Automatically Tagging Constructions of Causation and Their Slot-Fillers

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

April 6, 2017

[*] Carnegie Mellon University

Shallow semantic parsing tags words bearing predicates and those predicates' argument spans.

PropBank
(Palmer et al., 2005)

Even brief exposures **cause** symptoms decades later .
ARG0 CAUSE.01 ARG1

FrameNet
(Ruppenhofer et al., 2016)

… a coloured poster , **too** large for indoor display …
ITEM SUFFICIENCY SCALE ENABLED_SIT.

Varied linguistic expression is challenging for most shallow semantic parsers, as evidenced by 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)

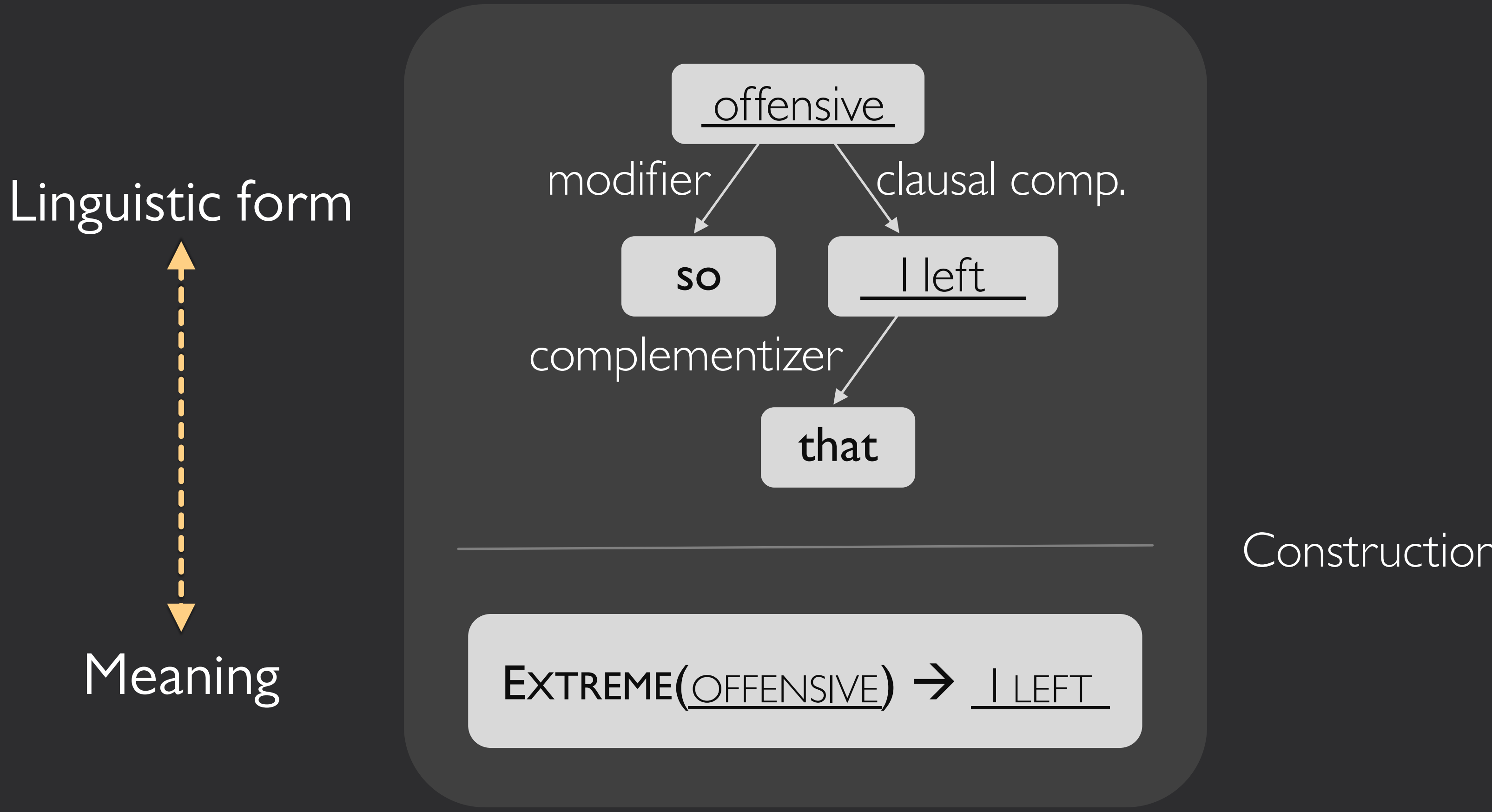
Shallow semantic parsers inherit the limitations of their representation schemes.

Semantic parser	Annotation scheme	Limitations
SENNA[1], ASSERT[2]	PropBank	Verb arguments only
End-to-end discourse parsers[3]	Penn Discourse Treebank (PDTB)[5]	Conjunctions and adverbials only
SEMAFOR[4], mateplus[6]	FrameNet	Triggers must be words or constituent MWEs

word ↔ **meaning**

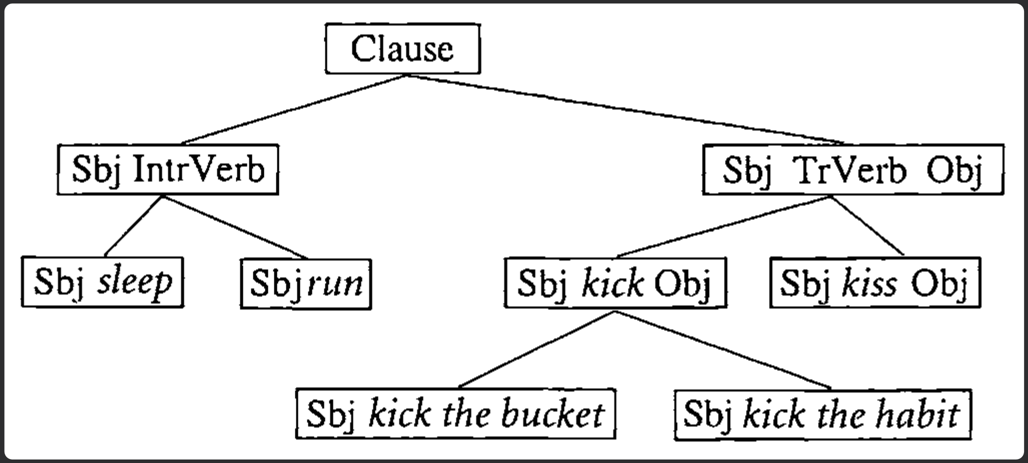
[1] Collobert et al., 2011
[2] Pradhan et al., 2004
[3] Xue et al., 2015
[4] Das et al., 2014
[5] Prasad et al., 2008
[6] Roth and Lapata, 2015

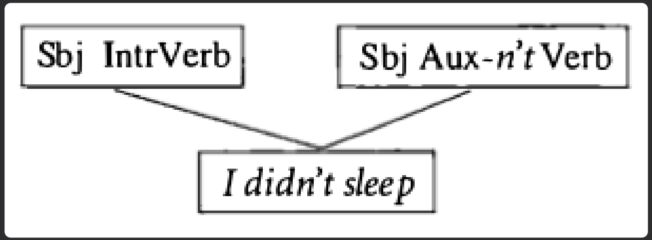
Construction Grammar (CxG) offers a way forward.



(Fillmore et al., 1988; Goldberg, 1995)

Full CxG theory entails a detailed hierarchy and complex interactions: "constructions all the way down."





(Croft, 2001)

The “constructions on top” approach

Tagging causal language
Construction recognition
POS tagging, syntactic parsing
Tokenization

Today's talk:

1. The BECauSE corpus
   of causal language
2. Causeway-L/Causeway-S: two simple systems
   for tagging causal constructions
3. Experiments & error analysis

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

(Dunietz et al., 2015)

Connective: arbitrarily complex fixed lexical cue indicating a causal construction

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

For markets **to** work, banks can't expect bailouts.

Ice cream consumption **causes** drowning.

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

Not "truly" causal

We have annotated a small corpus with this scheme.

Bank of **E**ffects and **Cau**ses **S**tated **E**xplicitly (BECauSE):

	Documents	Sentences	Causality annotations
New York Times Washington section (Sandhaus, 2014)	59	2004	529
Penn Treebank WSJ	47	1542	330
2014 NLP Unshared Task in PoliInformatics (Smith et al., 2014)	1	615	240
Total	107	4161	1099

Today's talk:

1. The BECauSE corpus
   of causal language
2. Causeway-L/Causeway-S: two simple systems
   for tagging causal constructions
3. Experiments & error analysis

Today's talk:

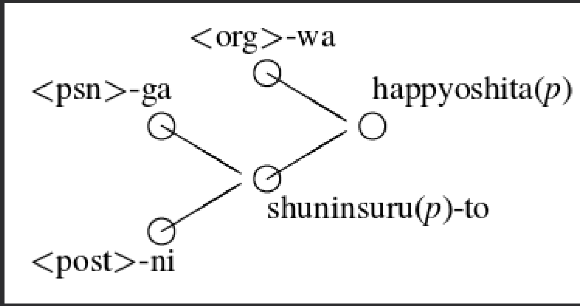
1. The BECauSE corpus of causal language
2. Causeway-L/Causeway-S: two simple systems for tagging causal constructions
3. Experiments & error analysis

Our tagging approach is rooted in information extraction patterns.

Lexical patterns
for hypernym discovery
(Hearst, 1992)

Y such as X
such Y as X…
X…and/or other Y
Y including X
Y, especially X

Dependency patterns for
general IE
(e.g., Sudo et al. 2001)

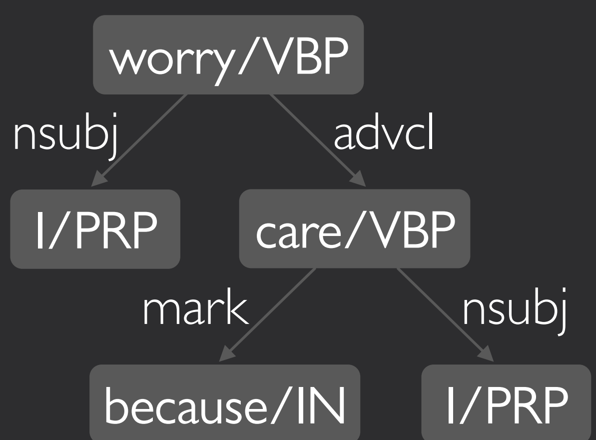


Lexico-syntactic patterns
for causal verbs
(Girju, 2003)

```
14 | hurricane | damage | ARG1+nsubj
   < cause > dobj+ARG2
11 | hiv | ads | ARG1+nsubj < cause
   > dobj+ARG2
```

Task definition: connective discovery + argument identification

I worry because
I care.



Connective discovery

Find lexical triggers of constructions

I worry **because**
I care.

Argument identification

Identify cause & effect spans for each connective
(fill slots)

Though simplified, this task is challenging.

Long tail of causal connectives
~1 per 2-3 new documents

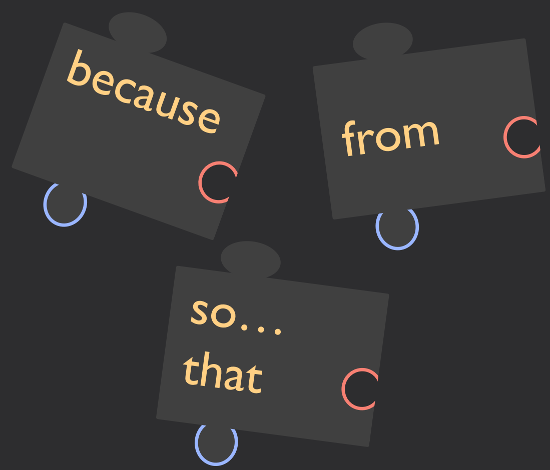
Requires sense disambiguation of connectives
e.g., "necessary **for** us **to** succeed" vs. "hard **for** me **to** do"

Combinatorial connective possibilities

1. Pattern-based connective discovery (tentative)

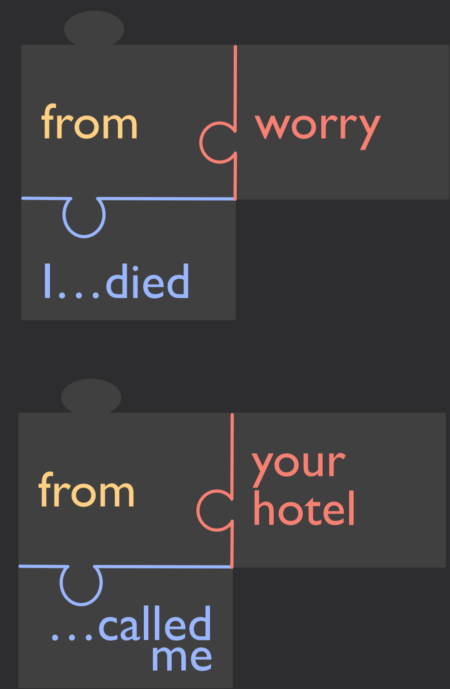
2. Argument identification (tentative)

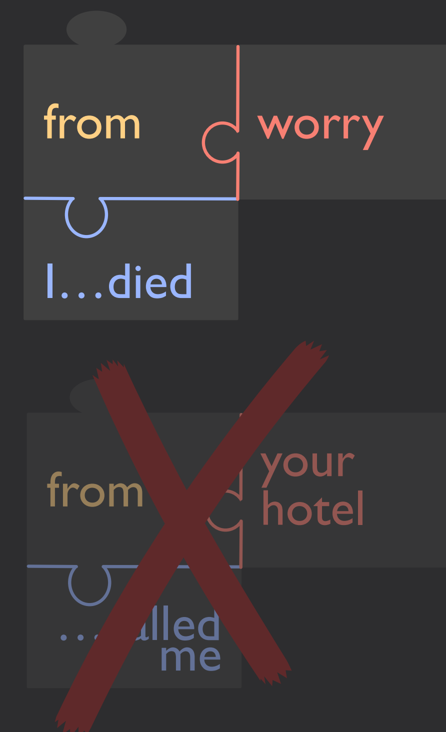
3. Statistical classifier to filter results



I nearly died **from** worry.

You could have called me **from** your hotel.





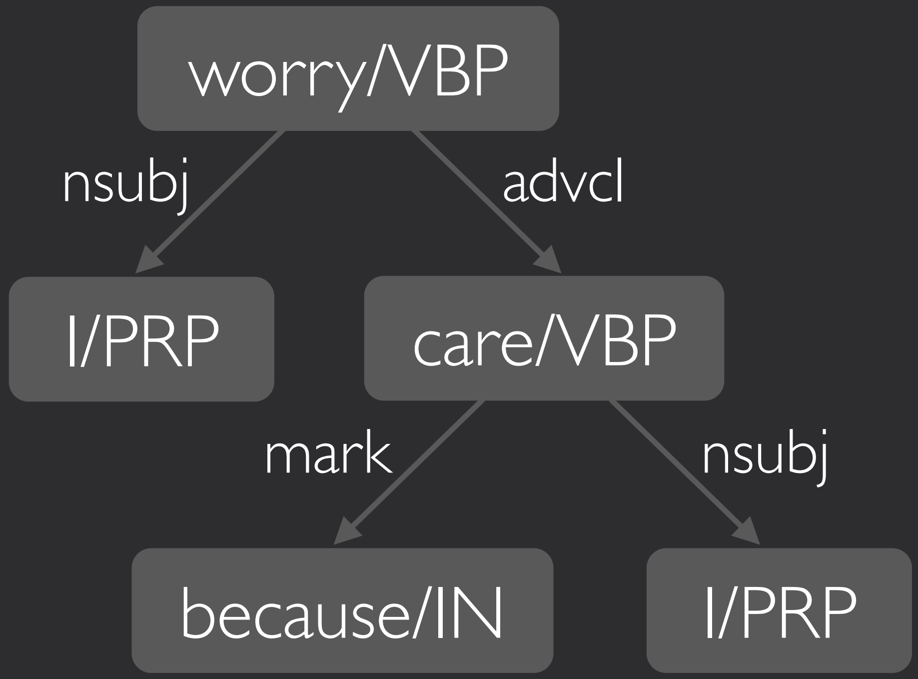
Approach 1: Syntactic patterns + head expansion rules

Approach 2: Lexical patterns + CRF sequence labeler

4. Remove duplicate connectives

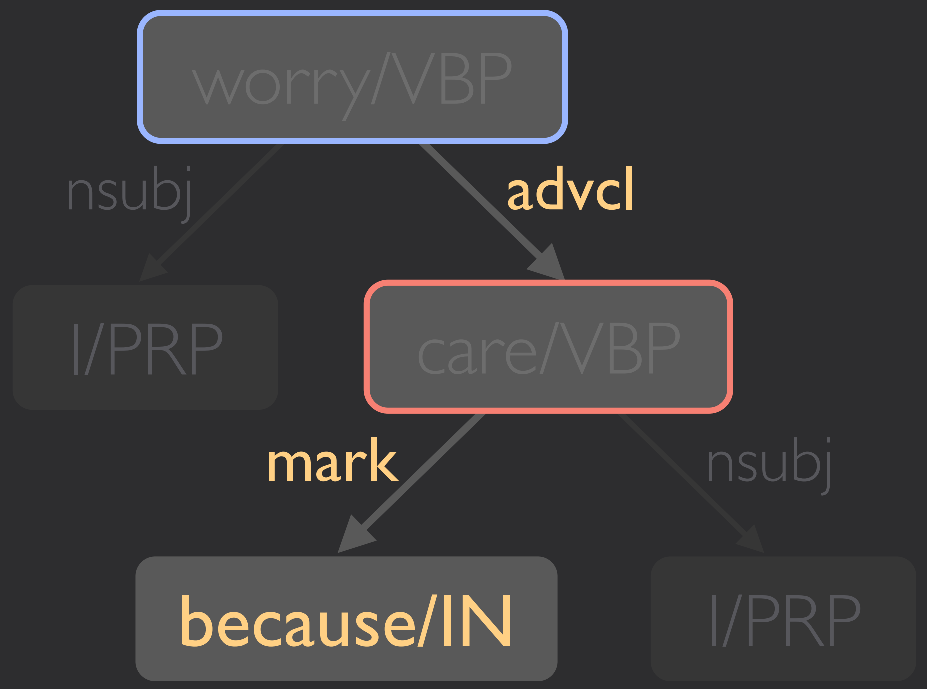
1. Causeway-L/Causeway-S: two simple systems for tagging causal constructions
   i. Causeway-S: Syntax-based pipeline
   ii. Causeway-L: Lexical pattern-based pipeline

Syntax-based connective discovery:
each construction is treated as a partially-fixed parse tree fragment.



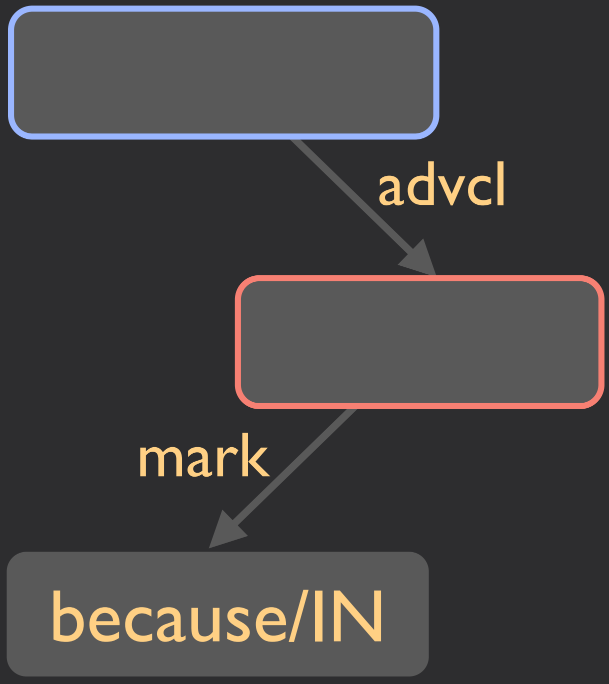
I worry because I care.

Syntax-based connective discovery:
each construction is treated as
a partially-fixed parse tree fragment.



I worry because I care.

Syntax-based connective discovery:
each construction is treated as a partially-fixed parse tree fragment.

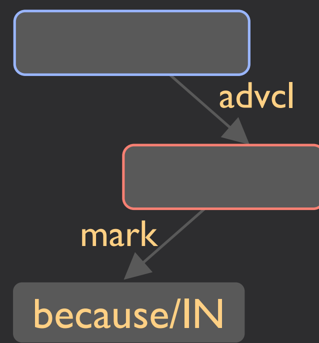


Syntax-based connective discovery:

TRegex patterns are extracted in training, and matched at test time.

Training:

I worry **because** I care.




```
(/^because_[0-9]+$/
   <2 /^IN.*/ <1 mark
   > (/.*_[0-9]+/
      <1 advcl
      > (/.*_[0-9]+/)))
```

Test:

I worry because I love you.

+

```
(/^because_[0-9]+$/
   <2 /^IN.*/ <1 mark
   > (/.*_[0-9]+/
      <1 advcl
      > (/.*_[0-9]+/)))
```

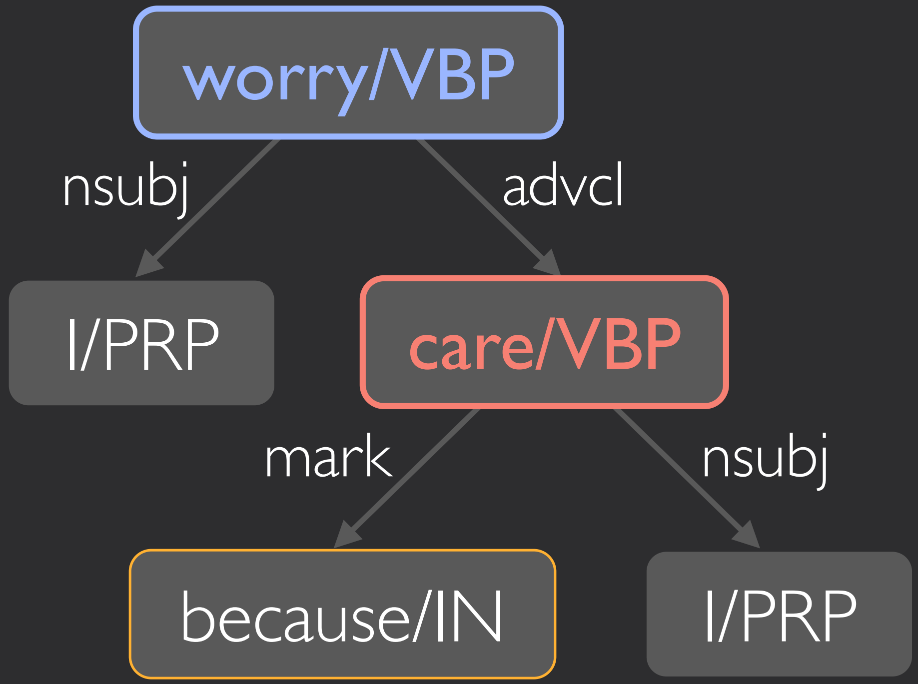
TRegex [1]

I worry **because** I love you.

[1] Levy and Andrew, 2006

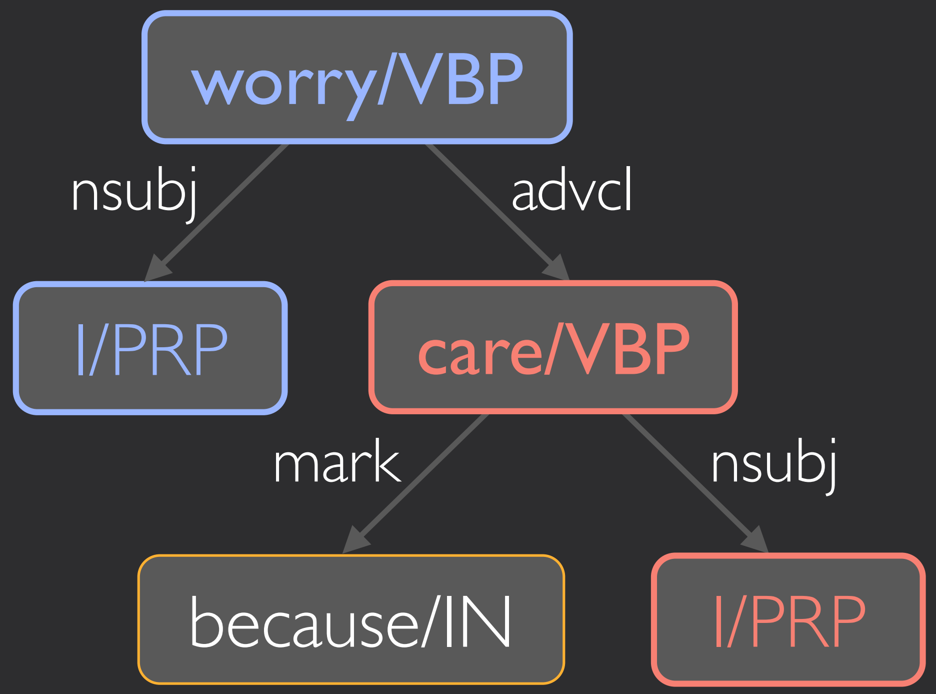
Syntax-based argument ID:

Argument heads are expanded to include most dependents.



Syntax-based argument ID:

Argument heads are expanded to include most dependents.



1. Causeway-L/Causeway-S: two simple systems for tagging causal constructions
   i. Causeway-S: Syntax-based pipeline
   ii. Causeway-L: Lexical pattern-based pipeline

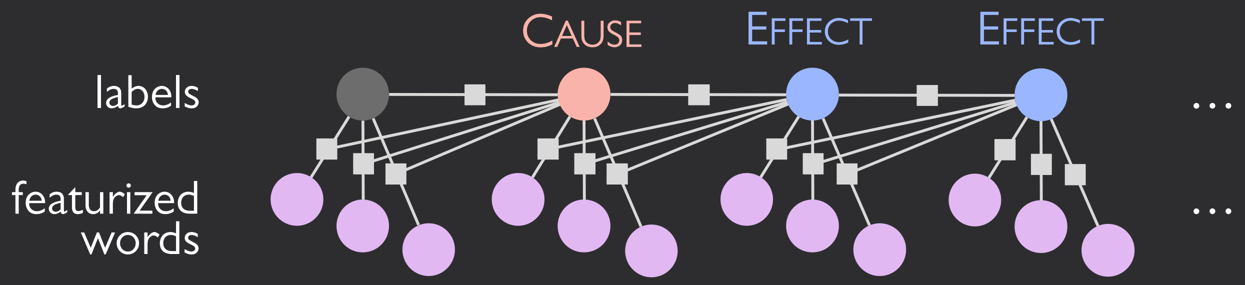
1. Causeway-L/Causeway-S: two simple systems for tagging causal constructions
   1. Causeway-S: Syntax-based pipeline
   2. Causeway-L: Lexical pattern-based pipeline

Lexical pattern-based connective discovery: **constructions are matched by regular expressions over word lemmas.**

Training: I worry **because** I care. → `(^ | )([ \ S]+ )+?(because/IN) ([ \ S]+ )+?`

Test: I worry because I love you. + `(^ | )([ \ S]+ )+?(because/IN) ([ \ S]+ )+?` regex → I worry **because** I love you.

Lexical pattern-based argument ID: **Arguments are labeled by a conditional random field.**



Features include information about:

- Word
- Connective
- Relationship between word & connective

Both approaches use a weighted soft vote of three classifiers as a filter.

Classifier 1

Global:

Classifiers 2 & 3

Connective X:

Connective Y:

Connective Z:

…

Example classifier features
(*c*=cause head, *e* = effect head):

- POS tags of *c* and *e*
- Number of words between *c* and *e*
- Domination relationship between *c* and *e*
- Matching connective pattern
- Pair of tense/aspect/modality modifier sets of *c* and *e*
- POS 1-skip-2-grams of cause and effect spans
- WordNet hypernyms

Our baseline is an argument-aware most-frequent-sense heuristic.

Connective	Parse paths to other tokens	Causal / Not causal
prevent from	nsubj, advcl	27/ 4
prevent from	nsubj, advmod	0 / 8
because of	case, case → nmod	14 / 1
…	…	…

Today's talk:

1. The BECauSE corpus
   of causal language

2. Causeway-L/Causeway-S: two simple systems
   for tagging causal constructions

3. Experiments & error analysis

Today's talk:

1. The BECauSE corpus
   of causal language
2. Causeway-L/Causeway-S: two simple systems
   for tagging causal constructions
3. Experiments & error analysis

Our results show the techniques are viable, but further work is needed.

	Connectives			Causes			Effects		
Pipeline [stages]	P	R	F_1	S_C	H_C	J_C	S_E	H_E	J_E
Causeway-S [1-2]	7.3	71.9	13.2	65.0	84.3	39.3	30.4	63.0	30.7
Causeway-S [1-4]	57.7	47.4	**51.8**	67.1	84.4	39.0	37.7	70.7	33.4
Causeway-L [1-2]	8.1	91.1	14.8	56.8	67.6	33.1	39.5	59.4	30.9
Causeway-L [1-4]	60.4	39.9	**47.9**	74.3	85.8	42.6	53.3	76.4	38.2
Baseline	88.4	21.4	**33.8**	74.1	94.7	43.7	48.4	83.3	38.4

The best performance comes from Causeway-S plus the baseline.

	Connectives			Causes			Effects		
Pipeline [stages]	P	R	F_1	S_C	H_C	J_C	S_E	H_E	J_E
Causeway-S [1-2]	7.3	71.9	13.2	65.0	84.3	39.3	30.4	63.0	30.7
Causeway-S [1-4]	57.7	47.4	51.8	67.1	84.4	39.0	37.7	70.7	33.4
Causeway-L [1-2]	8.1	91.1	14.8	56.8	67.6	33.1	39.5	59.4	30.9
Causeway-L [1-4]	60.4	39.9	47.9	74.3	85.8	42.6	53.3	76.4	38.2
Baseline	88.4	21.4	33.8	74.1	94.7	43.7	48.4	83.3	38.4
+ Causeway-S [1-4]	59.6	51.9	**55.2**	67.7	85.8	39.5	39.5	73.1	34.2
+ Causeway-L [1-4]	62.3	45.2	52.3	73.6	88.9	42.8	53.9	78.6	38.7

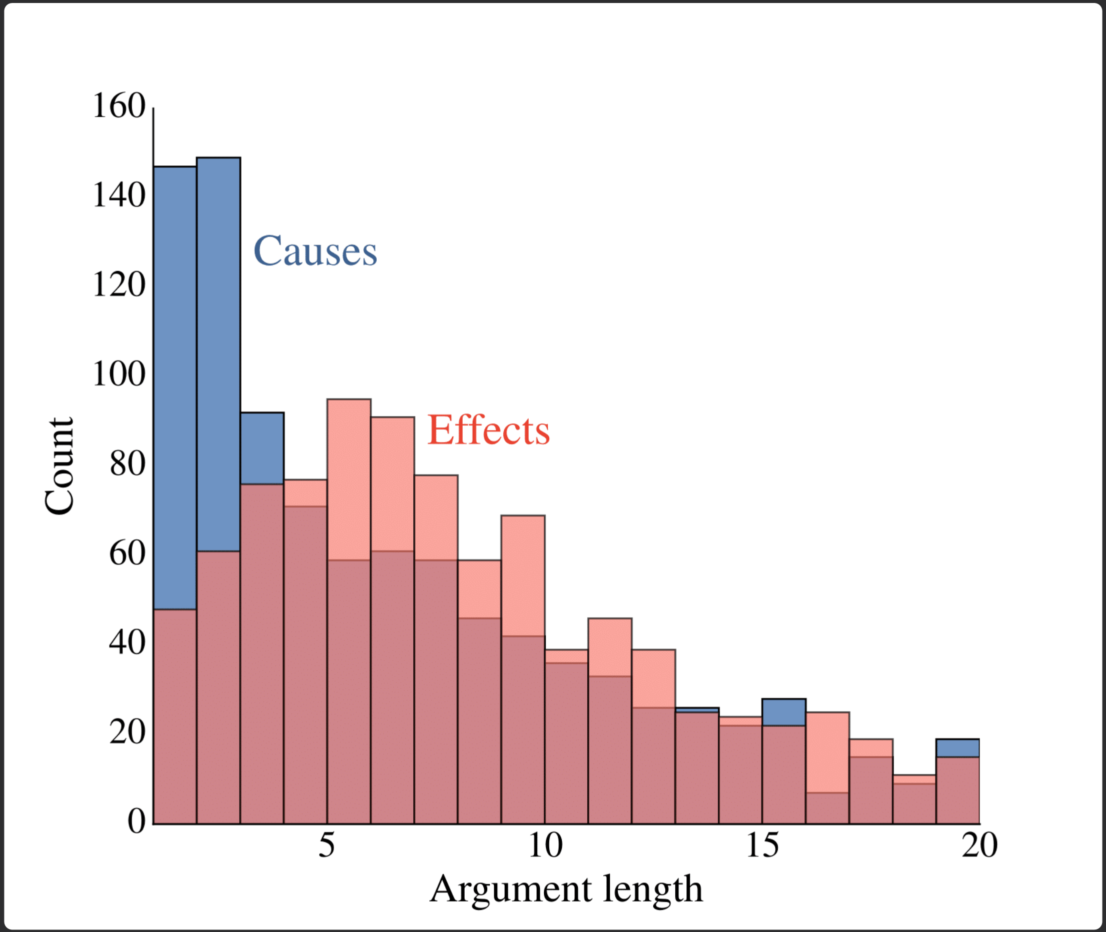
The classifier has the intended effect of balancing precision and recall for better F1.

	Connectives			Causes			Effects		
Pipeline [stages]	P	R	F_1	S_C	H_C	J_C	S_E	H_E	J_E
Causeway-S [1-2]	**7.3**	**71.9**	13.2	65.0	84.3	39.3	30.4	63.0	30.7
Causeway-S [1-4]	**57.7**	**47.4**	51.8	67.1	84.4	39.0	37.7	70.7	33.4
Causeway-L [1-2]	**8.1**	**91.1**	14.8	56.8	67.6	33.1	39.5	59.4	30.9
Causeway-L [1-4]	**60.4**	**39.9**	47.9	74.3	85.8	42.6	53.3	76.4	38.2
Baseline	88.4	21.4	33.8	74.1	94.7	43.7	48.4	83.3	38.4
+ Causeway-S [1-4]	59.6	51.9	55.2	67.7	85.8	39.5	39.5	73.1	34.2
+ Causeway-L [1-4]	62.3	45.2	52.3	73.6	88.9	42.8	53.9	78.6	38.7

Both systems score well on spans/heads, but effects seem to be harder than causes.

	Connectives			Causes			Effects		
Pipeline [stages]	P	R	F_1	S_C	H_C	J_C	S_E	H_E	J_E
Causeway-S [1-2]	7.3	71.9	13.2	65.0	84.3	39.3	30.4	63.0	30.7
Causeway-S [1-4]	57.7	47.4	51.8	**67.1**	**84.4**	**39.0**	**37.7**	**70.7**	**33.4**
Causeway-L [1-2]	8.1	91.1	14.8	56.8	67.6	33.1	39.5	59.4	30.9
Causeway-L [1-4]	60.4	39.9	47.9	**74.3**	**85.8**	**42.6**	**53.3**	**76.4**	**38.2**
Baseline	88.4	21.4	33.8	74.1	94.7	43.7	48.4	83.3	38.4
+ Causeway-S [1-4]	59.6	51.9	55.2	67.7	85.8	39.5	39.5	73.1	34.2
+ Causeway-L [1-4]	62.3	45.2	52.3	73.6	88.9	42.8	53.9	78.6	38.7

The culprit seems to be the difference in argument length.



Causeway-S improves significantly with gold-standard parses.

	Connectives			Causes			Effects		
Pipeline [stages]	P	R	F_1	S_C	H_C	J_C	S_E	H_E	J_E
				Automatically parsed					
Causeway-S [1-2]	14.9	73.3	**24.7**	63.6	90.9	40.3	18.1	72.7	25.3
Causeway-S [1-4]	54.7	40.2	**45.7**	78.7	98.4	44.6	46.0	78.4	36.7
Causeway-L [1-2]	9.3	84.6	16.7	59.4	68.5	33.1	43.2	62.1	31.8
Causeway-L [1-4]	52.4	37.2	43.2	72.9	84.5	40.0	52.3	73.4	35.7
				Gold-standard parses					
Causeway-S [1-2]	10.2	70.6	**17.7**	79.4	98.1	45.7	52.8	90.2	41.3
Causeway-S [1-4]	62.7	51.6	**56.0**	80.2	96.4	45.6	59.0	92.7	43.4
Causeway-L [1-2]	9.1	84.1	16.4	57.8	68.2	33.3	53.0	68.0	34.4
Causeway-L [1-4]	56.4	37.9	44.3	77.0	85.3	41.8	67.2	83.4	40.4

Contributions of this paper:

1. The BECauSE corpus
   covers many instances of causal language
   that other schemes do not
2. Causeway-L/Causeway-S: **two simple systems for tagging causal constructions**
3. Experiments & error analysis
   show that the systems achieve moderate performance,
   but more work is needed to filter false positives
   and to correctly tag long effect spans