The BECauSE Corpus 2.0:
Annotating Causality and Overlapping Relations

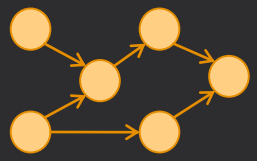
Jesse Dunietz*, Lori Levin*, & Jaime Carbonell*

LAW 2017
April 3, 2017

* Carnegie Mellon University

Recognizing causal assertions is critical to language understanding.

Ubiquitous in our mental models



Ubiquitous in language

12% of explicit discourse connectives
in Penn Discourse Treebank
(Prasad et al., 2008)

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

The prevention of FOXP3 expression
was not caused by interferences.

BECauSE draws on ideas from Construction Grammar (CxG) to annotate a wide variety of causal language.

Such swelling can **impede** breathing.	(Verbal)
They moved **because of** the schools.	(Prepositional)
Our success is **contingent on** your support.	(Adjectival)
We're running late, **so** let's move quickly.	(Conjunctive)
This **opens the way for** broader regulation.	(Multi-word expr.)
For markets **to** work, banks **can't** expect bailouts.	(Complex)

(Dunietz et al., LAW 2015)

Causal language is difficult to disentangle from overlapping semantic domains.

After a drink, she felt much better.	(Temporal)
They're **too** big **to** fail.	(Extremity)
The more I read his work, **the less** I like it.	(Correlation)
The police **let** his sister visit him briefly.	(Permission)
As voters get to know Mr. Romney, his poll numbers will rise.	(Temporal + Correlation)

Main contributions of this paper:

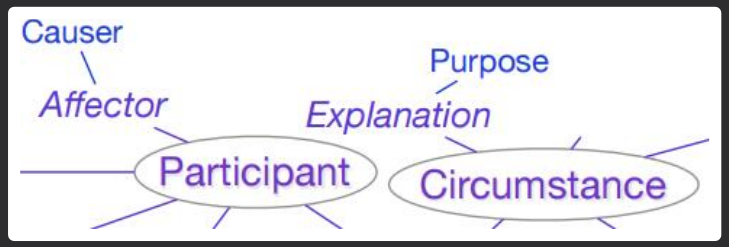
1. The BECauSE 2.0 annotation scheme including 7 overlapping relation types
2. The updated & expanded BECauSE 2.0 corpus
3. Evidence about how meanings compete for linguistic machinery

Several general-purpose schemes include some elements of causal language.

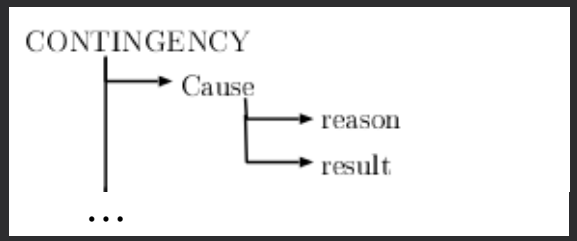
PropBank, VerbNet
(Palmer et al., 2005; Schuler, 2005)

Roleset id: prevent.01 , *stop, prevent, stopping in advance*

Prepositions
(Schneider et al., 2015, 2016)



Penn Discourse Treebank
(Prasad et al., 2008)

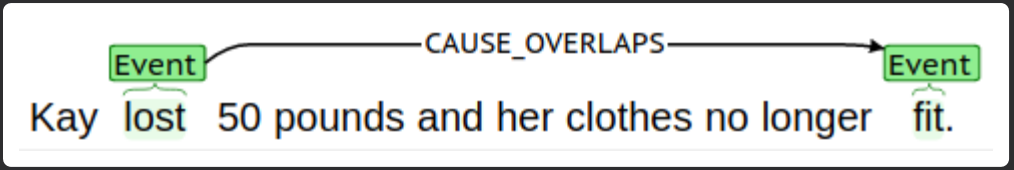


FrameNet
(Ruppenhofer et al., 2016)

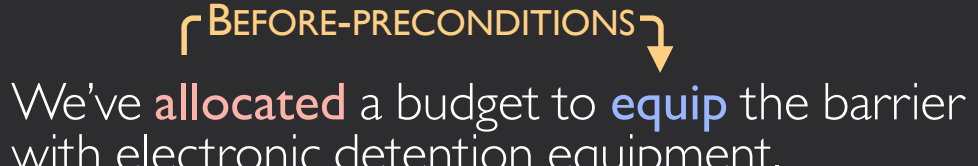
He **made** me bow to show his dominance .
[CAUSER] CAUSATION [EFFECT] [EFFECT] [PURPOSE]

Others have focused specifically on causality.

CaTeRS
(Mostafazadeh et al., 2016)

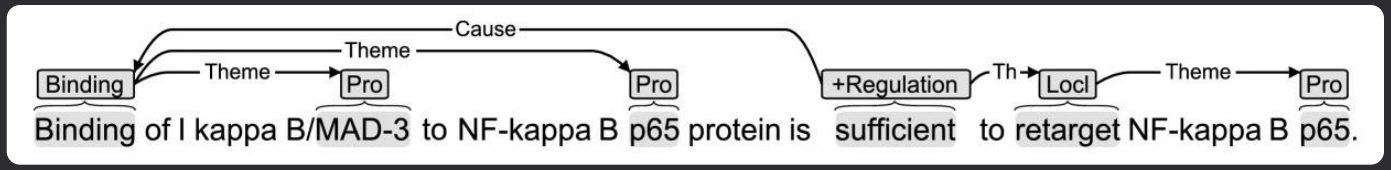


Richer Event Description
(O'Gorman et al., 2016)



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

BioCause
(Mihaila et al., 2013)



BECauSE 1.0 annotates causal language, expressed using arbitrary constructions.

Bank of **E**ffects and **Cau**ses **S**tated **E**xplicitly

Such swelling can **impede** breathing.	(Verbal)
They moved **because of** the schools.	(Prepositional)
Our success is **contingent on** your support.	(Adjectival)
We're running late, **so** let's move quickly.	(Conjunctive)
This **opens the way for** broader regulation.	(Multi-word expr.)
For markets **to** work, banks **can't** expect bailouts.	(Complex)

1. The BECauSE 2.0 annotation scheme including 7 overlapping relation types
    i. Practices retained from BECauSE 1.0
    ii. Improvements and extensions in BECauSE 2.0

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

Connective: fixed lexical cue indicating a causal construction

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
Cause: presented as producing 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.

Annotators were guided by a "constructicon."

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.

Causation can be positive or negative.

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

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

1. The BECauSE 2.0 annotation scheme including 7 overlapping relation types
    i. Practices retained from BECauSE 1.0
    ii. Improvements and extensions in BECauSE 2.0

1. The BECauSE 2.0 annotation scheme including 7 overlapping relation types
    i. Practices retained from BECauSE 1.0
    ii. Improvements and extensions in BECauSE 2.0

Update 1: Three 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~~

Update 2: Means arguments for cases with an agent and an action

My dad caused a commotion by shattering a glass.
CAUSE EFFECT MEANS

By altering immune responses,
inflammation can trigger depression.

Update 3: Overlapping semantic relations are annotated when they can be coerced to causal interpretations.

After last year's fiasco, everyone is being cautious.

MOTIVATION + TEMPORAL — ARGC — ARGE

After last year's fiasco, they've rebounded this year.

TEMPORAL — ARGC — ARGE

He won't be back until after Thanksgiving.

We annotate 7 different types of overlapping relations.

TEMPORAL	After; once; during
CORRELATION	As; the more…the more…
HYPOTHETICAL	If…then…
OBLIGATION/PERMISSION	Require; permit
CREATION/TERMINATION	Generate; eliminate
EXTREMITY/SUFFICIENCY	So…that…; sufficient…to…
CONTEXT	Without; when (non-temporal)

Annotators applied several tests to determine when an overlapping relation was also causal.

- Can the reader answer a "why" question?
- Does the cause precede the effect?
- Counterfactuality: would the effect have been just as probable without the cause?
- Ontological asymmetry: could the cause and effect be reversed?
- Can it be rephrased as "because?"

(see Grivaz, 2010)

Inter-annotator agreement remains high.

	Causal	Overlapping
Connective spans (F_1)	0.77	0.89
Relation types (κ)	0.70	0.91
Degrees (κ)	0.92	(n/a)
CAUSE/ARGC spans (%)	0.89	0.96
CAUSE/ARGC spans (Jaccard)	0.92	0.97
CAUSE/ARGC heads (%)	0.92	0.96
EFFECT/ARGE spans (%)	0.86	0.84
EFFECT/ARGE spans (Jaccard)	0.93	0.92
EFFECT/ARGE heads (%)	0.95	0.89

260 sentences; 98 causal instances; 82 overlapping relations

1. The BECauSE 2.0 annotation scheme including 7 overlapping relation types
2. The updated & expanded BECauSE 2.0 corpus
3. Evidence about how meanings compete for linguistic machinery

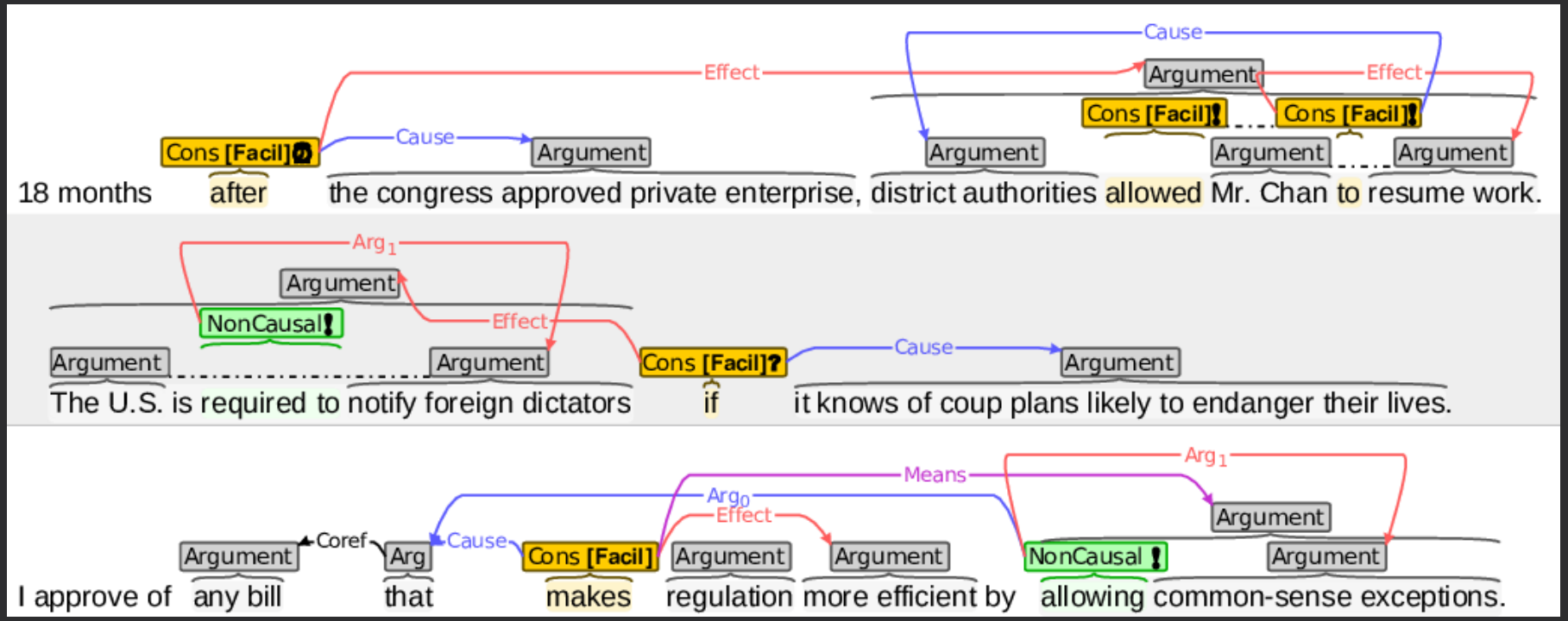
1. The BECauSE 2.0 annotation scheme including 7 overlapping relation types
2. The updated & expanded BECauSE 2.0 corpus
3. Evidence about how meanings compete for linguistic machinery

We have annotated an augmented corpus with this scheme.

	Documents	Sentences	Causal	Overlapping
New York Times Washington section (Sandhaus, 2014)	59	1924	717	519
Penn TreeBank WSJ	47	1542	534	340
2014 NLP Unshared Task in PoliInformatics (Smith et al., 2014)	3	695	326	149
Manually Annotated Sub-Corpus (Ide et al., 2010)	12	629	228	166
Total	121	4790	1805	1174

bit.ly/BECauSE

We have annotated an augmented corpus with this scheme.



1. The BECauSE 2.0 annotation scheme including 7 overlapping relation types
2. The updated & expanded BECauSE 2.0 corpus
3. Evidence about how meanings compete for linguistic machinery

1. The BECauSE 2.0 annotation scheme including 7 overlapping relation types
2. The updated & expanded BECauSE 2.0 corpus
3. Evidence about how meanings compete for linguistic machinery

Causality has thoroughly seeped into the temporal and hypothetical domains.

Of the causal expressions in the corpus:

> 14% are piggybacked on temporal relations

~7% are expressed as hypotheticals

Conditional hypotheticals don't have to be causal, but most are.

Non-causal: If he comes, he'll bring his wife.
Causal: If I told you, I'd have to kill you.

84% carry causal meaning

We seem to prefer describing causation in terms of agents' motivations.

~45% of causal instances
are MOTIVATION or PURPOSE

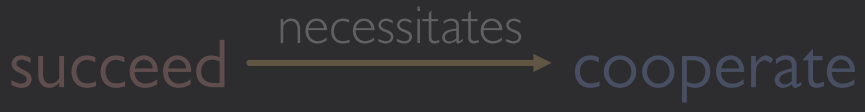
1. The BECauSE 2.0 annotation scheme including 7 overlapping relation types
2. The updated & expanded BECauSE 2.0 corpus
3. Evidence about how meanings compete for linguistic machinery

Lingering difficulties include other overlapping relations and bidirectional relationships.

Origin/destination	**toward** that goal
Topic	fuming **over** recent media reports
Component	**as part of** the liquidation
Evidentiary basis	went to war **on** bad intelligence
Having a role	**as** an American citizen
Placing in a position	**puts** us **at** risk

Lingering difficulties include other overlapping relations and bidirectional relationships.

For us to succeed, we all have to cooperate.



Main contributions of this paper:

1. The BECauSE 2.0 annotation scheme including 7 overlapping relation types
2. The updated & expanded BECauSE 2.0 corpus
3. Evidence about how meanings compete for linguistic machinery