

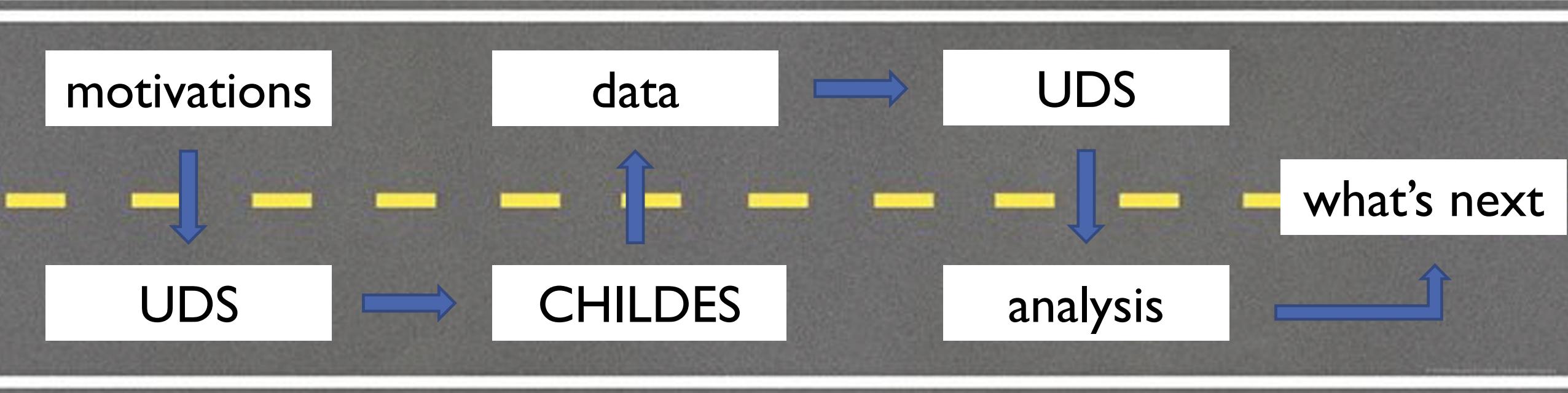
UNIVERSAL DECOMPOSITIONAL SEMANTICS (UDS) & CHILD SPEECH

CAROLINE GISH

APRIL 21, 2022

DATA SCIENCE FOR LINGUISTS

Project & presentation journey



Motivations

- lack of specific topic / dataset I was interested in
 - + personal interest in Semantics
 - + paper sent to Na-Rae
- = Data Science project focusing on Decomp / UDS

State of semantic annotation frameworks (before UDS)

- several independently-existing annotation frameworks
- prototypical approach to semantic representation and annotation
- require highly-trained annotators
- brittle in cases of
 - non-prototypical instances for a certain category
 - an instance fitting into multiple categories
 - capturing instances not considered at time of design

Motivations

- The overall purpose of this project was to
 - understand the make-up of the massive dataset
 - see where the researchers and developers are coming from and what they are ultimately aiming to do
 - explore how the UDS framework changes the state of semantic annotation frameworks
 - gain experience with a large dataset with its own toolkit designed for semantic research

What is UDS?

Universal Decompositional Semantics

What is UDS?



Decompositional Semantics Initiative

“Simple, commonsensical annotations of
meaning inspired by linguistic theory”

What is UDS?

The Universal Decompositional Semantics Dataset and Decomp Toolkit

Aaron Steven White[▷], Elias Stengel-Eskin[△], Siddharth Vashishtha[▷], Venkata Govindarajan[✉], Dee Ann Reisinger, Tim Vieira[△], Keisuke Sakaguchi[†], Sheng Zhang[‡], Francis Ferraro[○], Rachel Rudinger[†], Kyle Rawlins[△], Benjamin Van Durme[△]

[▷]University of Rochester, [△]Johns Hopkins University, [○]University of Maryland Baltimore County

[†]Allen Institute for Artificial Intelligence, [‡]Microsoft Research, [✉]University of Texas at Austin

{aaron.white, svashis3}@rochester.edu, {estenge2, kgr, vandurme}@jhu.edu, shezhan@microsoft.com,
{keisukes, rudinger}@allenai.org, ferraro@umbc.edu, {gvenkata1994, dee.ann.reisinger, tim.f.vieira}@gmail.com

Abstract

We present the Universal Decompositional Semantics (UDS) dataset (v1.0), which is bundled with the Decomp toolkit (v0.1). UDS1.0 unifies five high-quality, decompositional semantics-aligned annotation sets within a single semantic graph specification—with graph structures defined by the predicative patterns produced by the PredPatt tool and real-valued node and edge attributes constructed using sophisticated normalization procedures. The Decomp toolkit provides a suite of Python 3 tools for querying UDS graphs using SPARQL. Both UDS1.0 and Decomp0.1 are publicly available at <http://decomp.io>.

Keywords: semantics, semantic roles, factuality, genericity, temporal duration, entity typing

[The Universal Decompositional Semantics Dataset and Decomp Toolkit \(White et al., LREC 2020\)](#)

What is UDS?

Abstract

We present the Universal Decompositional Semantics (UDS) dataset (v1.0), which is bundled with the Decomp toolkit (v0.1). UDS1.0 unifies five high-quality, decompositional semantics-aligned annotation sets within a single semantic graph specification—with graph structures defined by the predicative patterns produced by the PredPatt tool and real-valued node and edge attributes constructed using sophisticated normalization procedures. The Decomp toolkit provides a suite of Python 3 tools for querying UDS graphs using SPARQL. Both UDS1.0 and Decomp0.1 are publicly available at <http://decomp.io>.

What is Decomp?

Decomp: A toolkit for decompositional semantics

Decomp is a toolkit for working with the Universal Decompositional Semantics (UDS) dataset, which is a collection of directed acyclic semantic graphs with real-valued node and edge attributes pointing into Universal Dependencies syntactic dependency trees.

<https://decomp.readthedocs.io/en/latest/index.html>

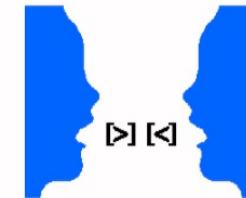
How Decomp / UDS work

- UDS dataset: 3 annotation layers built on English Web Treebank
 - i. syntactic graphs – UD parses of EWT
 - ii. semantic graphs – pred-arg structures
 - iii. semantic types – pred, arg, relationships

Goals of Decomp / UDS

- decompose more complex semantic classes into simple properties (that are still linguistically motivated)
- provide a semantic framework that can parse non-prototypical cases

CHILDES



- child language component of the TalkBank system
- transcriptions based on CHAT format (<https://talkbank.org/manuals/CHAT.pdf>)
 - CHAT files (.cha)
- Analyze with CLAN program (<https://talkbank.org/manuals/CLAN.pdf>)
- many different languages and types of corpora

Hicks corpus

CHILDES Narrative English Hicks Corpus

- narratives of different genre by 1st, 2nd, and 3rd graders
- asked to tell film's events in 3 different ways:
 - report – as a news reporter
 - event – as a sportscaster
 - story – as a storyteller

Hicks, D. (1990). Kinds of texts: Narrative genre skills among children from two communities. In A. McCabe (Ed.), *Developing narrative structure*. Hillsdale, NJ: Erlbaum.



Deborah Hicks
Partnership for Appalachian
Girls' Education

CHAT file format

1 RES: okay this is Deborah and David again .
%mor: co|okay pro:dem|this cop|be&3S n:prop|Deborah coord|and n:prop|David adv|again .
%gra: 1|3|COM 2|3|SUBJ 3|0|ROOT 4|3|PRED 5|4|CONJ 6|5|COORD 7|6|JCT 8|3|PUNCT

2 RES: and we're gonna be sportscasters .
%mor: coord|and pro:sub|we-aux|be&PRES part|go-PRESP~inf|to cop|be n|+n|sports+n|caster-PL .
%gra: 1|4|LINK 2|4|SUBJ 3|4|AUX 4|0|ROOT 5|6|INF 6|4|COMP 7|6|PRED 8|4|PUNCT

3 RES: and tell everything we see happening in the movie .
%mor: coord|and v|tell pro:indef|everything pro:sub|we co|see n:gerund|happen-PRESP prep|in det:art|the n|movie .
%gra: 1|2|LINK 2|0|ROOT 3|2|OBJ 4|6|SUBJ 5|6|COM 6|2|CJCT 7|6|JCT 8|9|DET 9|7|POBJ 10|2|PUNCT

4 RES: and I'm gonna start .
%mor: coord|and pro:sub|I~aux|be&1S part|go-PRESP~inf|to v|start .
%gra: 1|4|LINK 2|4|SUBJ 3|4|AUX 4|0|ROOT 5|6|INF 6|4|COMP 7|4|PUNCT

5 RES: the little boy is walking to a bakery shop .
%mor: det:art|the adj|little n|boy aux|be&3S part|walk-PRESP prep|to det:art|a n|bakery n|shop .
%gra: 1|3|DET 2|3|MOD 3|5|SUBJ 4|5|AUX 5|0|ROOT 6|5|JCT 7|9|DET 8|9|MOD 9|6|POBJ 10|5|PUNCT

10 CHI: well (.) these boys &-um are looking for the balloon too .
%mor: co|well det:dem|these n|boy-PL aux|be&PRES part|look-PRESP prep|for det:art|the n|balloon adv|too .
%gra: 1|5|COM 2|3|DET 3|5|SUBJ 4|5|AUX 5|0|ROOT 6|5|JCT 7|8|DET 8|6|POBJ 9|5|JCT 10|5|PUNCT
%cod: \$ind \$pr \$pg \$indef:kds

11 CHI: and then one boy points to the balloon .
%mor: coord|and adv:tem|then det:num|one n|boy v|point-3S prep|to det:art|the n|balloon .
%gra: 1|5|LINK 2|5|JCT 3|4|QUANT 4|5|SUBJ 5|0|ROOT 6|5|JCT 7|8|DET 8|6|POBJ 9|5|PUNCT
%cod: \$ind \$pr \$t:sq \$np:pkd \$np:bal

12 CHI: and they go .
%mor: coord|and pro:sub|they v|go .
%gra: 1|3|LINK 2|3|SUBJ 3|0|ROOT 4|3|PUNCT
%cod: \$ind \$pr \$pro:kds

13 CHI: and get it .

Hicks corpus CHAT files

text tier so then they catch [!] the balloon around the bakery .
mor tier: adv|so adv:tem|then pro:sub|they v|catch det:art|the n|balloon prep|around det:art|the n|bakery .
gra tier: 1|2|JCT 2|4|JCT 3|4|SUBJ 4|0|ROOT 5|6|DET 6|4|OBJ 7|4|JCT 8|9|DET 9|7|POBJ 10|4|PUNCT
cod tier: \$ind \$pr \$t:sq \$str \$pro:kds \$np:bal

text tier and then the boy comes out .
mor tier: coord|and adv:tem|then det:art|the n|boy v|come-3S adv|out .
gra tier: 1|5|LINK 2|5|JCT 3|4|DET 4|5|SUBJ 5|0|ROOT 6|5|JCT 7|5|PUNCT
cod tier: \$ind \$pr \$t:sq \$np:pas

text tier and looks for his balloon .
mor tier: coord|and v|look-3S prep|for det:poss|his n|balloon .
gra tier: 1|2|LINK 2|0|ROOT 3|2|JCT 4|5|DET 5|3|POBJ 6|2|PUNCT
cod tier: \$ind \$pr \$zero:pas \$np:bal

Hicks corpus idiosyncrasies

- many child utterances missing coding tier
- %cod tier annotations do not line up with the tokens
- %cod tier annotations difficult to read
- many codes not present in any coding tiers

```
*CHI: so      then      they      catch      the      balloon      around      the      bakery      .
%mor: adv|so  adv:tem|then  pro:sub|they  v|catch  det:art|the  n|balloon  prep|around  det:art|the  n|bakery  .
%gra: 1|2|JCT  2|4|JCT  3|4|SUBJ  4|0|ROOT  5|6|DET  6|4|OBJ  7|4|JCT  8|9|DET  9|7|POBJ  10|4|PUNCT
%cod: indpr t : sqstr pro : kdsnp:bal
```

[tiers_df.info\(\)](#)

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8992 entries, 0 to 8991
Data columns (total 4 columns):
 #   Column   Non-Null Count  Dtype  
 ---  -- 
 0   CHI_text  8992 non-null   object 
 1   mor_tier   8992 non-null   object 
 2   gra_tier   8992 non-null   object 
 3   cod_tier   8123 non-null   object 
dtypes: object(4)
```

Singling out a specific linguistic phenomenon

- dative alternation?
 - verbs like ‘give’ and ‘take’
- what else?
 - coding tier annotations



\$modv

MOD modals

MODV modal verbs

```
5th/report/rep01.cha:%cod:      $ind $pa $pg $modv
5th/report/rep01.cha:%cod:      $sadv $pg $modv
5th/report/rep02.cha:%cod:      $ind $pr $pf $pg $modv $a:adv $str
5th/report/rep03.cha:%cod:      $rel $pr $modv
5th/report/rep03.cha:%cod:      $ind $pr $modv
5th/report/rep03.cha:%cod:      $sadv $pg $modv
5th/report/rep03.cha:%cod:      $ind $pr $pg $modv $t:adv
5th/report/rep03.cha:%cod:      $sadv $pg $modv $str
```

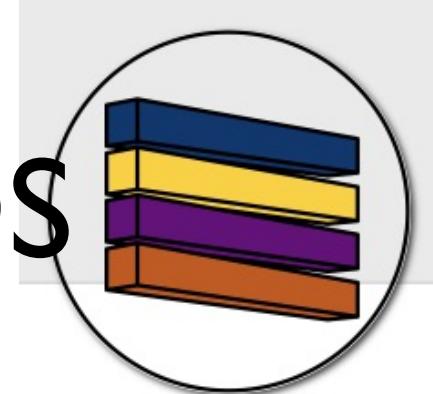
“try to”

```
evt027.cha-*CHI:      and they're trying to shoot it with a +...
evt027.cha-%mor:      coord|and pro:sub|they~aux|be&PRES part|try-PRESP inf|to v|shoot
evt027.cha-    pro:per|it prep|with det:art|a +...
evt027.cha-%gra:      1|4|LINK 2|4|SUBJ 3|4|AUX 4|0|ROOT 5|6|INF 6|4|COMP 7|6|OBJ 8|6|JCT
evt027.cha-    9|8|POBJ 10|4|PUNCT
evt027.cha:%cod:     $ind $pr $pg $modv $pro:kds $pro:bal
--
evt027.cha-%gra:      1|0|INCROOT 2|1|CONJ 3|7|QUANT 4|7|SUBJ 5|7|AUX 6|7|DET 7|2|COORD
evt027.cha-    8|7|NJCT 9|10|DET 10|8|POBJ 11|1|PUNCT
evt027.cha-%cod:     $mc $pr $t:sq $np:pkd
evt027.cha-*CHI:     trying to pop it .
evt027.cha-%mor:     part|try-PRESP inf|to v|pop pro:per|it .
evt027.cha-%gra:      1|0|INCROOT 2|3|INF 3|1|COMP 4|3|OBJ 5|1|PUNCT
evt027.cha:%cod:     $adv $pg $modv $emp:pkgd $pro:bal
```

“try” constructions

```
%cod: $ind $pr $pg $np:kds $pro:pas
*CHI: and then [/] and then he runs past a lady .
%mor: coord|and coord|and adv:tem|then pro:sub|he v|run-3S adj|past
det:art|a n|lady .
%gra: 1|5|LINK 2|1|CONJ 3|2|COORD 4|5|SUBJ 5|0|ROOT 6|8|MOD 7|8|DET 8|5|OBJ
9|5|PUNCT
%cod: $ind $pr $t:sq $pro:pas $indef:lad
*CHI: and then all the boys try to run past the lady .
%mor: coord|and adv:tem|then qn|all det:art|the n|boy-PL v|try inf|to
v|run adj|past det:art|the n|lady .
%gra: 1|6|LINK 2|6|JCT 3|5|QUANT 4|5|DET 5|6|SUBJ 6|0|ROOT 7|8|INF 8|6|COMP
9|11|MOD 10|11|DET 11|8|OBJ 12|6|PUNCT
%cod: $ind $pr $t:sq $modv $np:kds $np:lad
*CHI: but they couldn't .
```

Back to UDS



Importing UDS

In [1]:

```
# importing dataset  
  
from decomp import UDSCorpus
```

```
-----  
AttributeError                                 Traceback (most recent call last)  
/var/folders/wr/mkh_241s1697wrk7b4kxvm4r0000gn/T/ipykernel_38770/1048556902.py in <module>  
      1 # importing dataset  
      2  
----> 3 from decomp import UDSCorpus
```

Importing UDS + CRC

Installing on CRC done via two commands:

- `module load python/ondemand-jupyter-python3.8`
- `pip install --user git+https://github.com/decompositional-semantics-initiative/decomp.git`

additional:

- `pip install pathlib`
- `pip install ruamel-yaml`
- `pip install pyqtwebengine==5.13`
- `pip install pyqt5==5.13`

Visualizations

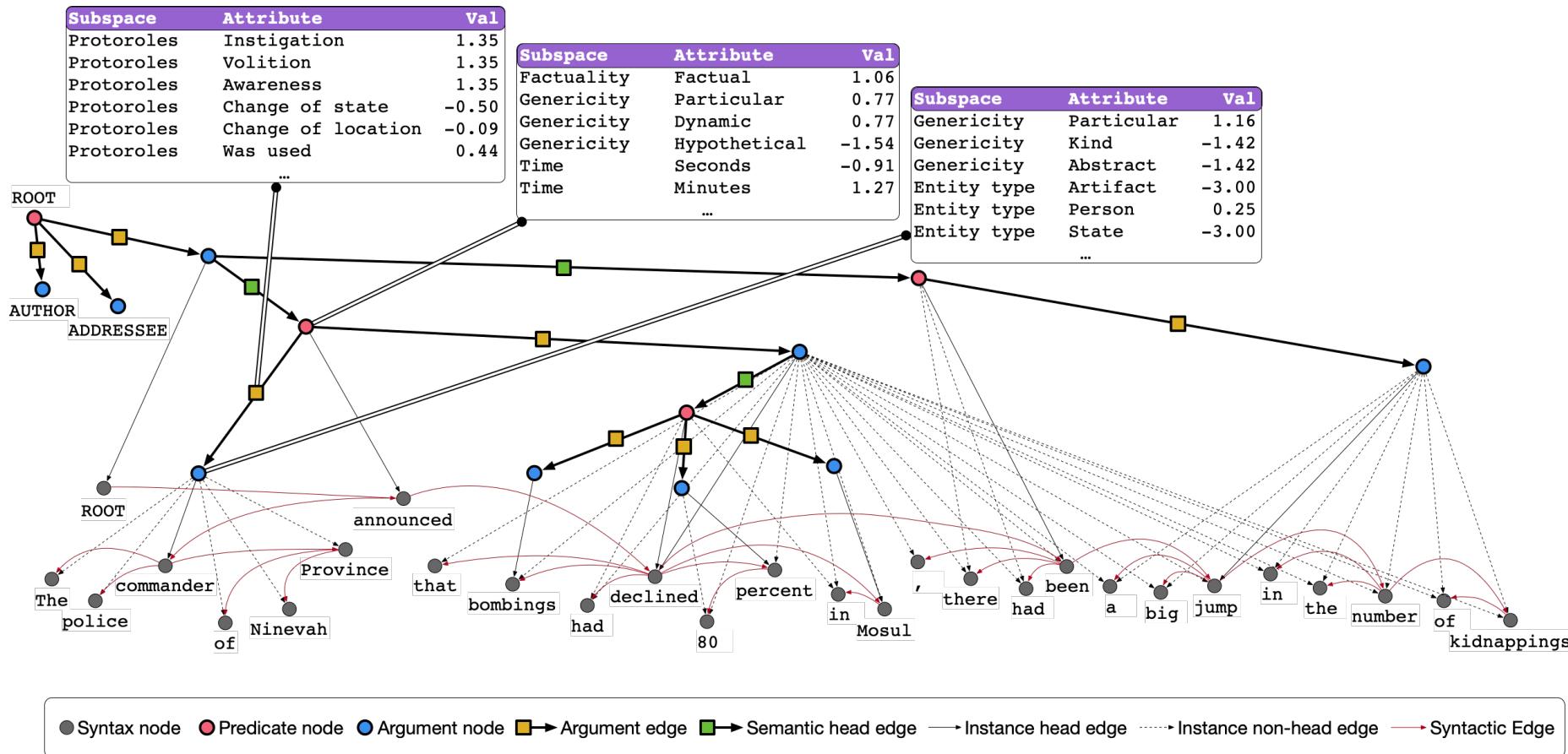
Visualizations:

```
In [ ]: from dash import Dash
```

```
In [2]: from decomp.vis.uds_vis import UDSVisualization
```

```
-----  
AttributeError                                 Traceback (most recent call last)  
/var/folders/wr/mkh_241s1697wrk7b4kxvm4r0000gn/T/ipykernel_38770/4186134077.py in <module>  
----> 1 from decomp.vis.uds_vis import UDSVisualization
```

So, what does a fancy Decomp visualization look like?



Semantic predicates and arguments

ewt-dev-539 “I tried to do it on the HRonline web-site, but the procedure is too complicated.”

arguments: entities that are participants in a situation

predicates: the type of event, state, situation in the sentence; take arguments

```
ewt-dev-539-semantics-pred-2  
(2, ['tried', 'try'])  
ewt-dev-539-semantics-pred-4  
(4, ['do', 'do'])  
ewt-dev-539-semantics-pred-18  
(18, ['complicated', 'complicated'])  
ewt-dev-539-semantics-arg-1  
(1, ['I', 'I'])  
ewt-dev-539-semantics-arg-4  
(4, ['do', 'do'])  
ewt-dev-539-semantics-arg-5  
(5, ['it', 'it'])  
ewt-dev-539-semantics-arg-11  
(11, ['site', 'site'])  
ewt-dev-539-semantics-arg-15  
(15, ['procedure', 'procedure'])
```

Modality

“system that allows speakers to express varying degrees of commitment to a proposition”

Sample sentence

ewt-dev-539

“I tried to do it on the HRonline web-site,
but the procedure is too complicated.”

CoNLL-U Format

ewt-dev-539 “I tried to do it on the HRonline web-site, but the procedure is too complicated.”

```
# sent_id = email-enronsent30_02-0029
# text = I tried to do it on the HRonline web-site, but the procedure is too complicated.
1   I       I       PRON   PRP   Case=Nom|Number=Sing|Person=1|PronType=Prs   2       nsubj
2:nsubj|4:nsubj:xsubj
2   tried   try    VERB   VBD   Mood=Ind|Number=Sing|Person=1|Tense=Past|VerbForm=Fin   0       root   0:root
-
3   to      to     PART   TO     4       mark   4:mark
4   do      do     VERB   VB     VerbForm=Inf   2       xcomp   2:xcomp
5   it      it     PRON   PRP   Case=Acc|Gender=Neut|Number=Sing|Person=3|PronType=Prs   4       obj    4:obj
-
6   on      on     ADP   IN     11      case   11:case
7   the    the    DET   DT     Definite=Def|PronType=Art   11      det    11:det
8   HRonline   HRonline  PROPN NNP   Number=Sing   11      compound
9   web    web    NOUN  NN     Number=Sing   11      compound
10  -      -      PUNCT HYPH   Number=Sing   11      punct  11:punct
11  site   site   NOUN  NN     Number=Sing   4       obl    4:obl:on
12  ,      ,      PUNCT ','    18      punct  18:punct
13  but    but    CCONJ CC     18      cc     18:cc
14  the    the    DET   DT     Definite=Def|PronType=Art   15      det    15:det
15  procedure   procedure NOUN  NN     Number=Sing   18      nsubj  18:nsubj
16  is      be     AUX   VBZ   Mood=Ind|Number=Sing|Person=3|Tense=Pres|VerbForm=Fin   18      cop    18:cop
-
17  too    too    ADV   RB     18      advmod 18:advmod
18  complicated   complicated ADJ   JJ     Degree=Pos   2       conj   2:conj:but
19  .      .      PUNCT .
-
```

Node type subspaces

- i. factuality
- ii. genericity
- iii. time
- iv. entity type

Factuality

<http://decomp.io/projects/factuality/>

whether an event happened or not

ewt-dev-539 “I tried to do it on the HRonline web-site, but the procedure is too complicated.”

“try”

```
        value : 1.0489},  
'factuality': {'factual': {'confidence': 1.0, 'value': 1.0489}},
```

“do”

```
'factuality': {'factual': {'confidence': 0.968, 'value': -0.336}},
```

Event structure <http://decomp.io/projects/event-structure/>

capture the structure of complex events

telicity: dealing with a clear endpoint to an action

“try”

```
-----  
'telic': {'confidence': 0.9999978542327881,  
           'value': 1.266179084777832}},
```

“do”

```
-----  
'telic': {'confidence': 0.9999988675117493,  
           'value': -1.076088786125183}},
```

ewt-dev-539 “I tried to do it on
the HRonline web-site, but the
procedure is too complicated.”

Genericity

<http://decomp.io/projects/genericity/>

capturing linguistic expressions of generalization

“try”

```
'genericity': {'pred-dynamic': {'confidence': 0.6365, 'value': 0.0092},  
               'pred-hypothetical': {'confidence': 0.6522, 'value': -0.0875},  
               'pred-particular': {'confidence': 0.5039, 'value': -0.0161}},
```

“do”

```
'genericity': {'pred-dynamic': {'confidence': 1.0, 'value': 1.1015},  
               'pred-hypothetical': {'confidence': 0.6153, 'value': -0.1813},  
               'pred-particular': {'confidence': 1.0, 'value': 1.1015}},
```

[ewt-dev-539](#) “I tried to do it on the HRonline web-site, but the procedure is too complicated.”

What's next

- more analysis of the different node type subspaces
- figure out what to do with the attribute values
- summary of what UDS and Decomp are good at, useful for

What I've learned

- how to ask for help when it is needed
- proper documentation is imperative
- plans change, and that is okay
- some tasks are best suited for a team of people

Thank you! + Questions?

