Interpretable Natural Language Processing
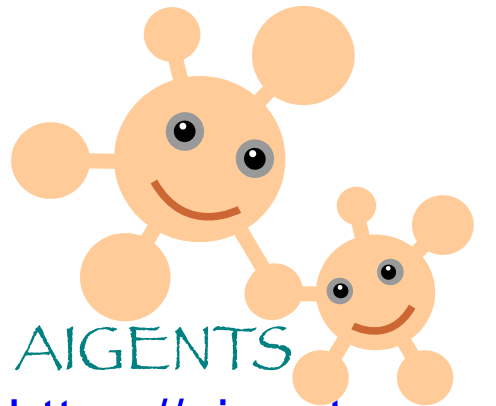
Anton Kolonin

akolonin@aigents.com
Facebook: akolonin
Telegram: akolonin

Novosibirsk State University
*THE REAL SCIENCE

AIGENTS
https://aigents.com

SingularityNET
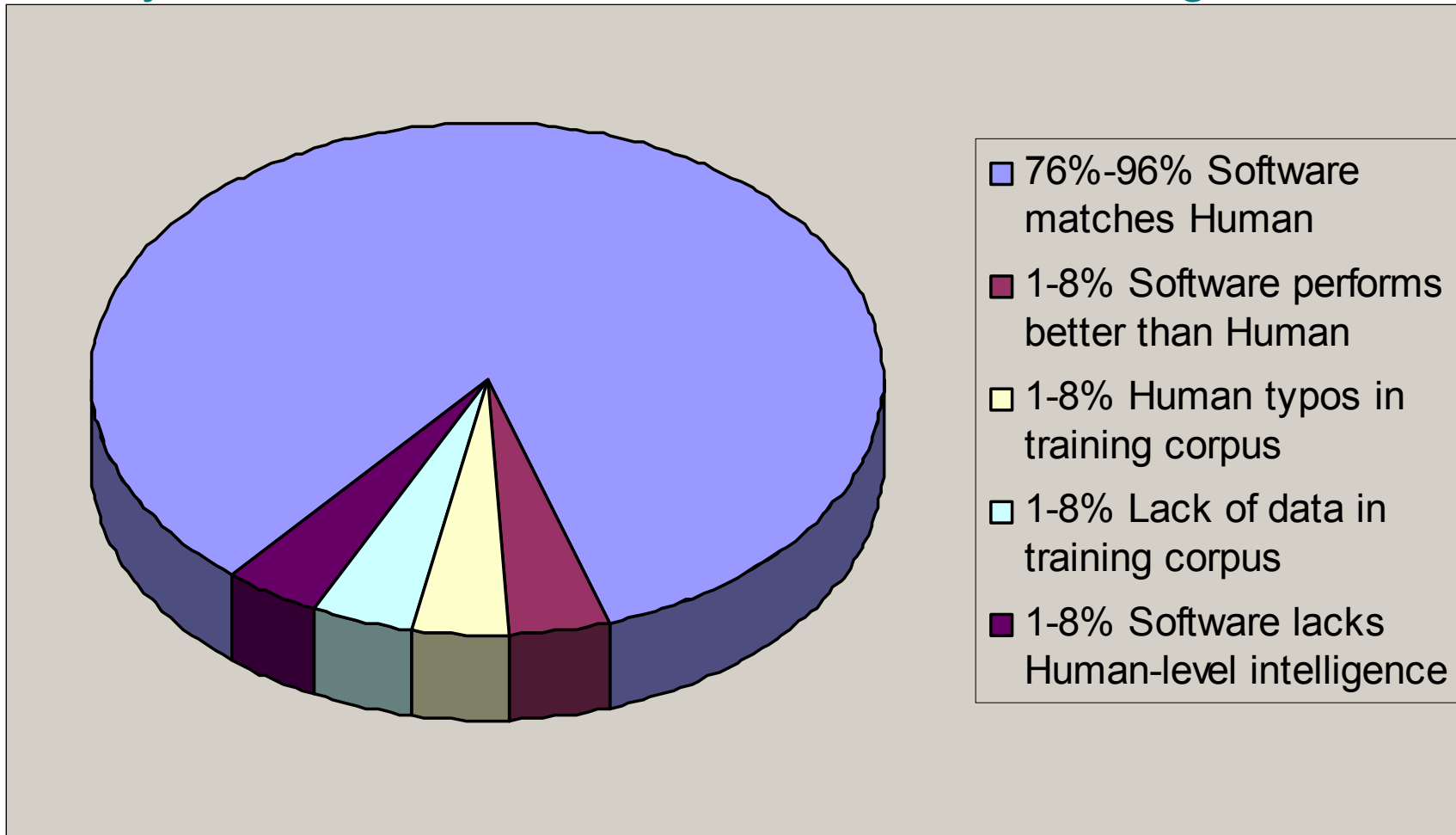https://singularitynet.io

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 brace it back to training data and fix that...

- 76%-96% Software matches Human
- 1-8% Software performs better than Human
- 1-8% Human typos in training corpus
- 1-8% Lack of data in training corpus
- 1-8% Software lacks Human-level intelligence

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

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.

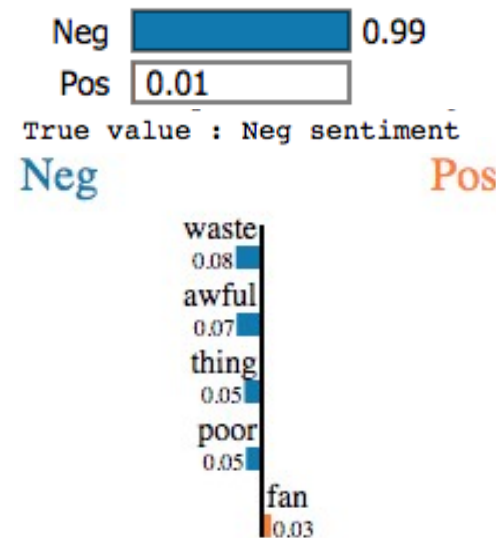## Explaining sentiment analysis model by LIME[1] method

LIME (Local interpretable model agnostic explanations)

- Model agnostic
- Learns local behaviour of black box model around an instance for a candidate linear interpretable model
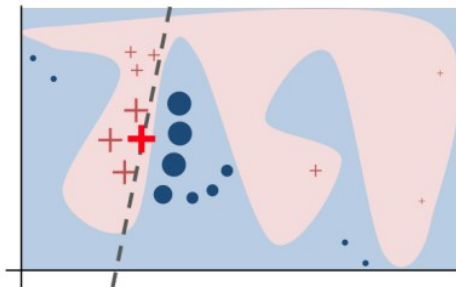
⬆ **Toy example showing LIME learning linear behaviour around class boundary**

Prediction Probabilities

Neg ▇▇▇▇▇▇▇ 0.99
Pos [0.01]

True value : Neg sentiment

Neg   Pos

waste
0.08
awful
0.07
thing
0.05
poor
0.05
fan
0.03

➤ **LIME** technique implemented on black box sentiment analysis classifier trained on IMDB movie dataset[2]. In the image, Lime estimates the model's decision making criteria for the prediction on a test instance.
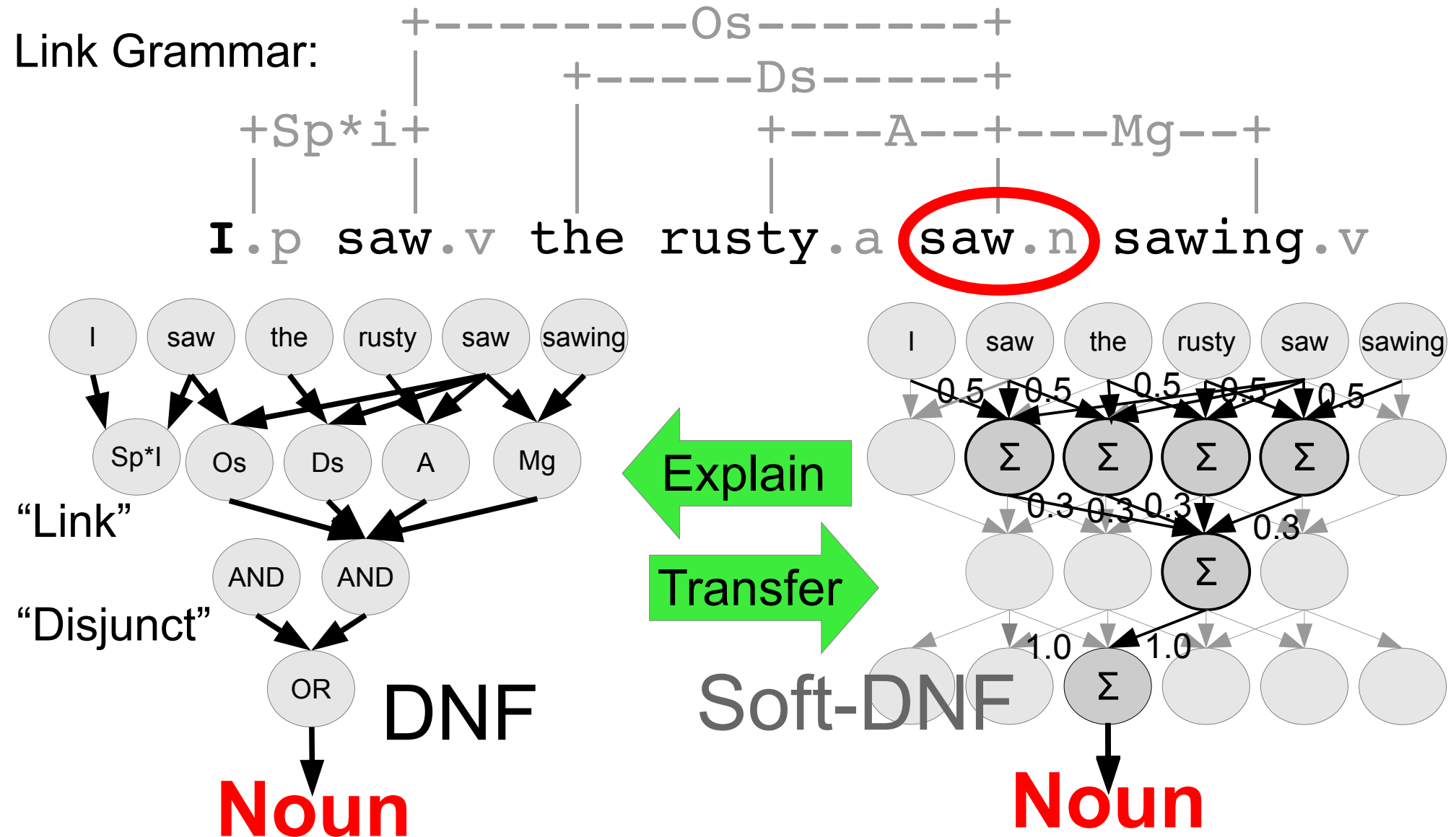
Text with highlighted words

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.|br /|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.|br /|lbr /|2/10

1 - Source: https://arxiv.org/abs/1602.04938
2 - Source: https://ai.stanford.edu/~amaas/data/sentiment/

Rohan K. Rathore, MS Student, Faculty of Mathematics & Mechanics, Novosibirsk State University

Bridging the Symbolic-Subsymbolic gap in NLP between distributed representations and formal grammars with ontologies

Link Grammar:

```
                +--------Os-------+
                |                 |
                |     +-----Ds----+
                |     |           |
      +Sp*i+    |     |   +---A--+---Mg--+
      |    |    |     |   |      |       |
   I.p saw.v the rusty.a saw.n sawing.v
```


"Link"

"Disjunct"

DNF

Soft-DNF

Explain

Transfer

Noun

Noun

Aigents® "Deep Patterns" - Language Model

Colored nodes may be thought as «activated» or «expressed»

acetaminophen may significantly reduce feelings of existential anxiety

Aigents® "Deep Patterns" - Text Mining

Classification

Category: "Healthcare"

tylenol
acetaminophen
placebo

Here's the Tylenol twist: Before they began writing, half of each group received acetaminophen while the other half swallowed a placebo. Even among those people who wrote about death, the Tylenol takers set bail at roughly $300—a sign that acetaminophen may significantly reduce feelings of existential anxiety, explains study lead author Daniel Randles, a PhD candidate in UBC's department of... psychology.

Case/Relationship Extraction

IS

Entity (Case): "Treatment: Healing anxiety with Tylenol"

HAS

significantly reduce feelings study

"acetaminophen may significantly reduce feelings of existential anxiety, explains study lead author Daniel Randles"

Property Attribution Entity Extraction

Brand: Tylenol
Substance: acetaminophen
Reliability: medium
Effect: positive
Diagnosis: Anxiety
Reporter: Daniel Randles

acetaminophen may reduce anxiety explains

acetaminophen may significantly reduce feelings of existential anxiety, explains study lead author Daniel Randles.

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.
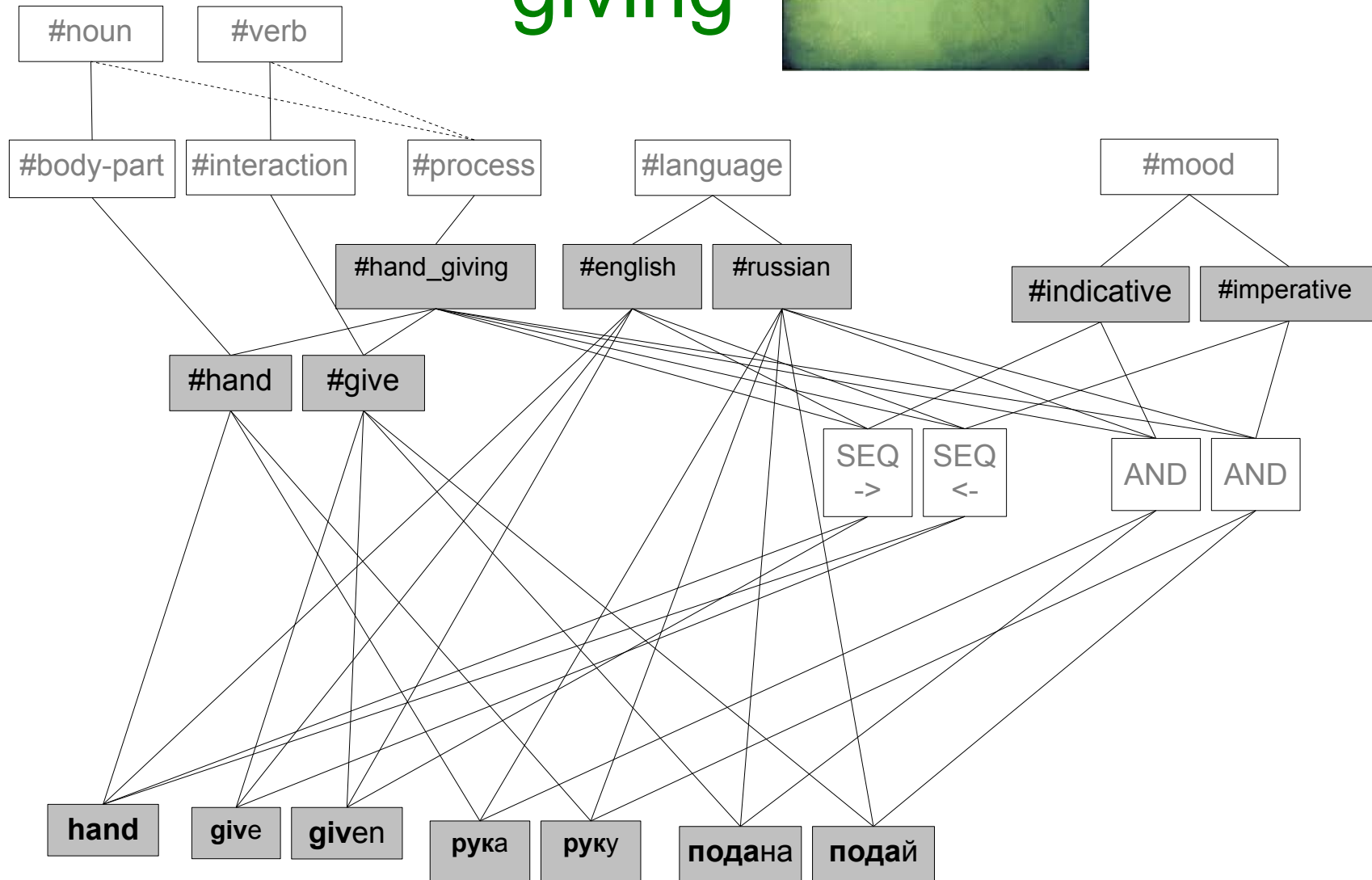   Small atraumatic soft tip. Ultra-thin 1 × 2 flat wire braid pattern"

=

{ **coating** : "hydrophillic", **description** : "convey guiding",
    pattern : "ultra-thin 1 × 2 flat wire braid", **tip** : "soft" }
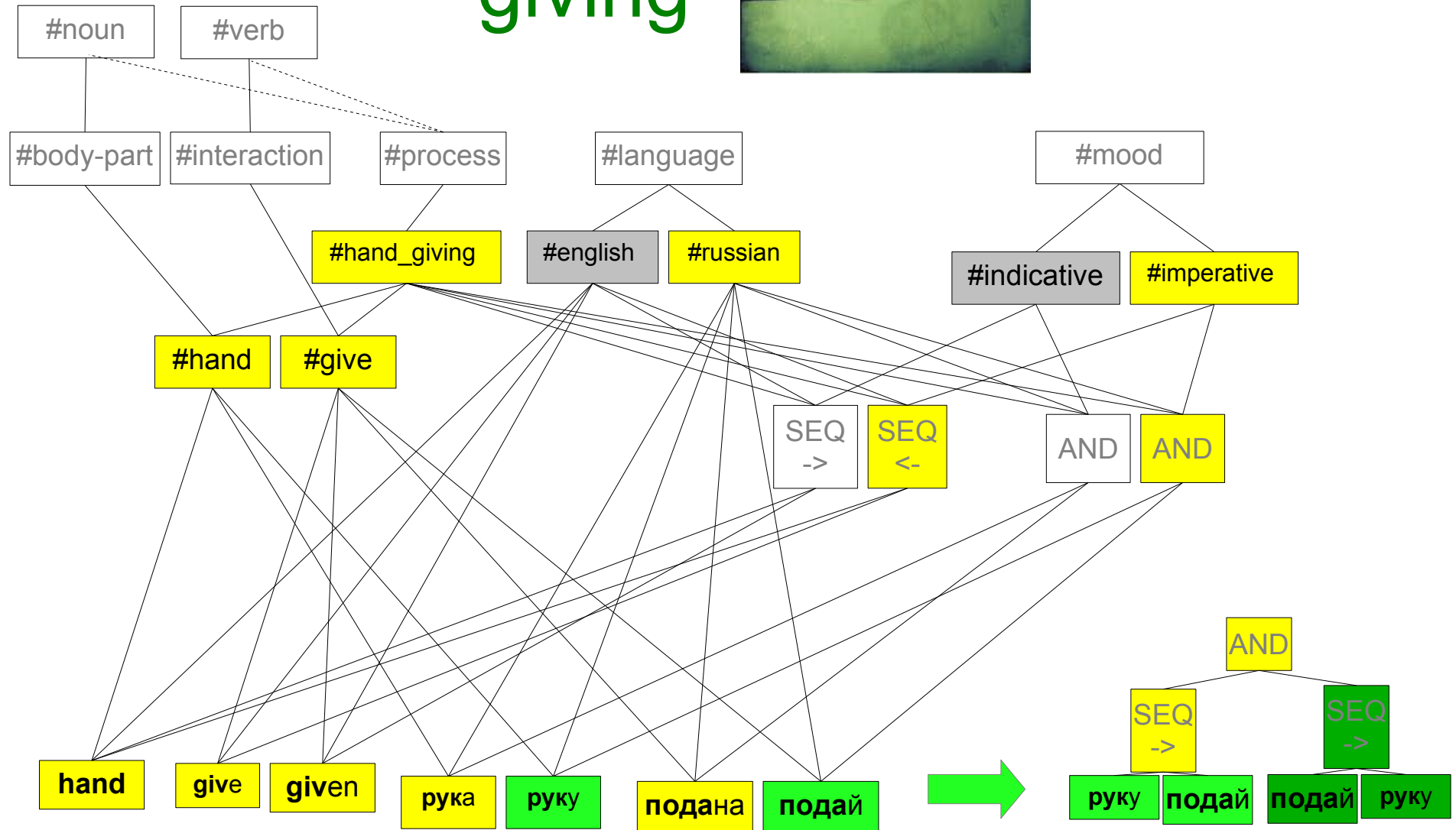
https://ieeexplore.ieee.org/document/7361868?arnumber=7361868
https://github.com/aigents/aigents-java

Grammar & Ontology Graph - Structure

Hand giving

#noun #verb

#body-part #interaction #process #language #mood

#hand_giving #english #russian #indicative #imperative

#hand #give

SEQ -> SEQ <- AND AND

hand give given рука руку подана подай

Grammar & Ontology Graph - Production

Hand giving

#noun #verb

#body-part #interaction #process #language #mood

#hand_giving #english #russian #indicative #imperative

#hand #give

SEQ -> SEQ <- AND AND

hand give given рука руку подана подай

AND

SEQ -> SEQ ->

руку подай подай руку

Grammar Learning from Scratch - Programmatically

I SAW THE SAW

PRONOUN **VERB** **ARTICLE** **NOUN**

SingularityNET

https://www.springerprofessional.de/unsupervised-language-learning-in-opencog/15995030
https://www.springerprofessional.de/en/programmatic-link-grammar-induction-for-unsupervised-language-le/17020348
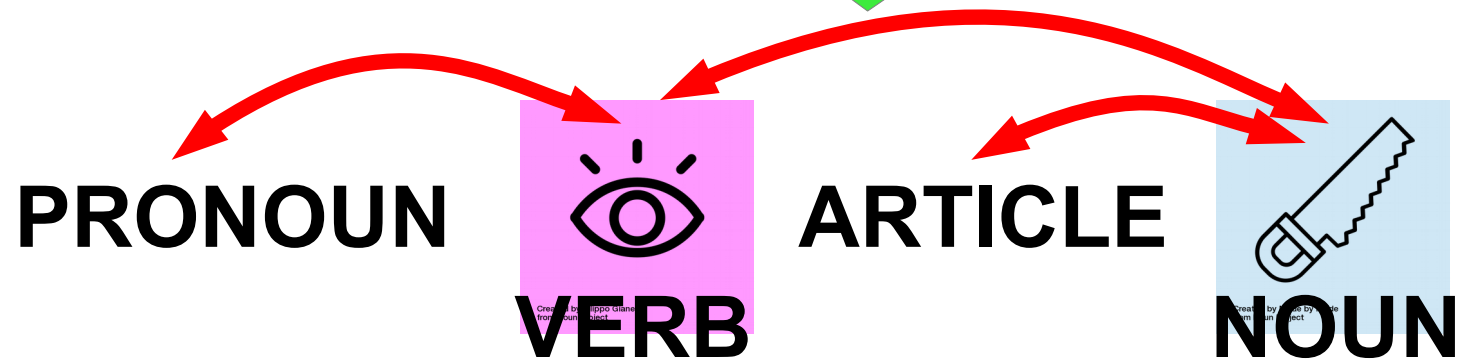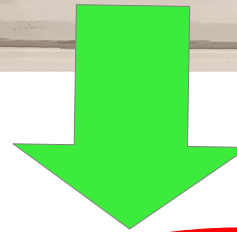
OpenCog Link Grammar Disjuncts & Connectors

SingularityNET
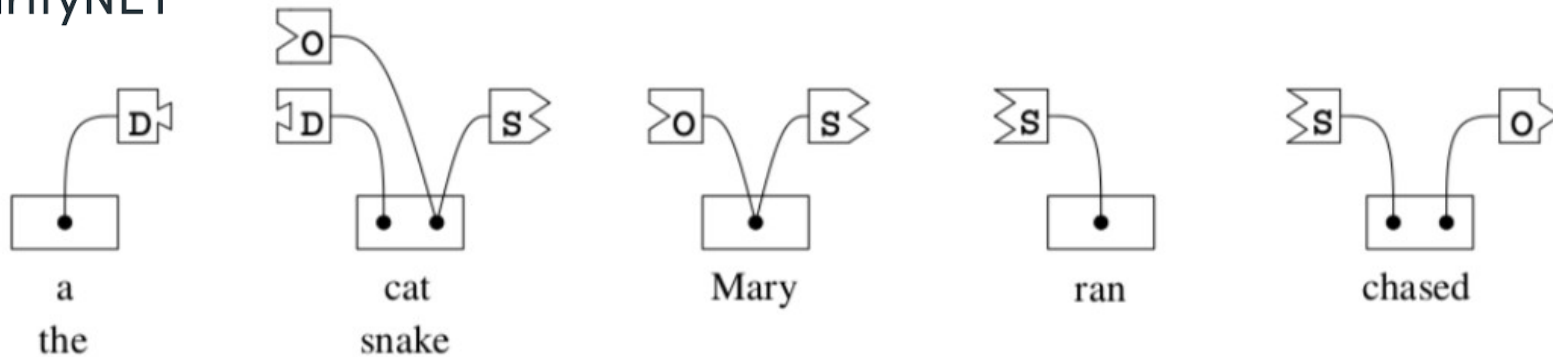
An illustration of Link Grammar connectors and disjuncts. The connectors are the jigsaw-puzzle-shaped pieces; connectors are allowed to connect only when the tabs fit together. A disjunct is the entire (ordered) set of connectors for a word. As lexical entries appearing in a dictionary, the above would be written as
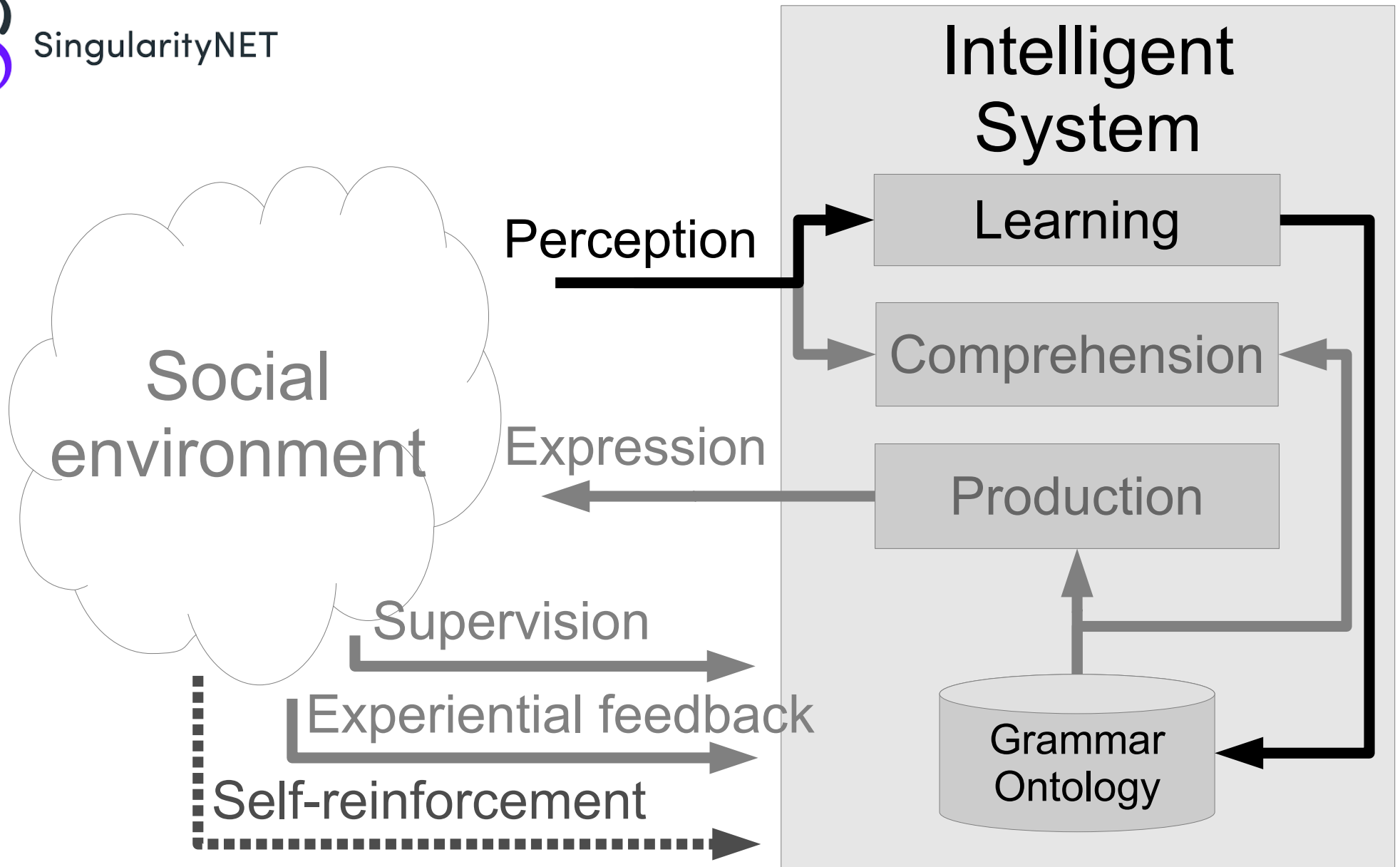
$$a \quad the: \quad D+;$$
$$cat \quad snake: \quad D- \,\& \,(S+ \,or \,O-);$$
$$Mary: \quad O- \,or \,S+;$$
$$ran: \quad S-;$$
$$chased \quad S- \,\& \,O+;$$

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:
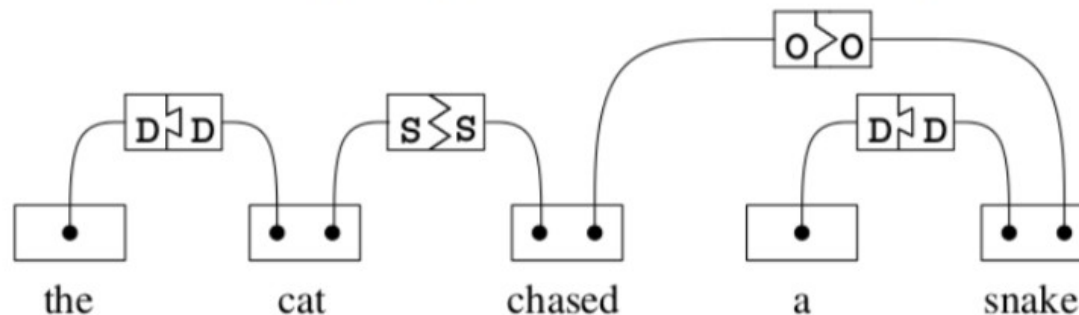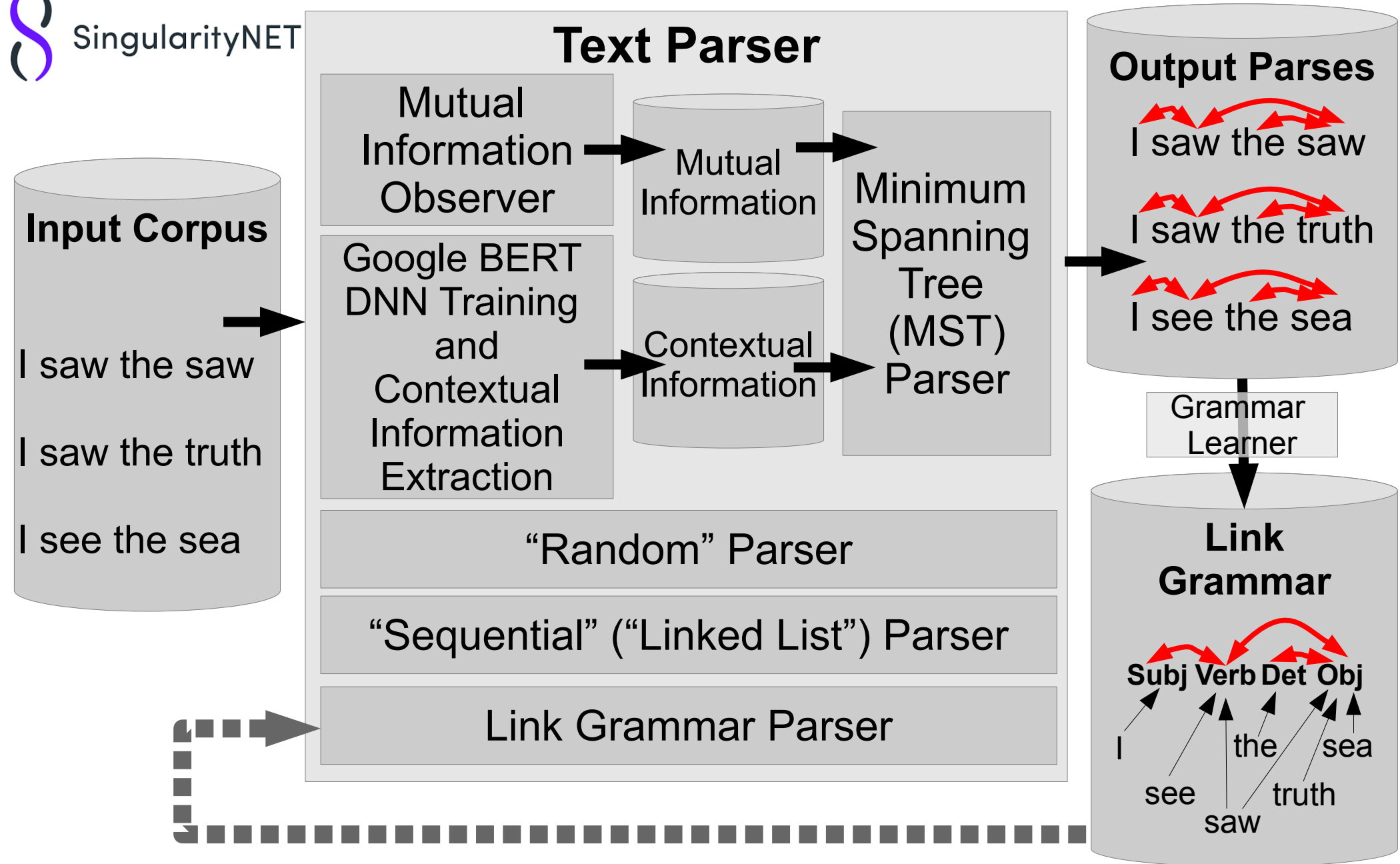
B. Goertzel,
L. Vepstas,
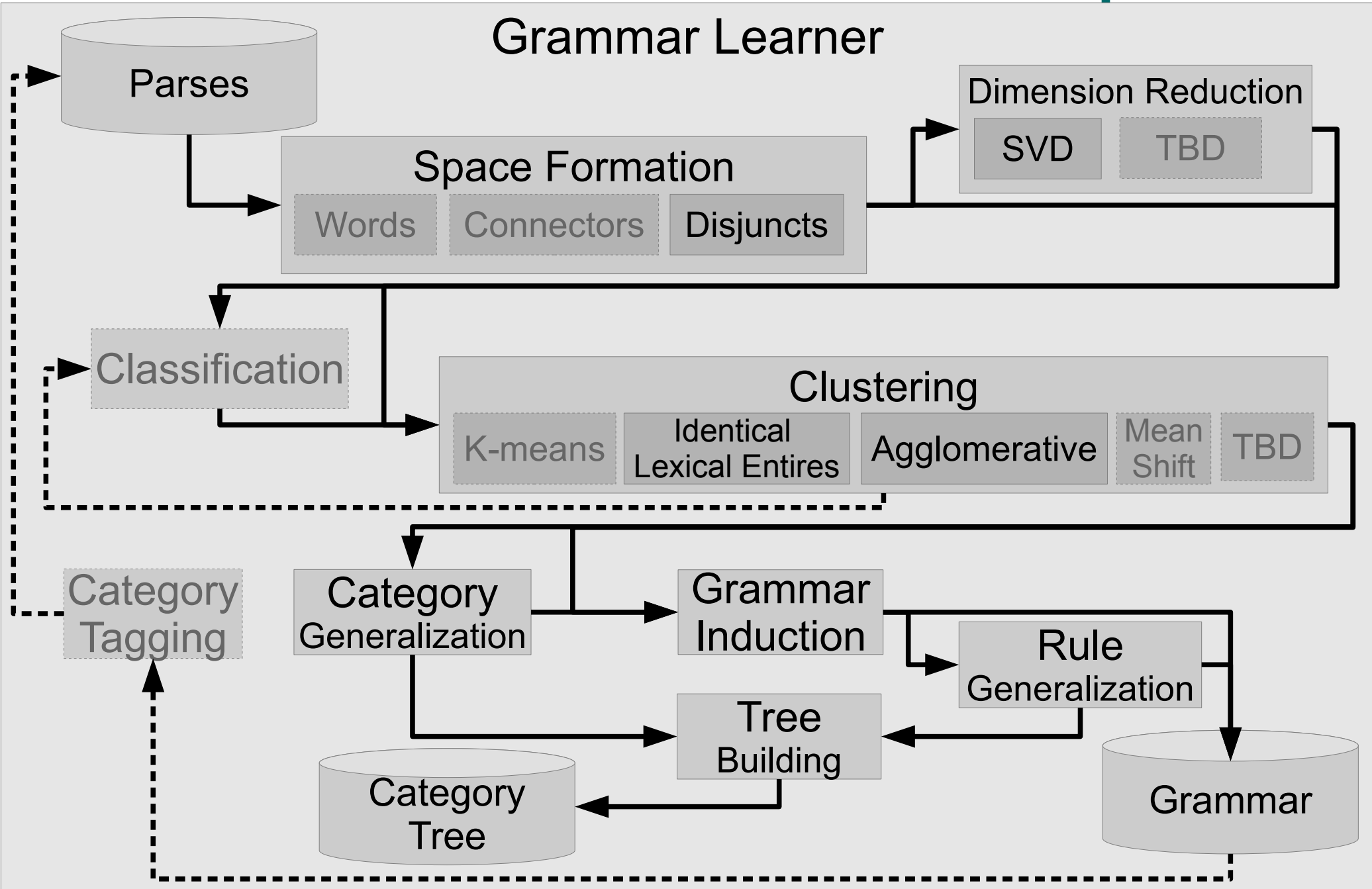2014

Language Learning Environment

SingularityNET

Intelligent System

Perception → Learning

Comprehension

Social environment

Expression ← Production

Supervision

Experiential feedback

Self-reinforcement

Grammar Ontology

Text Parsing for Link Grammar

SingularityNET

Input Corpus

I saw the saw

I saw the truth

I see the sea

Text Parser

Mutual Information Observer

Google BERT DNN Training and Contextual Information Extraction

Mutual Information

Contextual Information

Minimum Spanning Tree (MST) Parser

"Random" Parser

"Sequential" ("Linked List") Parser

Link Grammar Parser

Output Parses

I saw the saw

I saw the truth

I see the sea

Grammar Learner

Link Grammar

Subj Verb Det Obj

I   see   the   sea
      saw   truth

Link Grammar Learner Pipeline

Grammar Learner

Parses

Dimension Reduction
- SVD
- TBD

Space Formation
- Words
- Connectors
- Disjuncts

Classification

Clustering
- K-means
- Identical Lexical Entires
- Agglomerative
- Mean Shift
- TBD

Category Tagging

Category Generalization

Grammar Induction

Rule Generalization

Tree Building

Category Tree

Grammar

Corpora in Use

SingularityNET

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/down-loads/babi/

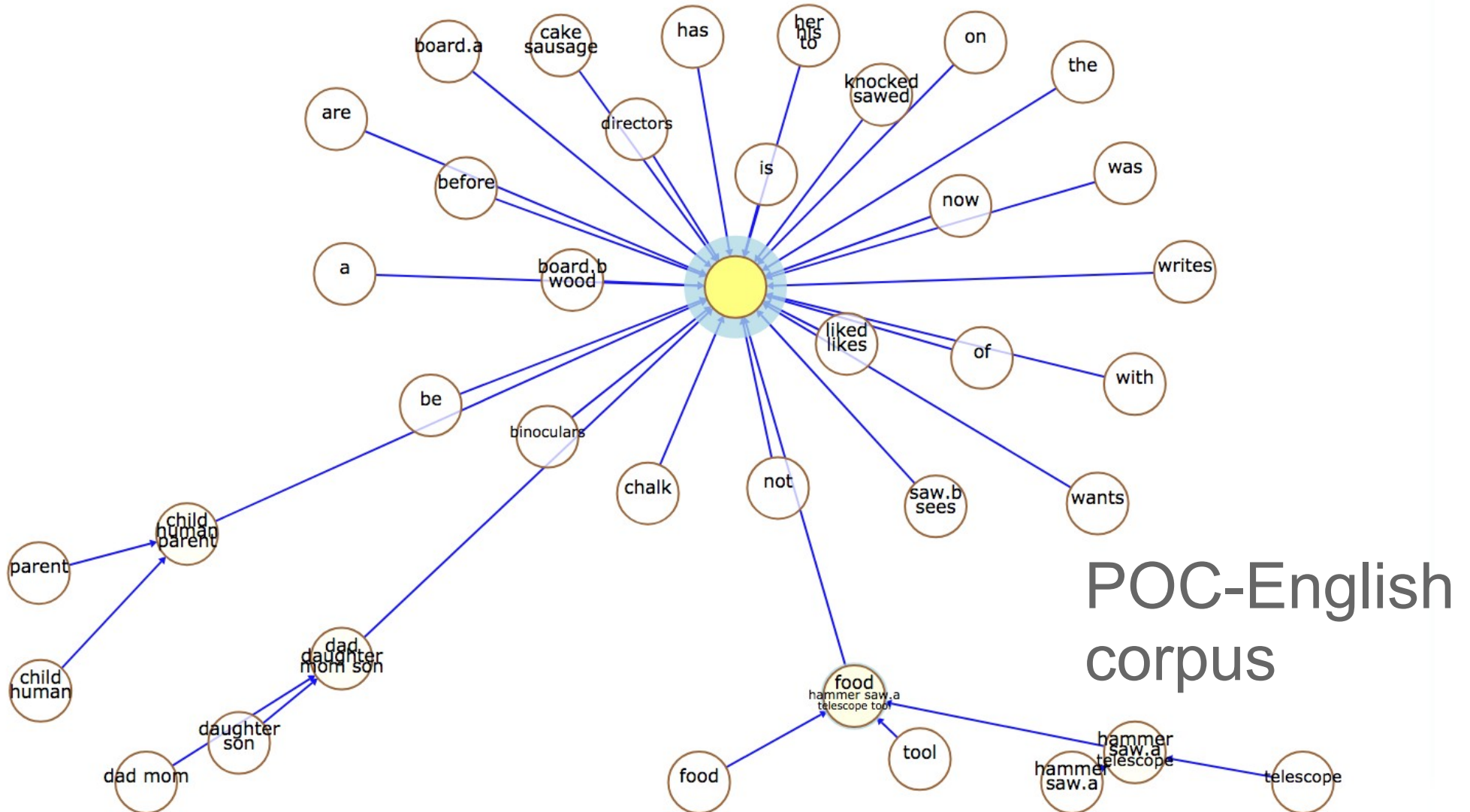# OpenCog Unsupervised Language Learning of Grammatical Categories and Link Grammar Dictionaries
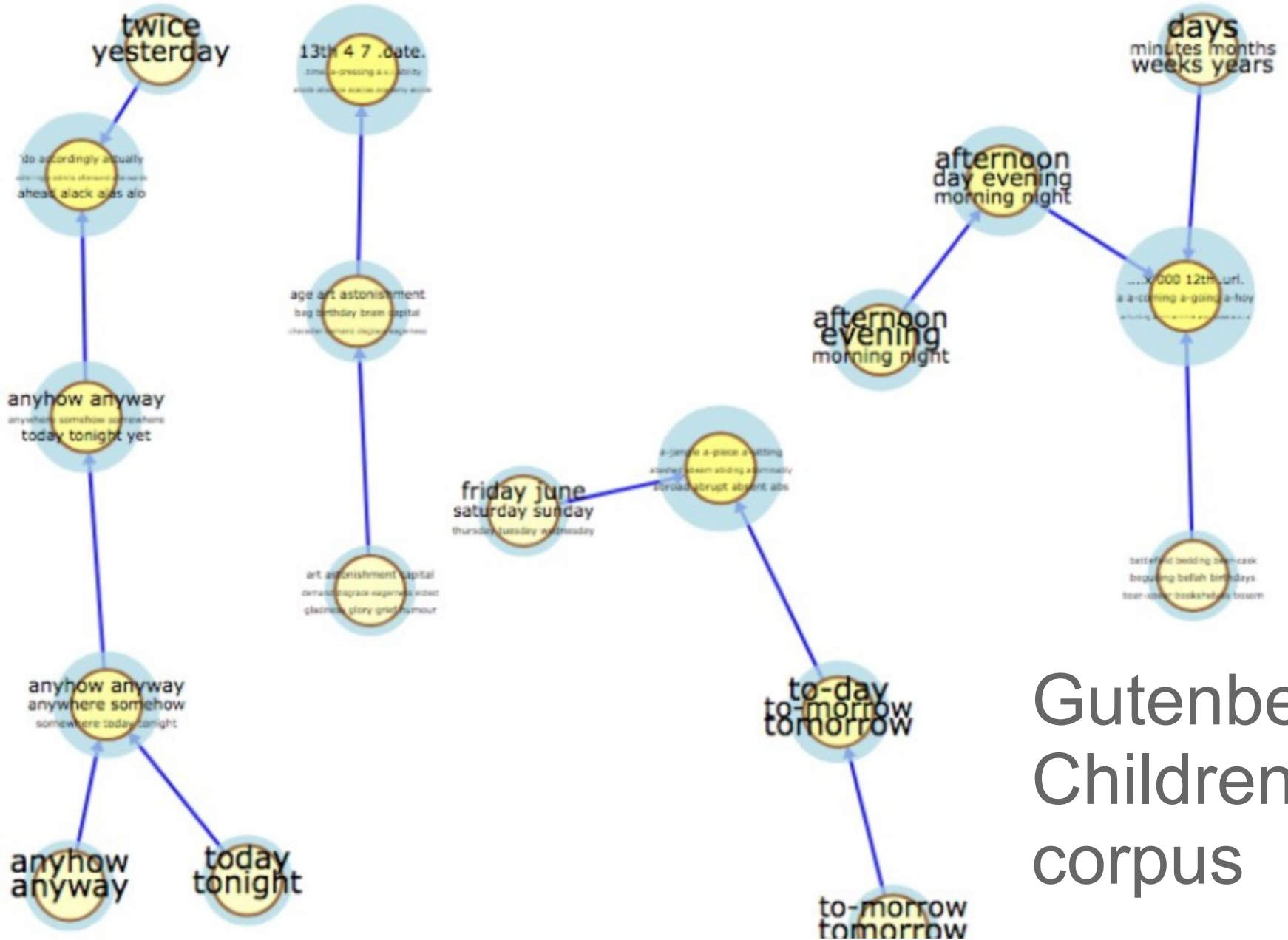

POC-English (Connectors)

POC-English (Disjuncts)
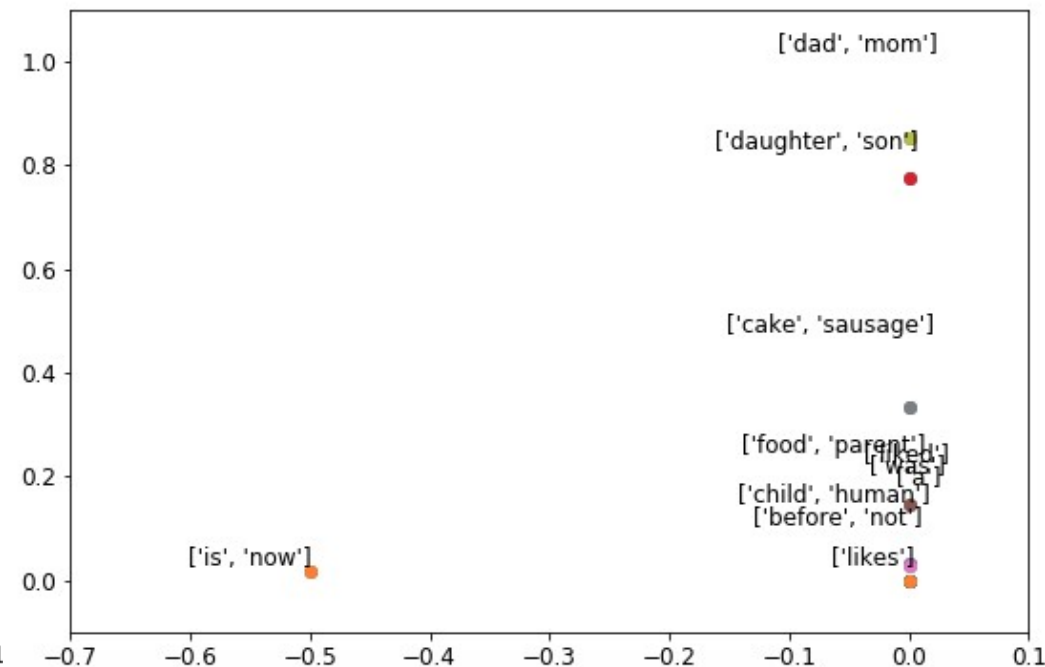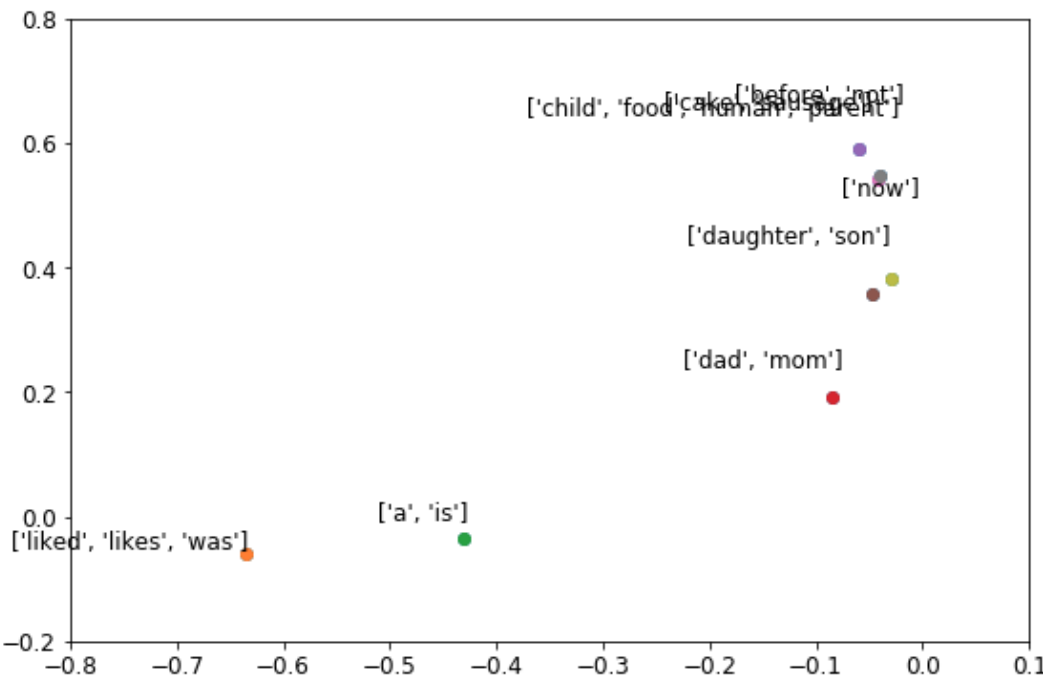
Grammar Ontology from Parses
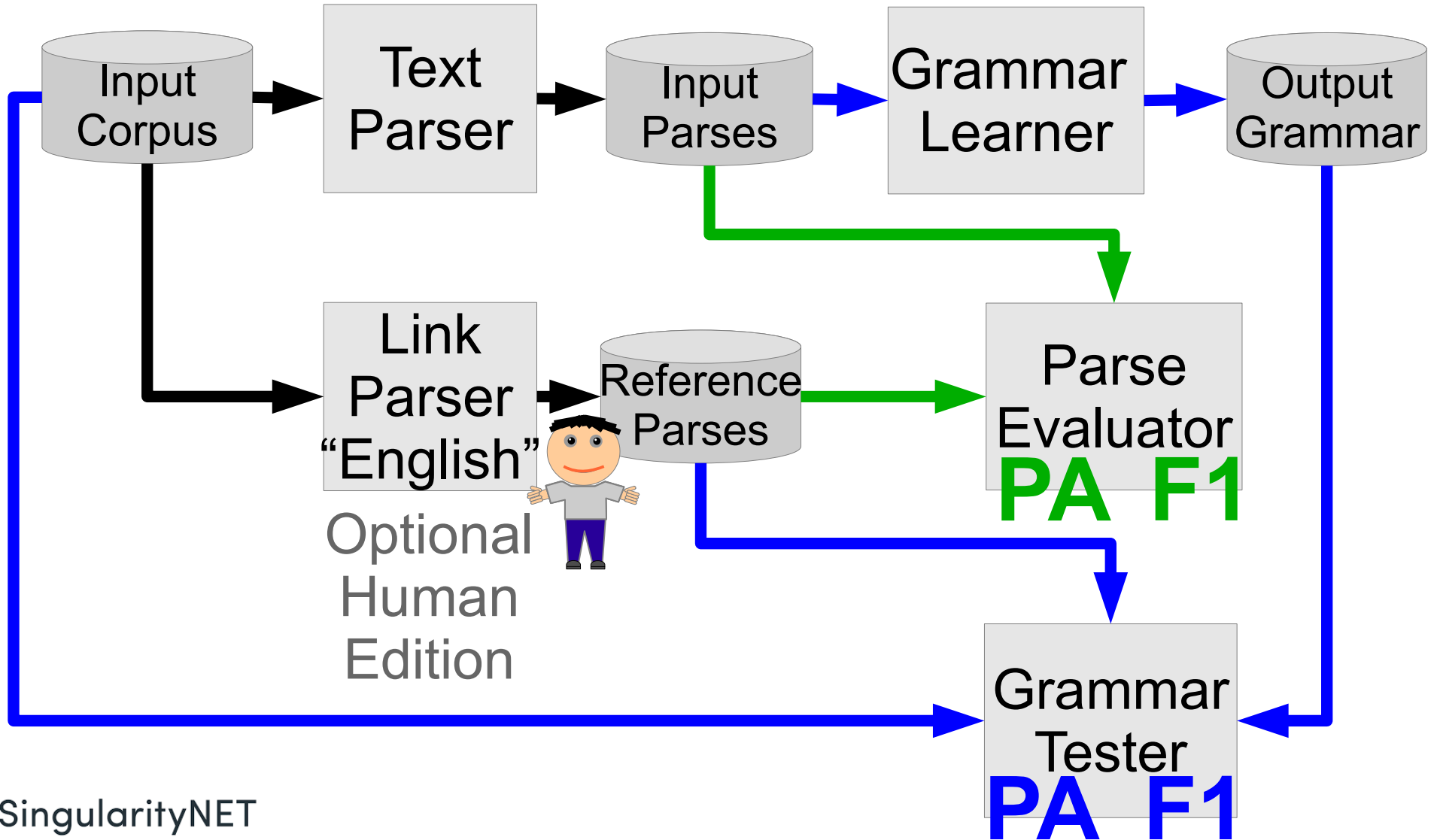

POC-English corpus

Grammar Ontology from Parses

Gutenberg Children corpus

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

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"

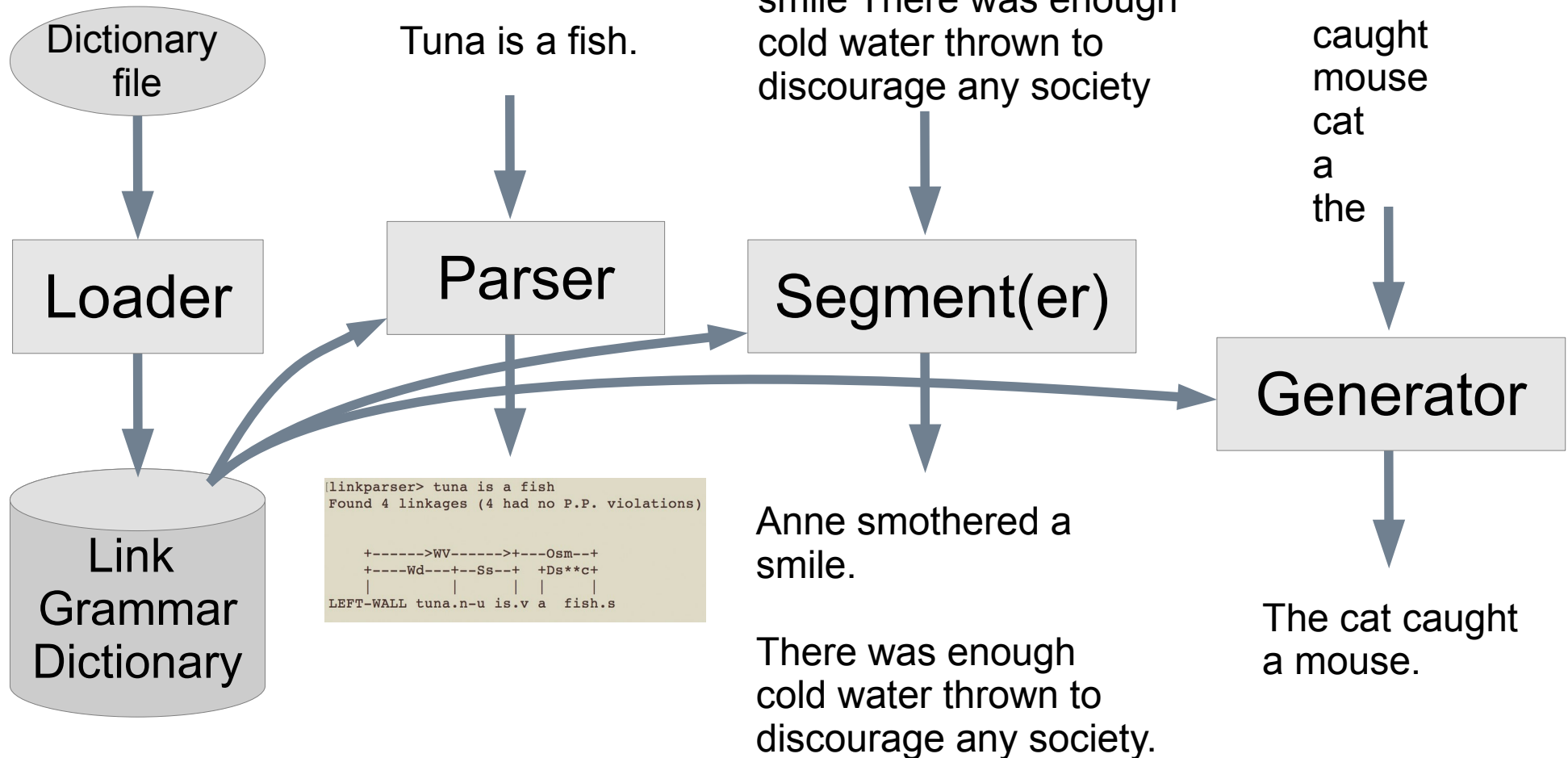 - Better grammars for larger corpora

  - Incremental Grammar Learning

SingularityNET

Link Grammar – beyond parsing

Parsing Segmentation Grammatical Generation

Dictionary file

Tuna is a fish.

Anne smothered a smile There was enough cold water thrown to discourage any society

caught
mouse
cat
a
the

Loader

Parser

Segment(er)

Generator

Link Grammar Dictionary

```
[linkparser> tuna is a fish
Found 4 linkages (4 had no P.P. violations)


    +------->WV------->+---Osm--+
    +----Wd---+--Ss--+  +Ds**c+
    |         |      |    |    |
LEFT-WALL tuna.n-u is.v a  fish.s
```

Anne smothered a smile.

There was enough cold water thrown to discourage any society.

The cat caught a mouse.

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

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
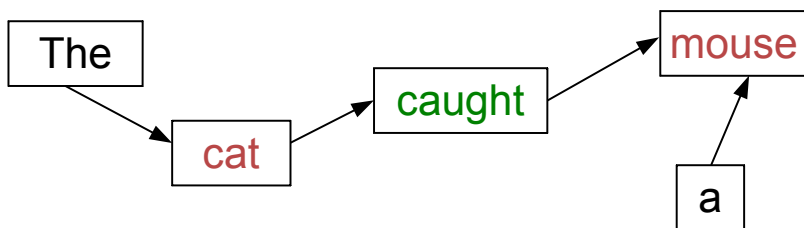
Algorithm 2: CONNECTS

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:

```
         +---Osm--+
         |  +Ds**c+
         |  |      |
       is.v  a  human.n
```
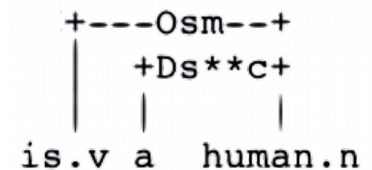
Grammatical Generation - Results

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/down-loads/babi/

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

Grammatical Generation Applications and Future Work

Applications

Question Answering

Natural Language Comprehension (NLC)

Parsing the question (input query) based on formal grammar (Link Grammar) → Performing semantic interpretation (extract concept represented by query and determine relationships between parts of the query)

Natural Language Generation (NLG)

Semantic query execution (determining answer to query based on relationships from the NLC step) → Construct grammatically valid sentences from the words associated with the semantic relationships derived during query execution

We are here

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

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).

N-grams for Sentiment

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

Aigents® for News Management

serving as a **crystal ball** in the world of social and online networks to recognize one's preferences, find what they need and help managing connections.

News

Aigents

Me

Ethereum
PayPal
Steemit
WWW
Golos.id
Google
VKontakte
Facebook
RSS
Discourse
Messenger
Twitter
Slack
Friends
Reddit
Telegram

https://github.com/aigents/aigents-java
https://medium.com/@aigents
https://www.youtube.com/aigents

Thank You and Welcome!

Anton Kolonin

akolonin@aigents.com
Facebook: akolonin
Telegram: akolonin

Novosibirsk State University
*THE REAL SCIENCE

AIGENTS®
https://aigents.com

SingularityNET
https://singularitynet.io