

Whodunit...and to Whom? Subjects, Objects, and Actions in Research Articles on American Labor Unions

Vilja Hulden

Department of History

University of Colorado

`vilja.hulden@colorado.edu`

Abstract

This paper investigates whether sentence structure analysis—examining who appears in subject versus object position—can illuminate who academic articles portray as having agency in labor relations. We extract subjects and objects from a corpus of 3,800 academic articles, and compare both the relative occurrence of different groups (workers, women, employers) in each position and the verbs that most commonly attach to each group. We conclude that agency, while elusive, can potentially be modeled by sentence structure analysis.

1 Introduction

In scholarship on grassroots movements and non-elite groups, the question of “agency” often looms large (Johnson, 2003). Who exactly do we, as scholars, portray as taking action, accomplishing historical change, doing rather than being done to?

With regard to scholarship on American labor unions, the main fault lines along which “doers” and “done-to” are split involve not only employers versus unions, but also the union leadership versus the rank and file, and unionized workers versus unaffiliated workers. This question has, indeed, informed some of the major shifts in the writing of labor history, as scholars have moved away from the “Wisconsin school” of John Commons, which focused on the institutional and organizational history of unions and toward a more inclusive and bottom-up social history of workers, affiliated or not (Isserman, 1976; Fink, 1988). More recently, perhaps spurred by the sorry state of American labor unions, interest in unions as institutions and organized movements has resurfaced (Currarino, 2011; Taillon, 2009). As many scholars have

noted, however, there seems to often be an excess of attention to the articulate leadership and the actions of the union as an institution, even if the rank and file (let alone unaffiliated workers) may not share those views or endorse those actions (Pierce, 2010).

The question of who gets to speak for social movements is hardly limited to the history of organized labor. Similar questions about whose actions command attention (as well as about who does the hard work and who gets the credit) have been raised about the Civil Rights Movement as well (Hall, 2005; Ransby, 2003). More recently, the efforts of the Black Lives Matter movement to remain multipolar and avoid focusing attention on “leaders” have raised both the question of whether that is a useful strategy vis-à-vis the media or the public’s perceptions of the significance or the message of the organization, and the question of the risks of one or a handful of “charismatic leaders” (Harris, 2015).

This paper investigates whether these problems of agency—fundamentally, who exercises some measure of power—can be perceived in scholarly writing using natural language processing (NLP) tools. A syntactic analysis has potential to go beyond bag-of-words models like topic modeling in illuminating power relations, as well as to capture more clearly who exactly is at the center of the analysis. Analysis of subjects and objects can also easily be combined with analysis of which actions are related to which subjects/objects, revealing interesting patterns about the ways different groups of actors are represented in the literature. In future work, we hope to expand the analysis by experimenting with Semantic Role Labeling in addition to syntactic analysis as well as with using FrameNet (Baker, 2008; Palmer, 2009) and VerbNet (Kipper et al., 2008) to discover patterns in the actions.

In what follows, we offer a preliminary analysis focusing on noun phrases (NP) that appear in either a subject (passive or active) or an object (indirect or direct) position, and of the actions they most commonly perform or are subjected to. Does this grammatical representation of the doers and the done-to reproduce the splits usually emphasized in scholarship? Who, in academic writing, appears as a doer, grammatically speaking? Do the actions associated with doers and done-to modify assumptions about who has agency in this corpus?

2 Dataset

The texts examined in this paper consist of the set of English-language research articles over 9 pages contained in the JSTOR article database answering the query (“american federation of labor”). The query was selected to weight attention toward “mainstream” organized labor rather than e.g. working-class culture or the Socialist movement, though naturally the dataset also contains articles on e.g. the radical Industrial Workers of the World (IWW).¹ This query produces a set of 4,183 articles, of which about 70 percent are published after 1945. The final set consists of a subset of 3,807 of these articles successfully processed using the Stanford CoreNLP parser (Manning et al., 2014).

3 Extracting subjects and objects

Extracting subjects and objects from the parsed articles was performed using the Stanford Tregex utility (Levy and Andrew, 2006).

The expressions used to extract subjects (active and passive) and objects (direct and indirect) are listed in table 1. The copula “to be” was excluded from consideration. As the main expressions capture rather long noun phrases (NPs), a constraining expression was used to further narrow those phrases down to more useful sub-NPs.

3.1 Most common entities

Disregarding for the moment whether an entity (NP) appears as subject, direct object, or indirect object, the list of most-frequent animate entities in the corpus reads like the cast of main characters

¹The American Federation of Labor was the dominant union umbrella group until the emergence of the Congress of Industrial Organizations (CIO) in the 1930s; the AFL and the CIO merged in 1955 to form the present-day AFL-CIO.

```

SUBJECT (ACTIVE) :
(NP [<<@/NN.?/])
  [$. (ADVP
    $. (VP < (@/VB.?/
      !< is|was|are|were
      !<< (be|been)))
    | $. (VP < (@/VB.?/
      !< is|was|are|were
      !<< (be|been))) ]
SUBJECT (PASSIVE) :
NP [<@/NN.?/
  >> (PP < (IN < by))
  !> (PP < (IN < !by))
INDIRECT OBJECT:
(NP [<<@/NN.?/])
  [> ((PP [< (IN < for)| < TO ]
    >> (VP < (@/VB.?/
      !< is|was|are|were|
      have|had|has)))) ]
  | [ > VP $+ NP]
DIRECT OBJECT:
NP << @/NN.?/
  > VP < @/VB.?/
    !< is|was|are|were
    |have|had|has
    !. NP
CONSTRAINT:
NP [!>> PP & !>> VP] [<@/NN.?/]

```

Table 1: Tregex expressions used

and issues of industrial relations, with e.g. *workers*, *employers*, and the *american federation* (of labor) as well as *legislation* and *wages* clearly represented (see table 2).²

Some trends can be extracted even from this basic count of subject/object NPs: for example, as figure 1 shows, women’s involvement in the labor movement has been of shifting scholarly interest, with the first peak coinciding roughly with the suffrage movement and the second upward trend beginning around the rise of second wave feminism in the 1970s. Although the topic model³ depicted in figure 2 finds a similar pattern in the data, the NP-based graph offers a much more fine-grained and more easily interpreted view.

3.2 Subjects and objects

But what about the question of agency? Is there any pattern in who appears as a subject and who appears as an object?

There is, though the results should be taken with some caution. Table 3 and figure 3 show selec-

²The count is the sum of the times the NP appeared as indirect, object, direct object, passive subject, and active subject.

³Topic model created using MALLET (McCallum, 2002), 50 topics, 1000 iterations, optimize-interval 20.

NP	TOTAL COUNT
workers	13703
members	6537
labor	8148
people	6145
union	9625
work	6014
unions	8334
state	5509
men	5316
president	5465
government	4808
congress	5853
law	5485
employers	5522
organization	3340
united states	4492
employees	5427
power	3577
committee	4126
women	7399
majority	2893
attention	2735
wages	2937
american federation	1961
legislation	2678
example	3117
history	2492
study	2290
board	5774
court	9389

Table 2: Selections from the 50 most common subject/object NPs. Generic terms (e.g. purpose, example) excluded.

tions of the most frequent human or human-like entities according to the entity’s degree of “subjectness.” The table and figure were constructed by first selecting the 1,000 most frequent NPs in the data and then calculating for each the ratio of how many times it appeared as a subject (passive or active) versus as an object (indirect or direct). From this was then deducted the overall ratio of subjects to objects in the dataset, and the resulting figure was used as a proxy for “subjectness.” Thus, negative ratios in table 3 indicate that the NP is found in object position more commonly than the average NP in the data (the count reflects the sum of mentions, each position being counted once per article). Of the 1,000 most frequent terms, few were of this “more-object-than-average” character; however, the spread of “subjectness” allows some preliminary conclusions.⁴

⁴We did not perform coreference resolution, and thus have no way of capturing repeated references to the same entity with different words. To mitigate this, we have used a count of how many *articles* an NP appears in as subject/object rather than allowing multiple counts per article. The order of the NPs in terms of subjectness if multiple instances per

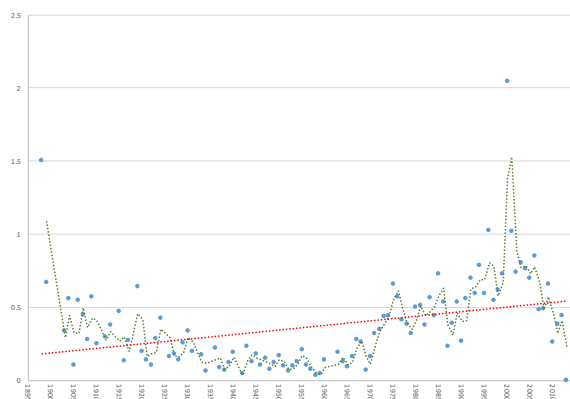


Figure 1: NP “women” (obj/subj) in the corpus

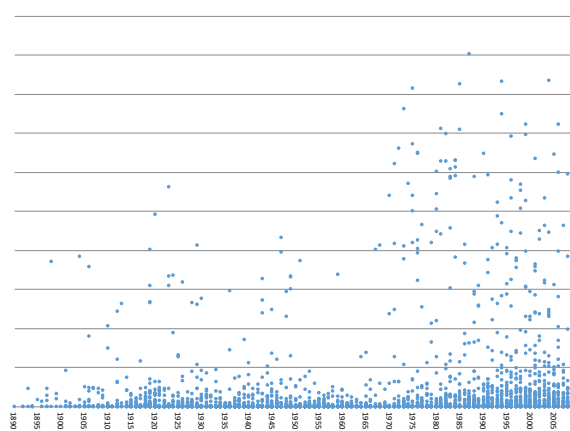


Figure 2: Topic “women” in the corpus (women women’s work men woman female family male children gender working workers equal sexual sex suffrage)

On the whole, workers (even strikers) appear quite commonly in an object position, whereas the government, unions as organizations, and employers appear clearly more commonly than average in a subject position. Partly the results are explainable by specificity: the AFL-CIO and the well-known AFL leader Samuel Gompers are more likely to appear as subjects, whereas “workers” is only barely above average in its “subjectness.” However, it is worth noting that “employers” and “manufacturers” are significantly above “workers” and variants thereof in subjectness. Even as “strikers,” workers’ subjectness is quite low—although as “unionists” their subjectness is slightly higher than that of manufacturers.

article are considered is very nearly the same as presented here.

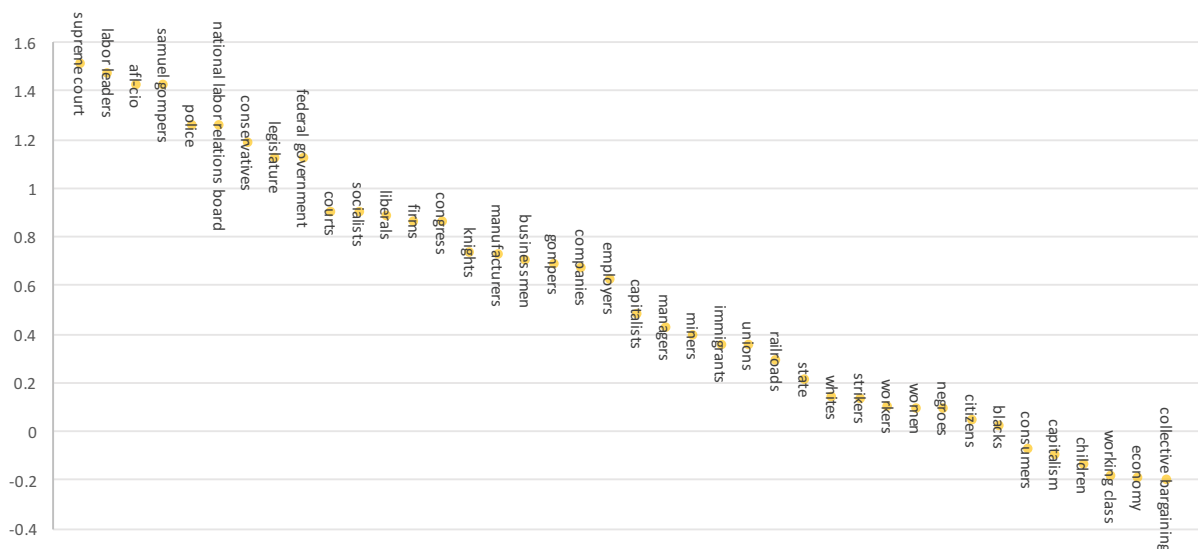


Figure 3: Selected NPs ordered from “subjectness” to “objectness.”

4 The actions of doers and done-to

To further investigate the meaning of a word appearing in a subject versus object position, the most common verbs for each word and each category (object, indirect object, subject) were extracted. The indirect object category was mostly too ambiguous to draw conclusions from, involving verbs like *send*, *provide*, *give*, *distribute*; the analysis below therefore focuses on the subject versus indirect object categories.⁵

4.1 “Unions”

The verbs associated with unions as actors (subjects) are the bread and butter of union activity: they *affiliate*, *represent*, *organize*, *seek*, *agree*, *refuse*, *grow*, *demand* — and encouragingly, *win* rather more often than *lose*.

As acted-upon (direct objects), unions seem to mainly reflect worker activity: *join* is by far the most common verb, followed by verbs like *form*, *organize*, and *build*. The third most common verb here is *recognize*, i.e., achieving union recognition by the employer. However, high on the list are also *destroy*, *break*, and *prohibit*, reflecting the contested history of labor.

4.2 “Workers,” “members,” and “strikers”

High on the list of verbs associated with workers as actors are *organize* and *strike*; interestingly,

⁵Excessively generic verbs like *do* or *make* are ignored in the analysis.

strike comes much higher on the list for “workers” than for “unions” (10th versus 73rd place).

“Members” as actors are clearly tied to the bureaucracy and process of union activity: they *appoint*, *vote*, *elect*, *represent*, and *participate*. Intriguingly, they also *grapple* and *adapt*.

As acted-upon, both members and workers are *organized*, *represented*, *recruited*, and *mobilized*, as well as *employed* and *hired*. However, workers are also *excluded* and *divided*, reflecting the divisions among workers and the not-always-inclusive nature of American labor unions. Meanwhile, members get *disciplined*, presumably reflecting conflicts between leadership and rank and file, and *forbidden*, possibly by police or courts. Both workers and members are the targets of someone’s efforts to *educate*.

When they appear as “strikers,” the main thing workers do is *return* (to work, presumably). They also *demand*, *remain* (on strike?), *refuse*, *vote*, and *win* or *lose*. As acted-upon, strikers most commonly get *replaced*. But they are also *supported*, *urged*, *aided*, *rehired*, and *reinstated* — as well as *restrained*, *arrested*, and *intimidated*.

4.3 “Employers”

Employers are not primarily the initiators of action in this corpus: rather, the two most common verbs for “employers” as actor are *refuse* and *agree*. In the top 25 are also *violate* (presumably agreements) and *resist* (presumably unions).

As acted-upon, employers in this corpus find

NP	RATIO	COUNT
experts	2.137	250
president roosevelt	2.133	292
supreme court	1.511	1330
labor leaders	1.470	391
american federation	1.439	1656
afl-cio	1.421	338
samuel gompers	1.420	316
authorities	1.397	292
union leaders	1.307	386
police	1.259	460
national labor relations board	1.255	332
conservatives	1.184	243
legislature	1.123	555
federal government	1.120	945
republicans	0.954	426
courts	0.898	1306
socialists	0.896	461
liberals	0.887	244
congress	0.856	2524
politicians	0.787	342
unionists	0.753	320
manufacturers	0.723	411
legislators	0.721	260
businessmen	0.704	251
democrats	0.684	438
companies	0.673	698
employers	0.625	2441
americans	0.617	998
reformers	0.601	268
government	0.594	2543
cio	0.567	676
communists	0.537	521
organized labor	0.518	1080
lawyers	0.511	333
iww	0.491	286
capitalists	0.483	263
labor unions	0.438	728
managers	0.423	307
businesses	0.387	241
immigrants	0.355	487
unions	0.354	3033
railroads	0.286	403
socialist party	0.229	283
state	0.214	2900
democratic party	0.209	282
communist party	0.176	328
whites	0.137	456
strikers	0.134	765
skilled workers	0.114	237
african americans	0.107	287
workmen	0.104	435
industry	0.102	1505
workers	0.099	4627
women	0.090	1781
negroes	0.090	518
black workers	0.074	287
laborers	0.068	414
citizens	0.044	857
blacks	0.018	589
industrial workers	-0.030	314
consumers	-0.075	330
capitalism	-0.096	313
children	-0.140	1147
working class	-0.187	376
economy	-0.191	513
collective bargaining	-0.199	856

Table 3: NPs and “subjectness.”

See section 3.2 for discussion.

themselves the target of efforts to *require*, *force*, *compel*, *prevent*, *prohibit*, and *coerce*, though also to *allow*, *permit*, and *induce*.

4.4 “Women”

The main thing that women do in this corpus is *work*; it seems that the main news about women as workers is that they exist. High on the list is also *enter*, probably from a phrase like “enter the workforce.” However, women also *participate*, *want*, *organize* and *negotiate*.

As acted-upon, women are *given*, *organized*, *employed*, and bafflingly, *ordained*. They are also *encouraged* and *excluded* (7th and 8th position).

5 Discussion

As the above analysis demonstrates, grammatical subjects and objects function as a rough proxy for examining agency, illuminating who tends to be the doer and who the done-to: the broad lines of which NPs have high “subjectness” coincide with one’s intuition of the prevailing power relations. At least as interesting, however, is that the verbs attached to each further demonstrate their different roles. Juxtaposing the subjectness and the common verbs is particularly interesting: for instance, it is intriguing that in a corpus where employers appear in a not-so-favorable light (as *resisting*, *refusing*, and *violating*, among other things), they are nevertheless as a group more likely than workers to occupy a position of agency as subjects. On the other hand, the tensions between union leadership and rank-and-file are also revealed in, for example, the fact that “members” find themselves the object of verbs like *discipline*.

6 Future research

In the future, we hope to investigate whether SRL analysis would offer greater clarity in distinguishing agents from non-agents. We also hope to refine the preliminary verb analysis presented here by using verb categories as defined in VerbNet and FrameNet. In addition, we plan to combine the type of analysis presented here with an analysis of named entities; this might allow us to investigate not only the prominence of well-known figures, but possibly also questions like whether the rise of bottom-up approaches in the 1970s or the cultural turn of the 1990s resulted in a greater variety of named entities.

References

- Collin Baker. 2008. FrameNet, present and future. In *The First International Conference on Global Interoperability for Language Resources*, pages 12–17.
- Roseanne Currarino. 2011. *The Labor Question in America: Economic Democracy in the Gilded Age*. University of Illinois Press, Urbana.
- Leon Fink. 1988. The new labor history and the powers of historical pessimism: Consensus, hegemony and the case of the Knights of Labor. *Journal of American History*, 75:115–136, June.
- Jacquelyn Dowd Hall. 2005. The long Civil Rights Movement and the political uses of the past. *Journal of American History*, 91(4):1233–1263.
- Fredrick C. Harris. 2015. The next Civil Rights Movement? *Dissent*, 63(3):34–40, Summer.
- Maurice Isserman. 1976. ‘God bless our American institutions’: The labor history of John R. Commons. *Labor History*, 17:309–328.
- Walter Johnson. 2003. On agency. *Journal of Social History*, 37(1):9–21.
- Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2008. A large-scale classification of English verbs. *Language Resources and Evaluation*, 42(1):21–40.
- Roger Levy and Galen Andrew. 2006. Tregex and Tsurgeon: tools for querying and manipulating tree data structures. *5th International Conference on Language Resources and Evaluation (LREC 2006)*.
- Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60.
- Andrew Kachites McCallum. 2002. MALLET: A machine learning for language toolkit. <http://mallet.cs.umass.edu>.
- Martha Palmer. 2009. Semlink: Linking PropBank, VerbNet and FrameNet. In *Proceedings of the Generative Lexicon Conference*, pages 9–15.
- Michael Pierce. 2010. *Striking with the Ballot: Ohio Labor and the Populist Party*. Northern Illinois University Press, DeKalb.
- Barbara Ransby. 2003. *Ella Baker and the Black Freedom Movement: A Radical Democratic Vision*. University of North Carolina Press.
- Paul Michel Taillon. 2009. *Good, Reliable, White Men: Railroad Brotherhoods, 1877–1917*. University of Illinois Press, Urbana.