# Construction Detection in a Conventional NLP Pipeline

Jesse Dunietz[1], Lori Levin[1], & Miriam R. L. Petruck[2]

March 28, 2017

[1] Carnegie Mellon University

[2] University of California, Berkeley

Frame-semantic parsing
identifies the words that evoke frames
and the argument spans for those frames.

**FrameNet**
(Ruppenhofer et al., 2016)

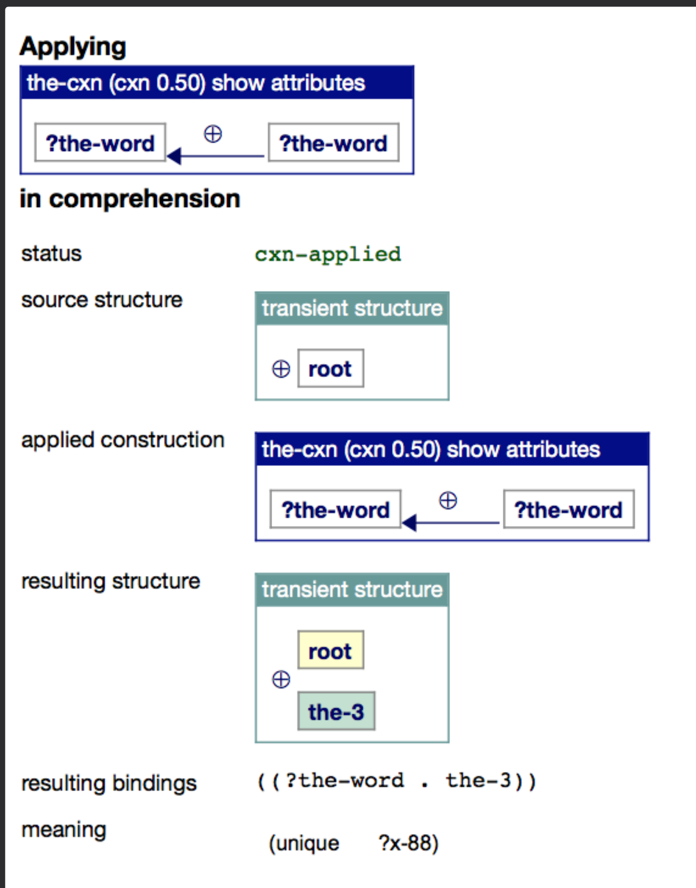
| He | **made** | me | bow | to show his dominance | . |
|---|---|---|---|---|---|
| CAUSER | CAUSATION | EFFECT | EFFECT | PURPOSE | |

# One domain that epitomizes the problem is causal language.

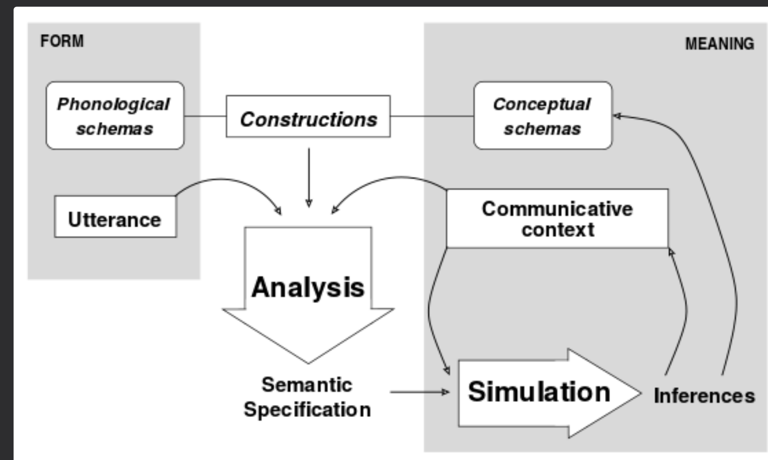
| | |
|---|---|
| This **opens the way for** broader regulation. | (Multi-word expr.) |
| **For** markets **to** work, banks **can't** expect bailouts. | (Complex) |
| Judy's comments were **so** offensive **that** I left. | (Complex) |
| These flavors complex**ify** the taste of the fruit. | (Morphological) |

# There are some powerful CxG-based NLP tools, but none can yet robustly parse, e.g., the NYT.

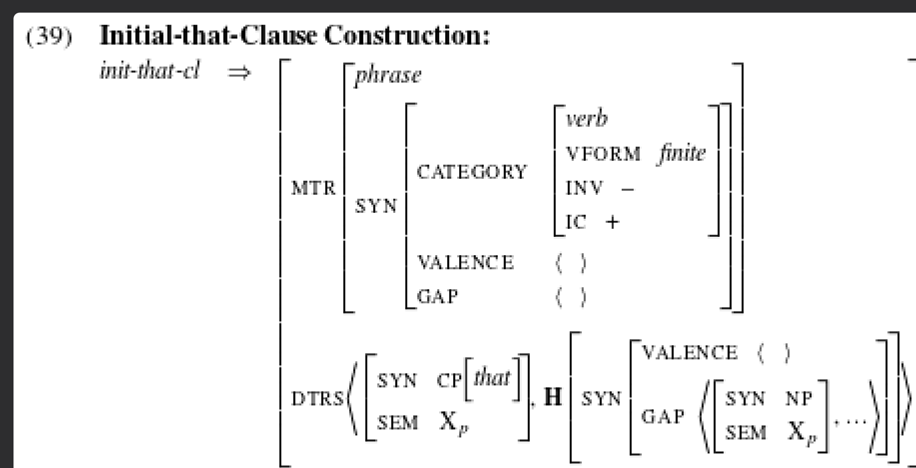
## Fluid Construction Grammar




## Embodied Construction Grammar



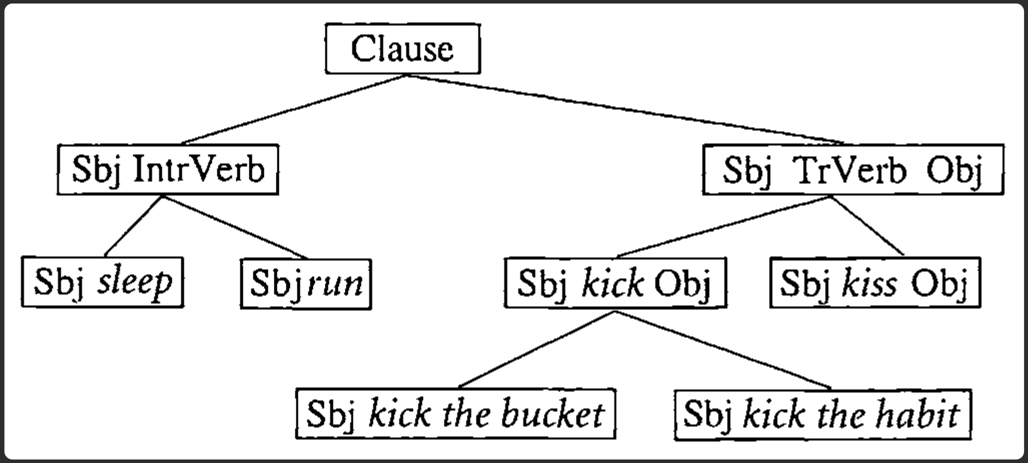

## Sign-Based Construction Grammar

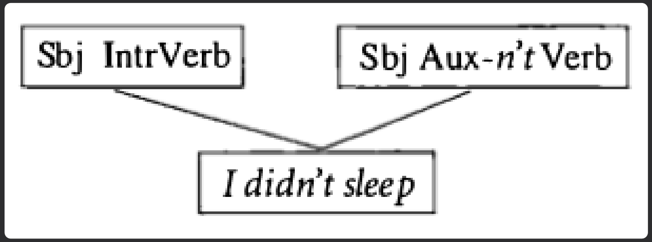


An intermediate step:
apply the key insights of CxG
**on top of** conventional NLP.

1. Morphemes, words, MWEs, and grammar are all on the same spectrum of linguistic forms.
2. Any aspect or combination of those forms is equally capable of being mapped to meanings.

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




(Croft, 2001)

# The "constructions on top" approach

Tagging causative frames

**Construction recognition**

POS tagging, syntactic parsing

Tokenization

# The “constructions on top” approach represents low-hanging fruit for CxG in NLP and lays the groundwork for further uptake.

Constructional analysis

- Tagging causative frames
- **Construction recognition**
- POS tagging, syntactic parsing
- Tokenization

# Today's talk:

1. The "constructions on top" approach
2. The BECauSE corpus
   of causal language
3. Causeway: a simple system
   for tagging causal constructions
4. Lessons learned

# FrameNet currently represents a relatively small number of non-lexical constructions.

The crane could buckle **due to** the heat .
⌊EFFECT⌋ CAUSATION ⌊ CAUSE ⌋

This **opens the way for** broader regulation .
⌊POTENTIAL_HINDRANCE⌋ PREVENTING_OR_LETTING ⌊ EVENT ⌋

# FrameNet currently represents a relatively small number of non-lexical constructions.

The crane could buckle **due to** the heat .
⌊EFFECT⌋ CAUSATION ⌊ CAUSE ⌋

This **opens the way for** broader regulation .
⌊POTENTIAL_HINDRANCE⌋ PREVENTING_OR_LETTING ⌊ EVENT ⌋

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

???

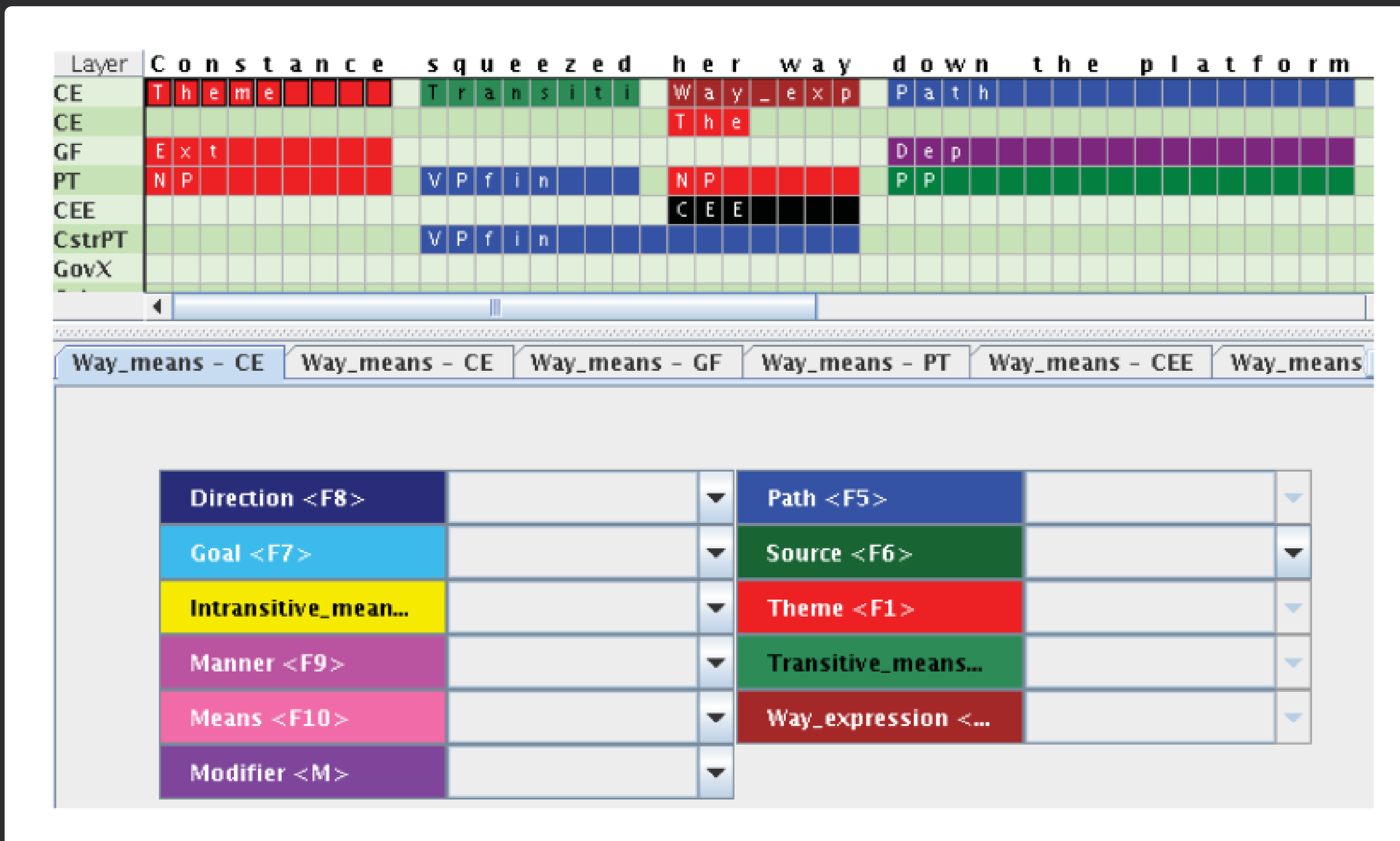
# The FrameNet Constructicon project has begun to fill this gap.

$\{^{\text{Motion verb}}\ [^{\text{Verb}}\ ]\ [^{\text{PossNP}}\ ]\ \}$

| | |
|---|---|
| Name | *verb-way* |
| M | Verb, evokes the `Motion` frame. Requires at least one SOURCE, PATH, or GOAL-related argument. |
| D1 | A verb with at least an ACTOR argument; any other arguments are suppressed and existentially interpreted. |
| D2 | An NP, headed by *way* and with a possessive pronoun coindexed to D1's external argument; able to be modified by ACTOR-modifying or PATH-modifying expressions. |
| Interpretation | the meaning of D1 (the verb) is incorporated into the `Motion` frame as a MANNER or MEANS of motion. This is clear in many cases but the distinction is not always clear. |

(Fillmore et al., 2012)

# The FrameNet Constructicon project has begun to fill this gap.




(Fillmore et al., 2012)

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: construction-evoking element 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

# Connective: arbitrarily complex construction-evoking element indicating a causal relationship

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

This **opens the way for** broader regulation.

# We have annotated a small corpus with this scheme.

**B**ank 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 |

# Constructional phenomena are frequent in the corpus.

16% of contiguous connective types…

62% of non-contiguous connective types…

Up to 20% of annotated instances…

…can't be represented as FN lexical units.

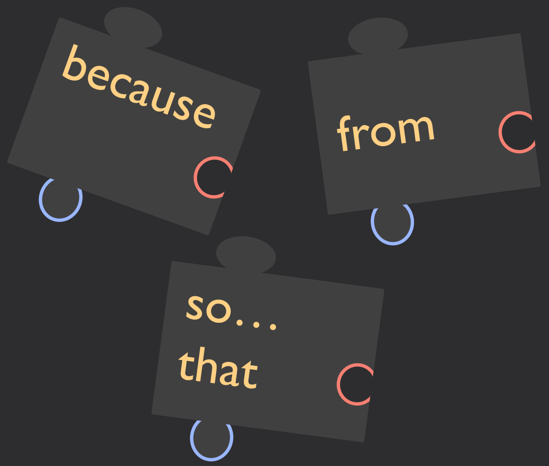
# Today’s talk:

1. The “constructions on top” approach
2. The BECauSE corpus of causal language
3. Causeway: a simple system for tagging causal constructions
4. Lessons learned

# Today's talk:

1. The "constructions on top" approach
2. The BECauSE corpus
   of causal language
3. Causeway: a simple system
   for tagging causal constructions
4. Lessons learned

# 1. Pattern-based connective discovery

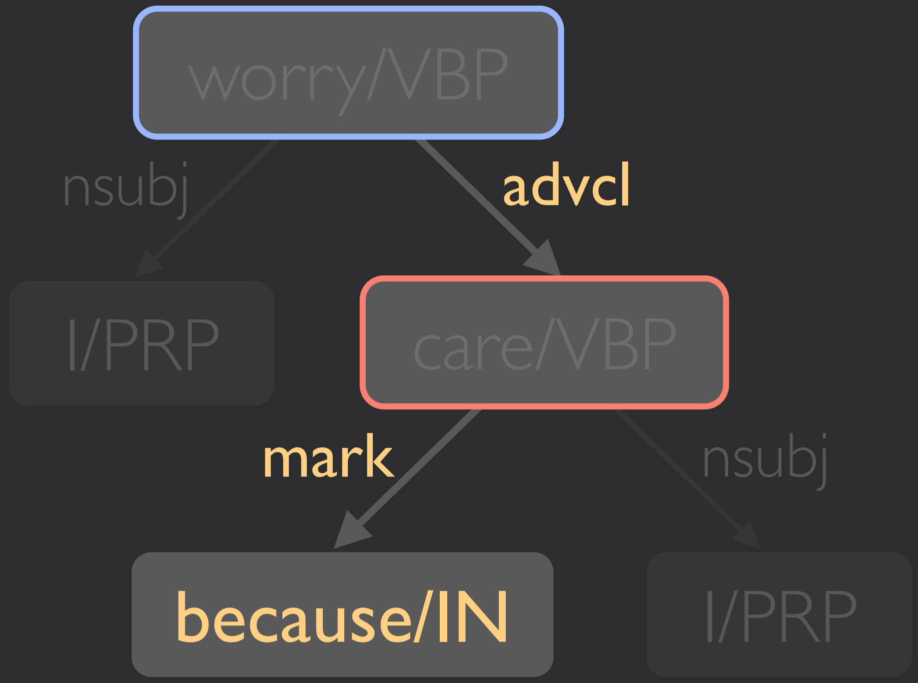



I nearly died **from** worry.

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

(Dunietz et al., 2017)

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

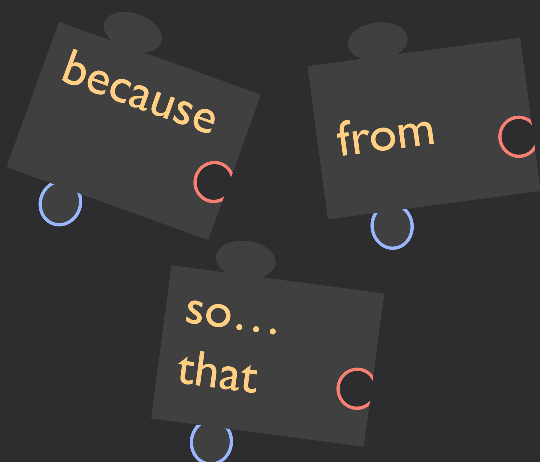



I worry because I care.

1. Pattern-based connective discovery (tentative)

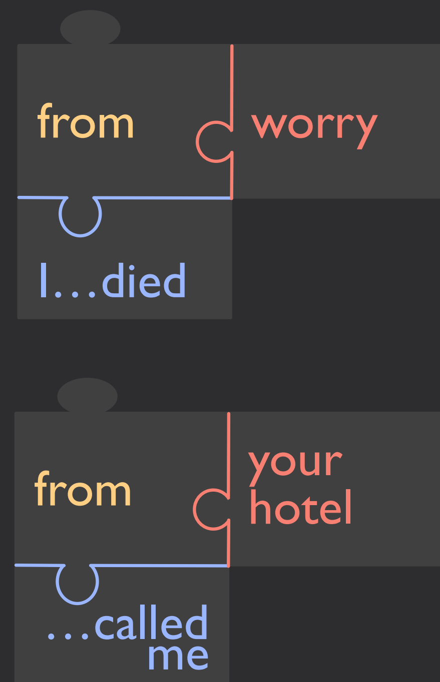
2. Argument identification (tentative)

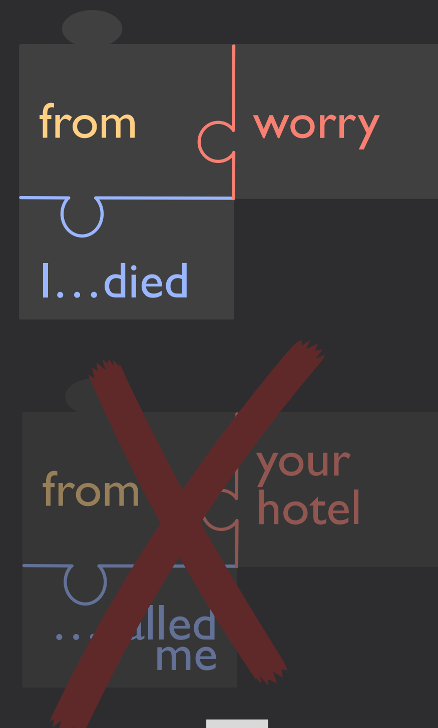
3. Statistical classifier to filter results




I nearly died **from** worry.

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




4. Remove duplicate connectives

(Dunietz et al., 2017)

# 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 [1-2] | 7.3 | 71.9 | 13.2 | 65.0 | 84.3 | 39.3 | 30.4 | 63.0 | 30.7 |
| Causeway [1-2] + MFS | 40.1 | 37.9 | 38.6 | 71.0 | 87.6 | 42.0 | 34.3 | 64.4 | 31.9 |
| Causeway [1-2] + MFS + 3b | 60.9 | 36.2 | 45.1 | 75.1 | 92.3 | 42.9 | 40.7 | 75.2 | 35.8 |
| Causeway [1-3] | 51.9 | 47.6 | **49.4** | 68.7 | 86.9 | 39.9 | 38.0 | 72.5 | 34.1 |
| Cauesway [1-3] + MFS | 57.7 | 47.4 | **51.8** | 67.1 | 84.4 | 39.0 | 37.7 | 70.7 | 33.4 |
| Baseline | 88.4 | 21.4 | **33.8** | 74.1 | 94.7 | 43.7 | 48.4 | 83.3 | 38.4 |
| + Causeway (full) | 59.6 | 51.9 | 55.2 | 67.7 | 85.8 | 39.5 | 39.5 | 73.1 | 34.2 |

# Today’s talk:

1. The “constructions on top” approach
2. The BECauSE corpus
   of causal language
3. Causeway: a simple system
   for tagging causal constructions
4. Lessons learned

# Today's talk:

1. The "constructions on top" approach
2. The BECauSE corpus
   of causal language
3. Causeway: a simple system
   for tagging causal constructions
4. Lessons learned

# Lesson 1:
# Constructions are hard to individuate.

too sweet to eat

too sweet for me to eat

sweet enough to eat

sweet enough for me to eat

sweet enough that I can eat it

so sweet that I can't eat it

so sweet I can't eat it

# Lesson 1:
# Constructions are hard to individuate.

too sweet to eat

too sweet for me to eat

→ sweet **enough to** eat

→ sweet **enough for** me **to** eat

→ sweet **enough that** I can eat it

so sweet that I can't eat it

so sweet I can't eat it

# Lesson 1: Constructions are hard to individuate.

→ too sweet to eat

→ too sweet **for me** to eat

→ sweet enough to eat

→ sweet enough **for me** to eat

sweet enough that I can eat it

so sweet that I can't eat it

so sweet I can't eat it

# Lesson 1:
# Constructions are hard to individuate.

→ **too** sweet **to** eat

→ **too** sweet **for** me **to** eat

→ sweet **enough** **to** eat

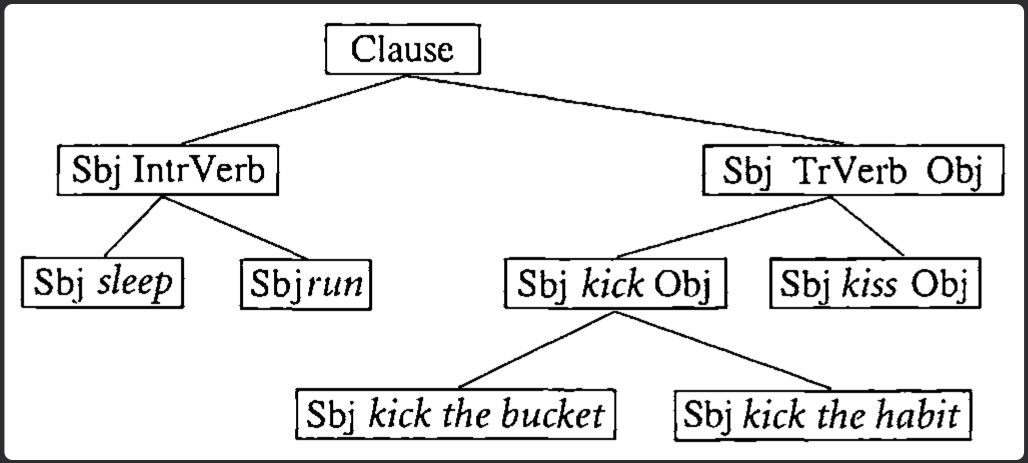
→ sweet **enough** **for** me **to** eat

sweet **enough** **that** I can eat it

**so** sweet **that** I can't eat it

**so** sweet I can't eat it

# The usual answer: construct a hierarchy to "capture the generalization."




…but when does a product of constructions become its own construction?

"As a result?"

For "constructions on top," decide where to draw the lines for computational convenience.

# Lesson 2:
# Constructions can simultaneously carry multiple semantic relations.

My head was hurting, but taking a drink **made** it feel much better.

My head was hurting, but **after** I took a drink it felt much better. (Temporal)

**If** you touch it, it will fall over. (Hypothetical)

These reports **create** a perception of higher risk. (Inception/termination)

# Contributions:

1. The “constructions on top” approach
2. The BECauSE corpus
   of causal language
3. Causeway: a simple system
   for tagging causal constructions
4. Lessons learned