Microsoft Research

Each year Microsoft Research hosts hundreds of influential speakers from around the world including leading scientists, renowned experts in technology, book authors, and leading academics, and makes videos of these lectures freely available.

2016 © Microsoft Corporation. All rights reserved.
Annotation Artifacts in Natural Language Inference Data
NW-NLP 2018
Suchin Gururangan & Swabha Swayamdipta
joint work with Omer Levy, Roy Schwartz, Sam Bowman, and Noah Smith

UW NLP
ARK
AT2
NYU DATA SCIENCE
Everyday NLP

Text → Model → Output
Everyday NLP

Annotators

Text -> Model -> Output
Everyday NLP

Annotators

Human Biases
Beliefs about the world
Personal Motives

Output

Text

Model
The Natural Language Inference (NLI) Task

Given a premise $p$ and a hypothesis $h$, the goal is to determine whether $p$ entails $h$. 
The Natural Language Inference (NLI) Task

Given a premise $p$ and a hypothesis $h$ the goal is to determine whether $p$ entails $h$.

Premise: The NW-NLP workshop is being held at Microsoft HQ.
The Natural Language Inference (NLI) Task

Given a premise $p$ and a hypothesis $h$ the goal is to determine whether $p$ entails $h$.

Premise: The NW-NLP workshop is being held at Microsoft HQ.

| Entailment | $h$ is definitely true given $p$ | The workshop is in Redmond. |
The Natural Language Inference (NLI) Task

Given a premise $p$ and a hypothesis $h$ the goal is to determine whether $p$ entails $h$.

**Premise:** The NW-NLP workshop is being held at Microsoft HQ.

<table>
<thead>
<tr>
<th>Entailment</th>
<th>$h$ is definitely true given $p$</th>
<th>The workshop is in Redmond.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contradiction</td>
<td>$h$ is definitely <strong>not</strong> true given $p$</td>
<td>Amazon is hosting the workshop.</td>
</tr>
</tbody>
</table>
The Natural Language Inference (NLI) Task

Given a premise \( p \) and a hypothesis \( h \) the goal is to determine whether \( p \) entails \( h \).

**Premise:** The NW-NLP workshop is being held at Microsoft HQ.

<table>
<thead>
<tr>
<th>Entailment</th>
<th>( h ) is definitely true given ( p )</th>
<th>The workshop is in Redmond.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contradiction</td>
<td>( h ) is definitely <strong>not</strong> true given ( p )</td>
<td>Amazon is hosting the workshop.</td>
</tr>
<tr>
<td>Neutral</td>
<td>( h ) might be true given ( p )</td>
<td>Satya Nadella is attending the workshop.</td>
</tr>
</tbody>
</table>
Strong performance on NLI tasks

<table>
<thead>
<tr>
<th>Authors</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parikh et al 2016</td>
<td>84.7 (SNLI)</td>
</tr>
<tr>
<td>Chen et al 2017</td>
<td>74.9 (MNLI)</td>
</tr>
<tr>
<td>Gong et al 2017</td>
<td>78.4 (MNLI)</td>
</tr>
</tbody>
</table>

SNLI [Bowman et al., 2015]
multiNLI [Williams et al., 2017]
Most NLI models

E  C  N

Premise  Classifier  Hypothesis
A surprising finding...

Classifier

Hypothesis

E C N
A surprising finding...

E  C  N

Classifier

Hypothesis

<table>
<thead>
<tr>
<th></th>
<th>SNLI</th>
<th>MultiNLI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyp-only (fastText*)</td>
<td>66.7</td>
<td>55.4</td>
</tr>
<tr>
<td>Majority Baseline</td>
<td>34.3</td>
<td>35.4</td>
</tr>
</tbody>
</table>

*Joulin et al., 2016*
A surprising finding...

We don’t need to learn entailment to perform ~2x above baselines

<table>
<thead>
<tr>
<th></th>
<th>SNLI</th>
<th>MultiNLI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyp-only</td>
<td>66.7</td>
<td>55.4</td>
</tr>
<tr>
<td>Majority Baseline</td>
<td>34.3</td>
<td>35.4</td>
</tr>
</tbody>
</table>

*Joulin et al., 2016*
Annotation Artifacts in NLI data
Annotation Artifacts in NLI data

- We show that elicited annotations induce lexical clues (annotation artifacts) that reveal the entailment class.
Annotation Artifacts in NLI data

- We show that **elicited annotations** induce lexical clues (**annotation artifacts**) that reveal the entailment class.

- We show that supervised models **exploit** annotation artifacts to make their predictions.
Annotation Artifacts in NLI data

- We show that **elicited annotations** induce lexical clues (**annotation artifacts**) that reveal the entailment class.

- We show that supervised models **exploit** annotation artifacts to make their predictions.

- We present the wider implications for model evaluation and task design.
Argument 1:
There exist annotation artifacts in NLI hypotheses that heavily correlate with the entailment class.
Building SNLI & multiNLI
Building SNLI & multiNLI

“A white dog is running through the snow”

“3 million in relief in the form of emergency housing.”
Building SNLI & multiNLI

“A white dog is running through the snow”

“8 million in relief in the form of emergency housing.”

MTurk
Building SNLI & multiNLI

“A white dog is running through the snow”

“8 million in relief in the form of emergency housing.”

MTurk

Entailment

Contradiction

Neutral
Building SNLI & multiNLI

Emphasis on scale
~1M high-agreement examples

"A white dog is running through the snow"

"8 million in relief in the form of emergency housing."

MTurk

Entailment

Contradiction

Neutral
Characterizing annotation artifacts
Entailment Artifacts
Entailment Artifacts

Generalizations

Premise: Some men and boys are playing frisbee in a grassy area.

Entailment: People play frisbee outdoors.
Entailment Artifacts

Generalizations

Premise: Some men and boys are playing frisbee in a grassy area.

Entailment: People play frisbee outdoors.

Shortening

Premise: A person in a red shirt is mowing the grass with a green riding mower.

Entailment: A person in red is cutting the grass on a riding mower.
Neutral Artifacts

Purpose Clauses

Premise: Two dogs are running through a field.

Neutral: Some puppies are running to catch a stick.
Neutral Artifacts

Purpose Clauses

Premise: Two dogs are running through a field.

Neutral: Some puppies are running to catch a stick.

Modifiers

Premise: A middle-aged man works under the engine of a train on rail tracks

Neutral: A man is doing work on a black Amtrak train.
Contradiction Artifacts

Negation

Premise: Older man with white hair and a red cap painting the golden gate bridge on the shore with the golden gate bridge in the distance.

Contradiction: Nobody has a cap.
Contradiction Artifacts

Negation

Premise: Older man with white hair and a red cap painting the golden gate bridge on the shore with the golden gate bridge in the distance.

Contradiction: Nobody has a cap.

Cats

Premise: Three dogs racing on racetrack

Contradiction: Three cats race on a track.
Argument 2:
Supervised models exploit annotation artifacts in the NLI task
Partitioning NLI benchmarks

Test Hypothesis \rightarrow \text{Hyp-only classifier} \rightarrow \begin{align*}
\text{Easy Example} & \quad \text{[green checkmark]} \\
\text{Hard Example} & \quad \text{[red cross]} 
\end{align*}
## Revisiting the benchmark NLI models

<table>
<thead>
<tr>
<th>Authors</th>
<th>Model</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parikh et al 2016</td>
<td>Decomposable Attention Model</td>
<td>84.7 (SNLI)</td>
</tr>
<tr>
<td>Chen et al 2017</td>
<td>Enhanced Sequential Inference Model</td>
<td>74.9 (MNLI)</td>
</tr>
<tr>
<td>Gong et al 2017</td>
<td>Densely Interactive Inference Network</td>
<td>78.4 (MNLI)</td>
</tr>
</tbody>
</table>
Easy examples drive NLI performance
Easy examples drive NLI performance

![SNLI bar chart showing accuracy for DAM, ESIM, and DIIN models. The chart compares Full, Hard, and Easy examples. DAM and DIIN models show similar performance, while ESIM has a slightly lower accuracy.](image-url)
Easy examples drive NLI performance

SNLI

Accuracy (%)

- DAM
- ESIM
- DIIN
Easy examples drive NLI performance

SNLI

Accuracy (%)

50  62.5  75  87.5  100

Model

DAM  ESIM  DIIN

Full  Hard  Easy
Closing thoughts
Annotation Artifacts as Sampling Bias

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Proportion of NLI data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose Clause</td>
<td></td>
</tr>
<tr>
<td>Generalization</td>
<td></td>
</tr>
<tr>
<td>Negation</td>
<td></td>
</tr>
<tr>
<td>Cats</td>
<td></td>
</tr>
<tr>
<td>Approximations</td>
<td></td>
</tr>
<tr>
<td>Modifiers</td>
<td></td>
</tr>
<tr>
<td>Inactivity</td>
<td></td>
</tr>
</tbody>
</table>
Lexically-driven models readily exploit artifacts
Lexically-driven models readily exploit artifacts
Lexically-driven models readily exploit artifacts
This is not just about NLI!

- CNN/DailyMail [Chen et al., 2017]
- VQA [Jabri et al., 2016]
- SQuAD [Jia and Liang et al., 2017]
- RoC Story [Schwartz et al., 2017, Cai et al., 2017]
- RTE [Levy et al., 2015]
Progress in NLI has been overestimated

- How do we direct annotators to avoid simple heuristics when generating examples?
- Can we better employ real-time rewards and batch reviews of annotated examples?
- Are there alternatives to human elicitation to building entailment datasets?
We released the **Hard benchmarks** on Kaggle and SNLI website.
Thanks!

@ssgrn

https://arxiv.org/abs/1803.02324
Simulating Action Dynamics with Neural Process Networks

Antoine Bosselut, Omer Levy, Ari Holtzman, Corin Ennis, Dieter Fox, Yejin Choi

Accepted to ICLR 2018
Making Apple Pie

Get some apples

Eat the pie!
Making Apple Pie

1. Get some apples
2. Saute the apples
3. Cut the apples
4. Set in pie sheet
5. Bake for 1 hr
6. Eat the pie!
Important Details

apples: cut
apples: cut
apples: cut
apples: cooked
Important Details

apples: cut
apples: cut
apples: cooked
pie = apples + sheet
Important Details

apples: cut
apples: cut
apples: cooked
pie = apples + sheet
pie: cooked
Recurrent Neural Networks

(Schmidhuber et al., 1997; Cho et al., 2014; Suskever et al., 2014)
Memory Networks

(Henaff et al., 2016)
The problem with **context representations** is they don’t actually track the **state of the world**.
-tracking the **state of the world** means identifying and reasoning about **entities** and the **state changes** they experience
Challenges

- Entity-aware state-change tracking
  - Selecting entities correctly
  - Predicting state changes
  - Retaining state change information

“Wash and cut the tomatoes”

- Cookedness($e_{tomato}$) = ?
- Cleanliness($e_{tomato}$) = clean
- Shape($e_{tomato}$) = separated
- Temperature($e_{tomato}$) = ?
Challenges

- Action-oriented neural structure

“Wash and cut the tomatoes”

- Cookedness( e_{tomato} ) = ?
- Cleanliness( e_{tomato} ) = clean
- Shape( e_{tomato} ) = separated
- Temperature( e_{tomato} ) = ?
Outline

- Introduction
- Crowdsourcing Action-state mappings
- Model
- Datasets and Training
- Evaluations
Actions

- Extract set of 384 actions from cooking recipes

lace, perch, pry, soften, snip, skin, skin, follow, stew, whack, simmer, brown, string, ladle, rise, damper, spoon, cook, slather, wipe, cool, stiffen, whisk, level, tear, pinch, try, sand, adjust, gut, skewer, dip, round, shave, force, fold, barbecue, bake, poke, peel, melt, crush, devenin, punch, water, dry, shave, scramble, leave, narrow, divide, lengthen, replace, plop, zip, paste, dangle, splash, strike, cram, sharpen, garnish, tenderize, warp, warm, stick, grill, join, squash, strain, hardboil, sweeten, cap, pour, thin, drill, pickle, scatter, wedge, debone, encircle, slit, bend, slip, dress, sit, tilt, enlarge, stir, stand, moisten, blacken, ley, drape, bind, smack, scoop, crumble, wind, ignite, move, shake, glide, braise, arrange, return, dab, flour, rake, shatter, break, bang, bread, drop, refresh, slide, roast, quarter, scorch, clamp, reduce, sprinkle, extract, whip, leave, settle, stud, correct, shut, grease, hull, clarify, put, bash, drench, spill, filter, turn, shred, place, suspend, wash, ring, open, paint, top, attach, store, shell, prune, shower, boil, coat, steep, clip, strengthen, saute, slap, dash, seed, churn, saw, blanch, screw, touchen, soak, knead, mix, build, singe, prick, puree, compress, siphon, evaporate, coil, refrigerate, clear, saturate, cover, face, pipe, clean, brew, reheat, blend, dot, thicken, overcook, mince, fillet, toast, dice, sugar, tape, cube, hit, get, beat, batter, submerge, dye, wring, smash, immerse, remove, patch, stuff, steam, set, dump, sear, light, freeze, dislodge, powder, close, marinate, glaze, pump, decorate, oxidize, spray, blot, unfold, broil, tug, tuck, contract, dissolve, load, widen, spear, smother, pop, scalp, tint, raise, stretch, pile, fry, empty, squash, carve, lift, spritz, dribble, chill, char, pierce, sculpt, wilt, straighten, harden, mount, loop, pack, bathe, pound, rip, caramelize, deep-fry, sear, pan-fry, pat, mold, make, parboil, baste, split, fill, oil, hang, deflate, dust, expand, butter, lower, drain, drip, grind, withdraw, twirl, position, poach, bruise, shape, cut, nudge, ease, add, spread, combine, crack, snap, knock, knot, press, insert, rotate, crisp, roll, husk, flatten, drizzle, scald, slice, preheat, microwave, inject, slash, hard, squint, prod, core, pepper, zest, rub, burn, stem, stew, bury, shuck, smear, stir-fry, surround, chop, bone, discard, float, brush, wrap, prepare, condense, solidify, douse, wax, tamp, funnel, form, defrost, frost, heat, heap, grate, line, squeeze, spice, pull, smooth, nestle, dirty, trim, pin, taste, toss, crumble, twist, curl, mash, pit, check, film, cream, tip, scrape, flood, swirl, tie, shrink, scrub, separate, scale, macerate, deglaze, salt, liquefy, crease, push, fasten, rinse
State Changes

- 6 state changes: composition, cleanliness, cookedness, temperature, shape, location

- absorption, removal, subtraction, addition were compressed into composition
Outline

- Introduction
- Crowdsourcing Action-state mappings
  - Model
  - Datasets and Training
  - Evaluations
Neural Process Network
Neural Process Network

Diagram showing the components and flow of information in a neural process network, including entities, actions, and state predictors.
Neural Process Network

![Diagram of Neural Process Network](image)

- **Enc.**
  - Cook: GRU
  - beef: GRU
  - in: GRU
  - pan: GRU
  - GRU

- **Simulation Module**
  - Recurrent Attention (Eq. 3)
  - Sequence Attention (Eq. 2)
  - Entity Updater (Eq. 2)

- **Applicator**
  - Location: pan
  - State Predictors: hot, clean, cooked

- **Action Selector**
  - MLP

- **Entity Selector**
  - MLP

- **Output**
  - $h_T$ (selectors)
Neural Process Network

Enc.

- Cook: GRU
- beef: GRU
- in: GRU
- pan: GRU

Entity Selector

- $h_t$
- Recurrent Attention (Eq. 3)
- Sequence Attention (Eq. 2)

Simulation Module

- Entity Updater (Eq. 2)
- Applicator

Action Selector

- $h_t$
- MP

State Predictors

- Location
- pan
- temp?
- hot
- Cooked?
- cooked

$\bar{e} = \mathcal{E}$
Neural Process Network

- **Enc.**
  - GRU
  - GRU
  - GRU
  - GRU
  - GRU
  - $h_T$ (to selectors)

- **Entity Selector**
  - Recurrent Attention (Eq. 3)

- **Action Selector**
  - MLP
  - $s_{cut}$
  - $s_{fry}$
  - $s_{wash}$
  - $s_{put}$

- **Simulation Module**
  - Entity Updater (Eq. 2)
  - Applicator
  - Location
  - Temp?
  - Clean?
  - Cooked?
  - pan
  - hot
  - $w_p$
Neural Process Network
Neural Process Network

Enc.: GRU → GRU → GRU → GRU → GRU

Entity Selector:
- Recurrent Attention (Eq. 3)
- Sequence Attention (Eq. 2)

Simulation Module:
- Entity Updater (Eq. 7)

Applicator:
- Location
- Temp?
- Clean?
- Cooked?
- pan
- hot
- ...
Neural Process Network

![Diagram of Neural Process Network with GRUs and attention mechanisms]
Outline

- Introduction
- Crowdsourcing Action-state mappings
- Model
- Datasets and Training
  - Evaluations
State Change Prediction

- **F1**
  - GRU (Cho et al., 2014): 41.16
  - REN (Henaff et al., 2016): 42.31
  - NPN (Ours): 44.65

- **Accuracy**
  - GRU (Cho et al., 2014): 52.59
  - REN (Henaff et al., 2016): 53.47
  - NPN (Ours): 55.07
Entity Selection

Raw Ingredient Recall

GRU (Cho et al., 2014)  REN (Henaff et al., 2016)  NPN

<table>
<thead>
<tr>
<th>GRU</th>
<th>REN</th>
<th>NPN</th>
</tr>
</thead>
<tbody>
<tr>
<td>67.69</td>
<td>71.88</td>
<td>74.88</td>
</tr>
</tbody>
</table>

Raw Recall
Entity Selection

Raw Ingredient Recall

"Cut the tomatoes."

- GRU (Cho et al., 2014)
- REN (Henaff et al., 2016)
- NPN
Entity Selection

Raw Ingredient Recall

- "Cut the tomatoes."
- "Bake for 20 minutes"
- "Bake them for 20 minutes"

GRU (Cho et al., 2014) | REN (Henaff et al., 2016) | NPN
67.69 | 71.88 | 74.88
Entity Selection

Compositional Ingredient Recall

- GRU (Cho et al., 2014)
- REN (Henaff et al., 2016)
- NPN

Comp Recall

- 7.74
- 9.87
- 20.45
“Combine the water and flour in a bowl.”

“Stir the mixture until a dough forms.”
Elided Example

Step 1. Melt butter in medium saucepan over heat.
  - Pred: butter
  - Gold: butter

Step 2. Remove from heat.
  - Pred: butter
  - Gold: butter

Step 3. Stir in oats, sugar, flour, corn syrup, milk, vanilla extract and salt
  - Pred: butter, oats, sugar, flour, corn_syrup, milk, vanilla_extract, salt
  - Gold: butter, oats, sugar, flour, corn_syrup, milk, vanilla_extract, salt
Composition Example

Step 1. Brown meat in oil until brown on all sides
  - Pred: oil, meat
  - Gold: oil, meat

Step 2. Pour off grease
  - Pred: oil
  - Gold: <NONE>
Composition Example

Step 1. Brown meat in oil until brown on all sides
  - Pred: oil, meat
  - Gold: oil, meat

Step 2. Pour off grease
  - Pred: oil
  - Gold: <NONE>

Step 3. Add tomato paste, beef broth, salt and pepper, garlic, chili powder, cumin, chile peppers, and water.
  - Pred: meat, garlic, chili_powder, tomato_paste, cumin, chiles, broth, water, pepper
  - Gold: Same + [oil]
Composition Example

Step 1. Brown meat in oil until brown on all sides
  - Pred: oil, meat
  - Gold: oil, meat

Step 2. Pour off grease
  - Pred: oil
  - Gold: <NONE>

Step 3. Add tomato paste, beef broth, salt and pepper, garlic, chili powder, cumin, chile peppers, and water.
  - Pred: meat, garlic, chili_powder, tomato_paste, cumin, chiles, broth, water, pepper
  - Gold: Same + [oil]

Step 4. Bring to boil, then turn very low, cover and simmer about one and one-half hours, or until meat is tender
  - Pred: meat, garlic, chili_powder, tomato_paste, cumin, chiles, broth, water, pepper
  - Gold: Same + [oil]
Modeling Composition

% Change in Entity Embedding Similarity when Combined

Percentage Change (%)

Step Number

1 2 3 4 5 6 7 8 9 10

Combined
Modeling Composition

Percentage Change in Entity Similarity when Combined

<table>
<thead>
<tr>
<th>Step Number</th>
<th>Percentage Change (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>

- **Not Combined**
- **Combined**
# Action Functions

<table>
<thead>
<tr>
<th>Action</th>
<th>Nearest Neighbors</th>
</tr>
</thead>
<tbody>
<tr>
<td>cut</td>
<td>slice, split, snap, slash, carve, slit, chop</td>
</tr>
<tr>
<td>boil</td>
<td>cook, microwave, fry, steam, simmer</td>
</tr>
<tr>
<td>add</td>
<td>sprinkle, mix, reduce, splash, stir, dust</td>
</tr>
<tr>
<td>grease</td>
<td>coat, rub, dribble, spray, smear, line</td>
</tr>
<tr>
<td>wash</td>
<td>rinse, scrub, refresh, soak, wipe, scale</td>
</tr>
<tr>
<td>mash</td>
<td>spread, puree, squeeze, liquefy, blend</td>
</tr>
<tr>
<td>place</td>
<td>ease, put, lace, arrange, leave</td>
</tr>
<tr>
<td>warm</td>
<td>reheat, ignite, heat, light, crisp, preheat</td>
</tr>
</tbody>
</table>
Recipe Step Generation

<table>
<thead>
<tr>
<th>Metric</th>
<th>S2S</th>
<th>Attn-S2S</th>
<th>REN-gen</th>
<th>NPN-gen</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLEU-4</td>
<td>4</td>
<td></td>
<td>34</td>
<td>31</td>
</tr>
<tr>
<td>Rouge-L</td>
<td>37</td>
<td></td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>Verb F1</td>
<td>20</td>
<td></td>
<td>45.25</td>
<td>40.5</td>
</tr>
<tr>
<td>State Change F1</td>
<td>50</td>
<td></td>
<td>35.75</td>
<td>31</td>
</tr>
</tbody>
</table>

Charts show performance metrics for different models.
## Recipe Step Generation

**Ingredients:** water, butter, ...

**Step 1:** Preheat oven to 425 degrees.

<table>
<thead>
<tr>
<th>Model</th>
<th>Continuation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence-2-Sequence</td>
<td>Lightly grease baking pan with oil.</td>
</tr>
<tr>
<td>(Suskever et al., 2014)</td>
<td></td>
</tr>
<tr>
<td>Attentive Seq2seq</td>
<td>Combine all ingredients and mix well.</td>
</tr>
<tr>
<td>(Cho et al., 2014)</td>
<td></td>
</tr>
<tr>
<td>Neural Process Network</td>
<td><strong>Melt butter in skillet</strong></td>
</tr>
<tr>
<td>Reference</td>
<td>Melt butter in small saucepan and mix in bourbon, thyme, pepper, and salt</td>
</tr>
</tbody>
</table>
Takeaways

- Text only gets you so far — need a model of the world
- World modeling priors can be integrated into neural network architecture
- Model of the world can be learned from data
Future Work

- New Domains
  - Stories
  - Dialogues
  - Scientific Processes
Future Work

- New Domains
  - Stories
  - Dialogues
  - Scientific Processes

- New Models of the World
Future Work

• New Domains
  - Stories
  - Dialogues
  - Scientific Processes

• New Models of the World
  - Verb Physics (ACL 2017)
  - Connotation Frames (ACL 2016)
  - Temporal Event Frames (ACL 2016)
  - Naive Psychology (ACL 2018)
  - Affect Event Frames (ACL 2018)
https://homes.cs.washington.edu/~antoineb/

Code available soon