

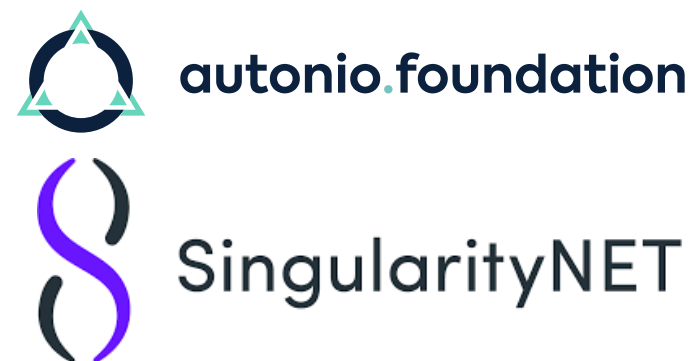
Interpretable Natural Language Processing

Anton Kolonin

akolonin@aigents.com

Facebook: [akolonin](#)

Telegram: [akolonin](#)



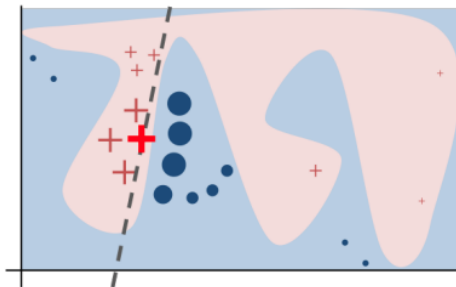
Why do we need Interpretable NLP (AI)?

Certain kinds of confusing classification errors due to errors in models may be better avoided in advance rather than found excuses later.

Explaining sentiment analysis model by LIME¹ method

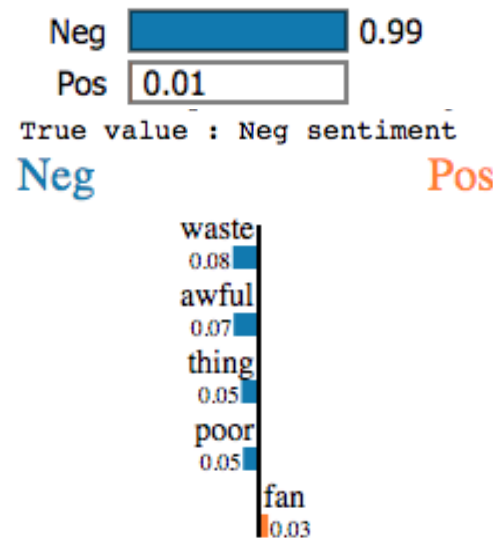
LIME (Local interpretable model agnostic explanations)

- Model agnostic
- Learns local behaviour of black box model around an instance for a candidate linear interpretable model



↑ Toy example showing LIME learning linear behaviour around class boundary

Prediction Probabilities



➤ **LIME** technique implemented on black box sentiment analysis classifier trained on IMDB movie dataset². In the image, Lime estimates the model's decision making criteria for the prediction on a test instance.

Text with highlighted words

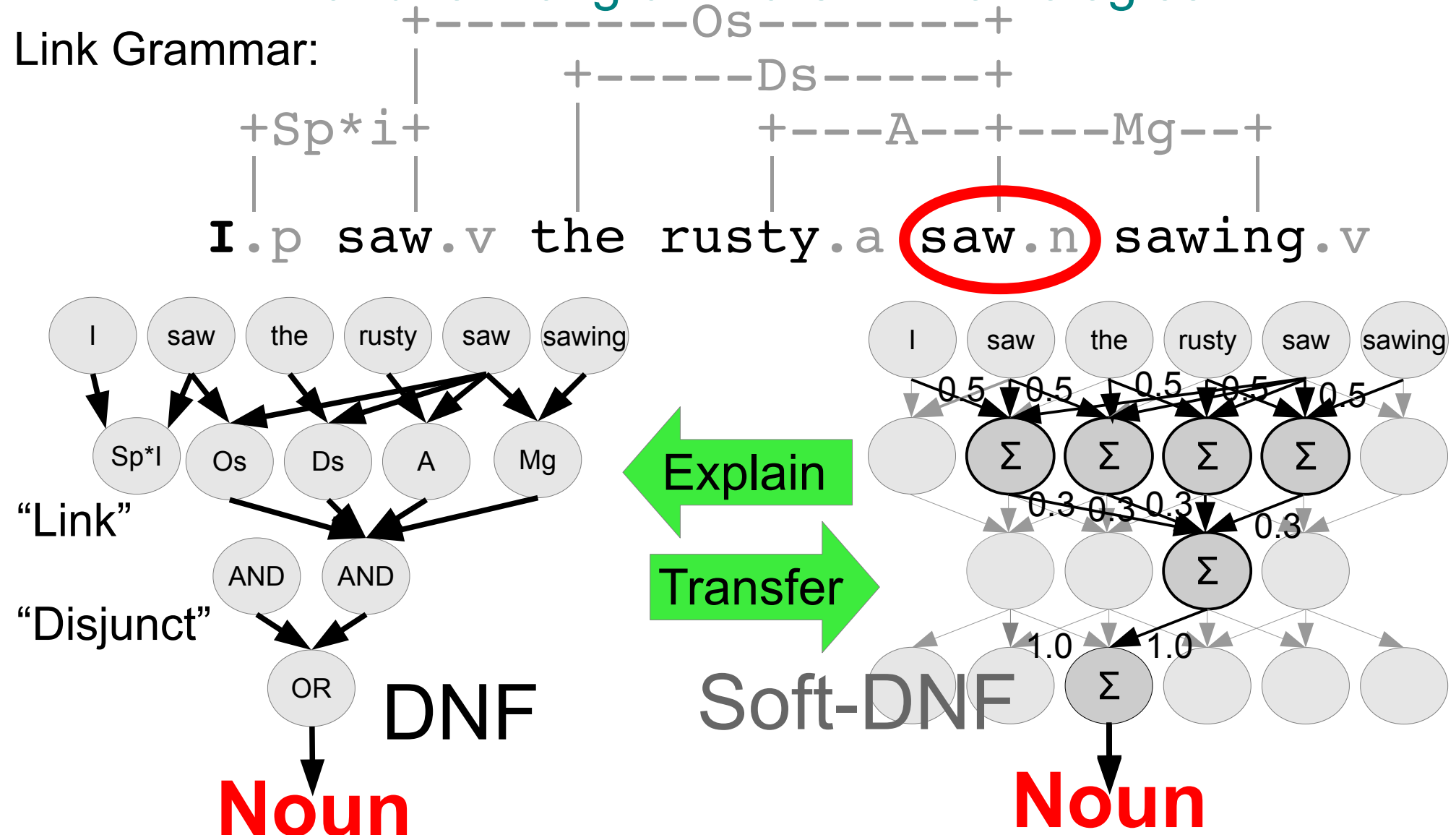
Being a long-time fan of Japanese film, I expected more than this. I can't really be bothered to write to much, as this movie is just so poor. The story might be the cutest romantic little something ever, pity I couldn't stand the awful acting, the mess they called pacing, and the standard "quirky" Japanese story. If you've noticed how many Japanese movies use characters, plots and twists that seem too "different", forcedly so, then steer clear of this movie. Seriously, a 12-year old could have told you how this movie was going to move along, and that's not a good thing in my book. lbr //lbr //Fans of "Beat" Takeshi: his part in this movie is not really more than a cameo, and unless you're a rabid fan, you don't need to suffer through this waste of film. lbr //lbr //12/10

1 - Source: <https://arxiv.org/abs/1602.04938>

2 - Source: <https://ai.stanford.edu/~amaas/data/sentiment/>

Bridging the Symbolic-Subsymbolic gap

in NLP between distributed representations
and formal grammars with ontologies



<https://arxiv.org/abs/1401.3372>

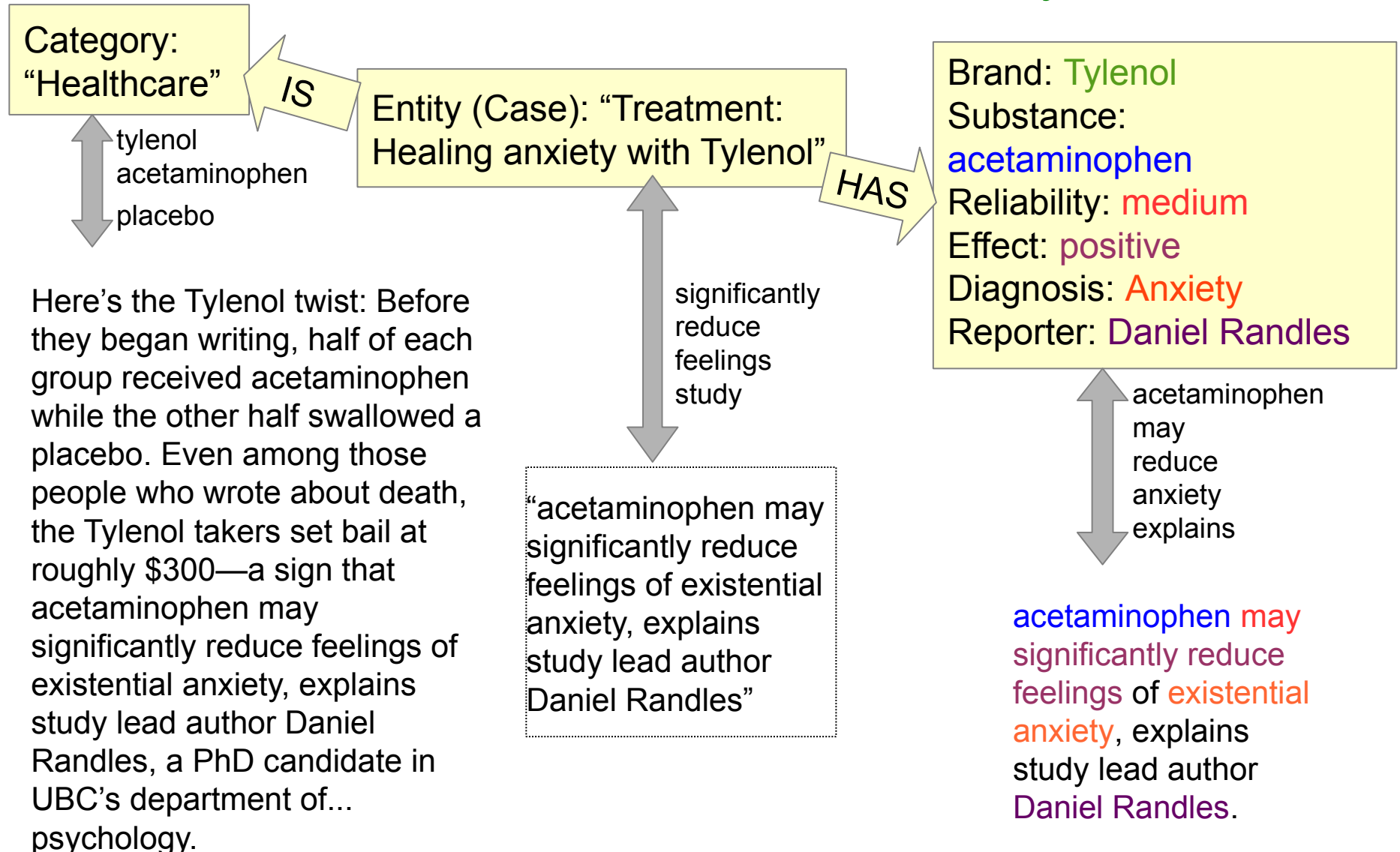
<https://arxiv.org/pdf/2006.00093.pdf>

Aigents® “Deep Patterns” - Text Mining

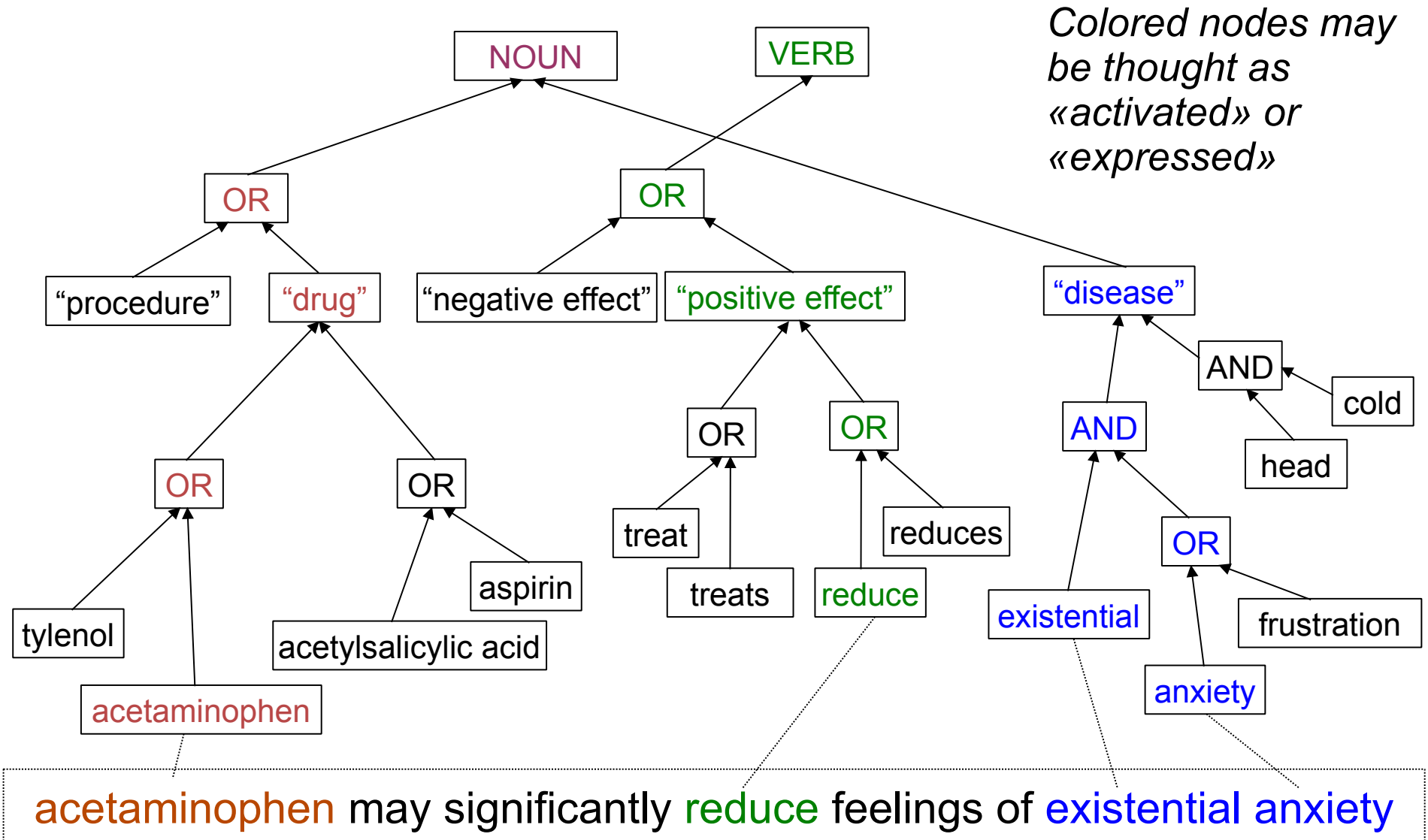
Classification

Case/Relationship
Extraction

Property Attribution
Entity Extraction



Aigents® “Deep Patterns” - Language Model



Aigents® “Deep Patterns” - Text Mining

`<set> := <disjunctive-set> | <conjunctive-set> | <M-skip-N-gram>`
`<disjunctive-set> := { <pattern> * }`
`<conjunctive-set> := (<pattern> *)`
`<N-gram> := [<pattern> *]`
`<pattern> := <token> | <regexp> | <variable> | <set>`

Variables may have domain restrictions
in ontology and/or refer to other
patterns as subgraphs

Example:

`[$description catheter] [$coating coating] [$inner-diameter
{diameter inner-diameter}] [$tip tip] [$pattern pattern]`

X

“Convey Guiding Catheter. Unique hydrophilic coating.
Small atraumatic soft tip. Ultra-thin 1 × 2 flat wire braid pattern”

=

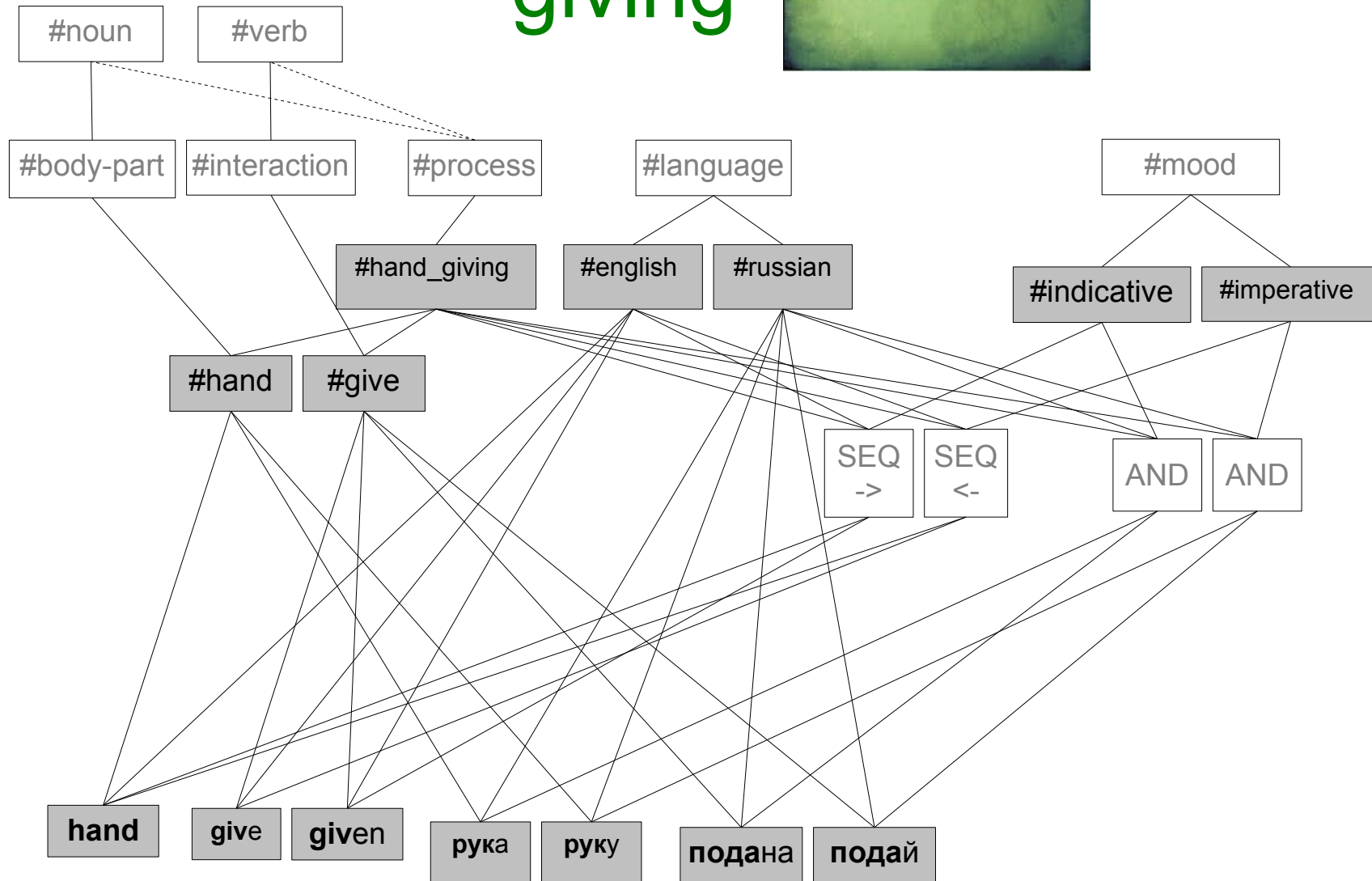
`{ coating : "hydrophilic", description : "convey guiding",
pattern : "ultra-thin 1 × 2 flat wire braid", tip : "soft" }`

<https://ieeexplore.ieee.org/document/7361868?arnumber=7361868>

<https://github.com/aigents/aigents-java>

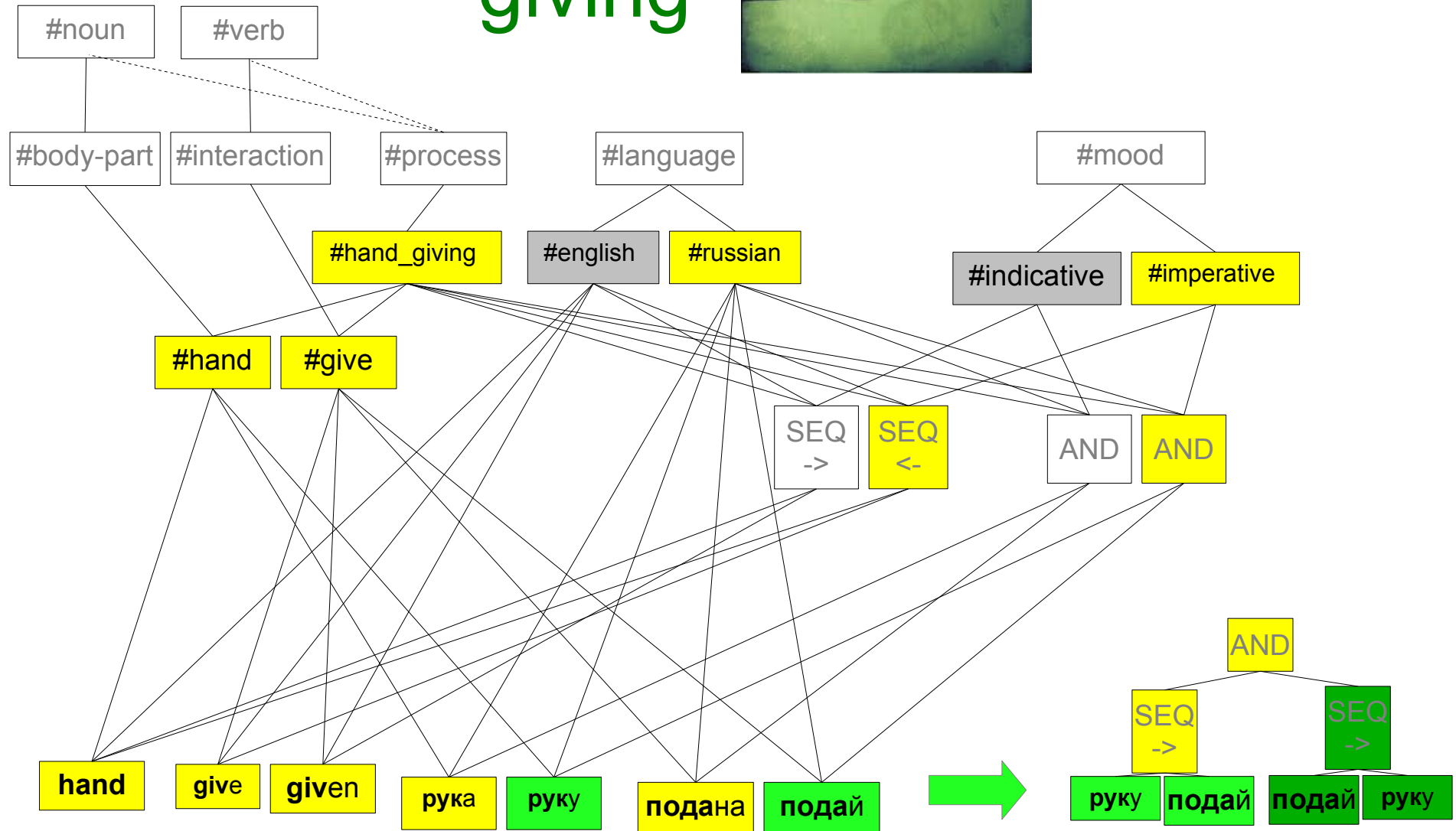
Grammar & Ontology Graph - Structure

Hand
giving

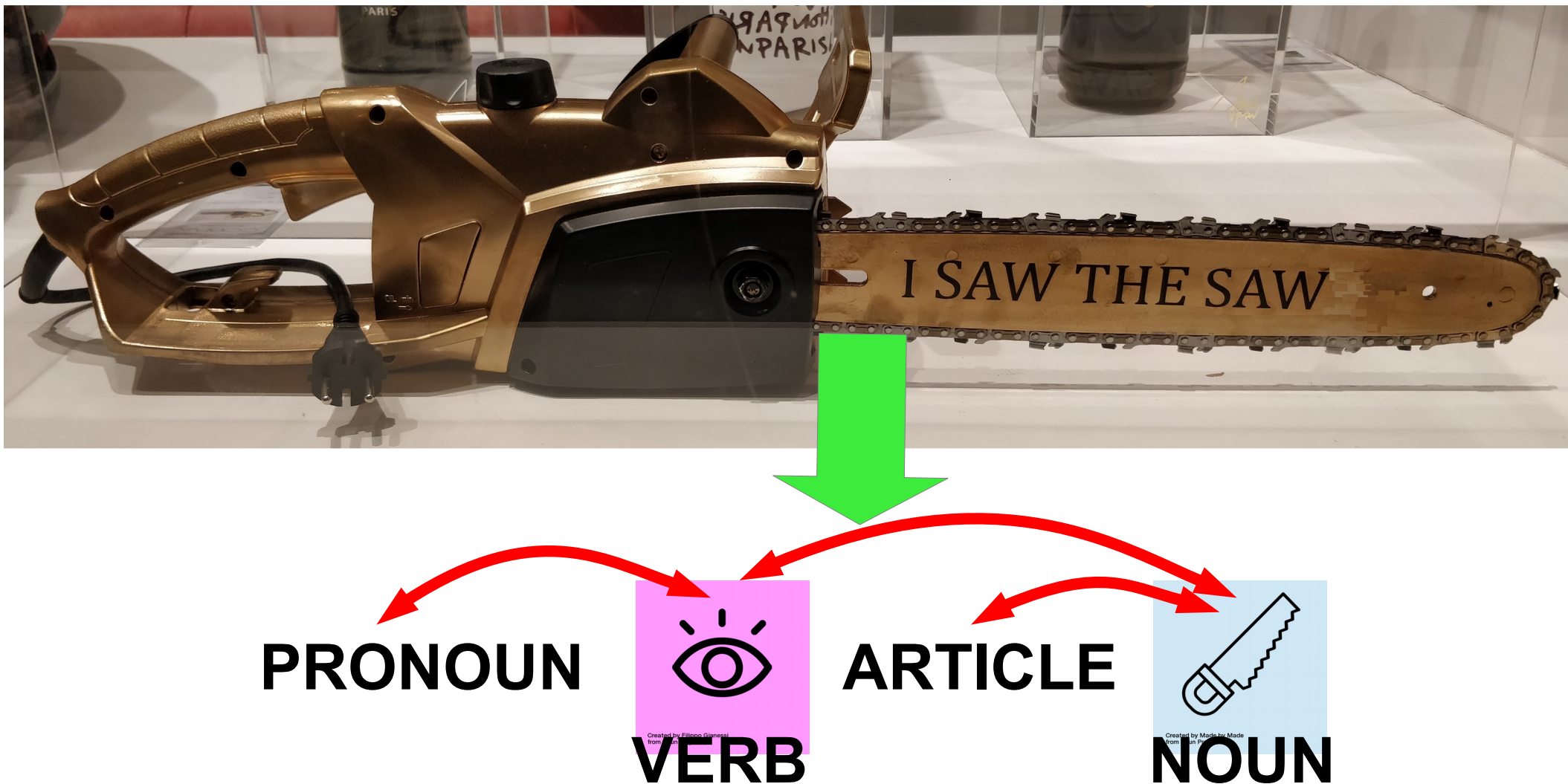


Grammar & Ontology Graph - Production

Hand
giving



Unsupervised Link Grammar Learning



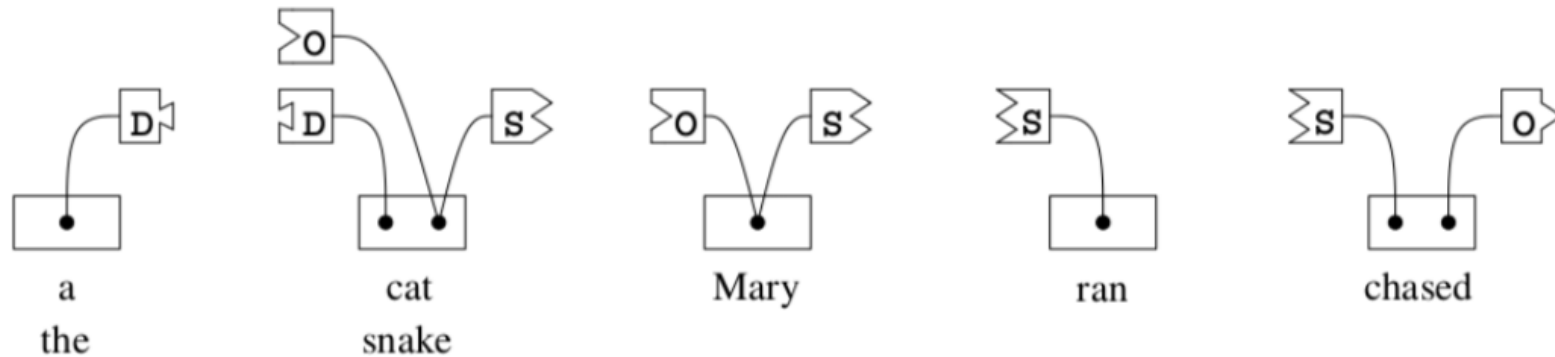
<https://www.springerprofessional.de/unsupervised-language-learning-in-opencog/15995030>

<https://www.springerprofessional.de/en/programmatic-link-grammar-induction-for-unsupervised-language-le/17020348>

Project goal and applications

- Grammar learning from scratch - programmatically
- Grammar extension/customization for specific domains
- Building dictionaries and patterns for NLP applications
- Parsing texts for NLP applications
- Grammar checking (more than spell checking)

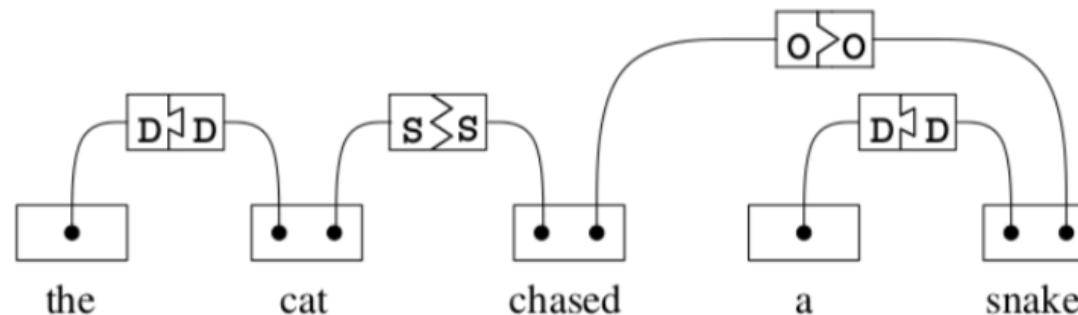
Link Grammar Connectors & Disjuncts as a fundamental particles



An illustration of Link Grammar connectors and disjuncts. The connectors are the jigsaw-puzzle-shaped pieces; connectors are allowed to connect only when the tabs fit together. A disjunct is the entire (ordered) set of connectors for a word. As lexical entries appearing in a dictionary, the above would be written as

a the: D+;
cat snake: D- & (S+ or O-);
Mary: O- or S+;
ran: S-;
chased S- & O+;

Note that although the symbols ‘&’ and ‘or’ are used to write down disjuncts, these are *not* Boolean operators, and do *not* form a Boolean algebra. They do form a non-symmetric compact closed monoidal algebra. The diagram below illustrates puzzle pieces, assembled to form a parse:

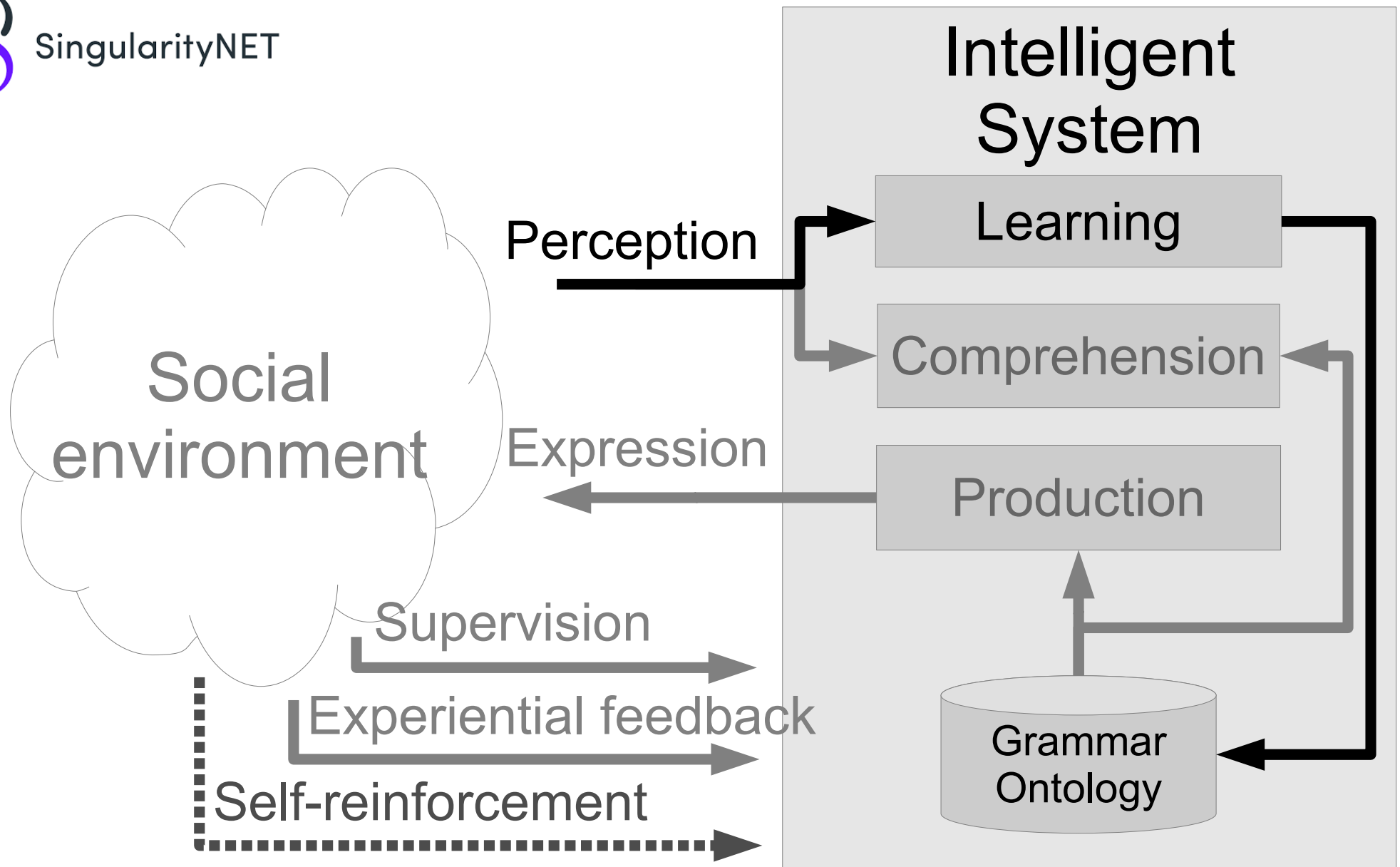


B. Goertzel,
L. Vepstas,
2014

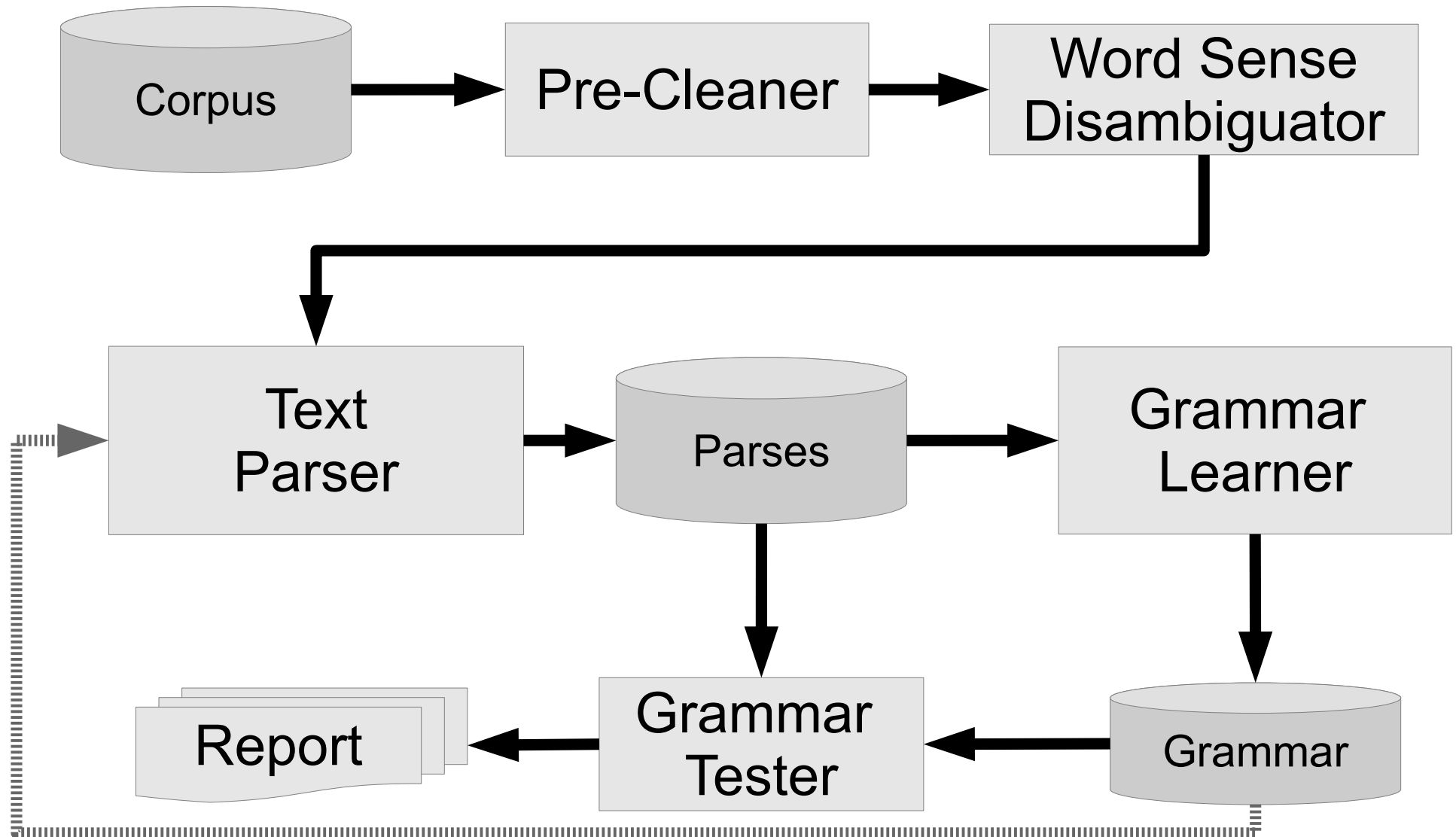
Constraints of the currently explored approach

- Controlled corpora
- Using Link Grammar formalism
- Relying on MST parses
- No account for morphology
- Self-reinforcement with F1 on parses
- Test against training data

Language Learning Environment



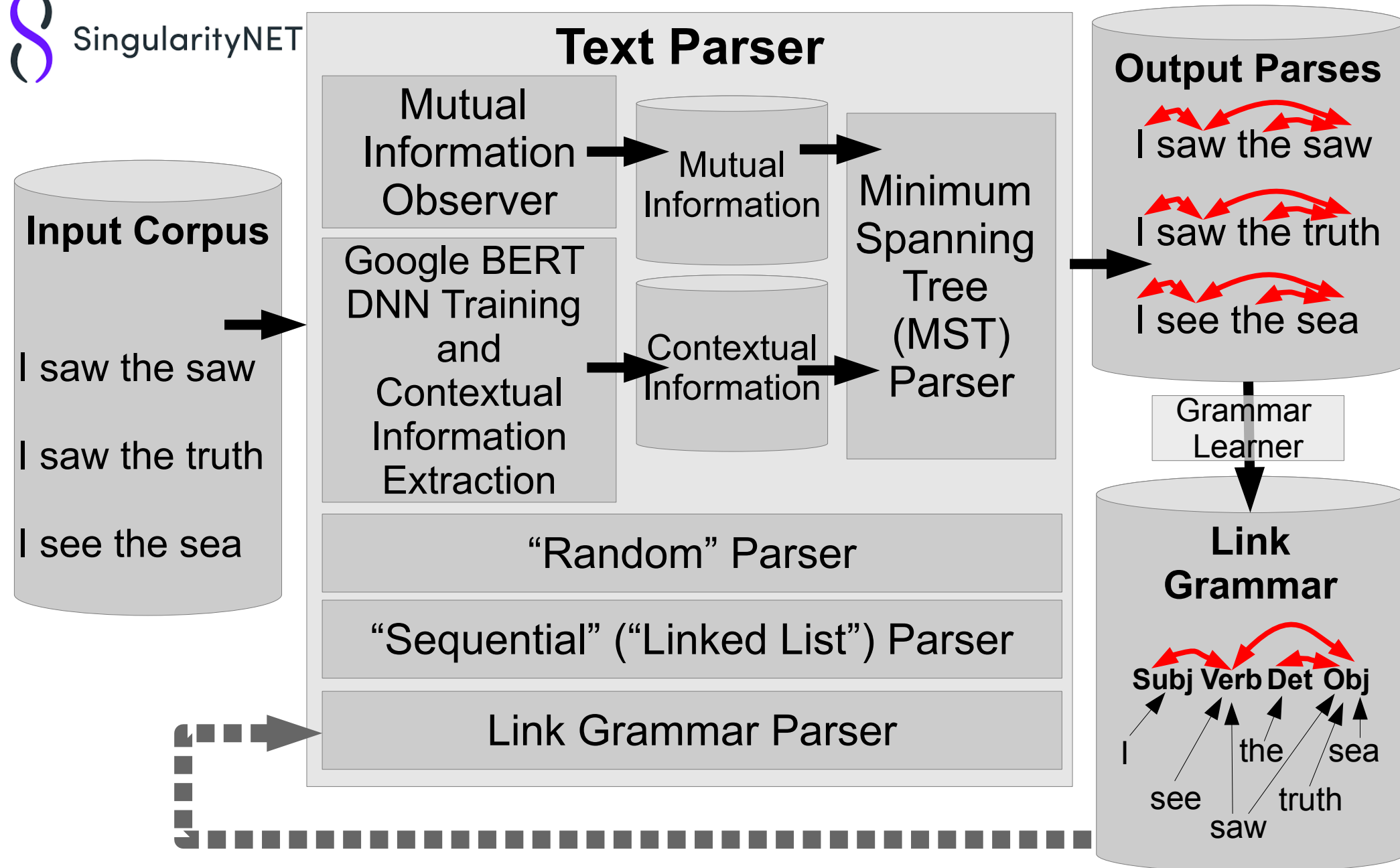
Unsupervised language learning pipeline with OpenCog



Text Parsing for Link Grammar



SingularityNET

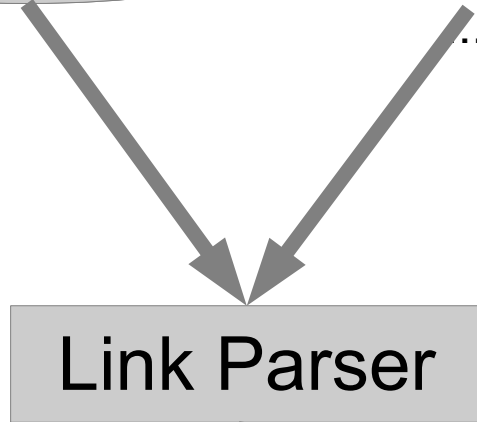


MST Parses vs. Link Parses

Corpus:

...
There is a snake
The boy saw a snake
The dog chased a snake
The cat chased a snake
...

English
Link Grammar



Link Parse:

```
[linkparser> the cat chased a snake  
Found 2 linkages (2 had no P.P. violations)  
Linkage 1, cost vector = (UNUSED=0 DIS= 0.00 LEN=9)
```

```
+----->WV----->+  
+-----Wd-----+ +-----Os-----+  
| +Ds**c+---Ss---+ | +Ds**c+ |  
| | | | | | | | |  
LEFT-WALL the cat.n chased.v- a snake.n
```

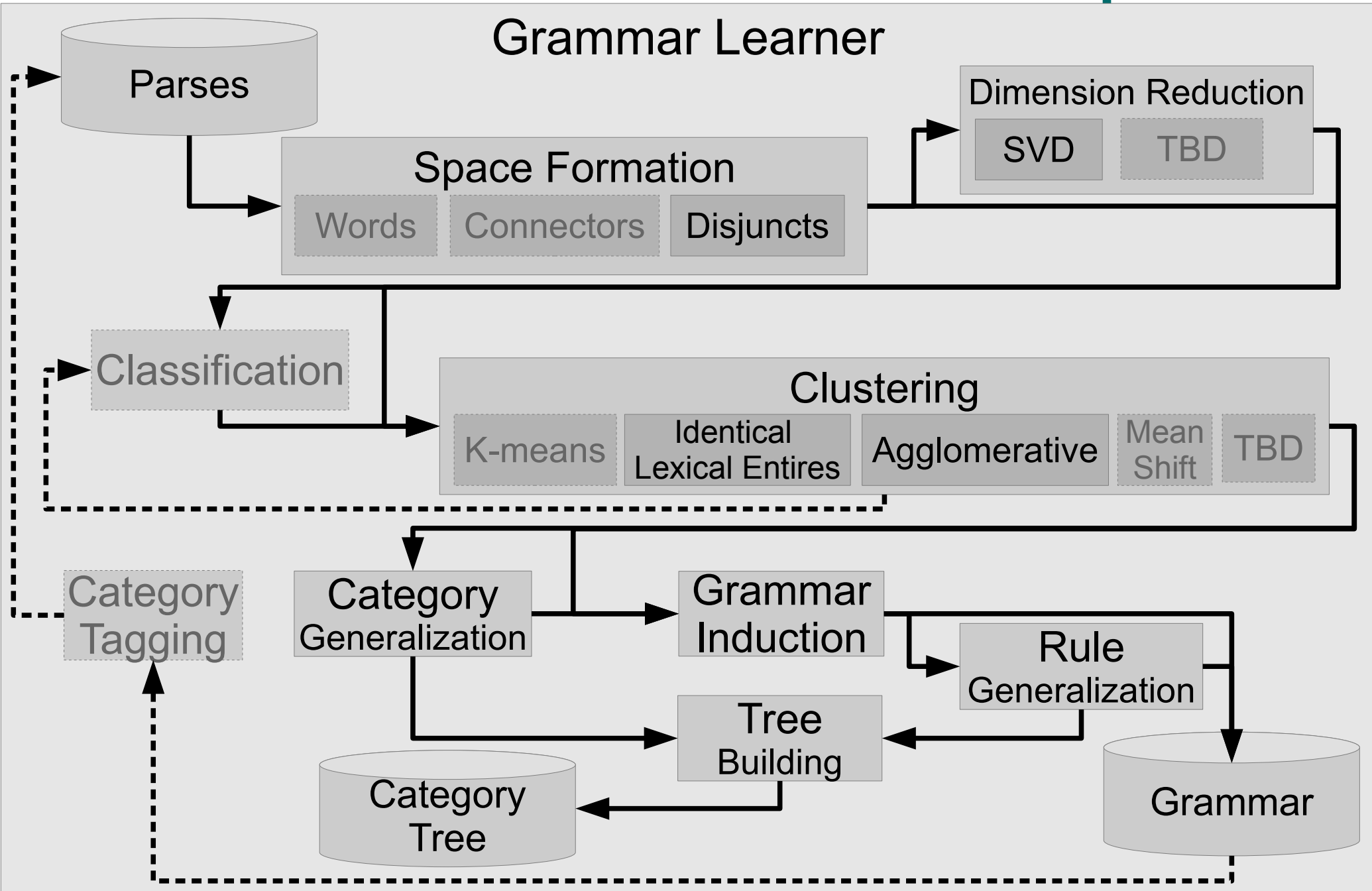


Mutual
Information

MST Parse:

```
LEFT-WALL the cat chased a snake  
0 ###LEFT-WALL### 2 cat  
0 ###LEFT-WALL### 3 chased  
1 the 2 cat  
2 cat 3 chased  
3 chased 5 snake  
4 a 5 snake
```

Link Grammar Learner Pipeline



Corpora in Use

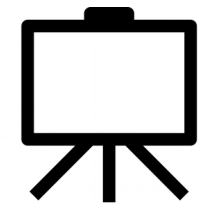
Corpus	Total words	Unique words	Occurrences per word	Total sentences	Average sentence length
POC-English	388	55	7	88	4
Child-Directed Speech	124185	3399	37	38181	4
Gutenberg Children	2695151	54054	50	207130	13

- POC-English – Proof-of-Concept corpus made of artificially selected sentences on limited number of topics (“small world”).
- Child Directed Speech (CDS) – corpus obtained from subsets of the CHILDES corpus – a collection of English communications directed to children with limited lexicon and grammar complexity
<https://childes.talkbank.org/derived/>
- Gutenberg Children (GC) - compendium of books for children contained within Project Gutenberg (<https://www.gutenberg.org>), following the selection used for the Children’s Book Test of the Babi CBT corpus
<https://research.fb.com/downloads/babi/>

Word-Sense Disambiguation

Using AdaGram¹ we disambiguate our POC-English corpus without supervision.

Two ambiguous words in corpus, with only two senses each:

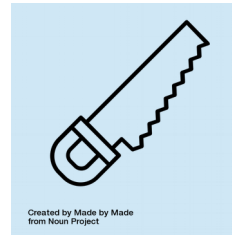


Created by Iconstock
from Noun Project



Created by b fariss
from Noun Project

board



Created by Made by Made
from Noun Project



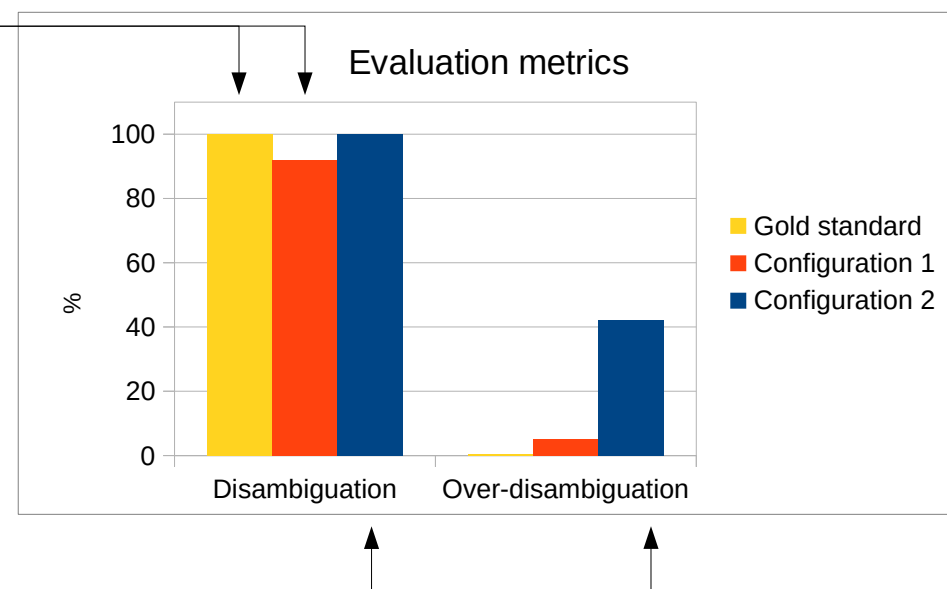
Created by Filippo Gianessi
from Noun Project

saw

After parameter tuning, we found two promising results:

mom saw@a dad with a saw@b .

mom@a saw@a dad@b with a@c saw@b .

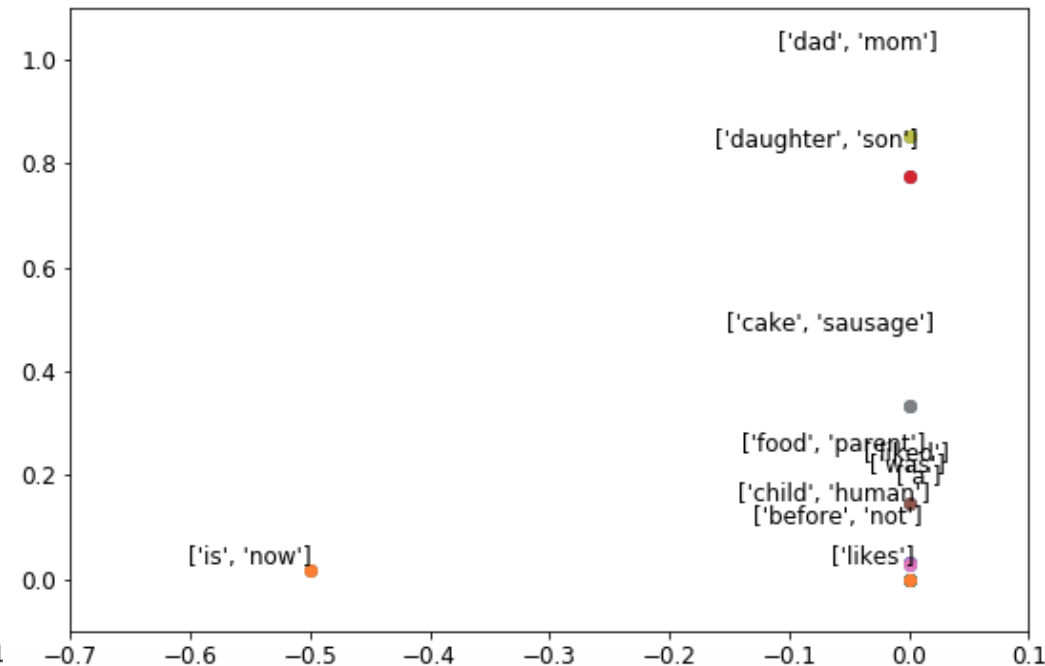
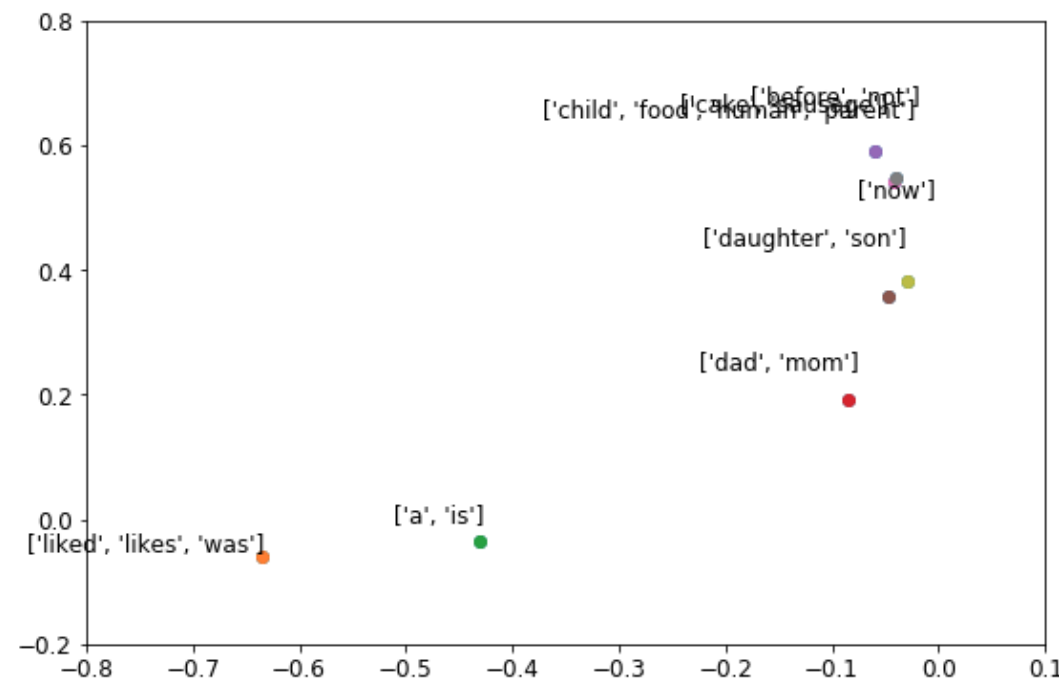


¹ https://github.com/glicerico/AdaGram/tree/take_sentences

Learning Grammatical and Semantic Categories in Vector Spaces of Connectors and Disjuncts

POC-English
(Connectors)

POC-English
(Disjuncts)

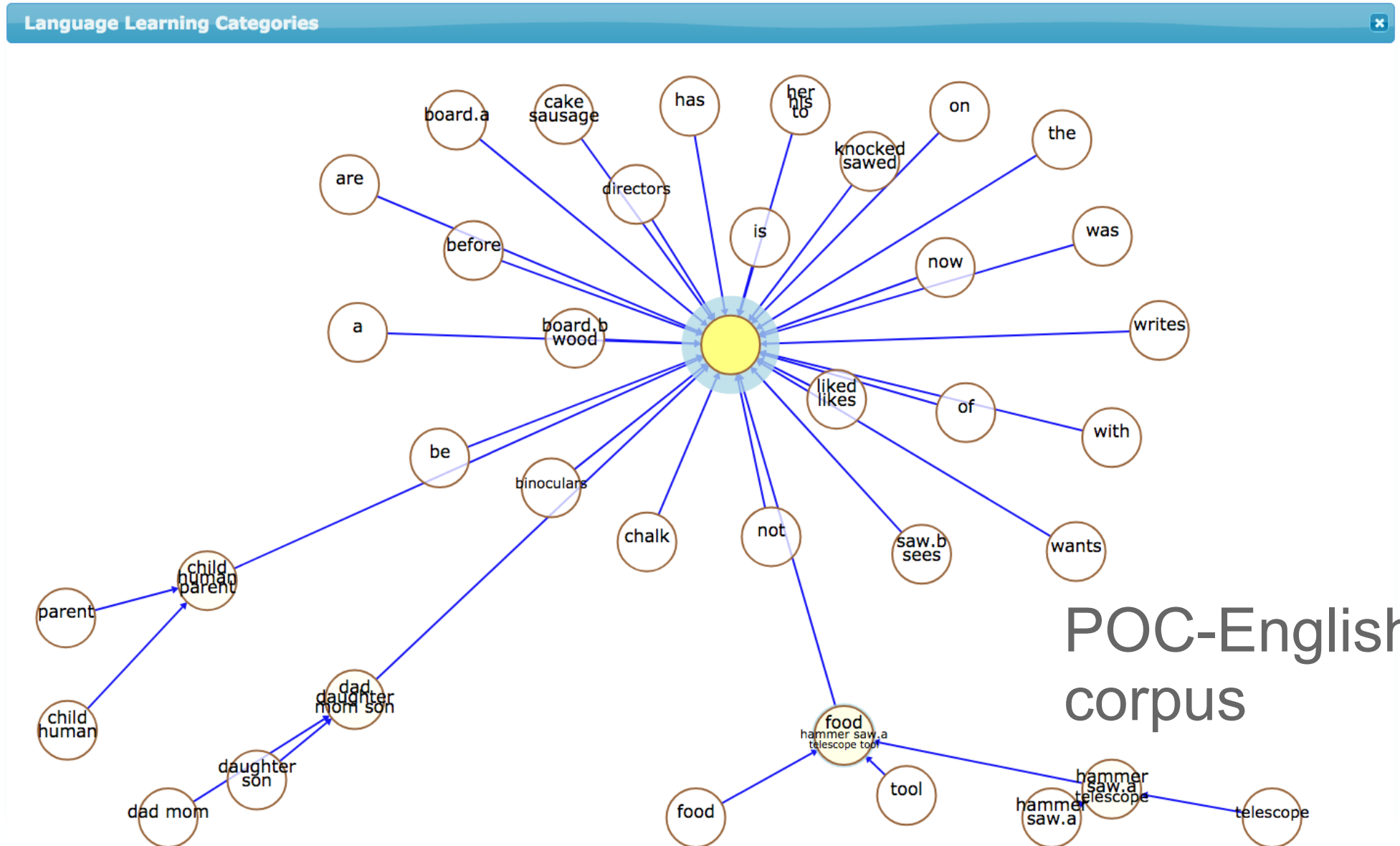


<https://github.com/singnet/language-learning>

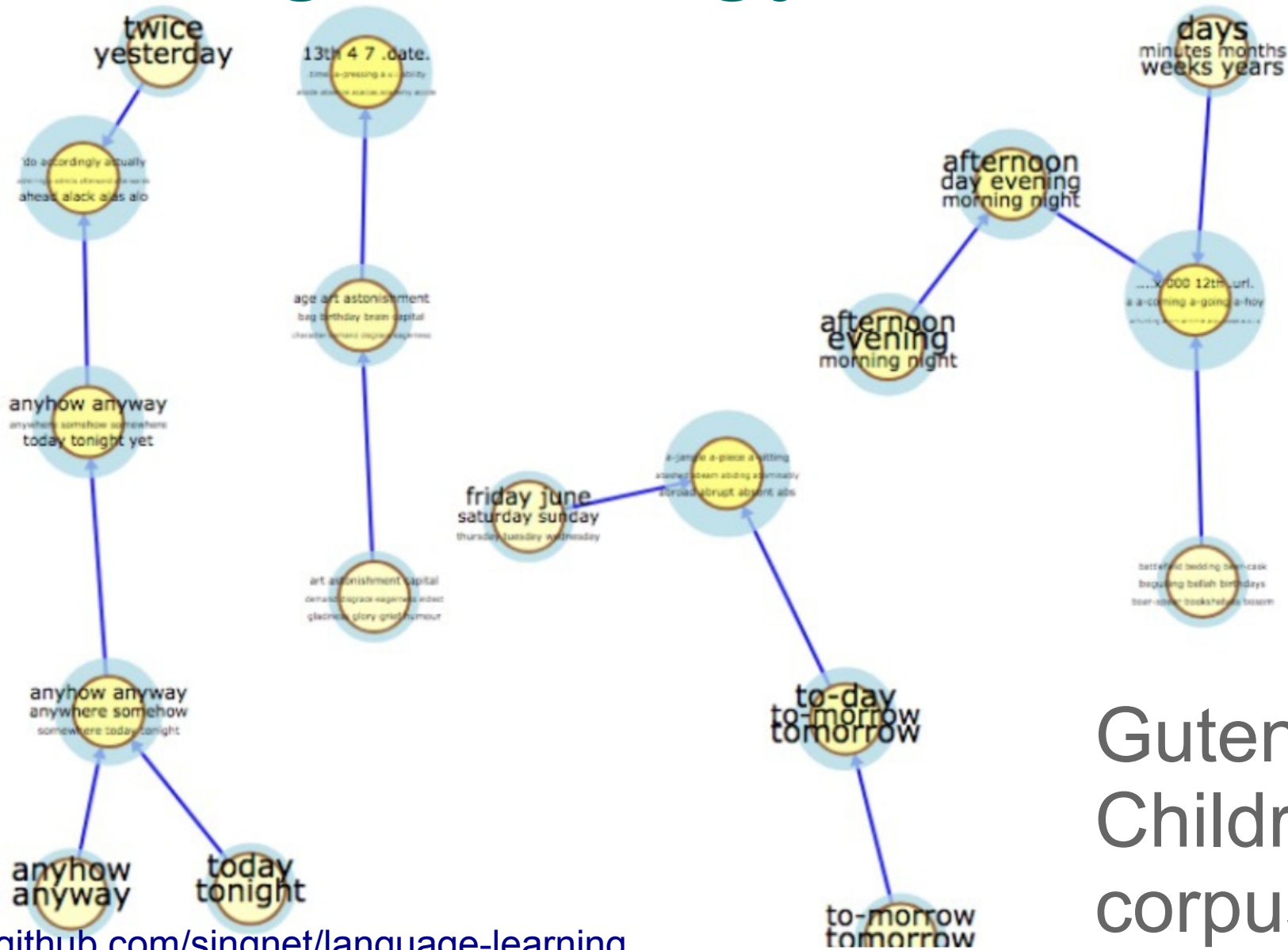
<https://www.youtube.com/watch?v=cwgtcOfA3KI>

<https://www.springerprofessional.de/en/programmatic-link-grammar-induction-for-unsupervised-language-le/17020348>

Grammar Ontology from Parses



Learning Ontology from Parses



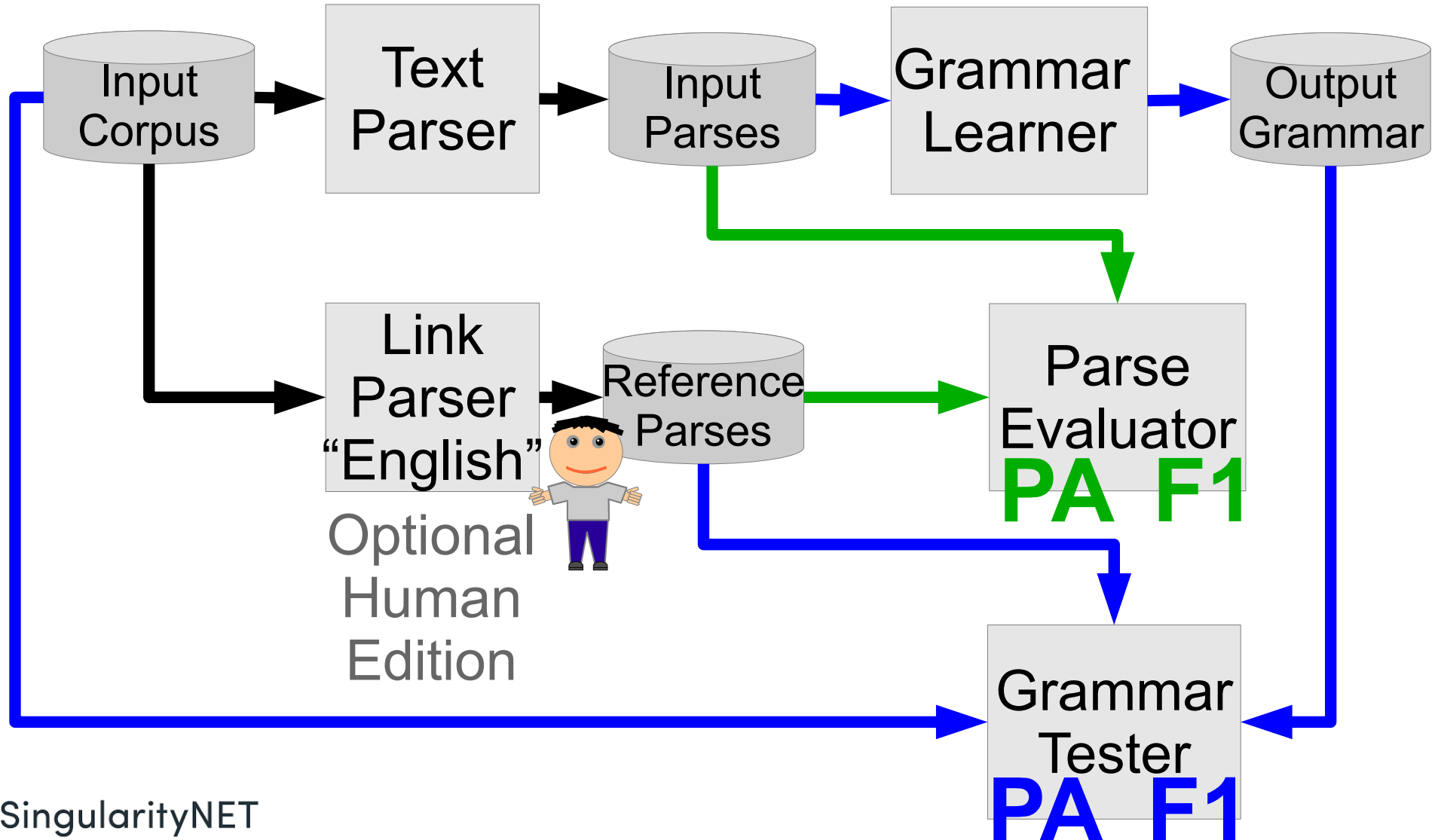
Gutenberg Children corpus

<https://github.com/singnet/language-learning>

<https://www.youtube.com/watch?v=cwgtcOfA3KI>

<https://www.springerprofessional.de/en/programmatic-link-grammar-induction-for-unsupervised-language-le/17020348>

Quality-Assessment with on Parses and Grammar



F1 Results Across the Corpora

Corpus	Parses	Parses F1	Clustering	Parse-Ability	Grammar F1
POC-English	Manual	1.00	ILE	100%	1.00
POC-English	Manual	1.00	ALE-400	100%	1.00
POC-English	MST	0.71	ILE	100%	0.72
POC-English	MST	0.71	ALE-400	100%	0.73
Child-Directed Speech	LG-English	1.00	ILE	99%	0.98
Child-Directed Speech	LG-English	1.00	ALE-400	99%	0.97
Child-Directed Speech	MST	0.68	ILE	71%	0.45
Child-Directed Speech	MST	0.68	ALE-400	82%	0.50
Gutenberg Children	LG-English	1.00	ILE	63%	0.65
Gutenberg Children	LG-English	1.00	ALE-500	69%	0.66
Gutenberg Children	MST	0.52	ILE	93%	0.50
Gutenberg Children	MST	0.52	ALE-500	99%	0.53

<https://www.springerprofessional.de/unsupervised-language-learning-in-opencog/15995030>

<https://www.springerprofessional.de/en/programmatic-link-grammar-induction-for-unsupervised-language-le/17020348>

Conclusions and Next Steps

- Grammars can be induced from parses
 - Better parses => better grammars (Pearson between F1 on parses and F1 on grammar ≥ 0.9)
- MST-Parsing and BERT-Milking can't get parses better than “sequential” (“linked list”)
- “Curriculum learning” is a next try for:
 - Parses better than “sequential”
 - Better grammars for larger corpora
 - Incremental Grammar Learning

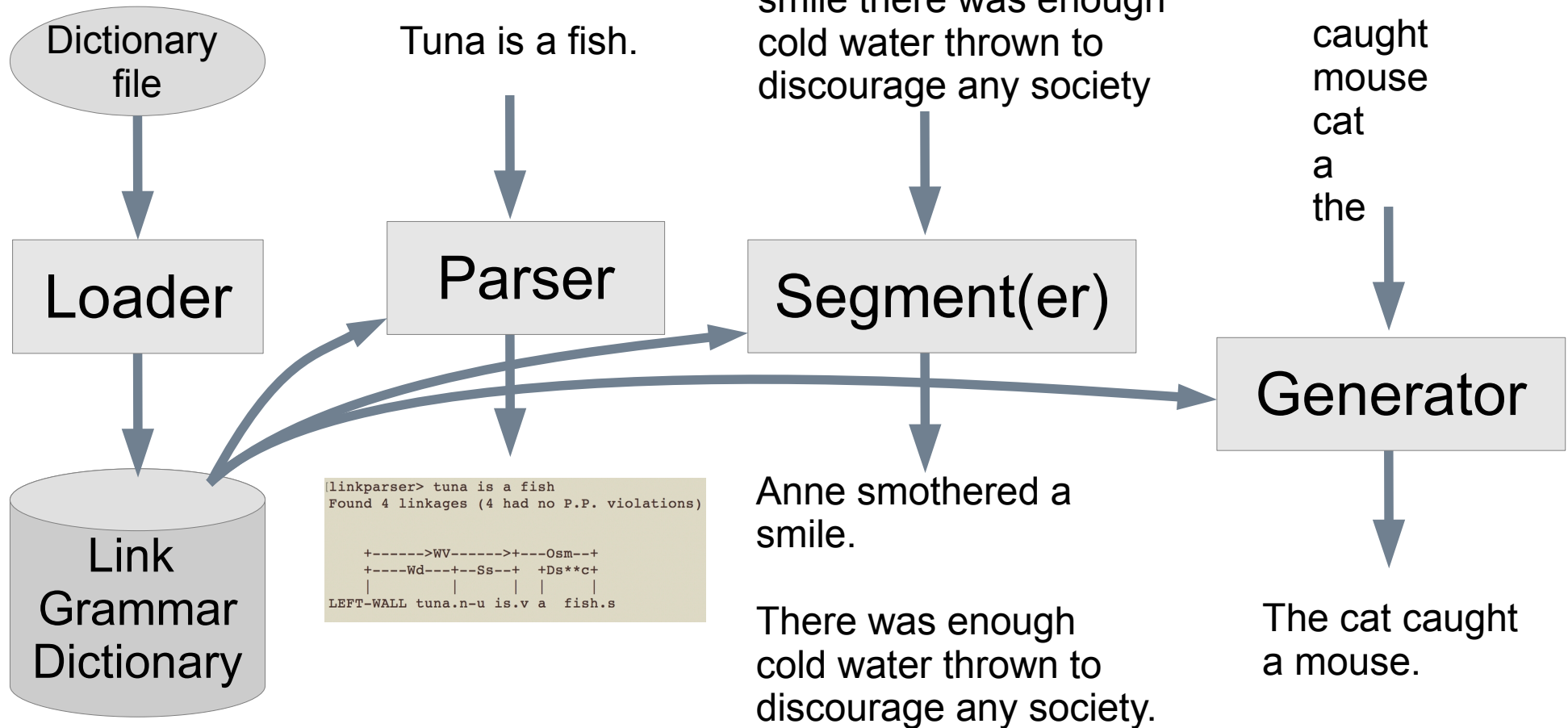


Link Grammar – beyond parsing

Parsing

Segmentation

Generation



Code: <https://github.com/aigents/aigents-java-nlp>

Segmentation: <https://ieeexplore.ieee.org/document/9303220>

Generation: <https://arxiv.org/abs/2105.00830>

Grammatical Generation - Methodology

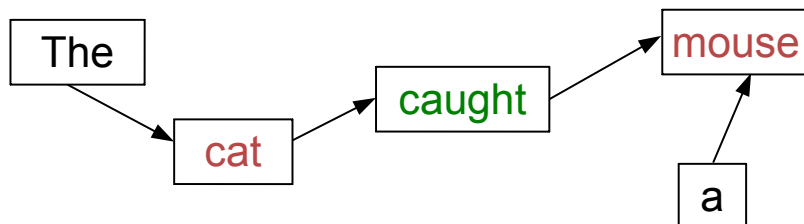
Generator determines what sentences can be formed from a given list of words via valid Link Grammar rules:

1) Given a list of words, the Generator determines a subset of all orderings of those words that satisfies initial checks of the planarity and connectivity metarules.

2) For each ordering in the subset, the Generator determines if that ordering is valid; specifically, it ensures that every pair of consecutive words can be connected via links part of the Dictionary objects. To do so, the Generator uses the `connects()` function, which returns a boolean value indicating whether its two parameters `left` and `right` can be linked together.

Planarity metarule: links do not cross

Connectivity metarule: links and words of a sentence must form a connected graph that can be completely traversed via one path



Algorithm 2: CONNECTS

Input : A pair of strings *left* and *right*, representing the two words to potentially be connected

Output: An boolean value indicating whether *left* and *right* can be connected via valid Link Grammar rules

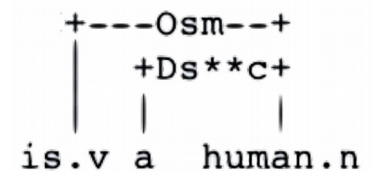
Obtain *leftList*, the list of rules corresponding with *left* (i.e. the rule when *left* is a verb, the rule when *left* is a gerund, etc.), from the global Dictionary variables *dict* and *hyphenated*

Obtain *rightList* in a similar manner

```

for leftRule in leftList do
  for rightRule in rightList do
    Split leftRule and rightRule into lists of Disjuncts ld and rd
    for l in ld do
      for r in rd do
        Replace all instances of '-' in l with '+' and vice versa
        if l = r then
          return true
        else
          continue
        end
      end
    end
  end
end
return false
  
```

`connects()` is not always applicable. For instance, when the determiner “a” is present in the phrase “is a human,” the links are not “is” → “a” and “a” → “human” but rather “is” → “human” and “a” → “human” as shown in this Link Grammar parse:



<https://arxiv.org/abs/2105.00830>

Grammatical Generation - Results

Our algorithm was primarily tested on 92 sentences with words all part of SingularityNET's "small world" POC-English corpus. For this purpose, we have used the Link Grammar dictionary (automatically inferred from high quality Link Grammar parses created by SingularityNET's ULL pipeline) containing 42 total words and 5 total word clusters.

When tested on the same 92 sentences while using the complete Link Grammar dictionary for English, the algorithm achieved the following results. The decrease in "Single correct generated sentence" and increase in "Multiple sentences with one correct" is a direct result of the increased grammatical and semantic ambiguity from using Link Grammar instead of "small world" grammar. Since the "small world" grammar was created from the "small world" corpus itself, each of the words in the corpus contains only a subset of the grammatical or semantic contexts that Link Grammar does.

Our NLG architecture was also tested on 54 sentences part of Charles Keller's production of Lucy Maud Montgomery's "Anne's House of Dreams" as found in the Gutenberg Children corpus and performed as follows.

Metric	Result
Single correct generated sentence	62/92
Multiple sentences with one correct	30/92
Multiple sentences with none correct	0/92
No generated sentences	0/92
Too many results*	0/92
Accuracy	1.000
Total runtime	18 min, 46 sec
Average runtime per sentence	0 min, 12 sec

Metric	Result
Single correct generated sentence	8/92
Multiple sentences with one correct	57/92
Multiple sentences with none correct	0/92
No generated sentences	0/92
Too many results*	27/92
Accuracy	0.707
Total runtime	115 min, 6 sec
Average runtime per sentence	1 min, 15 sec

* "Too many results" is defined as over 25 generated sentences.

Metric	Result
Single correct generated sentence	1/54
Multiple sentences with one correct	53/54
Multiple sentences with none correct	0/54
No generated sentences	0/54
Accuracy	1.000
Total runtime	141 min, 51 sec
Average runtime per sentence	2 min, 37 sec

- POC-English – Proof-of-Concept corpus made of artificially selected sentences on limited number of topics ("small world").
- Gutenberg Children (GC) - compendium of books for children contained within Project Gutenberg (<https://www.gutenberg.org>), following the selection used for the Children's Book Test of the Babi CBT corpus <https://research.fb.com/downloads/babi/>

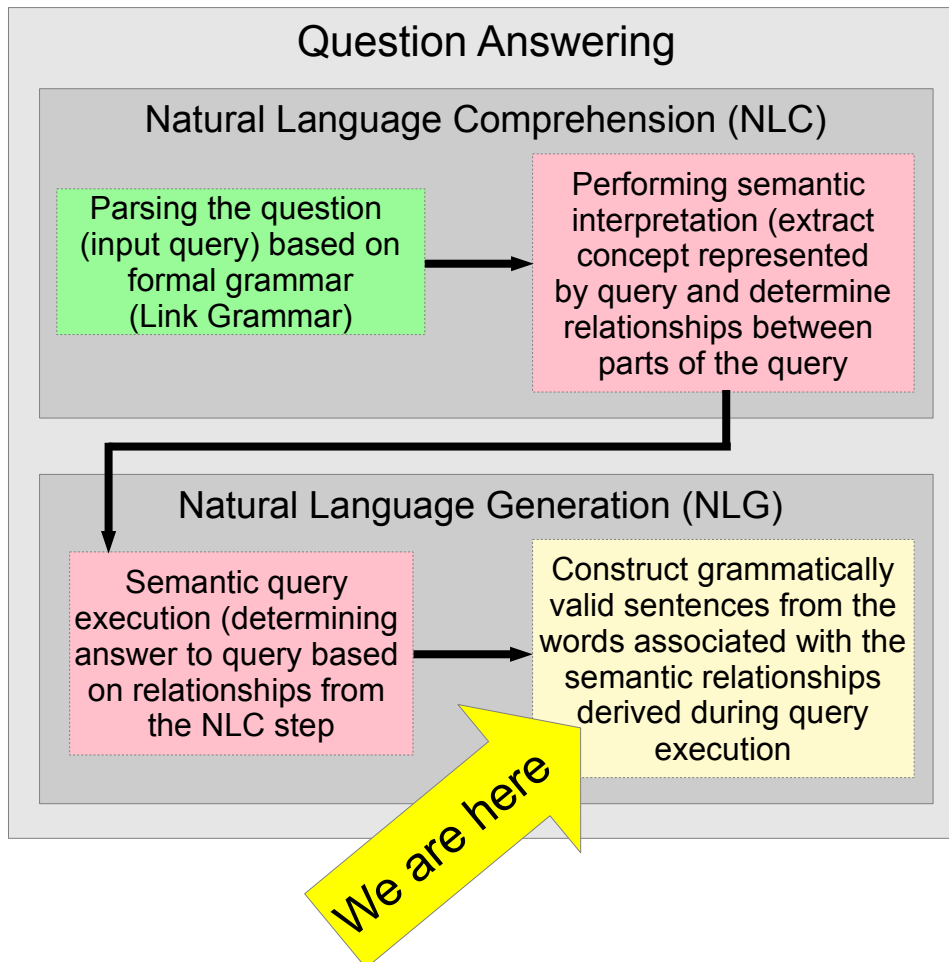
<https://github.com/aigents/aigents-java-nlp>

<https://arxiv.org/abs/2105.00830>

Grammatical Generation

Applications and Future Work

Applications



<https://github.com/aigents/aigents-java-nlp>
<https://arxiv.org/abs/2105.00830>

Limitations

Grammatical ambiguity: same word may have different roles in a sentence

"I saw the saw."

First "saw" – verb, second "saw" – noun (different sets of grammar rules for each instance of "saw" – semantic/word sense disambiguation)

Subject-object ambiguity: a specific case of grammatical ambiguity which refers to the potential interchangeability of the subject and object in a sentence

["mouse", "a", "the", "caught", "cat"]

Result 1: "The cat caught a mouse."

Result 2: "The mouse caught a cat."

Both results are grammatically valid, but "The cat caught a mouse" is more contextually valid. Implementing grammatical and semantic disambiguation to solve these issues will be a product of our future work, along with extending the algorithm's generation capabilities to languages other than English (including those that require heavy morphology usage, such as Russian).

Link Grammar – Question Answering

Context: Identity and Relationships

A **mom** is a **human**. A **dad** is a **human**. A **mom** is a parent. A **dad** is a parent. A son is a child. A **daughter** is a child. A son is a **human**. A **daughter** is a **human**. **Mom** is a **human** now. **Dad** is a **human** now. **Mom** is a parent now. **Dad** is a parent now. Son is a child now. **Daughter** is a child now. Son is a **human** now. **Daughter** is a **human** now. **Mom was a daughter before**. **Dad** was a son before. **Mom** was not a parent before. **Dad** was not a parent before.

Question: **mom daughter**

Answer: [**Mom was a daughter before**., **Daughter was a mom before**.]

Question: **dad human**

Answer: [**Dad is a human**.]

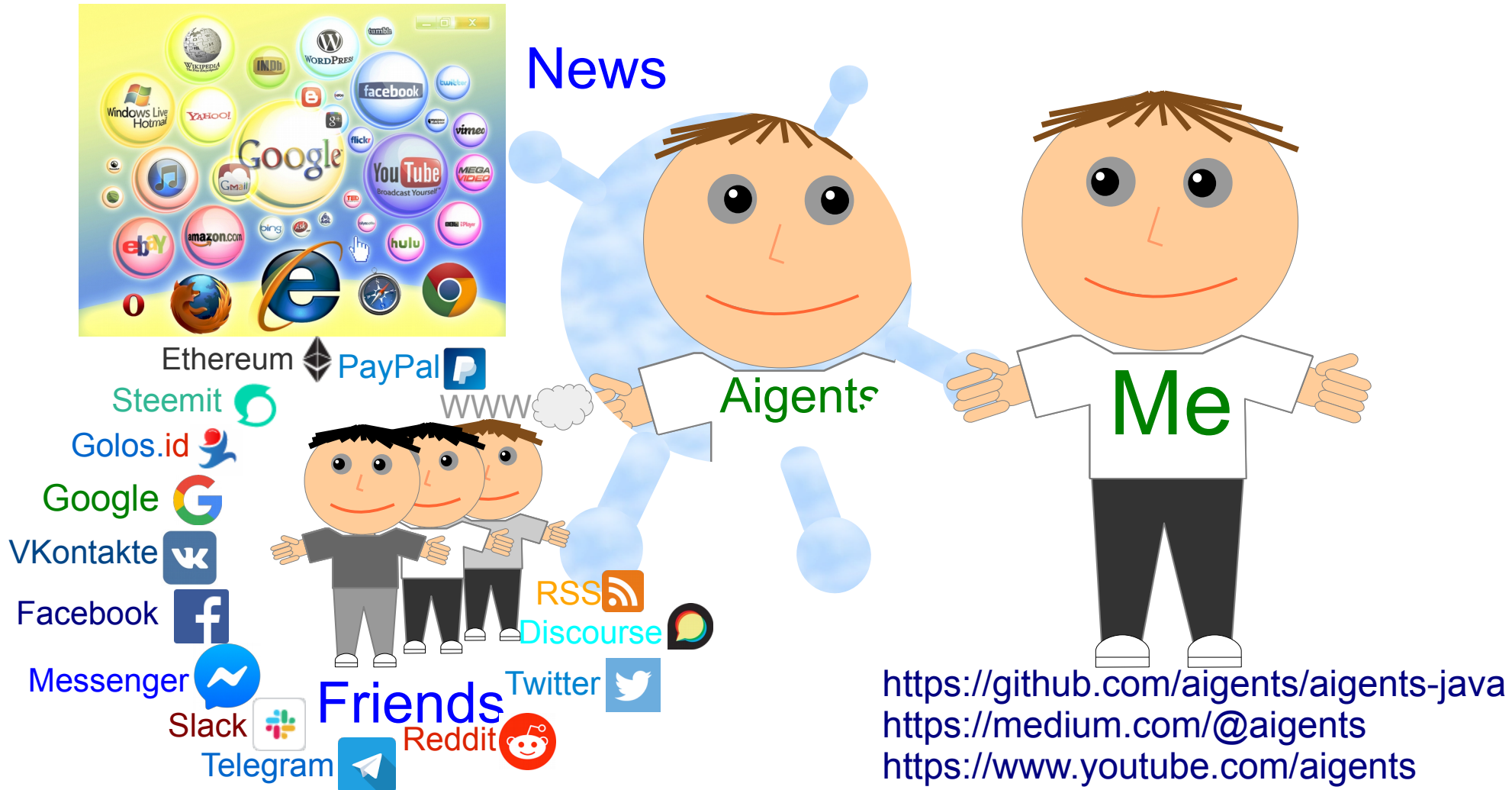
Table 1. Results when tested on 60 queries from SingularityNET's POC-English corpus.

Metric	Results				
	Ours	BERT	ELECTRA	DistilBERT	RoBERTa
BLEU	0.878	0.639	0.712	0.604	0.767
WVCS	0.944	0.606	0.741	0.595	0.799
WER	0.645	0.924	0.550	1.095	0.150
TER	0.166	0.381	0.342	0.457	0.245

<https://www.youtube.com/watch?v=MKIOqO9FRq0>

https://link.springer.com/chapter/10.1007/978-3-030-93758-4_22

Interpretable Sentiment Mining and Topic Matching in Aigents®



N-grams for Sentiment

The screenshot shows the Aigents website interface. At the top, there's a navigation bar with links to Aigents, Topics, Sites (26), News, Friends, Graph, and Chat. Below this is a search bar labeled "Input search text". The main content area displays a list of news items, each with a date filter (today, yesterday), a small image, a headline, and a URL. The items are as follows:

- today**: putin says he has noted joe biden's sharp anti-Russian rhetoric. https://www.reddit.com/r/JoeBiden/comments/j6zeex/putin_says_he_has_note
- today**: the united states of america is set for a pivotal election where joe bide current president donald trump. <https://moderndiplomacy.eu/2020/09/12/kamala-harris-as-vice-president-attractive-for>
- yesterday**: **learn more about blockchain domains:** introduction video f see you on thursday. https://www.reddit.com/r/CryptoCurrency/comments/j6rijh/ama_with_brad_ka
- yesterday**: The Russians may know more than we do : cia analyst says putin l doctors amid covid secrecy. https://www.reddit.com/r/politics/comments/j6t6yp/the_russians_may_know_more
- yesterday**: alexa-derived chatbot tech allegedly understands the british accent. https://www.theregister.co.uk/2020/09/17/amazon_lex_can_speak_british_english
- yesterday**: but not strongly enough to make a difference published: 16 mar 2 bernie sanders lost his last chance to take joe biden down. <https://www.theguardian.com/commentisfree/2020/mar/16/bernie-sanders-lost-h>
- yesterday**: it must also support in-space assembly manufacturing of hardware be tested and evaluated for future use. https://www.theregister.co.uk/2020/07/18/sierra_nevada_space_station

At the bottom, there is a copyright notice: "Copyright 2020 IP Anton Kolonin, Aigents®, Privacy Policy".

The screenshot shows the Aigents website interface. At the top, there's a navigation bar with links to Aigents, Topics, Sites (26), News, Friends, Graph, and Chat. Below this is a search bar labeled "Input new thing name or template". The main content area displays a list of news items, each with a date filter (today, yesterday), a small image, a headline, and a URL. The items are as follows:

- today**: putin says he has noted joe biden's sharp anti-Russian rhetoric. https://www.reddit.com/r/JoeBiden/comments/j6zeex/putin_says_he_has_note
- today**: the united states of america is set for a pivotal election where joe bide current president donald trump. <https://moderndiplomacy.eu/2020/09/12/kamala-harris-as-vice-president-attractive-for>
- yesterday**: **learn more about blockchain domains:** introduction video f see you on thursday. https://www.reddit.com/r/CryptoCurrency/comments/j6rijh/ama_with_brad_ka
- yesterday**: The Russians may know more than we do : cia analyst says putin l doctors amid covid secrecy. https://www.reddit.com/r/politics/comments/j6t6yp/the_russians_may_know_more
- yesterday**: alexa-derived chatbot tech allegedly understands the british accent. https://www.theregister.co.uk/2020/09/17/amazon_lex_can_speak_british_english
- yesterday**: but not strongly enough to make a difference published: 16 mar 2 bernie sanders lost his last chance to take joe biden down. <https://www.theguardian.com/commentisfree/2020/mar/16/bernie-sanders-lost-h>
- yesterday**: it must also support in-space assembly manufacturing of hardware be tested and evaluated for future use. https://www.theregister.co.uk/2020/07/18/sierra_nevada_space_station

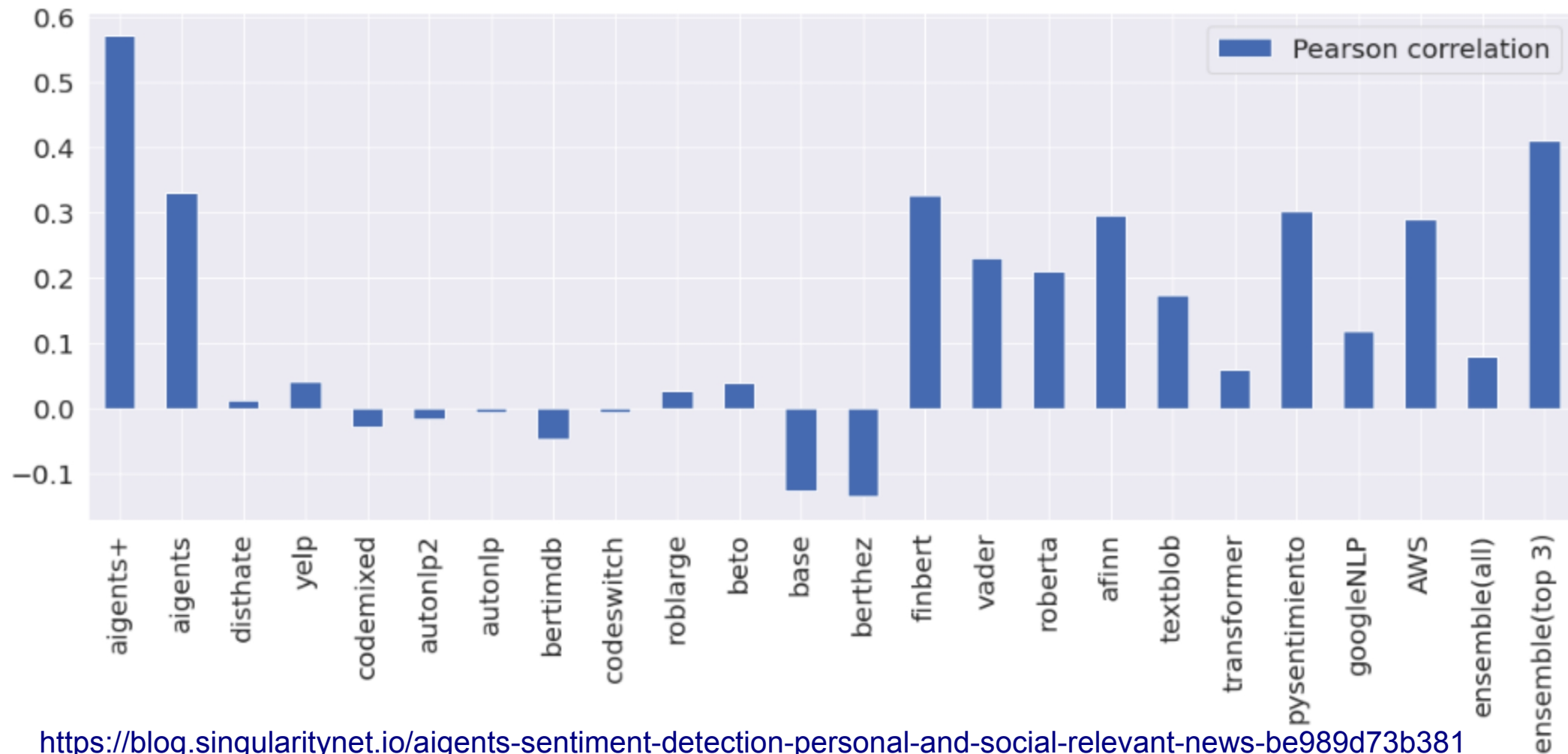
At the bottom, there is a copyright notice: "Copyright 2020 IP Anton Kolonin, Aigents®, Privacy Policy".

<https://blog.singularitynet.io/aigents-sentiment-detection-personal-and-social-relevant-news-be989d73b381>

Sentiment Analysis – Models' Fight

Out-of-the-box Aigents® “interpretable” model competes with Bert model fine-tuned on financial data

Average correlation across all models



<https://blog.singularitynet.io/aigents-sentiment-detection-personal-and-social-relevant-news-be989d73b381>

Sentiment and Behavioral Patterns Mining in Autonio Foundation and SingularityDAO



Following

Walmart to accept payments with cryptocurrencies using litecoin

cnb.cx/3A3cWuR



1:58 PM - 13 Sep 2021

CNBC deleted after 48 minutes

ID: 1437415272206450692

links in original tweet: <https://cnb.cx/3A3cWuR>

210 16:46



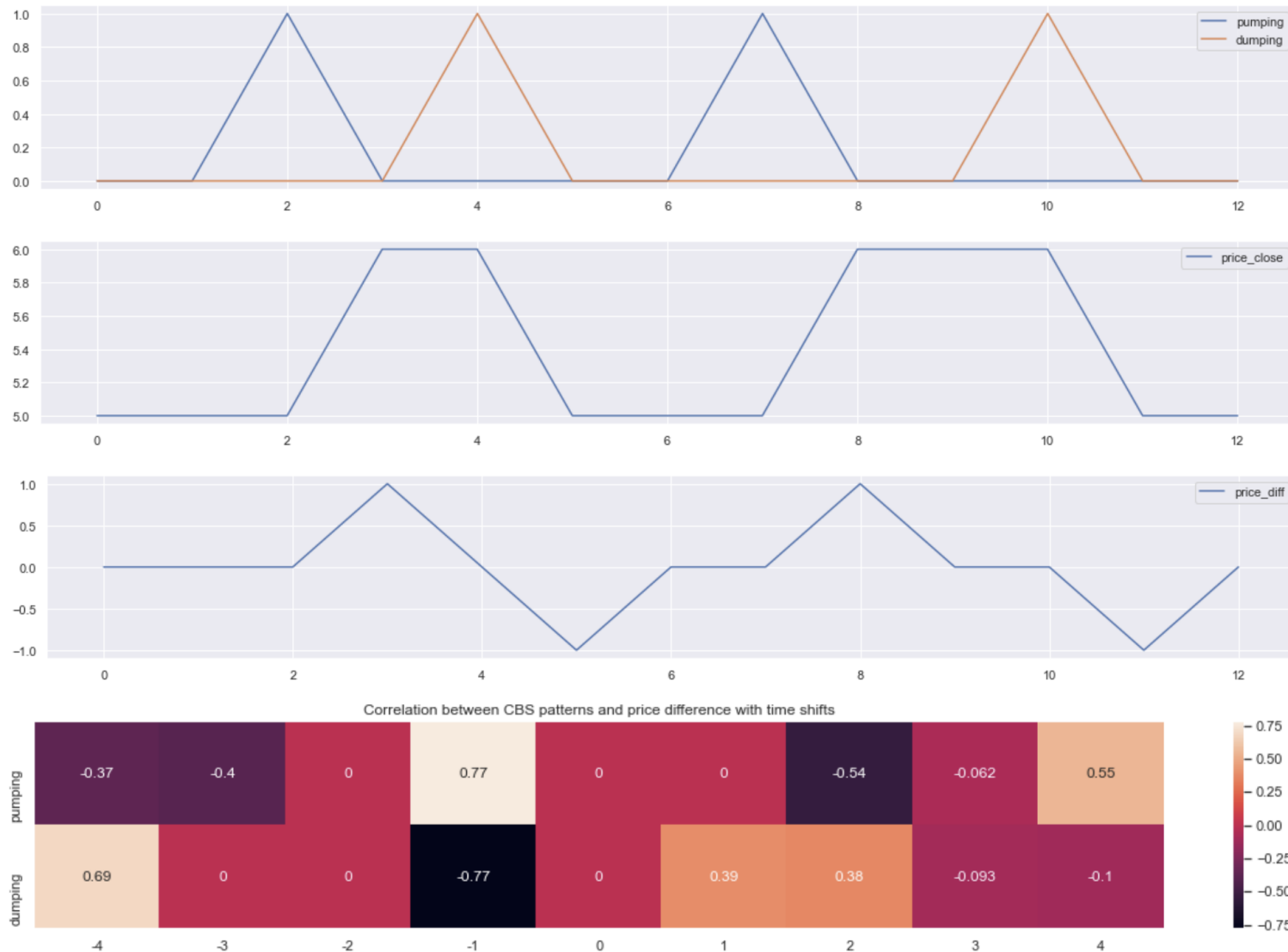
autonio.foundation



SingularityDAO

Sentiment for Market Price

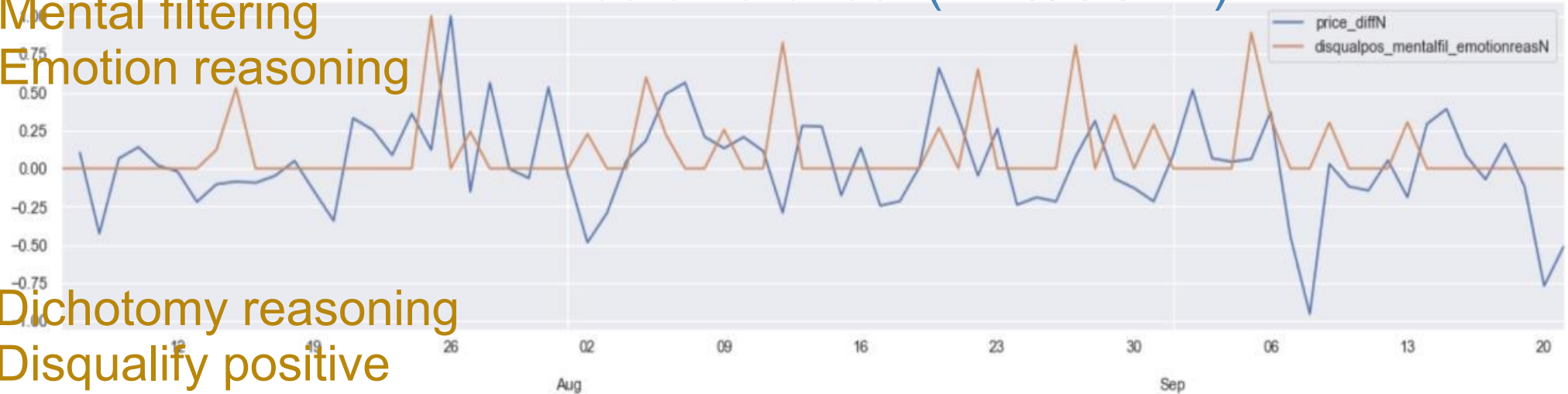
Reverse temporal correlation as causal connection between sentiment patterns and price change



Cognitive Distortions for FinTech

Disqualify positive
Mental filtering
Emotion reasoning

Price difference (BTC/USDT)



Dichotomy reasoning
Disqualify positive
Fortune telling
Mental filtering
Emotion reasoning

Trade count (BTC/USDT)

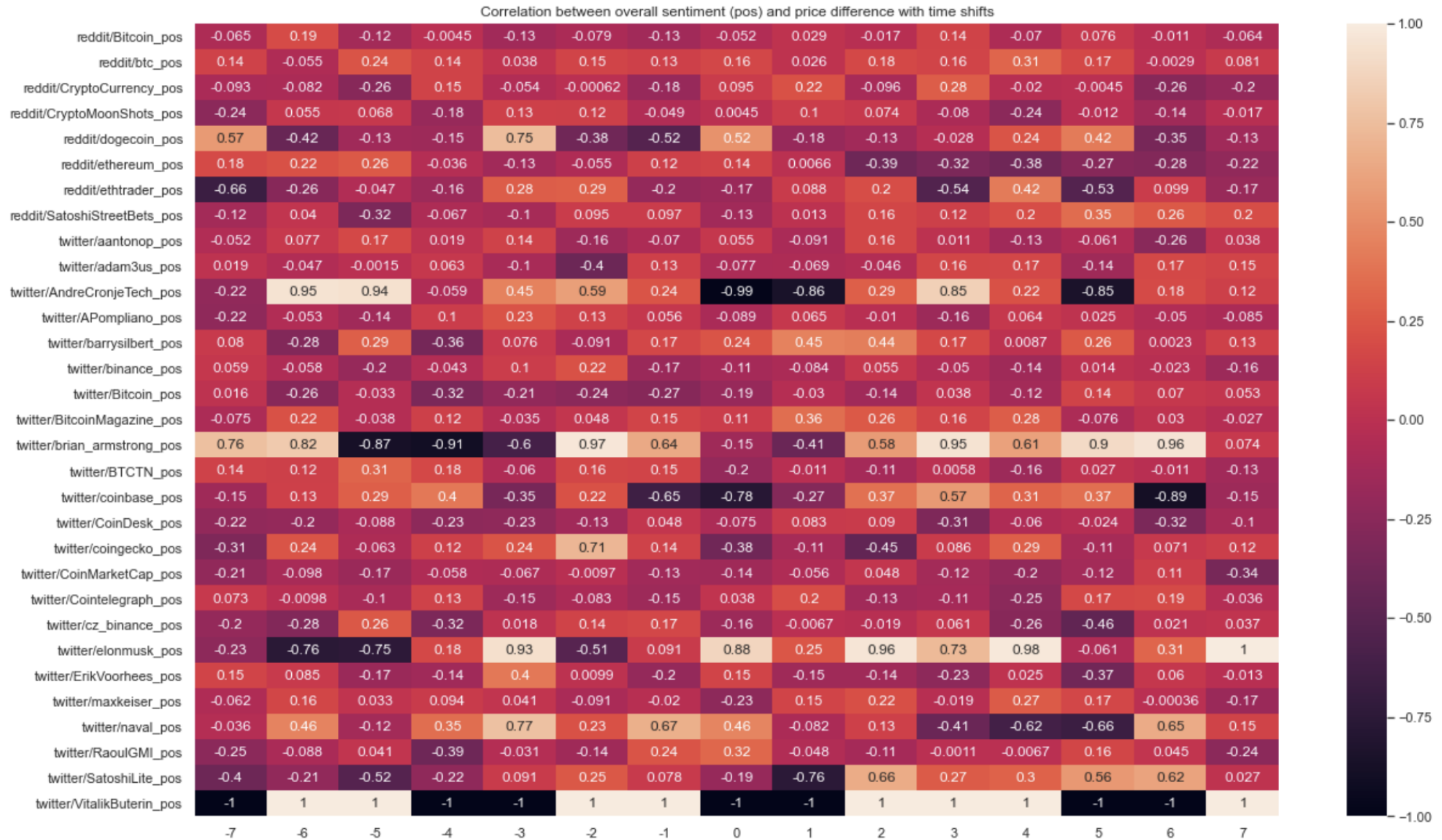


https://www.youtube.com/watch?v=5AoZ_hwPCko

<https://www.pnas.org/content/118/30/e2102061118>

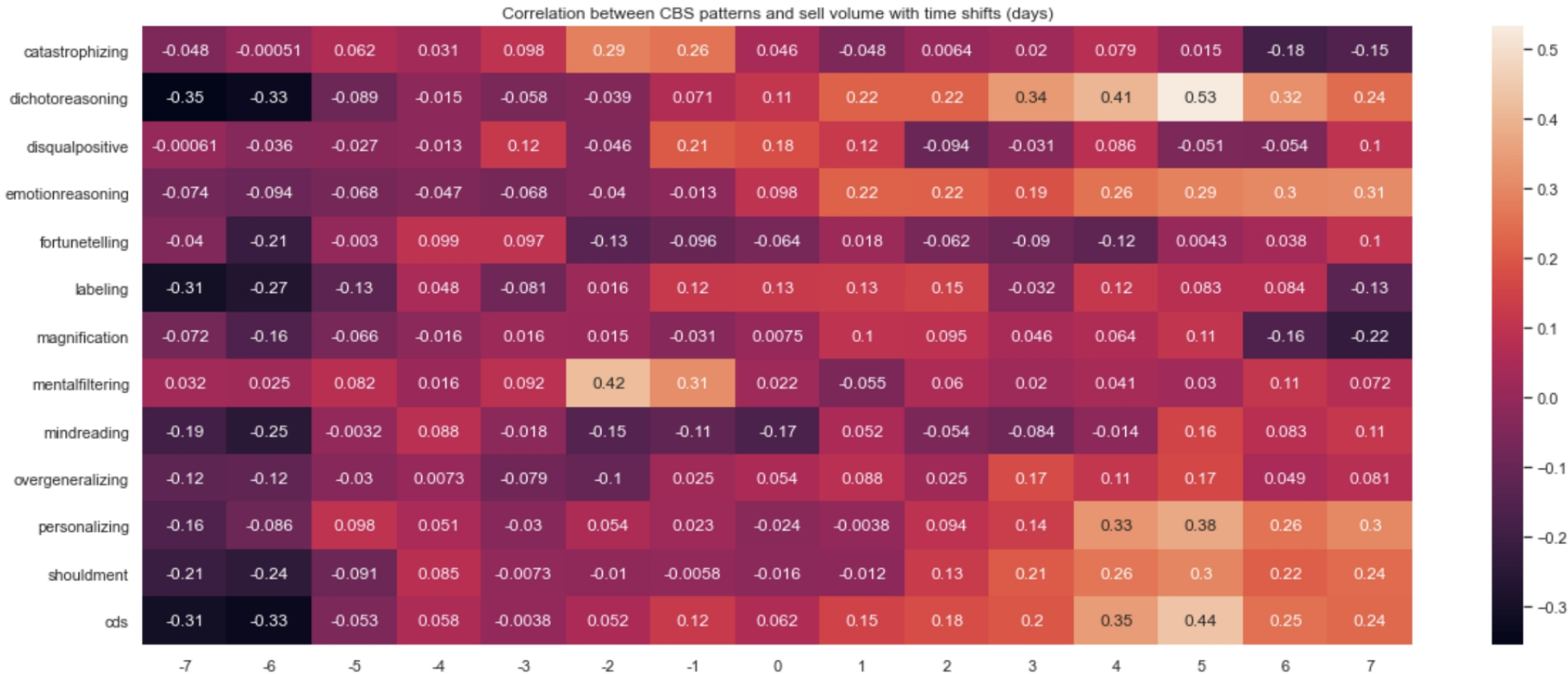
Sentiment for Market Price

Tracking Twitter and Reddit sentiment for BTC price



CBS patterns for Market Price

Tracking Twitter and Reddit Cognitive Behavioral Schemata patterns for BTC buy/sell volume



#Catastrophizing: Exaggerating the importance of negative events
distortions['catastrophizing'] = "will fail, will go wrong, will end, will be impossible, will not happen..."
#Mental Filtering: Paying too much attention to negative details instead of the whole picture
distortions['mentalfiltering'] = "I see only, all I see, all I can see, can only think, nothing good..."

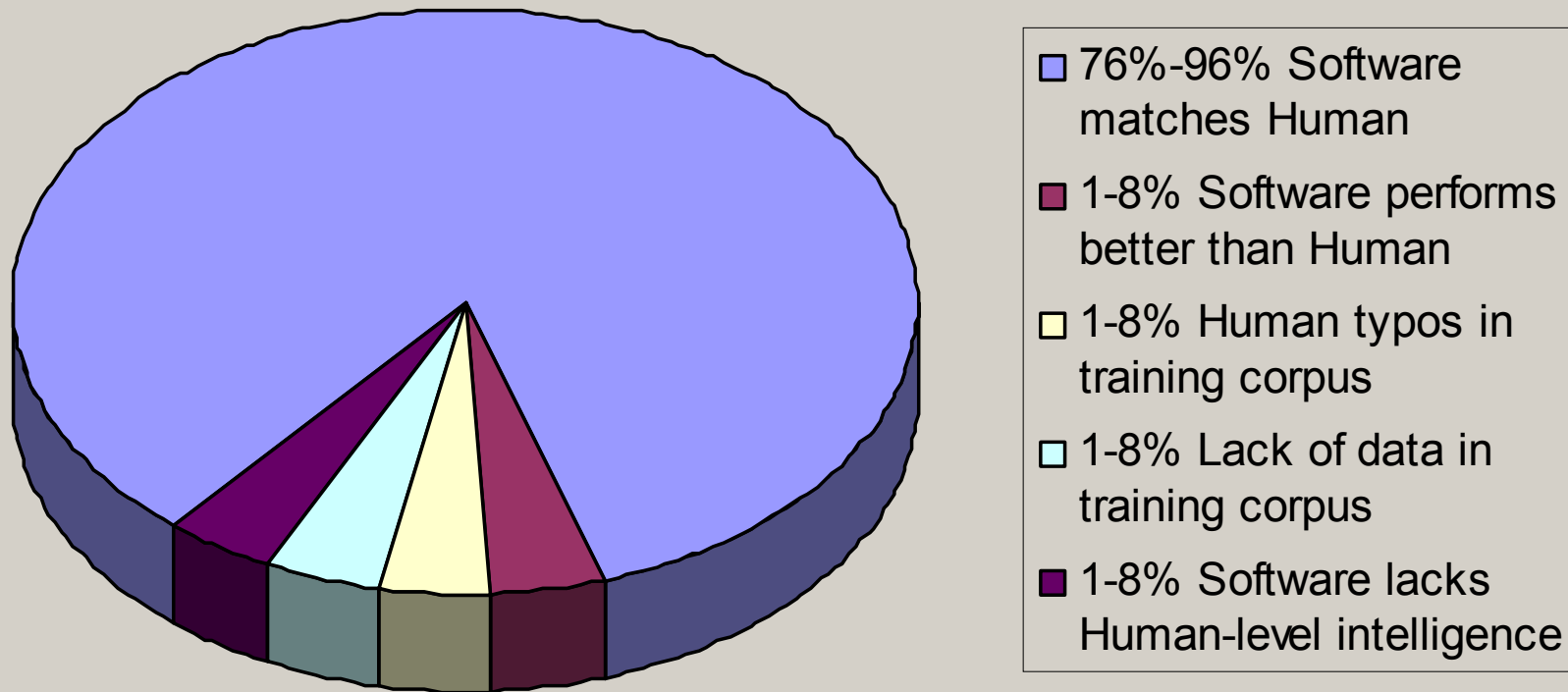
Why do we need Interpretable NLP (AI)?

Errors and blind spots in training data cause incorrect models.

Explainable AI enables to locate errors in the models.

Interpretable AI makes it possible to fix the models with no overtraining.

Ideally, we would like it to trace it back to training data and fix that...



<http://webstructor.net/papers/Kolonin-HP-ACA-IC-text.pdf>

Welcome to the Interpretable Natural Language Processing Community and Series of Workshops

Telegram English
<https://t.me/internlp>

Telegram Russian
<https://t.me/agibots>

Workshop 2022 (Seattle, WA, USA & Online)
<https://aigents.github.io/inlp/>

Workshop 2021
<https://aigents.github.io/inlp/2021/>

Thank You and Welcome!

Anton Kolonin

akolonin@aigents.com

Facebook: [akolonin](#)

Telegram: [akolonin](#)

