words and phrases, or embedded sequences
 - ▶ the only language structure we have used is local 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 we have used is local ordering of words.
- ▶ Consider these examples:
 - ▶ “The defendant was negligent”
 - ▶ “The negligent defendant”
 - ▶ “The defendant was not negligent”
 - ▶ “The defendant, a driver of limousines, was negligent”
- ▶ How to reliably identify an attribute assignment between “negligent” and “defendant”?

Beyond Word Order

- ▶ The models we have seen so far have counted words and phrases, or embedded sequences
 - ▶ the only language structure we have used is local ordering of words.
- ▶ Consider these examples:
 - ▶ “The defendant was negligent”
 - ▶ “The negligent defendant”
 - ▶ “The defendant was not negligent”
 - ▶ “The defendant, a driver of limousines, was negligent”
- ▶ How to reliably identify an attribute assignment between “negligent” and “defendant”?
 - ▶ Previous methods using only word order fail. Dictionary methods would be complex and inaccurate.

Beyond Word Order

- ▶ The models we have seen so far have counted words and phrases, or embedded sequences
 - ▶ the only language structure we have used is local ordering of words.
- ▶ Consider these examples:
 - ▶ “The defendant was negligent”
 - ▶ “The negligent defendant”
 - ▶ “The defendant was not negligent”
 - ▶ “The defendant, a driver of limousines, was negligent”
- ▶ How to reliably identify an attribute assignment between “negligent” and “defendant”?
 - ▶ Previous methods using only word order fail. Dictionary methods would be complex and inaccurate.
 - ▶ A good solution requires information on sentence grammar → linguistic annotations

Constructing “Memes” with Parts of Speech (e.g. Ash 2018)

- ▶ A: Adjective, N: Noun, V: Verb, P: Preposition, D: Determinant, C: Conjunction.
- ▶ 2-grams: AN, NN, VN, VV, NV, VP.
 - ▶ tax credit, magistrate judge
- ▶ 3-grams: NNN, AAN, ANN, NAN, NPN, VAN, VNN, AVN, VVN, VPN, ANV, NVV, VDN, VVV, NNV, VVP, VAV, VVN, NCN, VCV, ACA, PAN.
 - ▶ armed and dangerous, stating the obvious

Constructing “Mememes” with Parts of Speech (e.g. Ash 2018)

- ▶ A: Adjective, N: Noun, V: Verb, P: Preposition, D: Determinant, C: Conjunction.
- ▶ 2-grams: AN, NN, VN, VV, NV, VP.
 - ▶ tax credit, magistrate judge
- ▶ 3-grams: NNN, AAN, ANN, NAN, NPN, VAN, VNN, AVN, VVN, VPN, ANV, NVV, VDN, VVV, NNV, VVP, VAV, VVN, NCN, VCV, ACA, PAN.
 - ▶ armed and dangerous, stating the obvious
- ▶ 4-grams: NCVN, ANNN, NNNN, NPNN, AANN, ANNN, ANPN, NNPN, NPAN, ACAN, NCNN, NNCN, ANCN, NCAN, PDAN, PNPV, VDNN, VDAN, VVDN.
 - ▶ Beyond a reasonable doubt (preposition, article, adjective, noun)
 - ▶ Earned income tax credit (adjective, noun, noun, noun)

What refers to what: Co-Reference Resolution

Finding all expressions that refer to the same entity in a text.

“My sister has a cat. Her name is Roberta.”



[Cat's] name is Roberta \leftrightarrow [Sister's] name is Roberta

What refers to what: Co-Reference Resolution

Finding all expressions that refer to the same entity in a text.

“My sister has a cat. Her name is Roberta.”



[Cat's] name is Roberta \leftrightarrow [Sister's] name is Roberta

Different packages (unclear winner):

- ▶ `neuralcoref` or `coreferee` by spacy
- ▶ <https://corenlp.run/>
 - ▶ Implemented in python through stanze

Outline

Dependency Parsing

Linguistics

Application: Arold et al. (2024): Do Words Matter? The Value of Collective Bargaining Agreements

Semantic Role Labeling

Application: Ash, Gauthier, and Widmer (2023): Relatio

Outline

Dependency Parsing

Linguistics

Application: Arold et al. (2024): Do Words Matter? The Value of Collective Bargaining Agreements

Semantic Role Labeling

Application: Ash, Gauthier, and Widmer (2023): Relatio

Dependency Grammar

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

Dependencies: Binary Directed Relations Between Words (Head and Dependent)

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

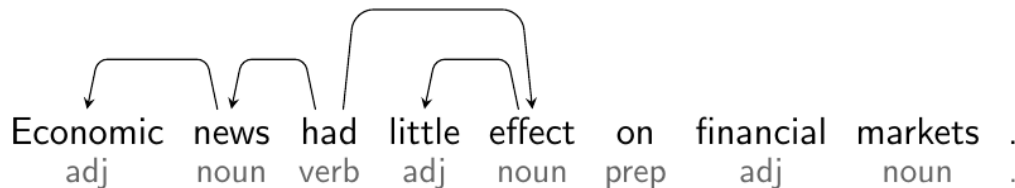
- ▶ dependency trees are mostly determined by the ordering of POS tags.

Dependencies: Binary Directed Relations Between Words (Head and Dependent)



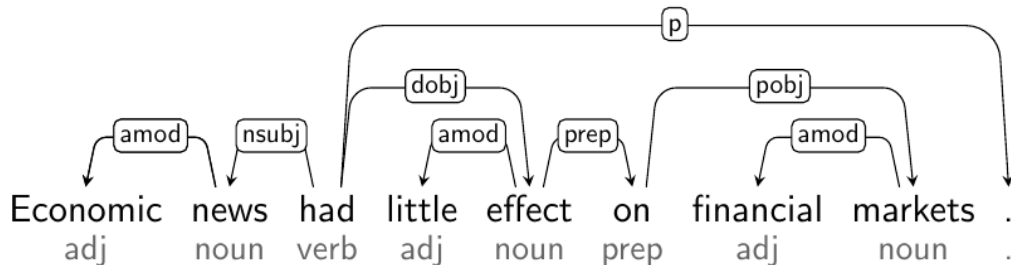
- ▶ the “root” of a sentence is the main verb (for compound sentences, the first verb).

Dependencies: Binary Directed Relations Between Words (Head and Dependent)



- ▶ directed arcs indicate dependencies: a one-way link from a “head” token to a “dependent” token.
- ▶ A word can be “head” multiple times, but “dependent” only one.

Dependencies: Binary Directed Relations Between Words (Head and Dependent)

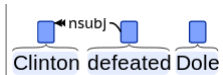


- ▶ arc labels indicate functional relations, e.g.:
 - ▶ nsubj: verb → subject doing the verb
 - ▶ dobj: verb → object targeted by the verb
 - ▶ amod: noun → attribute of the noun
- ▶ spaCy dependency visualizer: <https://explosion.ai/demos/displacy>

Subjects

- ▶ **nsubj: nominal subject**

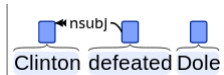
- ▶ non-clausal constituent in the subject position of an active verb.



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.



Objects

dobj: direct object

- ▶ noun phrase, the (accusative) object of the verb.

“She **gave** me a **raise**”

gave \xrightarrow{dobj} *raise*

Objects

dobj: direct object

- ▶ noun phrase, the (accusative) object of the verb.

“She **gave** me a **raise**”

gave \xrightarrow{dobj} *raise*

dative: dative or indirect object

- ▶ noun phrase, the (dative) object of the verb.

“She **gave** **me** a raise”

gave \xrightarrow{dative} *me*

Objects

dobj: direct object

- ▶ noun phrase, the (accusative) object of the verb.

“She **gave** me a **raise**”

gave \xrightarrow{dobj} *raise*

dative: dative or indirect object

- ▶ noun phrase, the (dative) object of the verb.

“She **gave** **me** a raise”

gave \xrightarrow{dative} *me*

pobj: object of a preposition

- ▶ noun phrase following a preposition

“I sat **on** the **chair**”

on \xrightarrow{pobj} *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 a saint**”: attr(is → saint)

Adjectives/Attributes

acomp: adjectival complement

- ▶ adjectival phrase which functions as object of verb.

“Bill **is honest**”: $\text{acomp}(\text{is} \rightarrow \text{honest})$

attr: attribute

- ▶ noun phrase that is a non-VP predicate usually following a copula verb.

“Bill **is a saint**”: $\text{attr}(\text{is} \rightarrow \text{saint})$

amod: adjectival modifier

- ▶ modifies the meaning of the noun phrase.

“Sam eats **red meat**”: $\text{amod}(\text{meat} \rightarrow \text{red})$

Adjectives/Attributes

acom: 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.

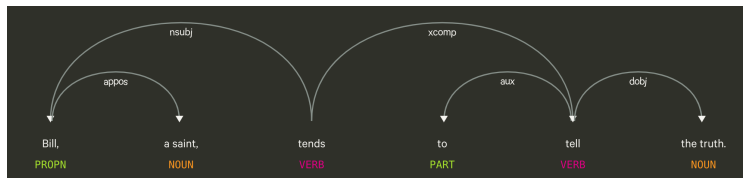
“Bill **is a saint**”: attr(is → saint)

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 verb it modifies.
 - ▶ “Bill **is not** a scientist”: $\text{neg}(\text{is} \rightarrow \text{not})$
 - ▶ “Bill **doesn't** drive”: $\text{neg}(\text{drive} \rightarrow \text{n't})$

Etc.

▶ **neg: negation modifier**

- ▶ captures negation and the verb it modifies.
- ▶ “Bill **is not** a scientist”: $\text{neg}(\text{is} \rightarrow \text{not})$
- ▶ “Bill **doesn't** drive”: $\text{neg}(\text{drive} \rightarrow \text{n't})$

▶ **poss: possession modifier**

- ▶ holds between noun phrase and its possessive determiner, or genitive's complement
- ▶ “**Bill's** clothes”: $\text{poss}(\text{clothes} \rightarrow \text{Bill})$

Etc.

- ▶ **neg: negation modifier**
 - ▶ captures negation and the verb it modifies.
 - ▶ “Bill **is not** a scientist”: $\text{neg}(\text{is} \rightarrow \text{not})$
 - ▶ “Bill **doesn't** drive”: $\text{neg}(\text{drive} \rightarrow \text{n't})$
- ▶ **poss: possession modifier**
 - ▶ holds between noun phrase and its possessive determiner, or genitive's complement
 - ▶ “**Bill's** clothes”: $\text{poss}(\text{clothes} \rightarrow \text{Bill})$
- ▶ **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...

Outline

Dependency Parsing

Linguistics

Application: Arold et al. (2024): Do Words Matter? The Value of Collective Bargaining Agreements

Semantic Role Labeling

Application: Ash, Gauthier, and Widmer (2023): Relatio

New corpus: 30,000 collective bargaining agreements from Canada, 1986-2015

2005 – 2006 calendar

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 cheques 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 rehire 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.

Parse Information on Subjects and Verbs

- ▶ Subject categories, assigned with lexicons
 - ▶ worker, union, owner, manager
- ▶ In contracts, modal verbs impose legal requirements:
 - ▶ strict (shall, will, must) modals express necessity.
 - ▶ permissive (may, can) modals express possibility.
- ▶ 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).

Table 2: Contract Clause Typology

Categorization Logic	Examples
<u>Obligations</u>	
Positive & Strict Modal & Active Verb	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
<u>Prohibitions</u>	
Negative & Any Modal & Active Verb	shall not exceed, shall not use, shall not discriminate
Negative & Permission Verb	shall not be allowed, is not permitted
Positive & Strict Modal & Constraint Verb	shall be prohibited, shall be restricted
<u>Permissions</u>	
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
<u>Rights</u>	
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

Note: Typology of clause types, word patterns, and corresponding examples.

What do Clauses Contain?

Table 3: Summary Statistics: What do Union Contracts Consist of?

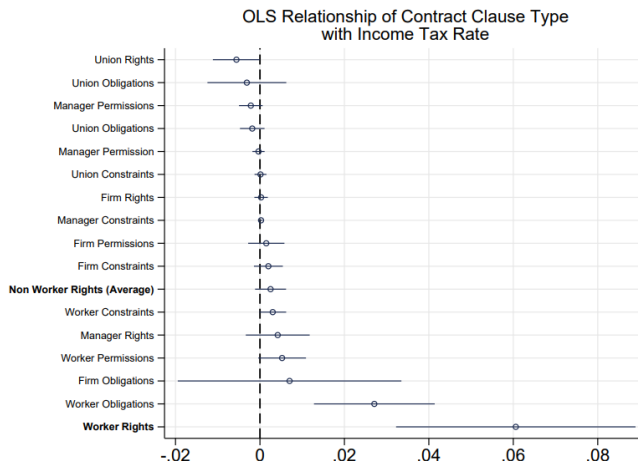
Agent	<u>Obligations</u>		<u>Rights</u>		<u>Permissions</u>		<u>Prohibitions</u>		<u>Total</u>	
	Count	%	Count	%	Count	%	Count	%	Count	%
Worker	666021	20.9	730327	22.9	266278	8.4	99795	3.1	1,762,421	55.3
Firm	788857	24.7	27382	0.9	108114	3.4	49099	1.5	973,452	30.5
Union	224119	7.0	68410	2.1	62912	2.0	17786	0.6	373,227	11.7
Manager	55518	1.7	7512	0.2	12537	0.4	3107	0.1	78,674	2.5
Total	1,734,515	54.4	833,631	26.2	449,841	14.1	169,787	5.3	3,187,774	100.0

Note: Counts and frequencies (%) of clause types in the contracts corpus, by agent (rows) and statement type (columns).
Data source: Employment and Social Development Canada.

- ▶ clauses are mostly about workers (55.3%), followed by the firm/employer (30.5%)
- ▶ most important clause: firm obligations and worker obligations

Clauses are used to Compensate Workers

Figure 4: Effect of Labor Income Tax Rates on Contract Terms



Note: Figure presents coefficients and 95% confidence intervals of effect of labor tax rate on contract clause types. Each coefficient is from a separate OLS regression. Outcome: Clause type share (number of clauses of type in question over the number of all clauses). Treatment: Labor tax rate, defined as logarithmized implicit personal income tax rate. Controls: Province-by-sector fixed effects and year-by-sector fixed effects. Inference: Standard errors clustered at the province-by-sector level. Data sources: Employment and Social Development Canada, Center for the Study of Living Standards.

Outline

Dependency Parsing

Linguistics

Application: Arold et al. (2024): Do Words Matter? The Value of Collective Bargaining Agreements

Semantic Role Labeling

Application: Ash, Gauthier, and Widmer (2023): Relatio

Who



did what to whom



at where?



The police officer detained the suspect at the scene of the crime

Agent

Predicate

Theme

Location

Source: Jurafsky-Martin slides.

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 (ARG0)
 - ▶ Volitional/sentient involvement in event or state
 - ▶ Causes an event or change of state in another participant
- ▶ Patient (ARG1)
 - ▶ Causally affected by an agent/action
 - ▶ Undergoes change of state
- ▶ ARG2 has three functions:
 - ▶ instrument for an action (“Pat opened the door with a crowbar.”)
 - ▶ attribute assigned to a patient (“Pat is an agent”).
 - ▶ benefactive: the dative/indirect object (“Sasha gave the crowbar to Pat.”)

Outline

Dependency Parsing

Linguistics

Application: Arold et al. (2024): Do Words Matter? The Value of Collective Bargaining Agreements

Semantic Role Labeling

Application: Ash, Gauthier, and Widmer (2023): Relatio

Extracting Narratives with Relatio

- ▶ Narratives defined as *Agent-Verb-Patient* tuples

Extracting Narratives with Relatio

- ▶ Narratives defined as *Agent-Verb-Patient* tuples
- ▶ Extracted combining SRL algorithm from AllenNLP (*discontinued*) and Spacy dependency parsing

Extracting Narratives with Relatio

- ▶ Narratives defined as *Agent-Verb-Patient* tuples
- ▶ Extracted combining SRL algorithm from AllenNLP (*discontinued*) and Spacy dependency parsing



Pipeline

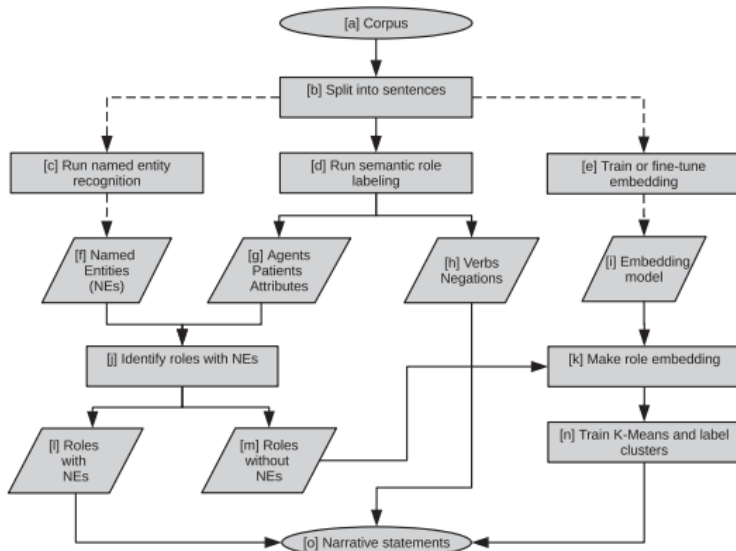


Figure 2. Flowchart for RELATIO. Code flowchart for programmatic implementation, open-sourced as the Python RELATIO package (github.com/relatio-nlp/relatio). Circles represent the start and the end of the pipeline. Rectangles represent arithmetic operations and data manipulations. Parallelograms represent inputs and outputs.

Results

Narrative	Freq.	Spoken sentence(s)
People lose job	1,627	<ul style="list-style-type: none">– President, people are losing jobs every day.– People shouldn't have to lose their jobs to pay for the New York fund.
Citizen abide law	1,176	<ul style="list-style-type: none">– I felt kicked around and ignored by the very system the government has in place to protect law-abiding citizens.– The vast majority of these private security officers are dedicated, hard-working, law-abiding citizens of this country [...].
American lose job	1,168	<ul style="list-style-type: none">– [...] Americans across this country have lost jobs, unemployment is at a high rate, people are having to make decisions.– These funds will go a long way in supporting American workers who have lost their jobs [...].
Government run healthcare	1,089	<ul style="list-style-type: none">– I don't understand what they are talking about: "socialized medicine," "Cuban-style, government-run health care."– Is that what you are talking about where you all of a sudden shift from people who figure out you can get the government to pay for everything, a government-run health care program?
American have healthcare	1,080	<ul style="list-style-type: none">– [...] [W]e can contain costs and help enable every American to have access to health insurance coverage.– [...] Americans already have universal health care because the emergency rooms cannot legally refuse to treat patients.

DAG visualization: <https://sites.google.com/view/trump-narratives>