

Explainable Language Processing with Link Grammar and Deep Patterns

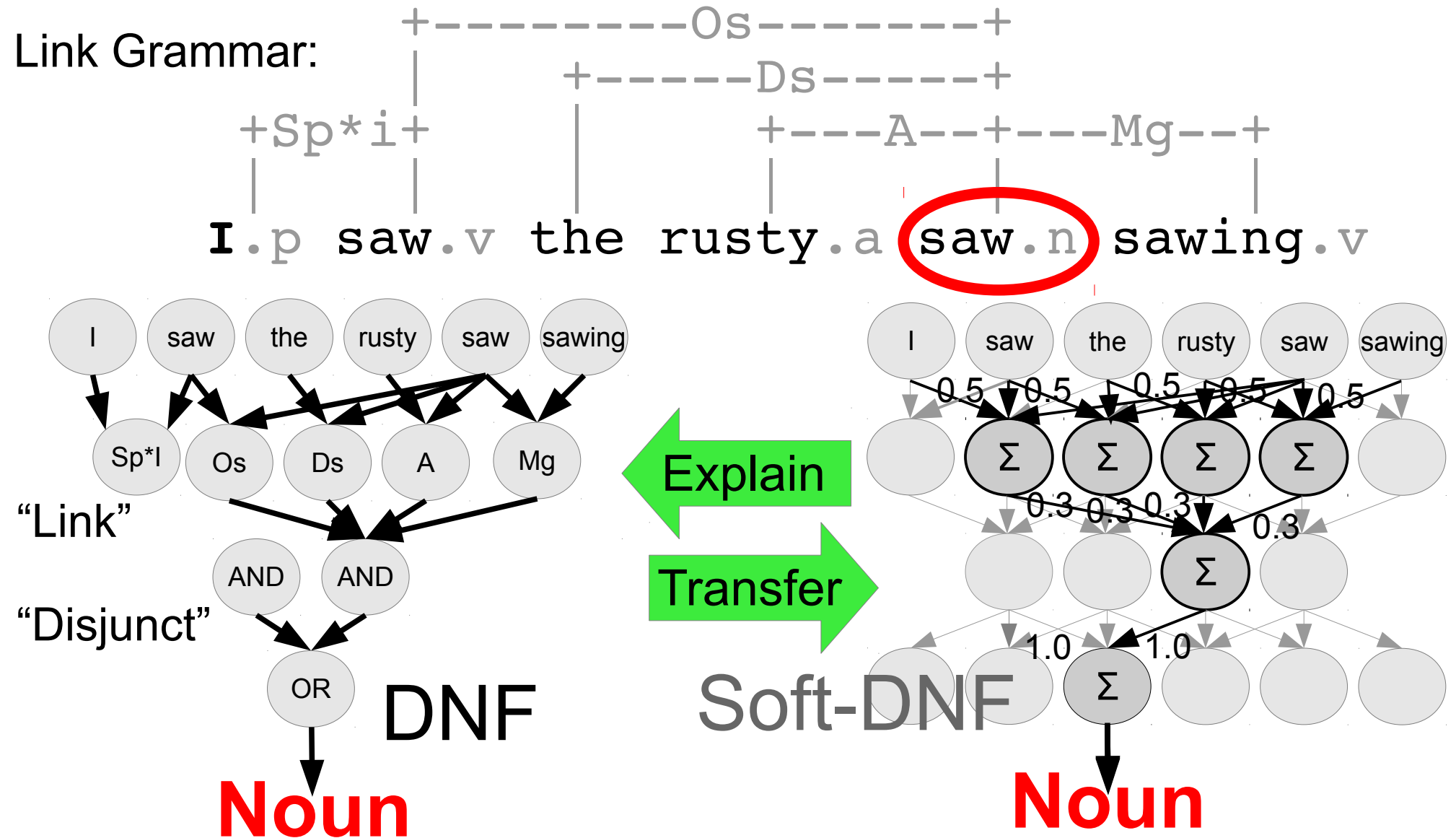
Anton Kolonin
akolonin@aigents.com

N* Novosibirsk
State
University
*THE REAL SCIENCE

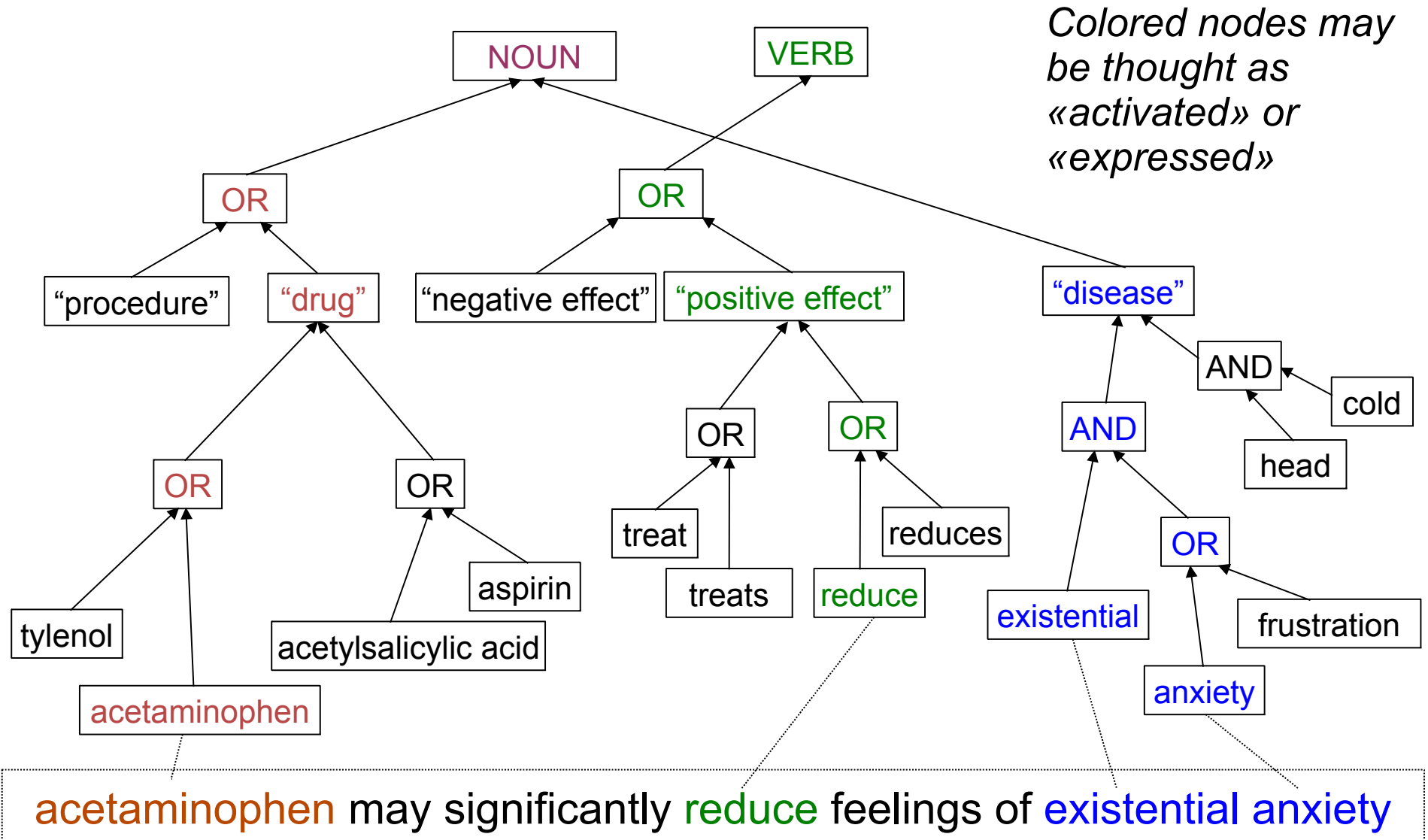


SingularityNET
<https://singularitynet.io>

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



Aigents® “Deep Patterns” - Language Model



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

Entity (Case): “Treatment:
Healing anxiety with Tylenol”

↑ 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.

IS

HAS

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

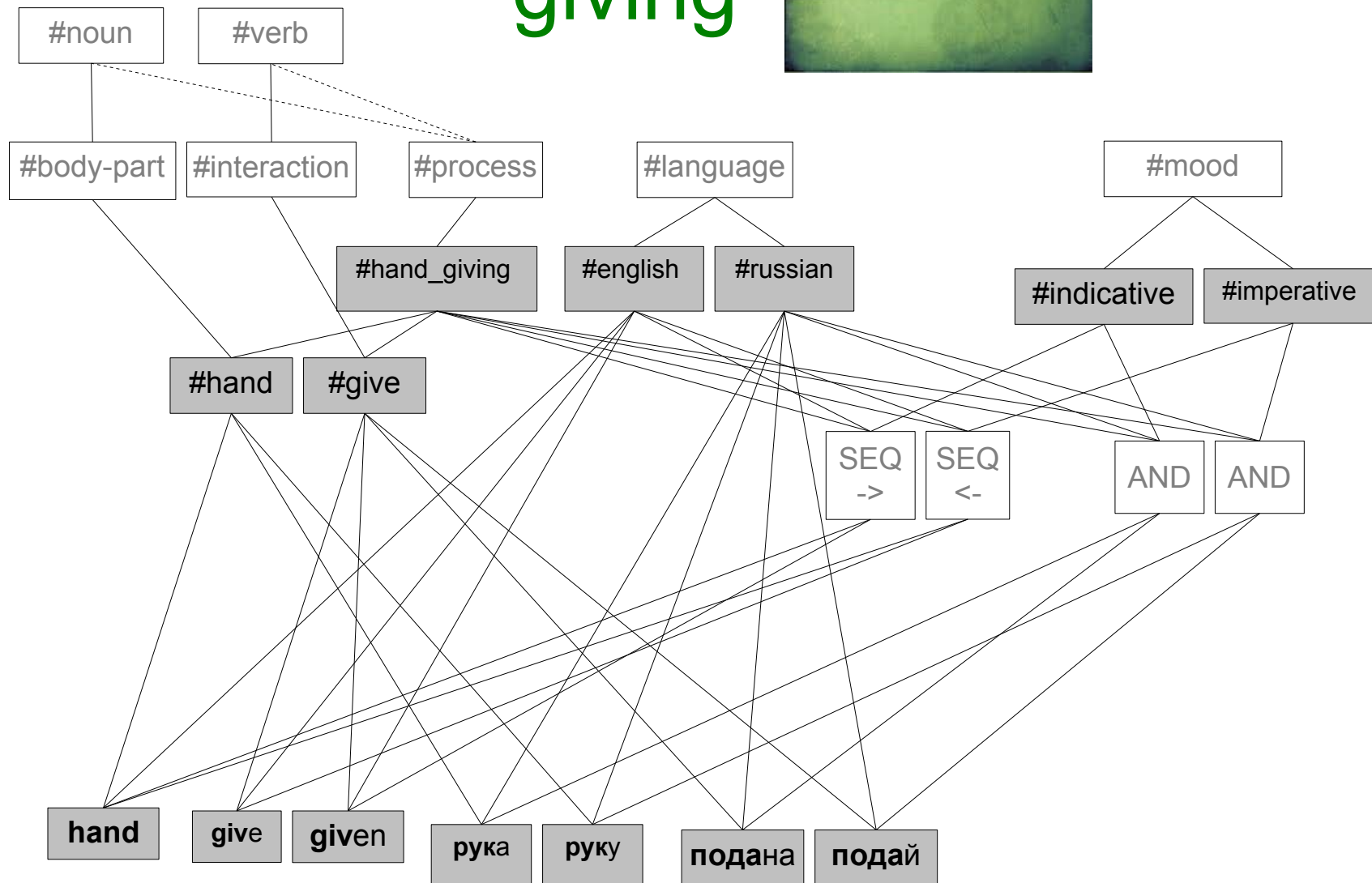
“Convey Guiding Catheter. Unique hydrophilic coating.
Small atraumatic soft tip. Ultra-thin 1 × 2 flat wire braid pattern”

=

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

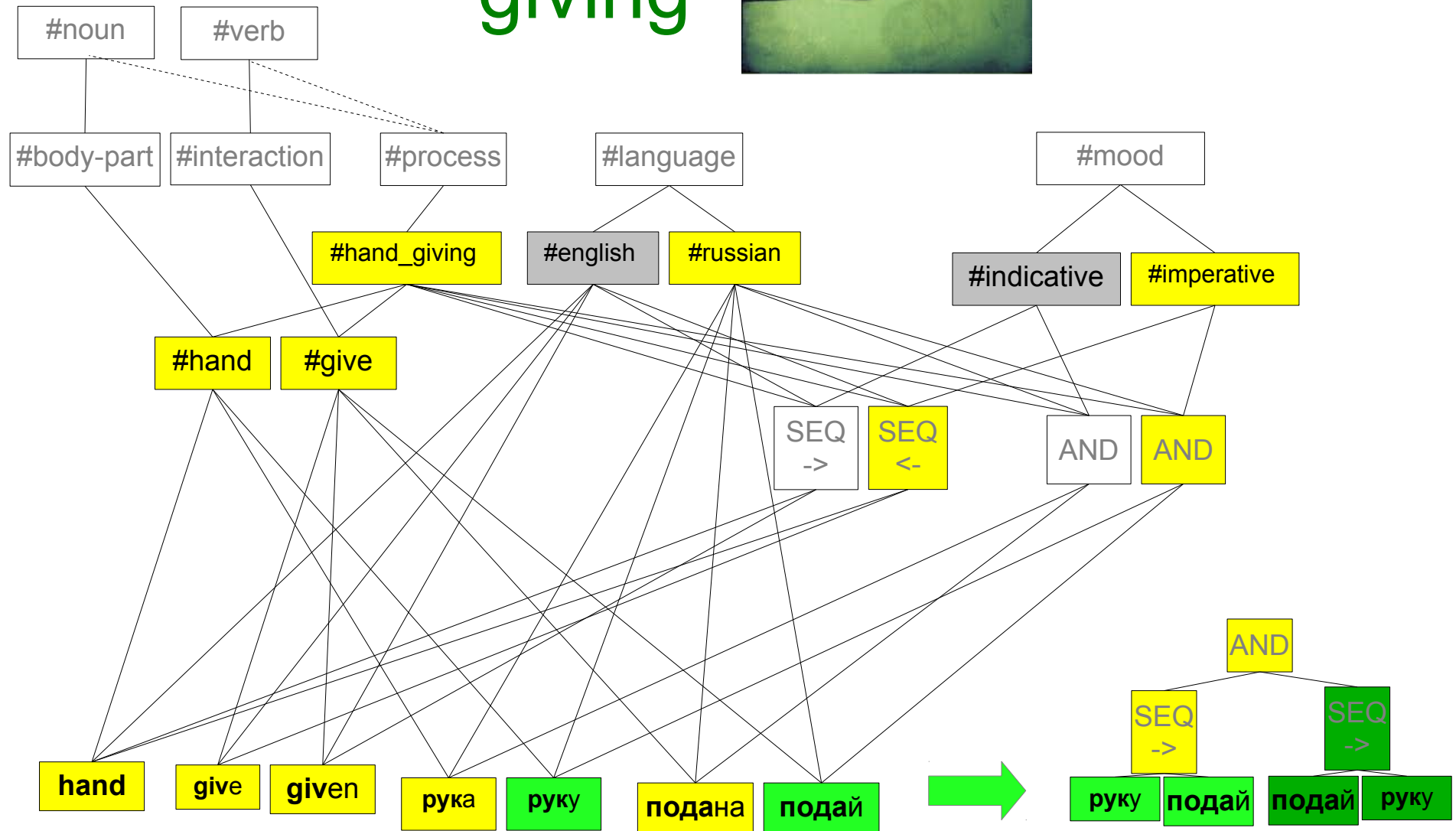
Grammar & Ontology Graph - Structure

Hand giving



Grammar & Ontology Graph - Production

Hand giving



Grammar Learning from Scratch - Programmatically



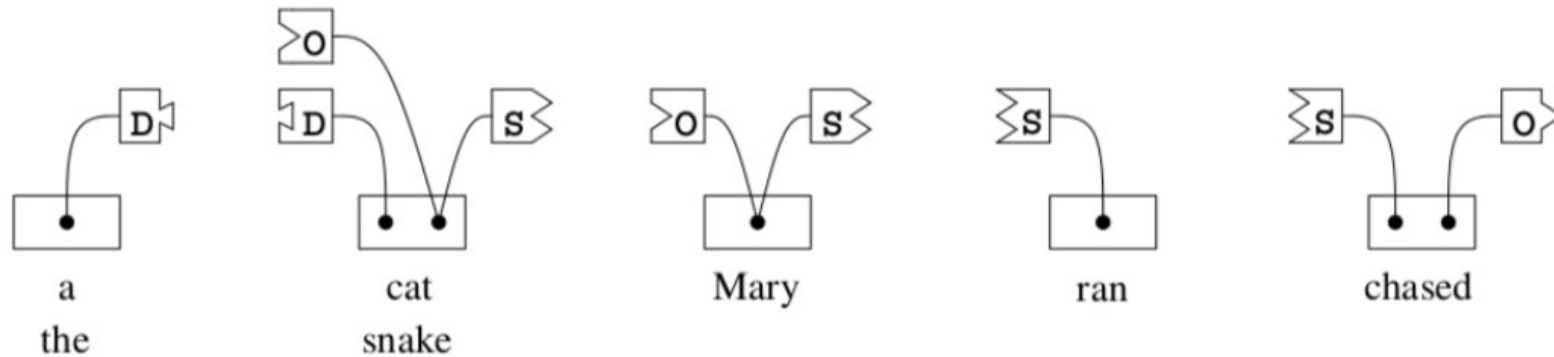
PRONOUN



ARTICLE



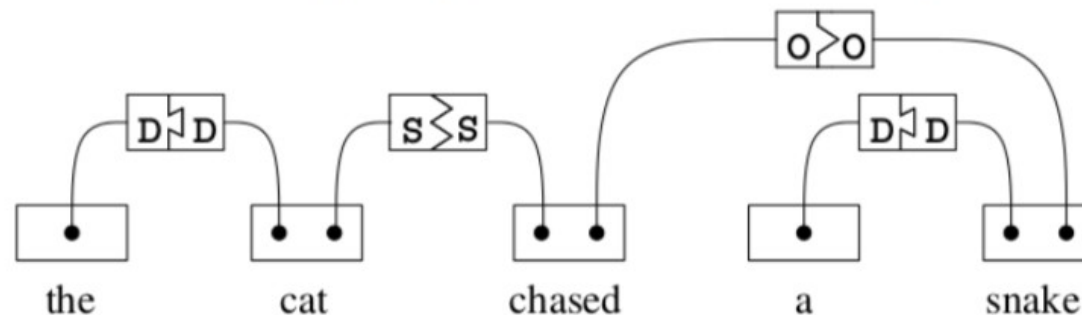
OpenCog Link Grammar Disjuncts & Connectors



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

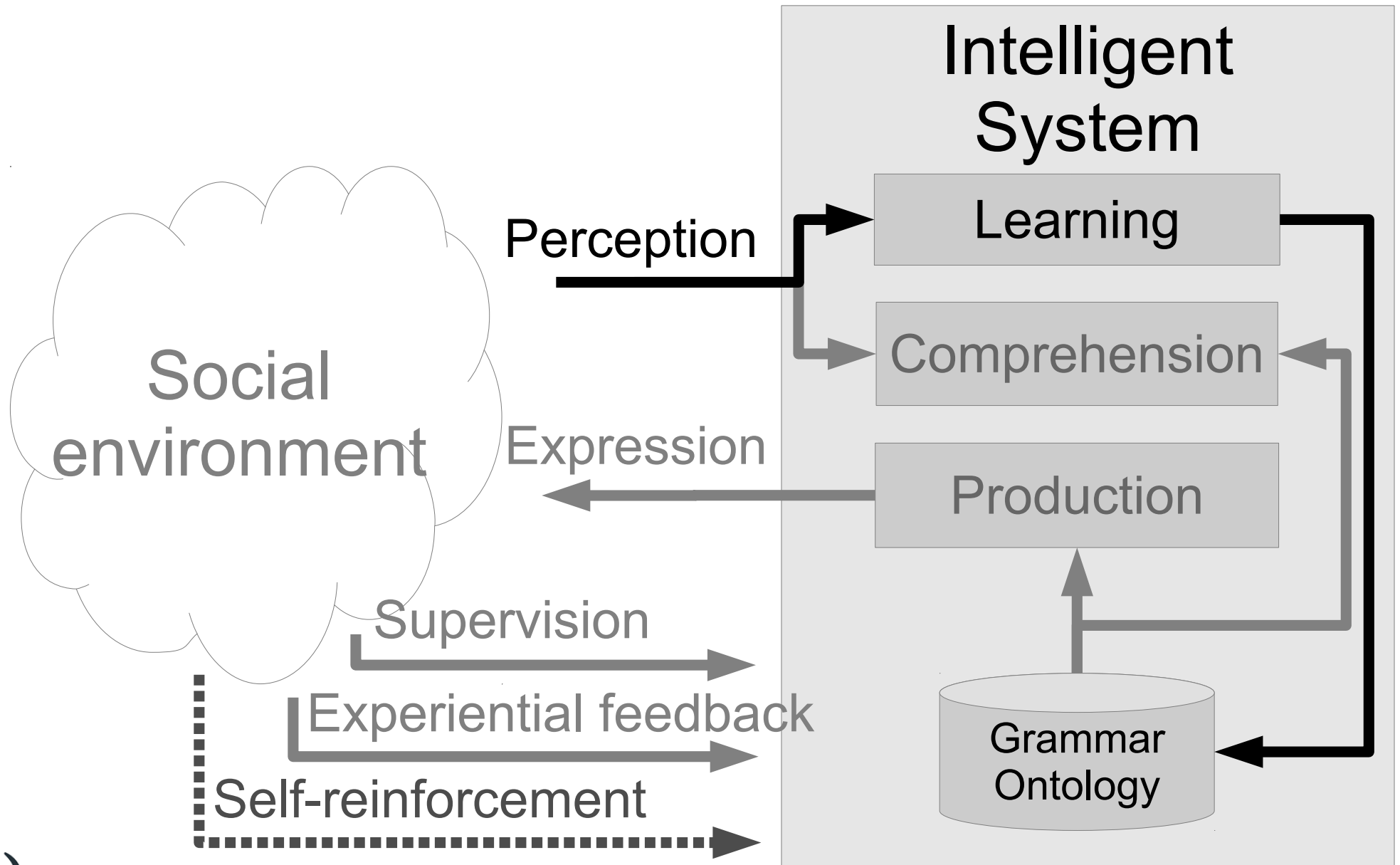
a the: D+;
 cat snake: D- & (S+ or O-);
 Mary: O- or S+;
 ran: S-;
 chased S- & O+;

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:

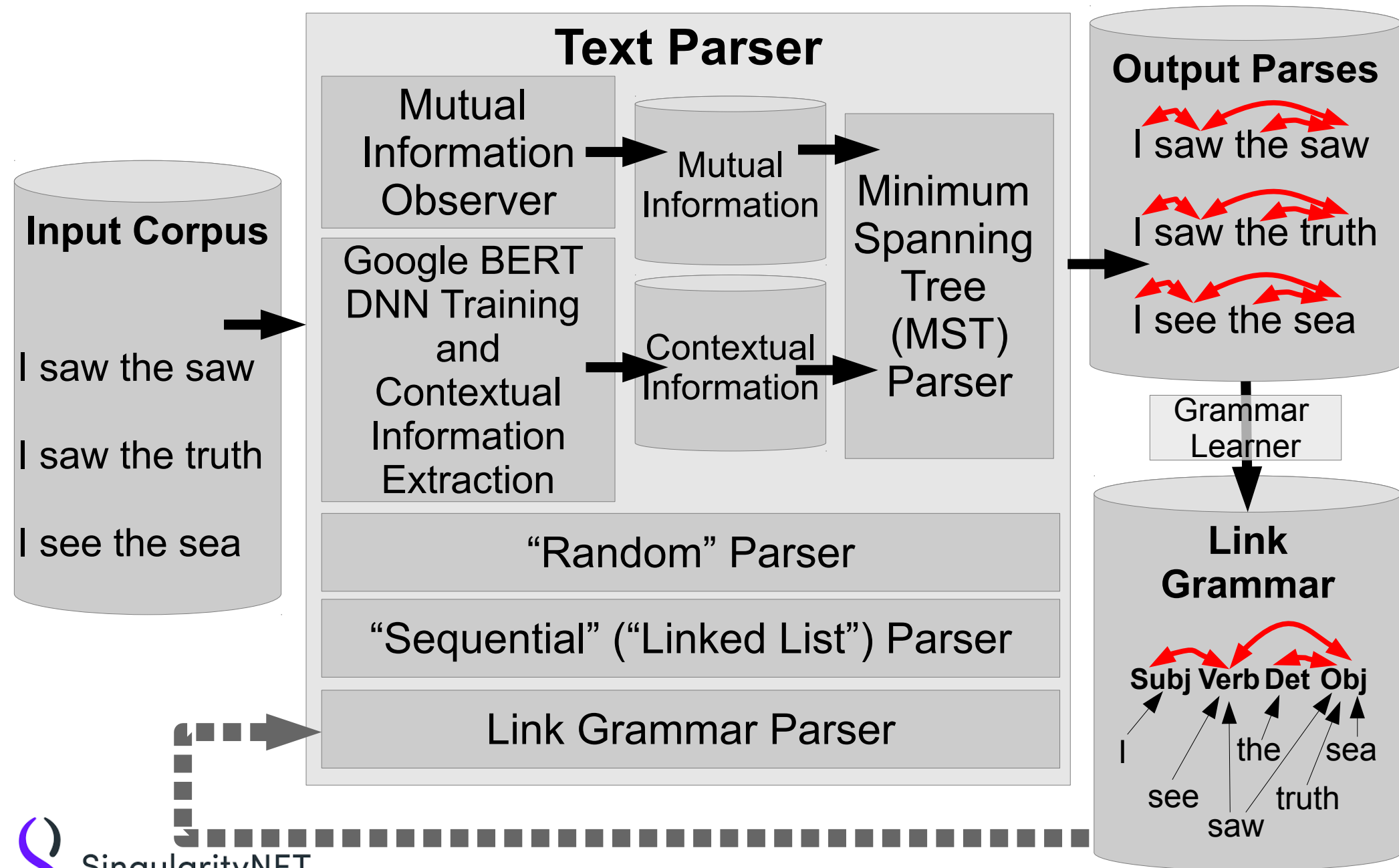


B. Goertzel,
 L. Vepstas,
 2014

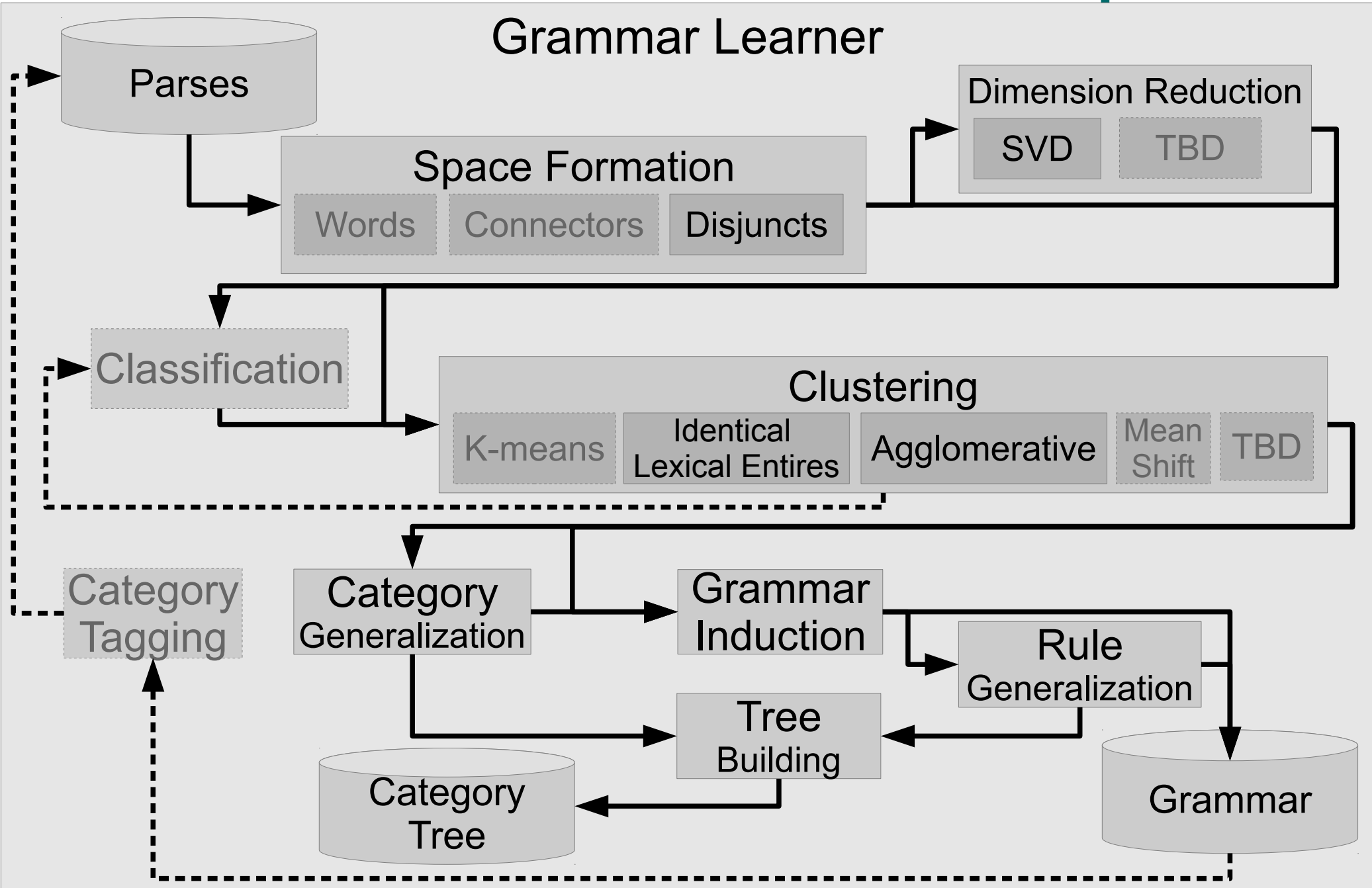
Language Learning Environment



Text Parsing for Link Grammar



Link Grammar Learner Pipeline



Corpora in Use

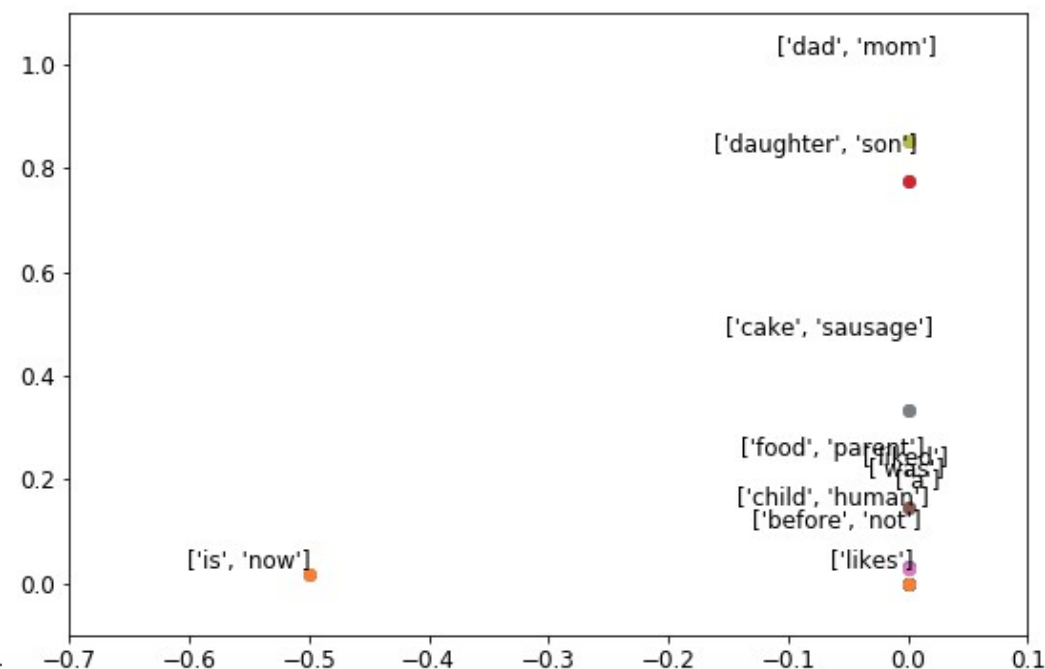
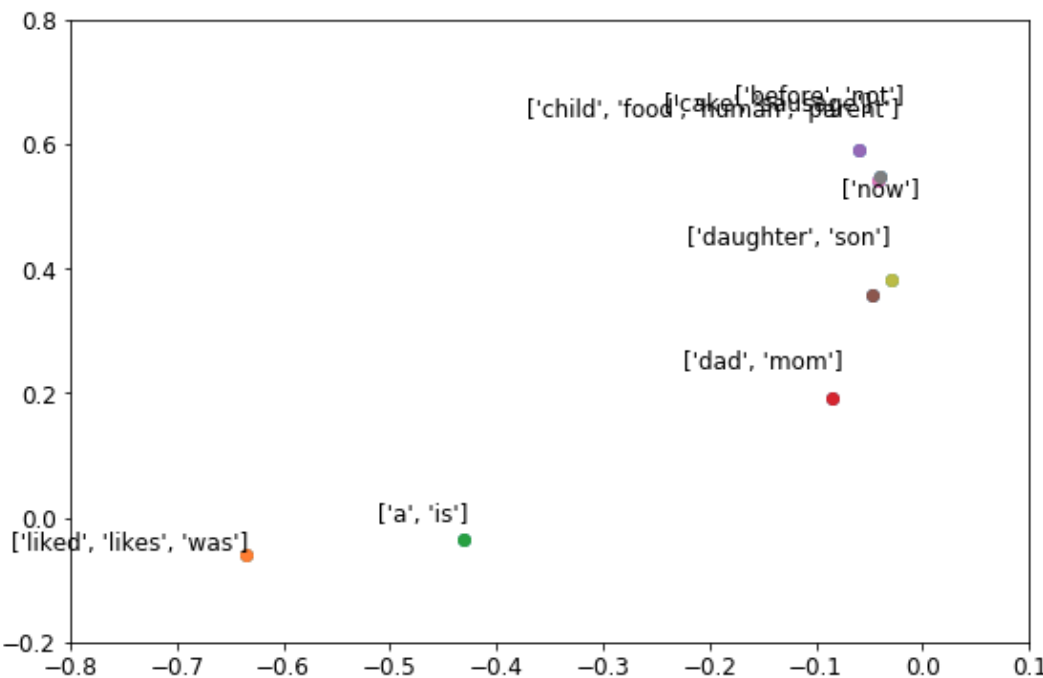
Corpus	Total words	Unique words	Occurrences per word	Total sentences	Average sentence length
POC-English	388	55	7	88	4
Child-Directed Speech	124185	3399	37	38181	4
Gutenberg Children	2695151	54054	50	207130	13

- 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/>

OpenCog Unsupervised Language Learning of Grammatical Categories and Link Grammar Dictionaries

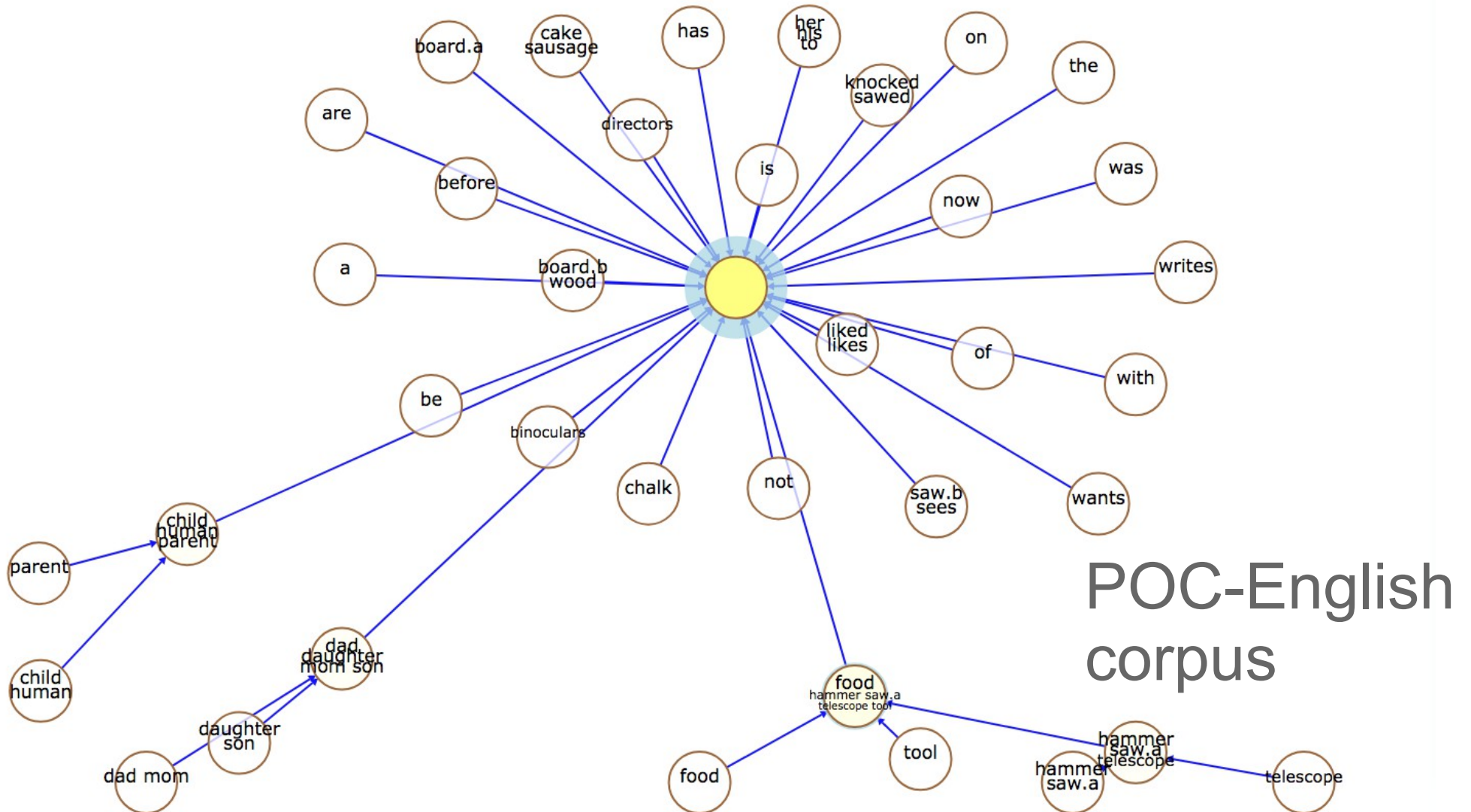
POC-English
(Connectors)

POC-English
(Disjuncts)

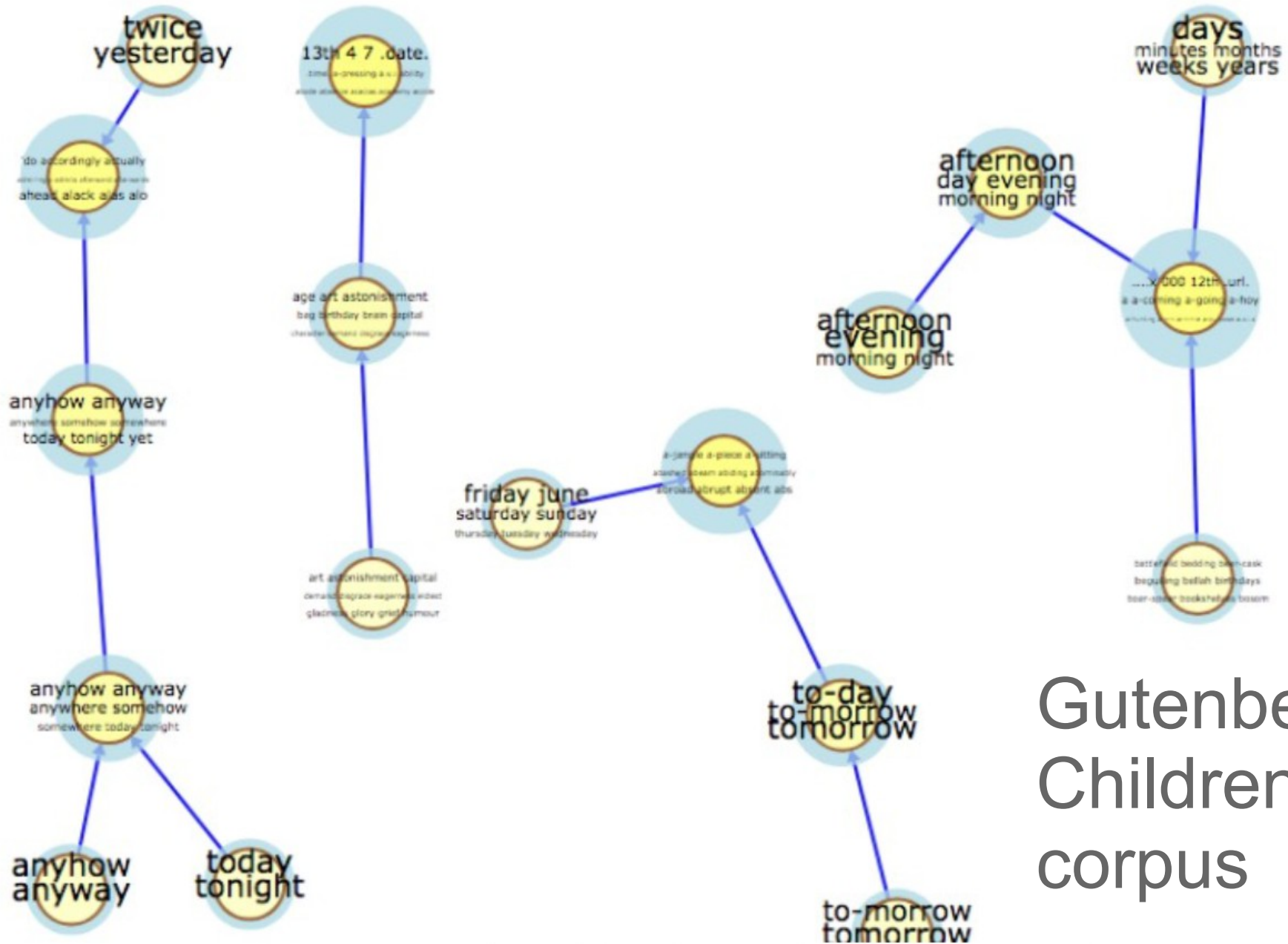


Grammar Ontology from Parses

Language Learning Categories

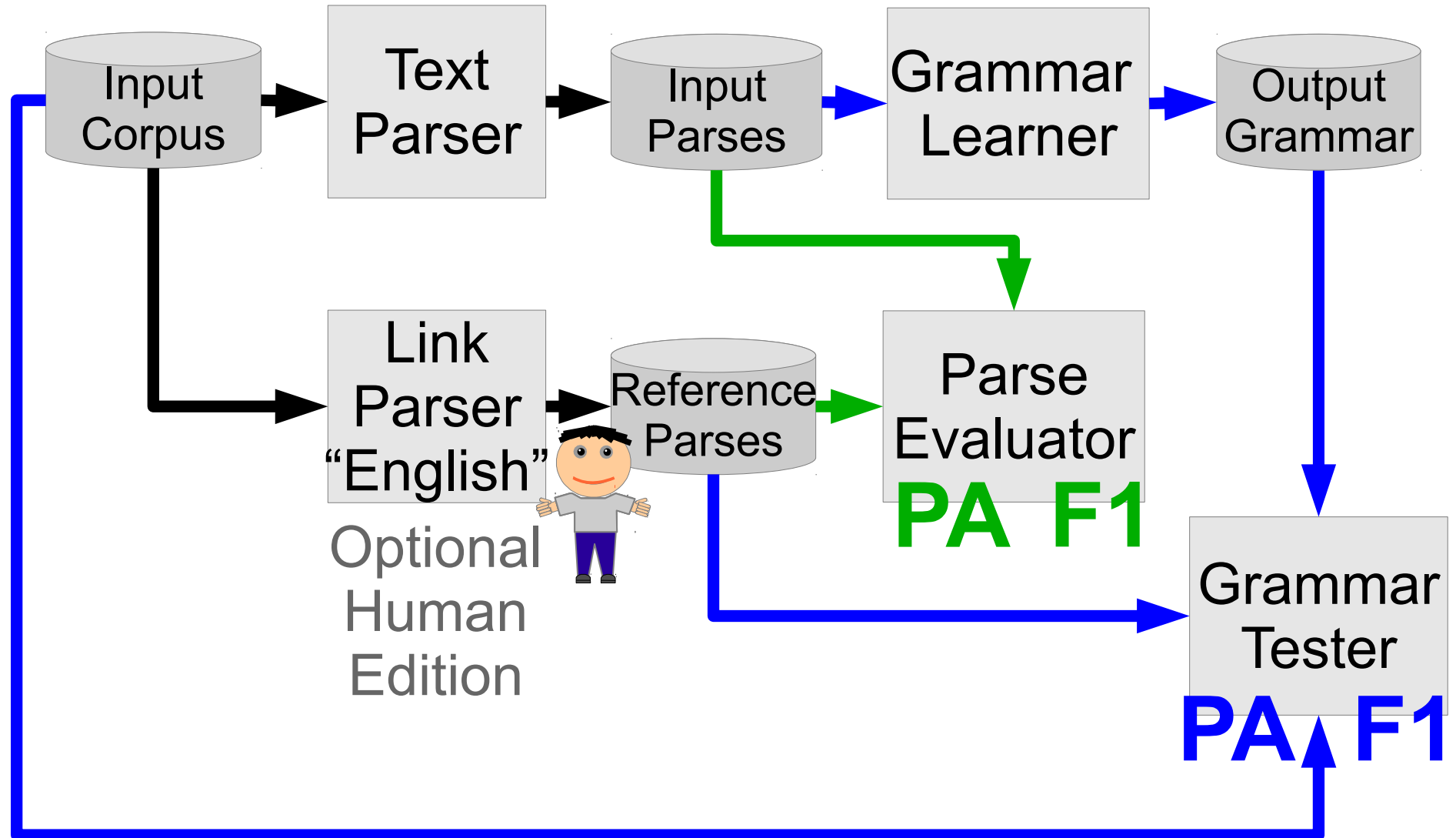


Grammar Ontology from Parses



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

Thank You and Welcome!

Anton Kolonin
akolonin@aigents.com

N* Novosibirsk
State
University
*THE REAL SCIENCE

