Annotating Causal Language Using Corpus Lexicography of Constructions

Jesse Dunietz, Lori Levin, and Jaime Carbonell

LAW 2015

June 5, 2015

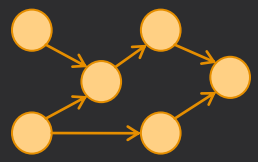
Contributions of this paper

- Raising issues about corpus annotation:
  - Low agreement among non-experts
- Methodology for annotation projects
  - Lexicon driven annotation: as in PropBank and FrameNet
- An annotation scheme for causal language in English
- A constructicon of causal language in English
- A small annotated corpus of causal language in English
- All still in progress

Causal relations would be useful to annotate well…

Ubiquitous in our mental models

- Medical symptoms
- Political events
- Interpersonal actions



Ubiquitous in language

- 2nd most common relation between verbs (33%; Conrath et al. 2011)

Useful for downstream applications (e.g., information extraction)

- The prevention of FOXP3 expression was not caused by interferences.

…but annotating them raises difficult annotation issues.

Varied linguistic expression

Smoking **causes** cancer.	(Verbal)
They came **because of** the schools.	(Prepositional)
For reasons of privacy, I can't tell you.	(Complex)

Tricky to circumscribe and agree on

The rules **forbid** me **to** leave.	(Permission)
I **convinced** him **to** go.	(Additional information)
They're **too** big **to** fail.	(Comparative)

Intertwined with other phenomena

If I had told you, I'd have to kill you.	(Counterfactuals)
After a drink, she felt much better.	(Temporal relations)
Don't do it **because of** the money!	(Negation)

1. A detailed, construction-based representation

Several projects have attempted to annotate **real-world causality**.

SemEval 2007 Task 4

(Girju et al., 2007)

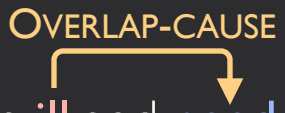
"A person infected with a <e1>flu</e1> <e2>virus</e2> strain develops antibodies against it."

`Cause-Effect(e2, e1) = "true"`

Richer Event Descriptions

(Ikuta et al., 2014)

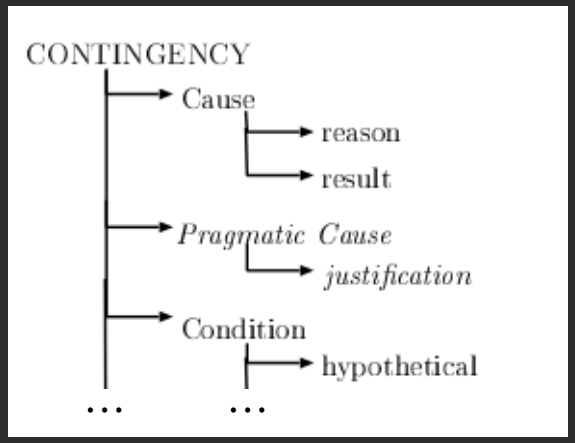
A budget was **allocated** for the barrier to be **equipped** with electronic detention equipment.



Heroin users are **ill** and **need** treatment.

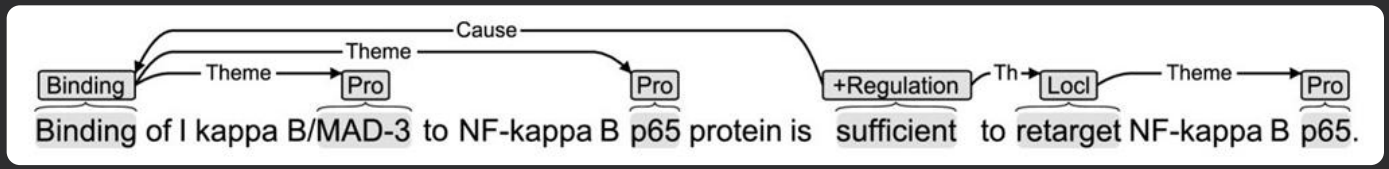
Others have focused on causal **language**.

Penn Discourse Treebank
(Prasad et al., 2008)



Causality in TempEval-3
(Mirza et al., 2014)

BioCause
(Mihaila et al., 2013)



Causal language:
a clause or phrase in which
one event, state, action, or entity
is **explicitly presented**
as promoting or hindering
another

Connective: fixed construction indicating a causal relationship

John killed the dog **because**
it was threatening his chickens.

John **prevented** the dog **from**
eating his chickens.

Ice cream consumption **causes** drowning.

She must have met him before, **because**
she recognized him yesterday.

Not "truly" causal

Effect: presented as outcome/inferred conclusion
Cause: presented as producing/indicating effect

John killed the dog because
it was threatening his chickens.

John prevented the dog from
eating his chickens.

Ice cream consumption causes drowning.

She must have met him before, because
she recognized him yesterday.

We exclude language that does not encode **pure, explicit causation**:

Relationships with no lexical trigger

John killed the dog. It was threatening his chickens.

Connectives lexicalizing a means or result

John killed the dog.

Unspecified causal relationships

The treatment is linked to better outcomes.

Temporal language

After he took a drink, he felt much better.

Four types of causation

The system failed **because of** a loose screw. → CONSEQUENCE

Mary left **because** John was coming. → MOTIVATION

Mary left **in order to** avoid John. → PURPOSE

The engine is still warm, **so** it must have been driven recently. → INFERENCE

Not all causal relationships are of equal strength or polarity.

~~ENTAIL~~

This has often **caused** problems elsewhere. → FACILITATE

Only by collaborating **can** we succeed. → ~~ENABLE~~

Without regulation, the problem will persist. → ~~DISENTAIL~~

He **kept** the dog **from** leaping at her. → INHIBIT

~~PREVENT~~

2. Comparison of two annotation approaches

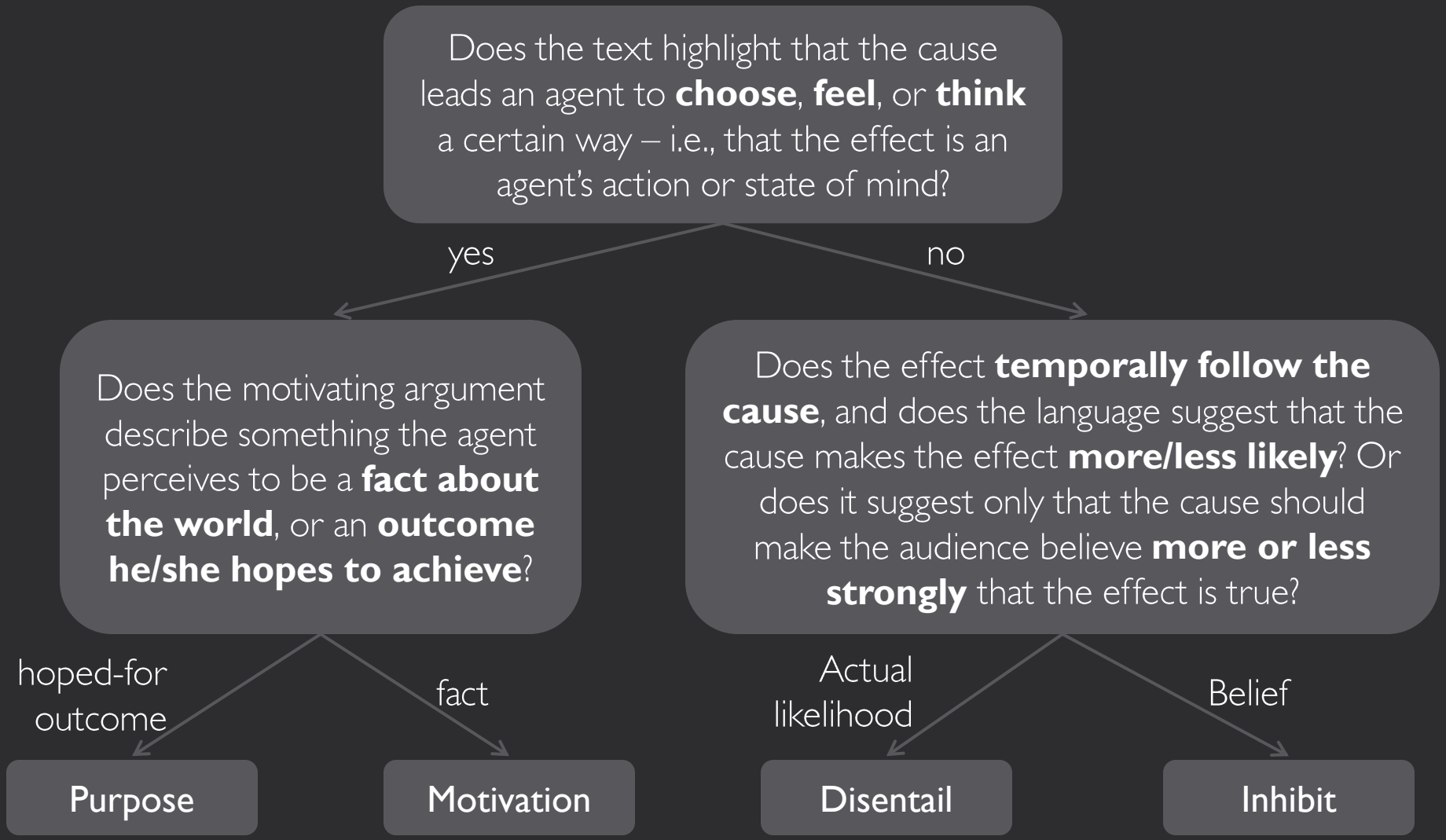
First Try

- Dunietz and three annotators (A1, A2, A3)
  - A1, A2, and A3 are recently graduated linguistics majors.
  - A1 had more than one year annotation experience.
  - A2 and A3 did not have annotation experience.

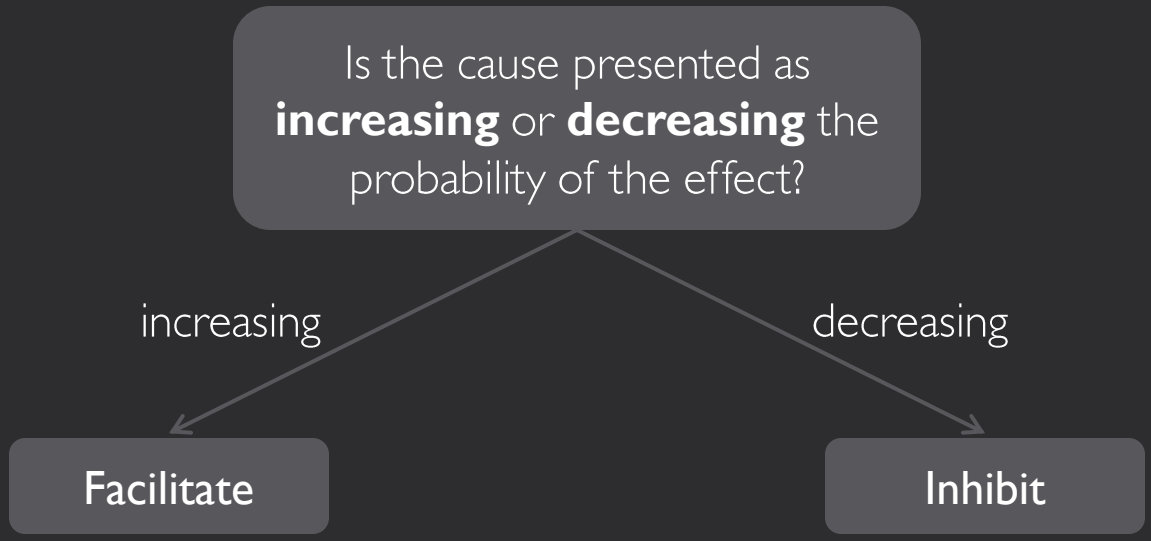
First try (Continued)

- Rounds of annotation and reconciliation
- Produced a coding manual
- Annotator A4
  - Masters in linguistics plus 30 years experience with corpus annotation and NLP

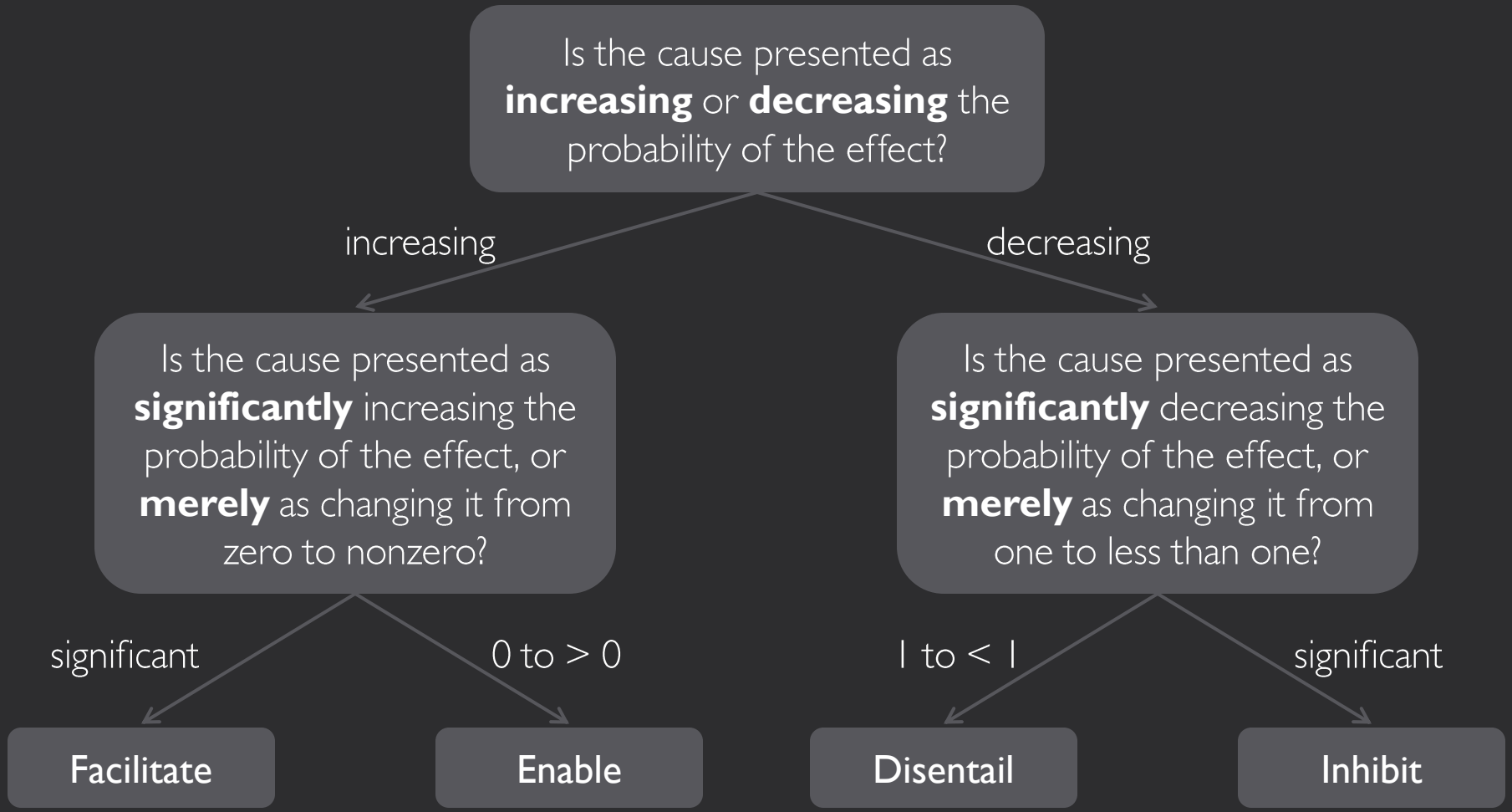
Annotators determined the causation type using a decision tree.



Annotators determined the causation degree using another decision tree.

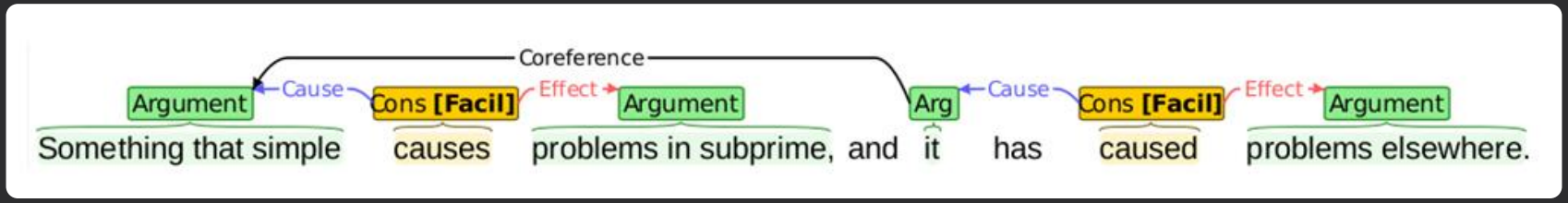


Annotators found a more fine-grained decision tree too difficult to apply.



We have annotated a small corpus with this scheme.

	Documents	Sentences	Causality annotations
New York Times Washington section (Sandhaus, 2014)	44	1407	400
Penn TreeBank WSJ	47	1542	289
2014 NLP Unshared Task in PoliInformatics (Development corpus; Smith et al., 2014)	2	384	156
Total	93	3333	845



We computed intercoder agreement between Dunietz and A4 after 3 weeks of training.

201 sentences from randomly selected documents in the NYT subcorpus

Causation types:

CONSEQUENCE	66
MOTIVATION	18
PURPOSE	4
INFERENCE	0
Total	88

Initial agreement between Dunietz and A4 was just moderate for connectives, and abysmal for causation types.

	Partial overlap:	
	Allowed	Excluded
Connectives (F_1)	0.70	0.66
Degrees (κ)	0.87	0.87
Causation types (κ)	0.25	0.29
Argument spans (F_1)	0.94	0.83
Argument labels (κ)	0.92	0.94

Very unhappy annotators!

To eliminate difficult, repetitious decision-making, we compiled a "**constructicon**."

- Constructicon:
  - Fillmore, Lee-Goldman, and Rhodes, 2012
  - Lee-Goldman and Petruck, ms.
- Our English causal language constructicon:
  - 79 lexical head words
  - 166 construction types
    - Counting *prevent* and *prevent from* as the same lexical head word but different constructions.

Connective pattern	<cause> prevents <effect> from <effect>	<enough cause> for <effect> to <effect>
Annotatable words	prevent, from	enough, for, to
WordNet verb senses	prevent.verb.01 prevent.verb.02	
Type	Verbal	Complex
Degree	INHIBIT	FACILITATE
Type restrictions	Not PURPOSE	
Example	His actions prevented disaster.	There's enough time for you to find a restroom.

Additional examples from the causal language constructicon

- For <effect> to <effect>, <cause>
- As a result, <effect>
- Enough <cause> to <effect>
- <effect> on grounds of <cause>
- <cause> is the reason to <effect>
- <effect> results from <cause>

Dunietz and a new annotator, A5, annotated a similarly-sized dataset using the constructicon.

< 1 day of training

260 sentences: annotated by Dunietz and A5

Causation types:

CONSEQUENCE	33
MOTIVATION	11
PURPOSE	21
INFERENCE	4
Total	69

A5 has a masters degree in language technologies and had no prior annotation experience.

Constructicon-based annotation improved results dramatically.

	Partial overlap:	
	Allowed	Excluded
Connectives (F_1)	0.78	0.70
Degrees (κ)	1.0	1.0
Causation types (κ)	0.82	0.80
Argument spans (F_1)	0.96	0.86
Argument labels (κ)	0.98	0.97

Annotators reported no difficulty!

Lexicography helps when, without it,
annotators must make
the same decisions repeatedly

3. Broader implications of low non-expert agreement

Expertise

Baseball players use physics, but they don't have to know physics.

What can we expect from people who speak languages but are not trained in metalinguistic awareness?

When they have trouble with our annotation schemes, we start to worry.

Is it something real that only experts are aware of?

Are we, the experts, just making things up?

What lends validity to an annotation scheme?

- Riezler (2014)
  - Reproducibility by non-experts
  - Improvement of an independent task
- Chomsky's notion of explanatory adequacy and predictive power

- This annotation scheme will be validated by independent task

Thank you for listening