GRAMMAR PARSING AS A FOUNDATION OF CONTEXT-BASED INFORMATION EXTRACTION AND CYBERBULLYING DETECTION

Gniewosz Leliwa | Samurai Labs
Director of AI Research & Co-founder
gniewosz@samurailabs.ai

SEMANTIC ROLE LABELING

samurai LABS

Semantic Role Labeling: discovering the predicate–argument structure of each predicate

Example:

Sentence: *I left my pearls to my daughter-in-law in my will.*

Output: *[A0: I] [V: left] [A1: my pearls] [A2: to my daughter-in-law] [AM-LOC: in my will].*

Where: *A0 = leaver, A1 = thing left, A2 = beneficiary, AM-LOC = location of the action*

Table 3
The overall system performance when argument boundaries are known

	Full Parsing			Shallow Parsing		
	Prec	Rec	F_1	Prec	Rec	F_1
Gold	91.58	91.90	91.74 ± 0.51	91.14	91.48	91.31 ± 0.51
Auto	90.71	91.14	90.93 ± 0.53	90.50	90.88	90.69 ± 0.53

Table 6
The overall system performance using the output from the pruning heuristics, applied on the gold standard full parse trees

	Full Parsing			Shallow Parsing		
	Prec	Rec	F_1	Prec	Rec	F_1
Gold	86.22	87.40	86.81 ± 0.59	84.14	85.31	84.72 ± 0.63
Auto	84.21	85.04	84.63 ± 0.63	86.17	84.02	85.08 ± 0.63

Conclusion: The full parsing information helps in argument identification. However, when the automatic parsers are used, using the full parsing information may not have better overall results compared to using shallow parsing.

Source: http://cogcomp.org/papers/PunyakanokRoYi07.pdf (2008)

INFORMATION / RELATION EXTRACTION

samurai LABS

Open Information Extraction: extracting triples (Argument_1, Relation, Argument_2) from text

Example:
Sentence: *Vigo is the largest city in Galicia and is located in the northwest of Spain.*
Triplet 1: *(Vigo, is the largest city in, Galicia)*
Triplet 2: *(Vigo, is located in, northwest of Spain)*

ReVerb: shallow syntactic features; DepOE: deep syntactic features

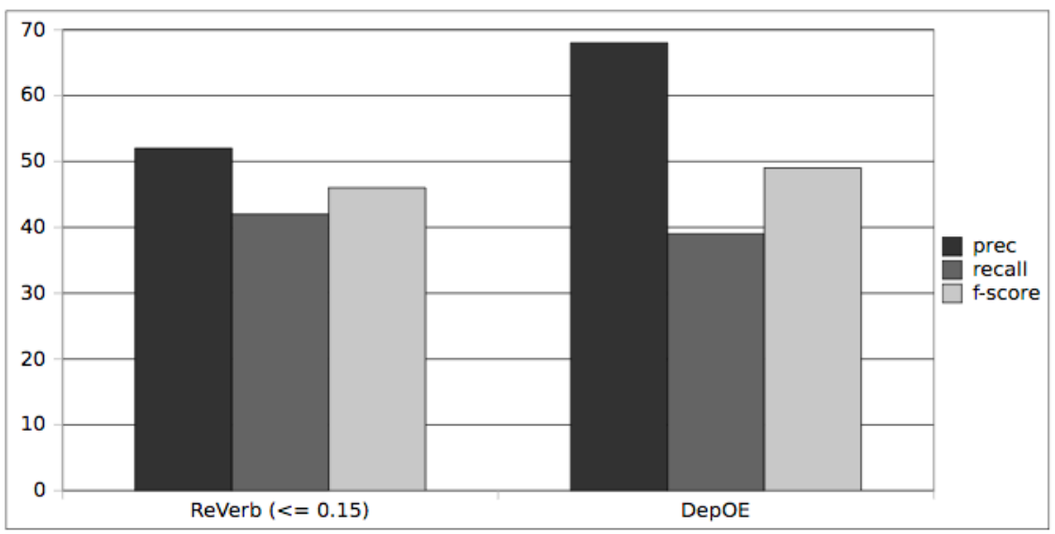


Figure 2: Evaluation of the extraction of triples (both relation and its arguments) performed by DepOE and ReVerb (with a confidence score $>= 0.15$).

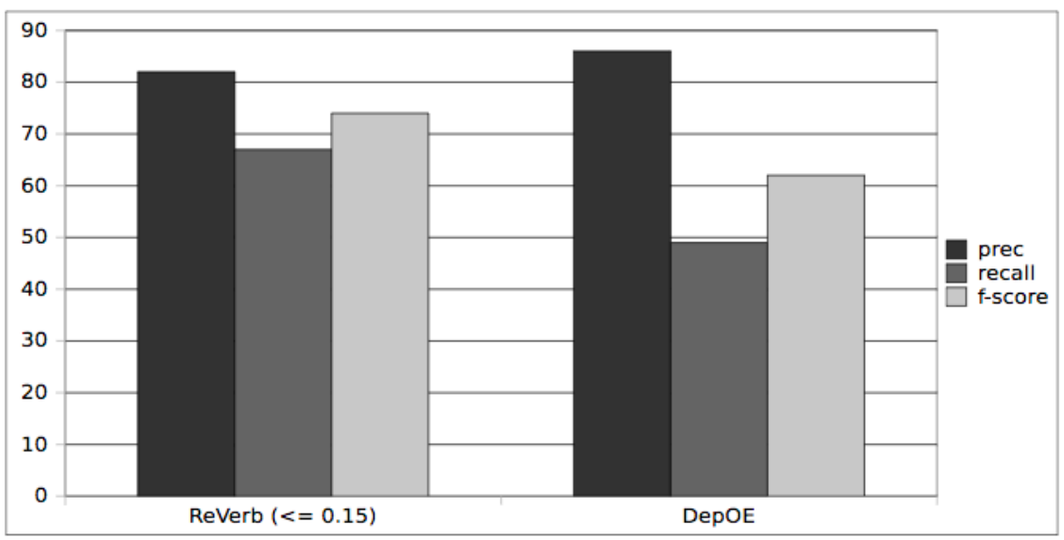


Figure 3: Evaluation of the relation extraction performed by DepOE and ReVerb (with a confidence score $>= 0.15$).

Source: http://www.anthology.aclweb.org/W/W12/W12-0702.pdf (2012)

SENTIMENT ANALYSIS

samurai LABS

Sentiment Analysis: determining the attitude of a subject with respect to a topic, document, etc.

Example:

Sentence: *The display is awesome, but the battery life is too short.*

Output: *display: positive; battery life: negative; overall: neutral*

SENTIMENT ANALYSIS

samurai LABS

Sentiment Analysis: determining the attitude of a subject with respect to a topic, document, etc.

Example:

Sentence: *The display is awesome, but the battery life is too short.*

Output: *display: positive; battery life: negative; overall: neutral*

"The best modern sentiment systems don't use parsers anymore, actually. They seem to be doing the task via end-to-end deep neural nets, these days."

– Dave Orr, PM for NLP in Google Research (2016)

Source: https://www.quora.com/What-can-I-do-with-the-result-of-an-NLP-parser-Other-than-sentiment-analysis

QUESTION ANSWERING

samurai LABS

Question Answering (QA): the ability to read text and then answer questions about it

Example:

Text: *In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under gravity. The main forms of precipitation include drizzle, rain, sleet, snow, graupel and hail.*

Question: *What causes precipitation to fall?*

Answer: *gravity*

Feature Groups	Description	Examples
Matching Word Frequencies	Sum of the TF-IDF of the words that occur in both the question and the sentence containing the candidate answer. Separate features are used for the words to the left, to the right, inside the span, and in the whole sentence.	Span: $[0 \leq \text{sum} < 0.01]$ Left: $[7.9 \leq \text{sum} < 10.7]$
Matching Bigram Frequencies	Same as above, but using bigrams. We use the generalization of the TF-IDF described in Shirakawa et al. (2015).	Span: $[0 \leq \text{sum} < 2.4]$ Left: $[0 \leq \text{sum} < 2.7]$
Root Match	Whether the dependency parse tree roots of the question and sentence match, whether the sentence contains the root of the dependency parse tree of the question, and whether the question contains the root of the dependency parse tree of the sentence.	Root Match = False
Lengths	Number of words to the left, to the right, inside the span, and in the whole sentence.	Span: $[1 <= \text{num} < 2]$ Left: $[15 \leq \text{num} < 19]$
Span Word Frequencies	Sum of the TF-IDF of the words in the span, regardless of whether they appear in the question.	Span: $[5.2 \leq \text{sum} < 6.9]$
Constituent Label	Constituency parse tree label of the span, optionally combined with the wh-word in the question.	Span: NP Span: NP, wh-word: "*what*"
Span POS Tags	Sequence of the part-of-speech tags in the span, optionally combined with the wh-word in the question.	Span: [NN] Span: [NN], wh-word: "*what*"
Lexicalized	Lemmas of question words combined with the lemmas of words within distance 2 to the span in the sentence based on the dependency parse trees. Separately, question word lemmas combined with answer word lemmas.	Q: "*cause*", S: "*under*" $\xleftarrow{\text{case}}$ Q: "*fall*", A: "*gravity*"
Dependency Tree Paths	For each word that occurs in both the question and sentence, the path in the dependency parse tree from that word in the sentence to the span, optionally combined with the path from the wh-word to the word in the question. POS tags are included in the paths.	VBZ $\xrightarrow{\text{nmod}}$ NN what $\xleftarrow{\text{nsubj}}$ VBZ $\xrightarrow{\text{advcl}}$ + VBZ $\xrightarrow{\text{nmod}}$ NN

Table 4: Features used in the logistic regression model with examples for the question "*What causes precipitation to fall?*", sentence "*In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under gravity.*" and answer "*gravity*". Q denotes question, A denotes candidate answer, and S denotes sentence containing the candidate answer.

	F_1	
	Train	Dev
Logistic Regression	91.7%	51.0%
– Lex., – Dep. Paths	33.9%	35.8%
– Lexicalized	53.5%	45.4%
– Dep. Paths	91.4%	46.4%
– Match. Word Freq.	91.7%	48.1%
– Span POS Tags	91.7%	49.7%
– Match. Bigram Freq.	91.7%	50.3%
– Constituent Label	91.7%	50.4%
– Lengths	91.8%	50.5%
– Span Word Freq.	91.7%	50.5%
– Root Match	91.7%	50.6%

Table 6: Performance with feature ablations. We find that lexicalized and dependency tree path features are most important.

Source: https://arxiv.org/pdf/1606.05250.pdf (2016)

PARSEY MCPARSEFACE RELEASE: MAY 2016

samurai LABS

Google just open sourced something called 'Parsey McParseface,' and it could change AI forever

by NATE SWANNER — May 12, 2016 in DESIGN & DEV

Technology Intelligence

Has Google's Parsey McParseface just solved one of the world's biggest language problems?

share

By Michael Wilkinson

17 MAY 2016 • 5:00PM

Save

SCI-TECH

Don't laugh: Google's Parsey McParseface is a serious IQ boost for computers

Google offers free use of SyntaxNet technology, a boon to anyone trying to get computers to understand natural human language.

BY STEPHEN SHANKLAND / MAY 13, 2016 3:25 AM PDT

AI: Google AI Tool 'Parsey McParseface' Could Detect Lies, Eliminate Problems Of Human Language With Artificial Intelligence Language Program

Sally Elliott

Source: https://thenextweb.com/dd/2016/05/12/google-just-open-sourced-something-called-parsey-mcparseface-change-ai-forever/
https://www.telegraph.co.uk/technology/2016/05/17/has-googles-parsey-mcparseface-just-solved-one-of-the-worlds-big/
https://www.cnet.com/news/google-offers-parsey-mcparseface-and-syntaxnet-ai-software-for-free/
https://www.inquisitr.com/3107832/ai-google-ai-tool-parsey-mcparseface-could-detect-lies-eliminate-problems-of-human-language-with-artificial-intelligence-language-program/

PARSEY MCPARSEFACE RELEASE: MAY 2016

samurai LABS

May 26 2016 at 8:59 AM
Updated May 26 2016 at 8:59 AM

Save article My Saved Articles Print License article

Google's robot parser, Parsey McParseface, doesn't speak money

by Michael Bailey

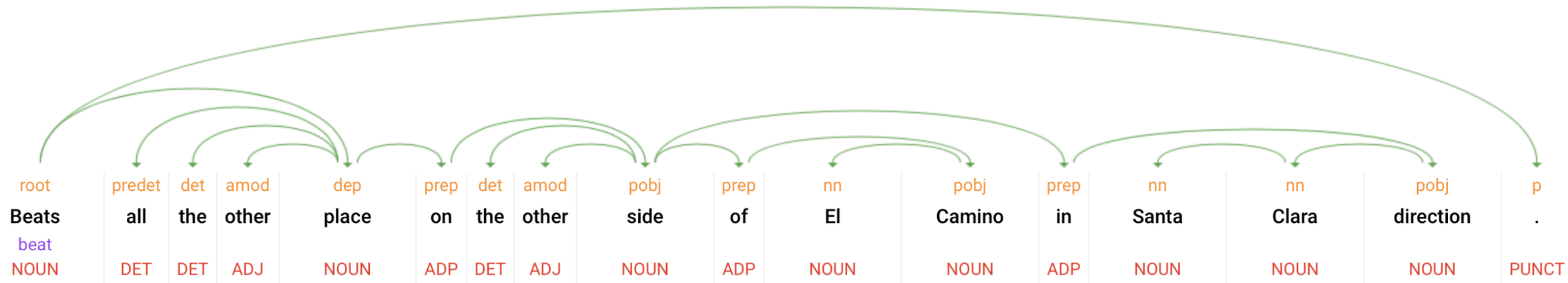
Google's new contribution to the cause is named Parsey McParseface. It has been trained both on the standard corpus of manually parsed English documents – the 20-year-old Penn Treebank, which was drawn from sentences in news wires and literature – as well as a corpus Google built itself in 2011 called the Google Web Treebank, which drew from the grammatically wilder worlds of blogs, discussion forums and social media posts.

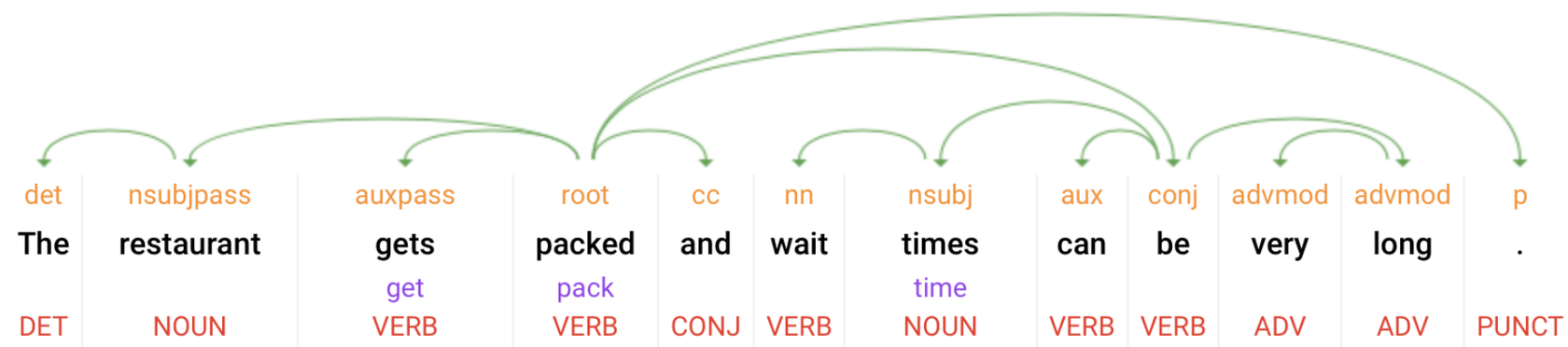
However, commercially viable products for natural language processing (NLP) require a higher level of accuracy than even 94 per cent, said the founder of text-mining start-up Health Language Analytics, Jon Patrick.

Source: https://www.afr.com/technology/googles-robot-parser-parsey-mcparseface-doesnt-speak-money-20160520-goznyx

LAS / UAS / PER-TOKEN ACCURACY IS NOT ENOUGH

samurai LABS

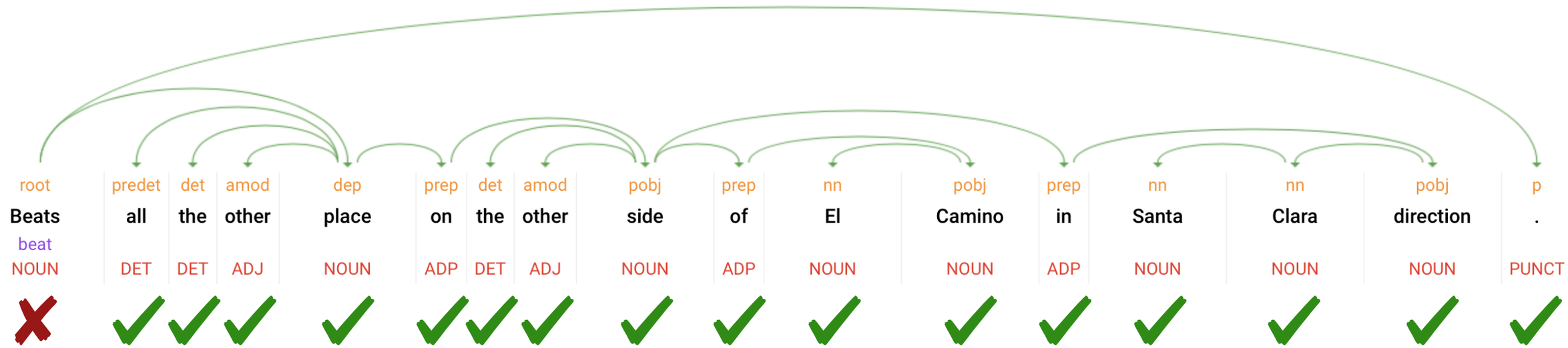


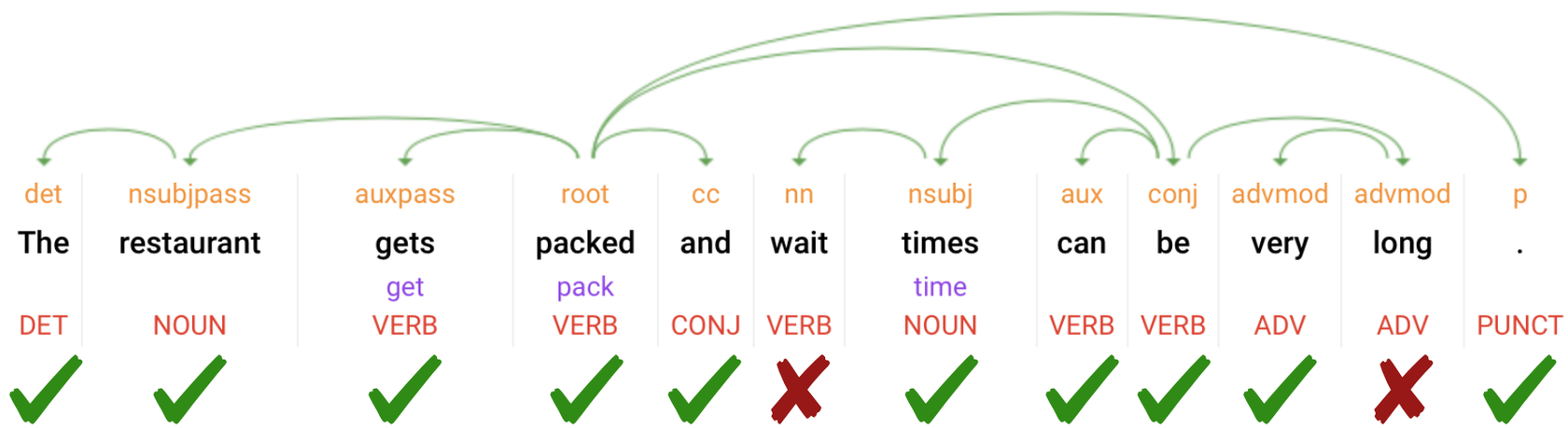


Source: https://cloud.google.com/natural-language/

LAS / UAS / PER-TOKEN ACCURACY IS NOT ENOUGH

samurai LABS

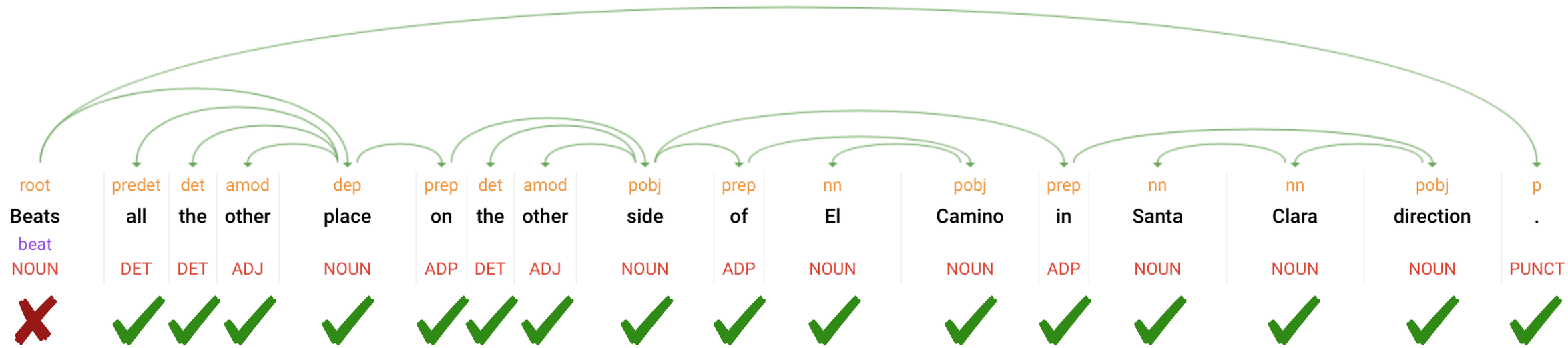


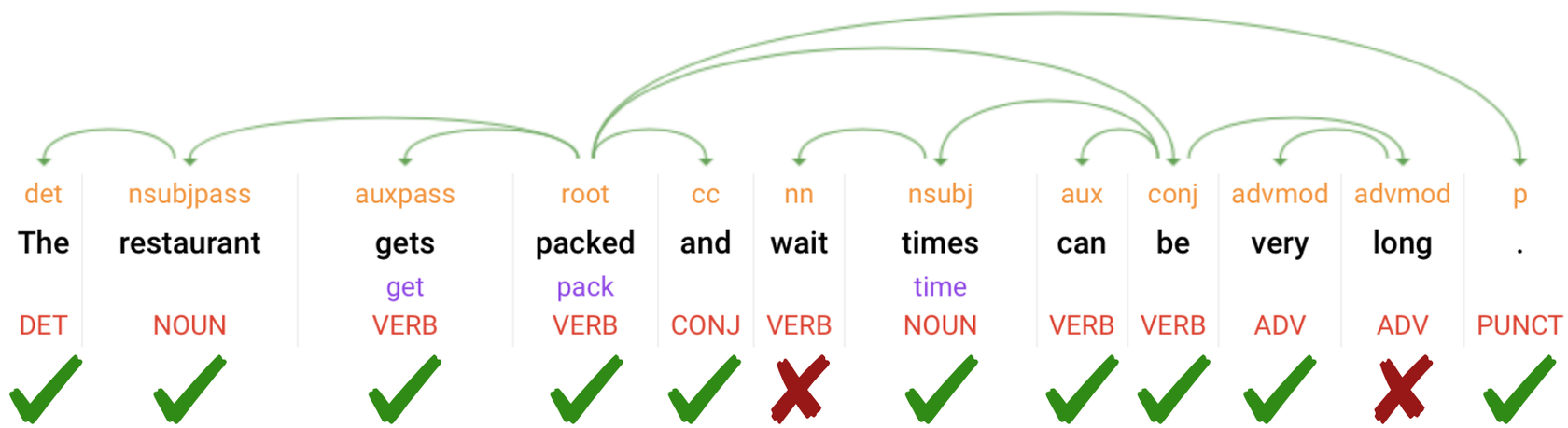


Source: https://cloud.google.com/natural-language/

LAS / UAS / PER-TOKEN ACCURACY IS NOT ENOUGH

samurai LABS





Number of tokens	29
Correct tokens	26
Incorrect tokens	3
Per-token accuracy	90%

Number of sentences	2
Correct sentences	0
Incorrect sentences	2
Per-sentence accuracy	0%

Source: https://cloud.google.com/natural-language/

PART OF SPEECH TAGGING

samurai LABS

Part-of-Speech Tagging from 97% to 100%: Is It Time for Some Linguistics?

Christopher D. Manning

Departments of Linguistics and Computer Science
Stanford University
353 Serra Mall, Stanford CA 94305-9010
manning@stanford.edu

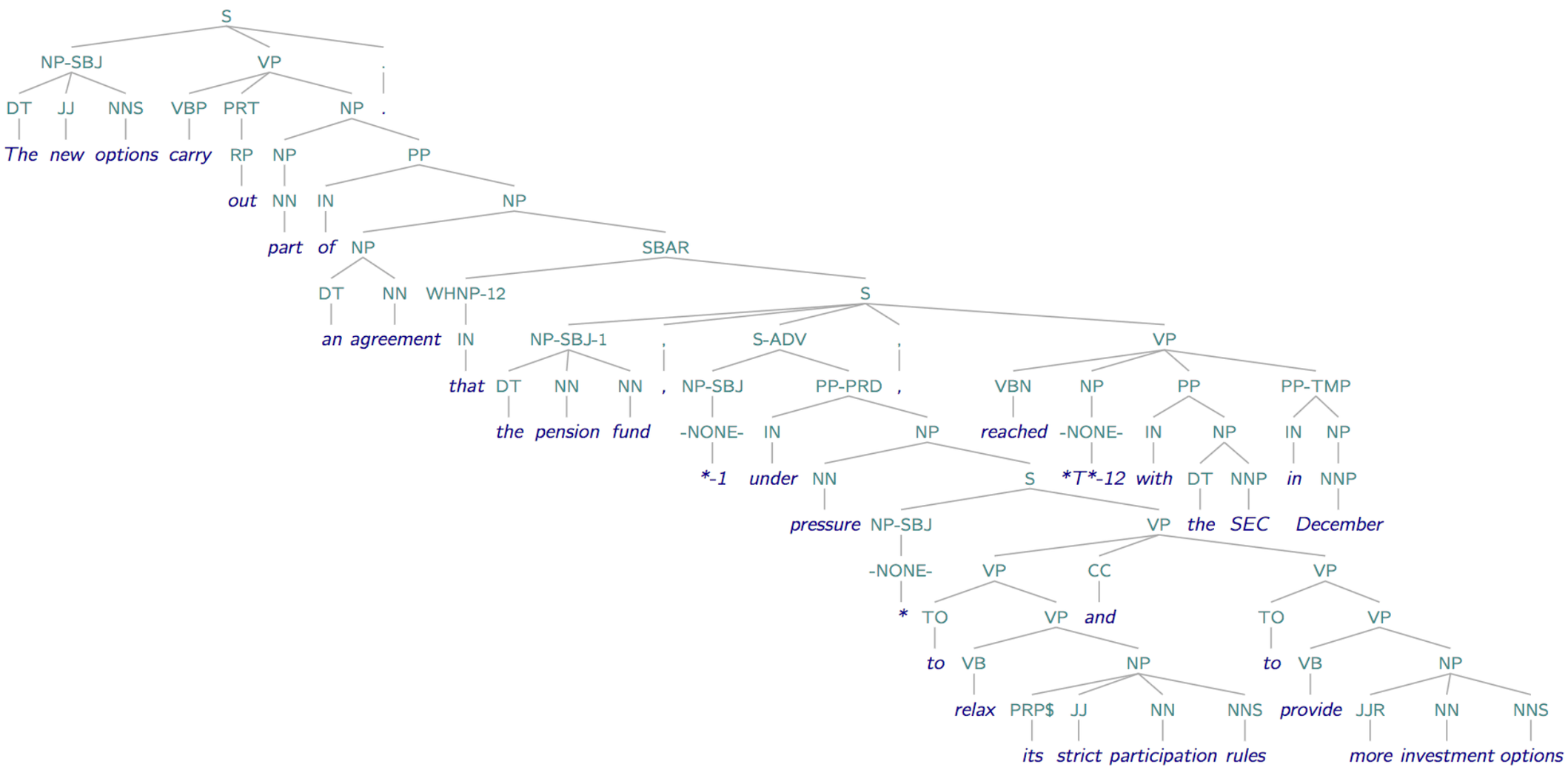
Abstract. I examine what would be necessary to move part-of-speech tagging performance from its current level of about 97.3% token accuracy (56% sentence accuracy) to close to 100% accuracy. I suggest that it must still be possible to greatly increase tagging performance and examine some useful improvements that have recently been made to the Stanford Part-of-Speech Tagger. However, an error analysis of some of the remaining errors suggests that there is limited further mileage to be had either from better machine learning or better features in a discriminative sequence classifier. The prospects for further gains from semi-supervised learning also seem quite limited. Rather, I suggest and begin to demonstrate that the largest opportunity for further progress comes from improving the taxonomic basis of the linguistic resources from which taggers are trained. That is, from improved descriptive linguistics. However, I conclude by suggesting that there are also limits to this process. The status of some words may not be able to be adequately captured by assigning them to one of a small number of categories. While conventions can be used in such cases to improve tagging consistency, they lack a strong linguistic basis.

Source: https://nlp.stanford.edu/pubs/CICLing2011-manning-tagging.pdf (2011)

REAL SENTENCES

samurai LABS

Syntactic parses of real sentences



- State-of-the-art parsers have accuracies of over 90%

⇒ *Most parses contain at least one error*

Source: http://web.science.mq.edu.au/~mjohnson/papers/Johnson15ParsingSurvey.pdf

LAST DECENT COMPARISON FROM 2015

samurai LABS

	With Punctuation						Without Punctuation					
	Overall			Exact Match			Overall			Exact Match		
	LAS	UAS	LS	LAS	UAS	LS	LAS	UAS	LS	LAS	UAS	LS
ClearNLP$_g$	**89.19**	**90.63**	**94.94**	**47.65**	**53.00**	**61.17**	**90.09**	**91.72**	**94.29**	**49.12**	**55.01**	**61.31**
GN13	87.59	89.17	93.99	43.78	48.89	56.71	88.75	90.54	93.32	45.44	51.20	56.88
LTDP$_g$	n/a	85.75	n/a	n/a	46.38	n/a	n/a	87.16	n/a	n/a	48.01	n/a
SNN	86.42	88.15	93.54	42.98	48.53	55.87	87.63	89.59	92.70	43.96	49.83	55.91
spaCy	87.92	89.61	94.08	43.36	48.79	55.67	88.95	90.86	93.32	44.97	51.28	55.70
Yara$_g$	85.93	87.64	92.99	42.94	47.77	54.79	87.39	89.32	92.24	44.25	49.44	54.96
ClearNLP	89.87	91.30	95.28	49.38	55.18	63.18	90.64	92.26	**94.67**	50.61	56.88	63.24
LTDP	n/a	88.18	n/a	n/a	51.62	n/a	n/a	89.17	n/a	n/a	53.54	n/a
Mate	**90.03**	**91.62**	**95.29**	49.66	**56.44**	62.71	**90.70**	**92.50**	**94.67**	50.83	**58.36**	62.72
RBG	89.57	91.45	94.71	46.49	55.49	58.45	90.23	92.35	94.01	47.64	56.54	58.07
Redshift	89.48	91.01	95.04	49.71	55.82	62.70	90.27	92.00	94.42	50.88	57.28	62.78
Turbo	89.81	91.50	95.00	48.08	55.33	60.49	90.49	92.40	94.34	49.29	57.09	60.52
Yara	89.80	91.36	95.19	**50.07**	56.18	**63.36**	90.47	92.24	94.57	**51.02**	57.53	**63.42**

Table 4: Overall parsing accuracy. The top 6 rows and the bottom 7 rows show accuracies for greedy and non-greedy parsers, respectively.

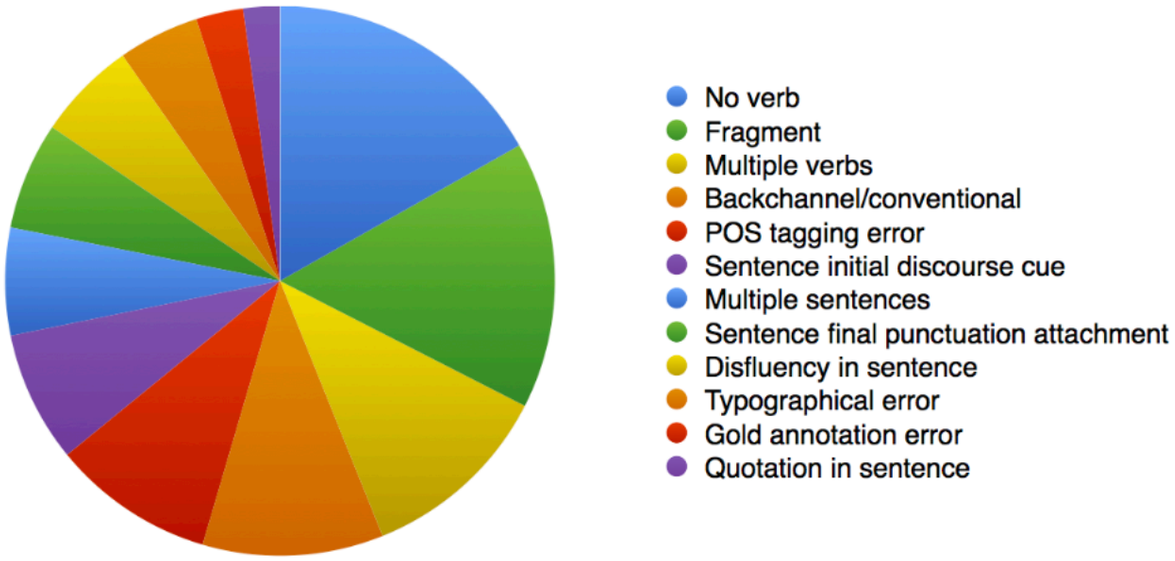
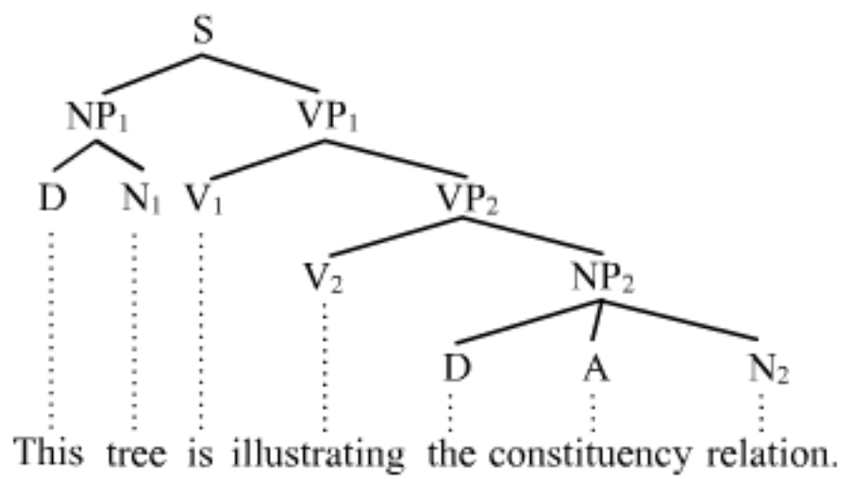


Figure 6: Common error types in erroneous trees.

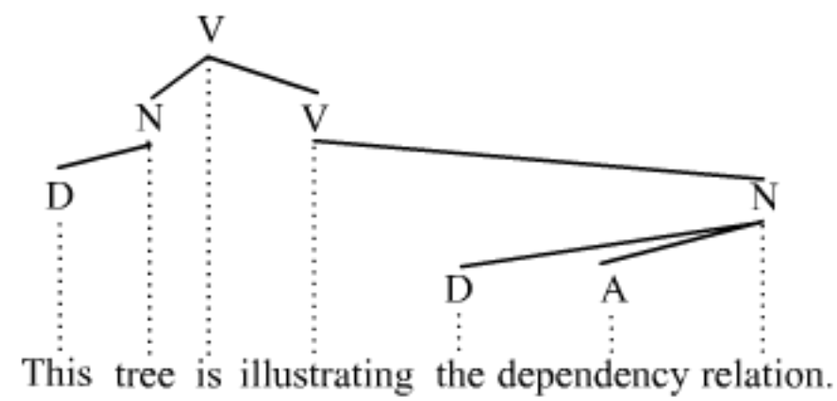
Source: http://www.aclweb.org/anthology/P15-1038

DEPENDENCY OR CONSTITUENCY (PHRASE STRUCTURE)?

samurai LABS



Constituency relation (PSG)

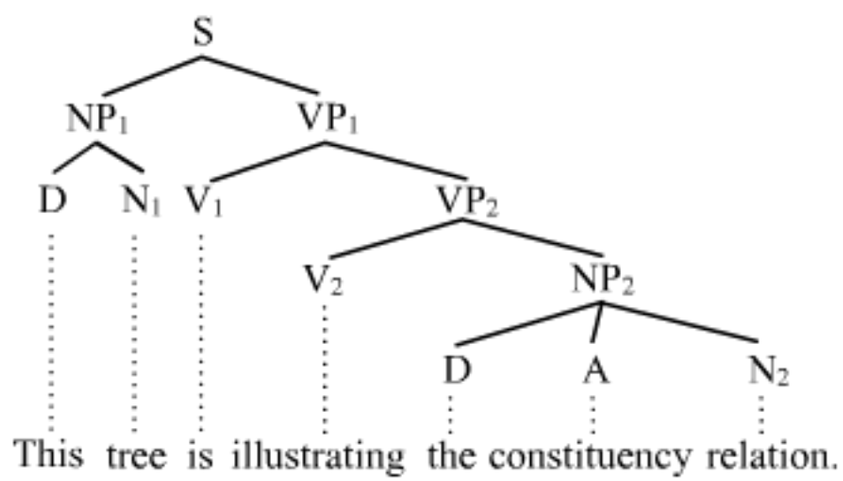


Dependency relation

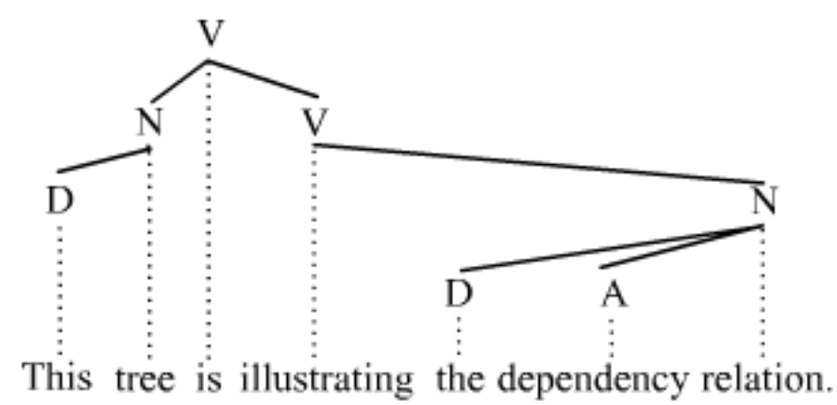
Source: https://en.wikipedia.org/wiki/Phrase_structure_grammar

DEPENDENCY OR CONSTITUENCY (PHRASE STRUCTURE)?

samurai LABS



Constituency relation (PSG)



Dependency relation

Example of a Samurai Labs' Syntactic Parser output

Source: https://en.wikipedia.org/wiki/Phrase_structure_grammar

IT STARTED WITH OPINIONS

samurai LABS

"Best place to watch the sunset and have a bonfire party in San Francisco. Mind your glasses!

They don't belong to the beach! Bring your dogs and surfboards!"

"Keep walking, you are under no obligation to give anybody money, don't flash your wallet,

and you will be fine. Beware that after dark even MORE homeless show up to the area."

"A fantastic place to visit. I'd recommend hiring a bike and riding across. Take a sweater

though as it can get quite chilly with the breeze."

IT STARTED WITH OPINIONS

samurai LABS

neutral:

"Best place to watch the sunset and have a bonfire party in San Francisco. Mind your glasses!

They don't belong to the beach! Bring your dogs and surfboards!"

neutral:

"Keep walking, you are under no obligation to give anybody money, don't flash your wallet, and you will be fine. Beware that after dark even MORE homeless show up to the area."

positive:

"A fantastic place to visit. I'd recommend hiring a bike and riding across. Take a sweater though as it can get quite chilly with the breeze."

THERE WERE TONS OF INTERESTING USER QUESTIONS

samurai LABS

When should I come?

What is this place recommended for?

What should I bring with me?

What kind of people can I expect there?

What should I watch out for?

What things should be avoided?

AND EVEN MORE FROM DIFFERENT PERSPECTIVES

samurai LABS

What do my users want me to add?

What do my users want me to change?

When should I come?

What is this place recommended for?

What should I bring with me?

What kind of people can I expect there?

What should I watch out for?

What things should be avoided?

Why do people change my product to another?

Why do people resign from my product?

What do my users have problems with?

HOW TO FIND ANSWERS FOR ALL THESE QUESTIONS?

samurai LABS

Make sure to take _____ with you as it can be chilly.

I wish there was _____ and _____, and then this app would have it all.

I've been taking _____ for 3 years and have experienced _____ and _____.

HOW TO FIND ANSWERS FOR ALL THESE QUESTIONS?

samurai LABS

Make sure to take _____ with you as it can be chilly.

I wish there was _____ and _____, and then this app would have it all.

I've been taking _____ for 3 years and have experienced _____ and _____.

Fill in the gaps with following terms:

an item to be brought

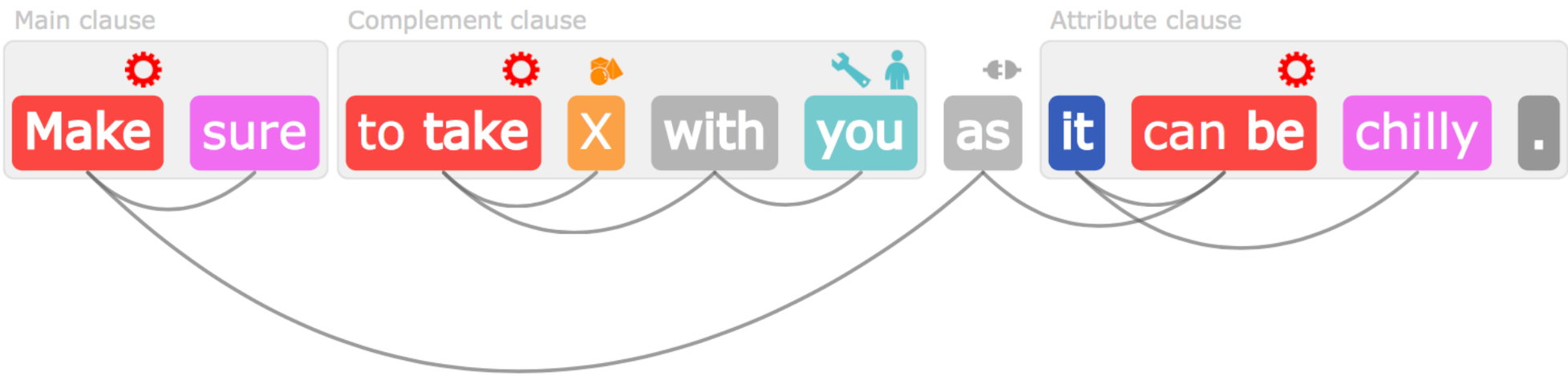
an adverse drug reaction (x2)

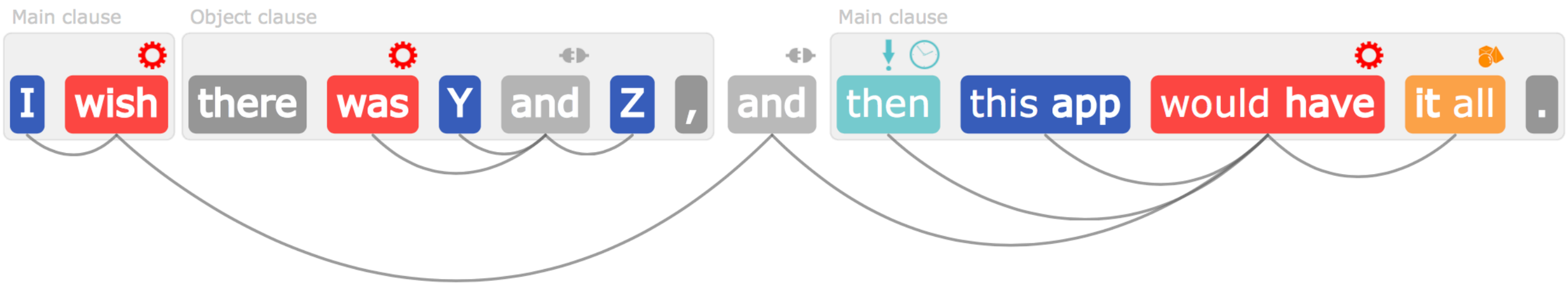
a drug name

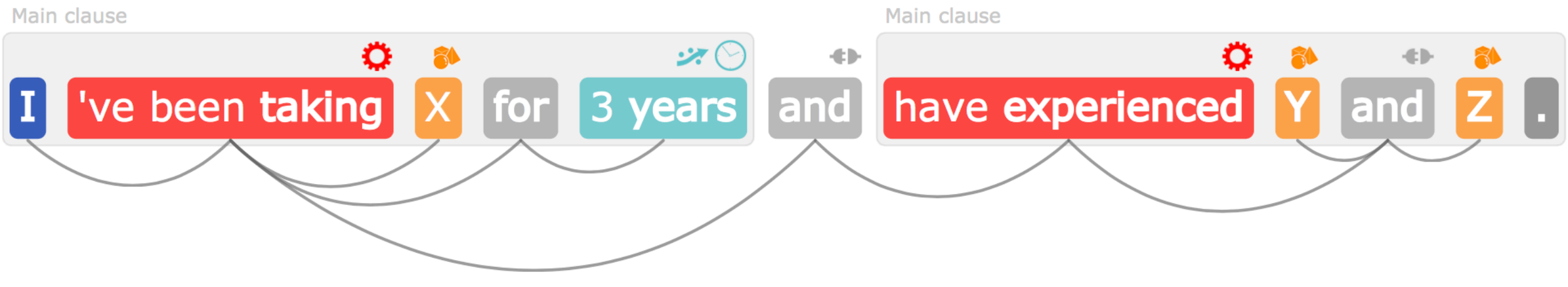
a feature to be added (x2)

WE HAVE A MAP

samurai LABS







All of the graphics are examples of a Samurai Labs' Syntactic Parser output

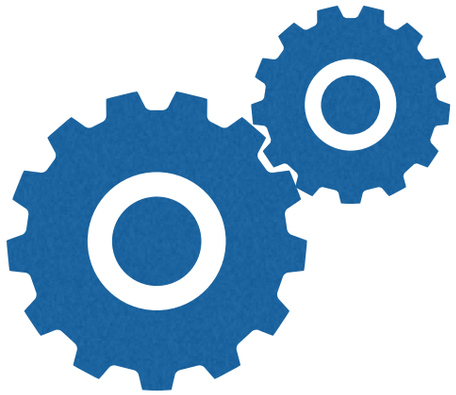
CONTEXT-BASED INFORMATION EXTRACTION

samurai LABS

```
1   SELECT
2   P:predicate AS DO,
3   P:target AS WHAT
4
5   WHERE
6   (
7       request(predicate)
8       [illegible]
9       [illegible]
10      [illegible]
11  )
12  AND target-predicate-connection(predicate, target)
13  AND NOT contrafactual-clause(predicate, target)
14  [illegible]
15  [illegible]
```

Example of an extraction model built with Samurai Labs' IE Query Language

Extraction Engine



What users ask you to add

Integration with evernote 14 | Calendar feature 4 | Geo fencing capabilities 3

Tags 2 | Pin code 2 | Search window 2 | more

What users ask you to fix

Font 4 | Microphone option 2 | Chrome gmail extension 2

Push notifications 2 | List panel on ipad version 1 | more

Example of a part of App Analytics proof of concept

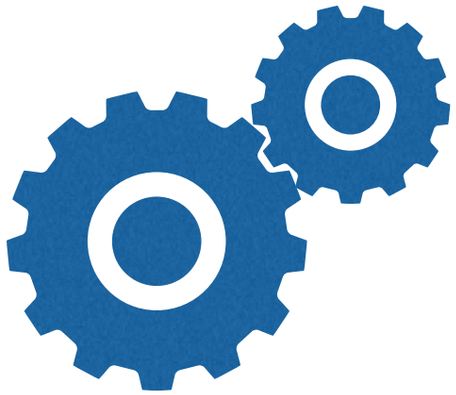
CONTEXT-BASED INFORMATION EXTRACTION

samurai LABS

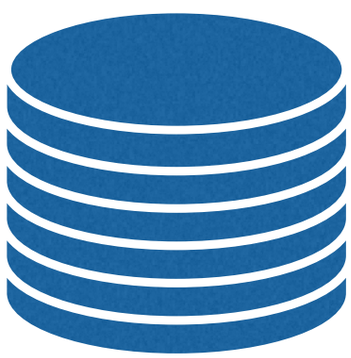
```
SELECT
P:predicate AS DO,
P:target AS WHAT

WHERE
(
    request(predicate)
    [illegible]
    [illegible]
    [illegible]
)
AND target-predicate-connection(predicate, target)
AND NOT contrafactual-clause(predicate, target)
[illegible]
[illegible]
```

Example of an extraction model built with Samurai Labs' IE Query Language

Extraction Engine



Library



What users ask you to add

Integration with evernote 14 | Calendar feature 4 | Geo fencing capabilities 3

Tags 2 | Pin code 2 | Search window 2 | more

What users ask you to fix

Font 4 | Microphone option 2 | Chrome gmail extension 2

Push notifications 2 | List panel on ipad version 1 | more

Example of a part of App Analytics proof of concept

WHY DO PEOPLE CHANGE ONE DRUG TO ANOTHER?

samurai LABS

ADHD DRUG#1

REASON FOR SWITCHING	ADDITIONAL CONDITION
get agitated easily	without it
sudden mood swings	
agitation	
dry mouth (3)	
loss of appetite (2)	
hard to get to sleep at night	sometimes
teeth grinding	
the stunting of my growth	
extreme appetite suppressment	
weight loss (2)	
grinding my teeth	as it wears off
the lack of appetite	
a fairly expensive medication	with no insurance
hard to fall asleep a night	if i take to late in the day
insomnia	
mild headaches	
i urinate like crazy	
excessive weight gain	
loss of weight	due to lack of appetite

ADHD DRUG#2

REASON FOR SWITCHING	ADDITIONAL CONDITION
dry mouth (3)	
not feeling hungry	
an increase in appetite	when i don't take it
become pretty lethargic	
loss of appetite (2)	
headaches (2)	from not eating
irritability	
keeps me awake at night	if i don't take it early
can no longer smoke weed	
expensive	if you're not covered
the dehydration	
the moodiness	when coming off from it
begin to get moody again	when it starts to wear off
bruising easily	
rapid heart beats	
dizziness	
increased agitation	when coming off of
the desire for a cigarette	for the first 2 hrs
a mild headache	around 9 pm

Raw data extracted from drug reviews and posts from various healthcare forums

WHAT PEOPLE ARE AFRAID OF?

samurai LABS

Source: https://www.reddit.com/r/dataisbeautiful/comments/3u8s3r/sneak_peek_into_redditors_minds_linguistic/

HOW PEOPLE EXPRESS…?

samurai LABS

HOW PEOPLE EXPRESS…?

samurai LABS

Recommendations

Vocative Case

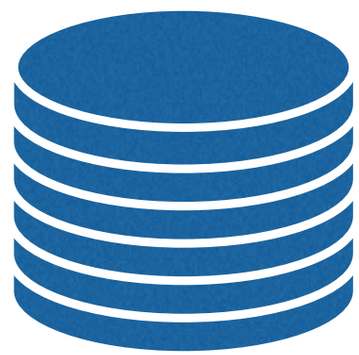
Greetings

Temporal Relations

Requests

Counter-factual Expressions

Wishes



Comparisons

Regrets

Persuasion

Reported Speech

Intentions

Negations

Linking Verb Relations

Conditionals

Coaxes

Possession

Assignments

Generalizations

Warnings

Clamors

Opinions

TECHNOLOGY MATCH: ONLINE VIOLENCE DETECTION

samurai LABS

Online Violence is a huge and very complex problem.

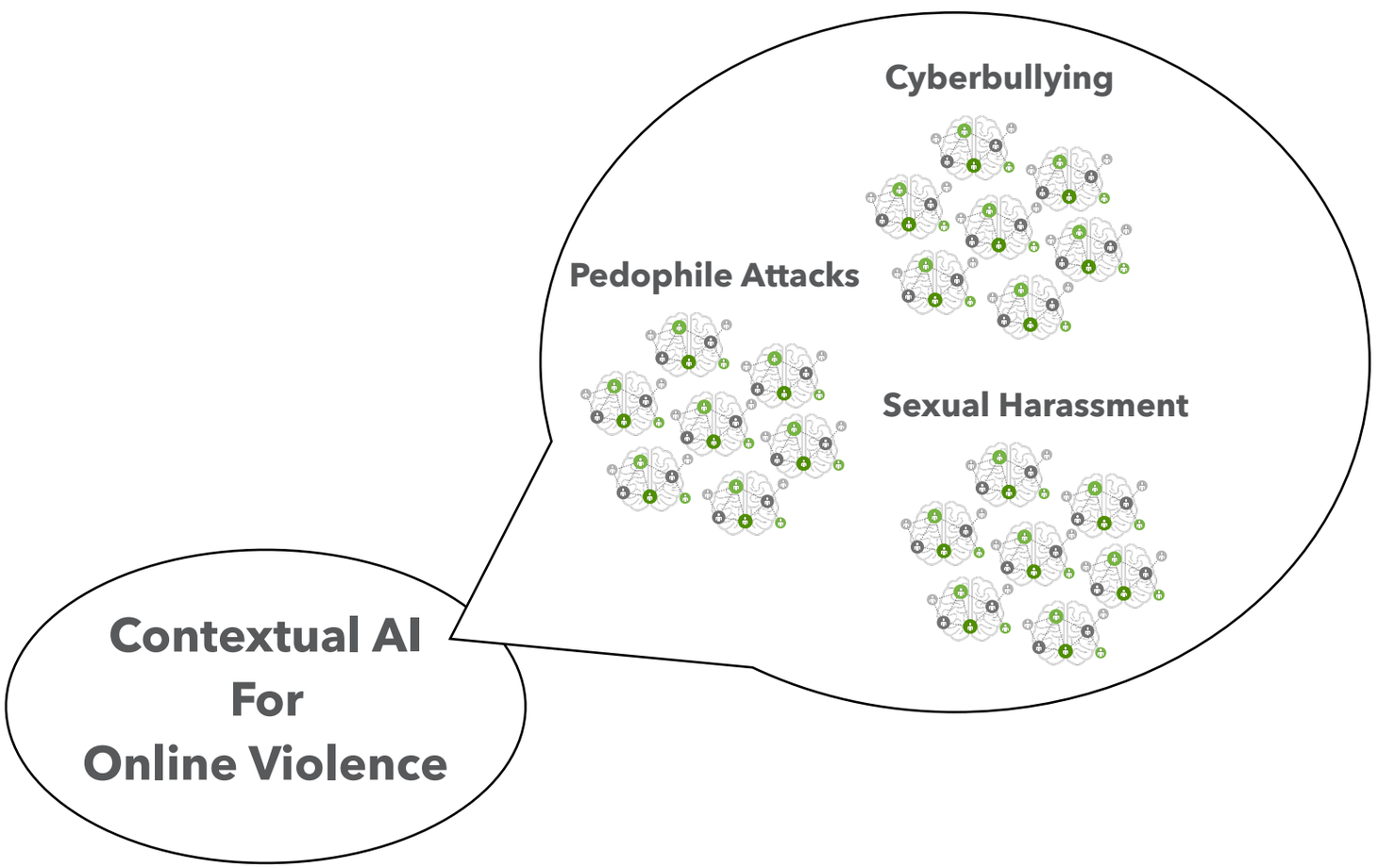
Contextual AI For Online Violence

TECHNOLOGY MATCH: ONLINE VIOLENCE DETECTION

samurai LABS

Online Violence is a huge and very complex problem.

That's why it needs to be divided into a set of smaller problems.

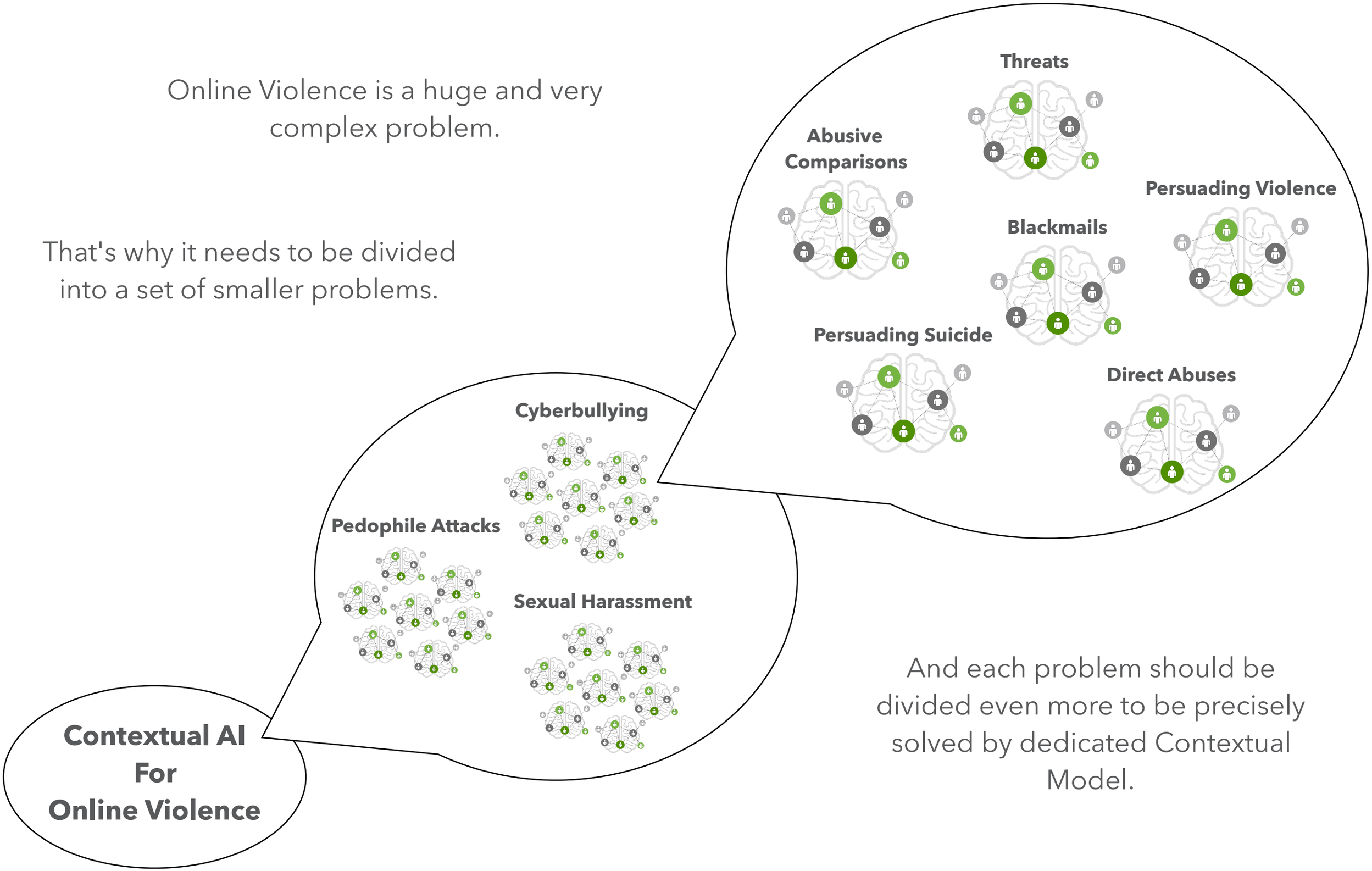


INTRODUCING CONTEXTUAL MODELS

samurai LABS

Online Violence is a huge and very complex problem.

That's why it needs to be divided into a set of smaller problems.



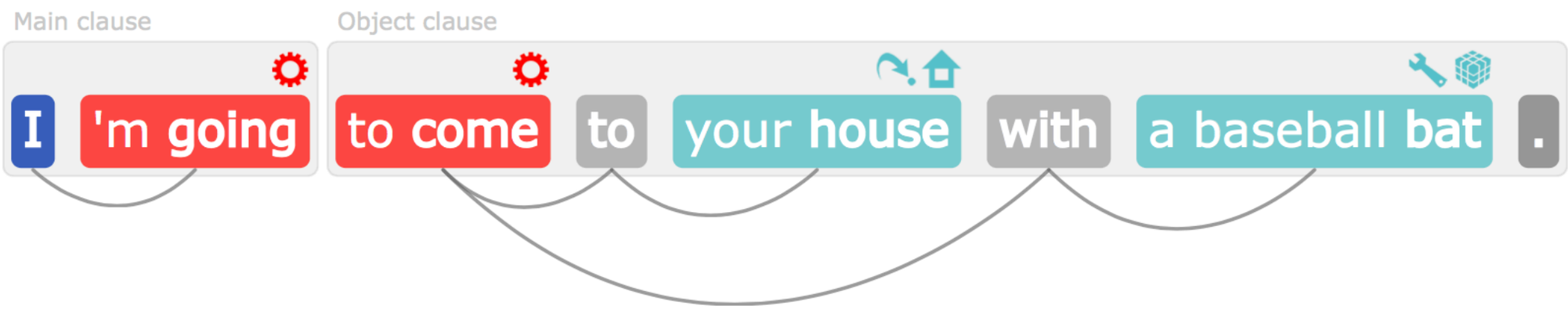
And each problem should be divided even more to be precisely solved by dedicated Contextual Model.

FEW EXAMPLES OF ONLINE VIOLENCE

samurai LABS

Personal Attack

Threat



Blackmail

All of the graphics are examples of a Samurai Labs' Syntactic Parser output

WHAT IS A BLACKMAIL?

samurai LABS

All of the graphics are examples of a Samurai Labs' Syntactic Parser output

WHAT IS A BLACKMAIL?

samurai LABS

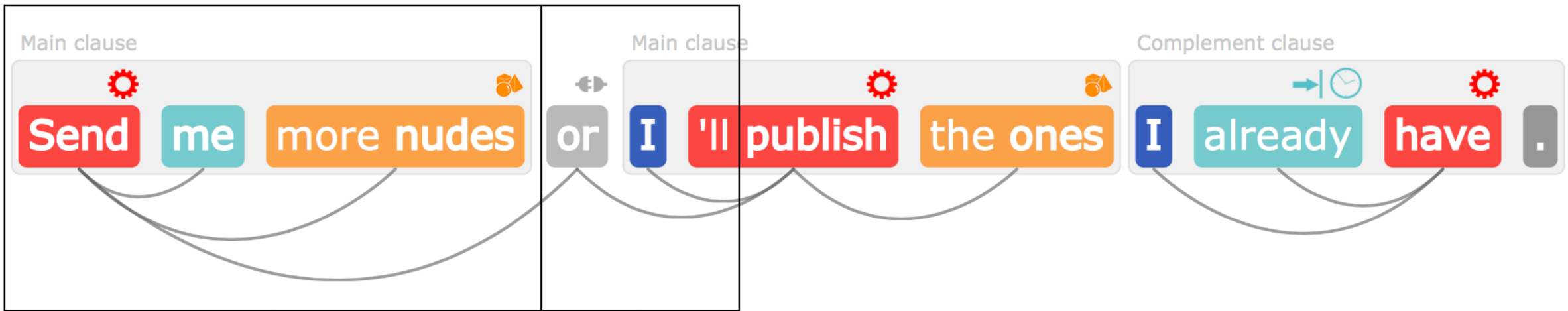
Request

Here: Using an imperative to express the request.

All of the graphics are examples of a Samurai Labs' Syntactic Parser output

WHAT IS A BLACKMAIL?

samurai LABS



Request

Here: Using an imperative to express the request.

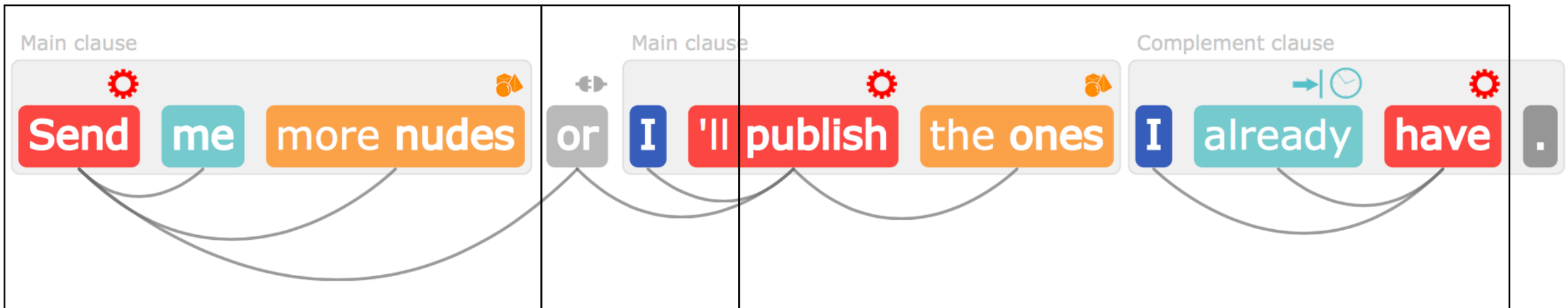
Alternative

Here: Using connector "or" along with future tense to express the alternative.

All of the graphics are examples of a Samurai Labs' Syntactic Parser output

WHAT IS A BLACKMAIL?

samurai LABS



Request

Here: Using an imperative to express the request.

Alternative

Here: Using connector "or" along with future tense to express the alternative.

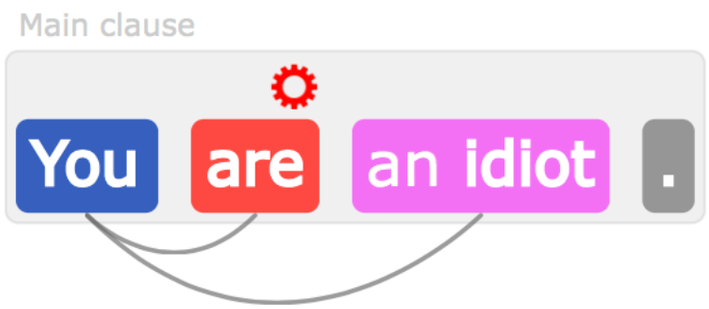
Negative Consequences

Here: Revealing something that the interlocutor wants to keep in secret as a form of negative consequence.

All of the graphics are examples of a Samurai Labs' Syntactic Parser output

GRAMMAR AS A KEY FOR ACHIEVING HIGH PRECISION

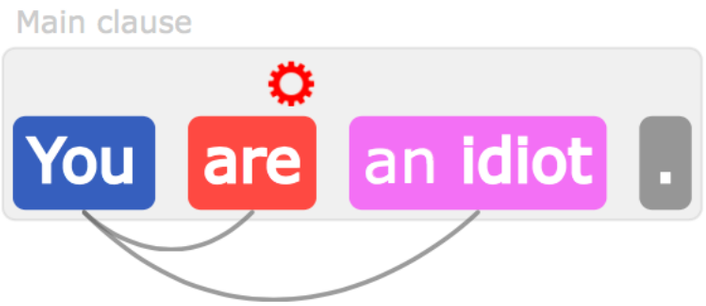
samurai LABS

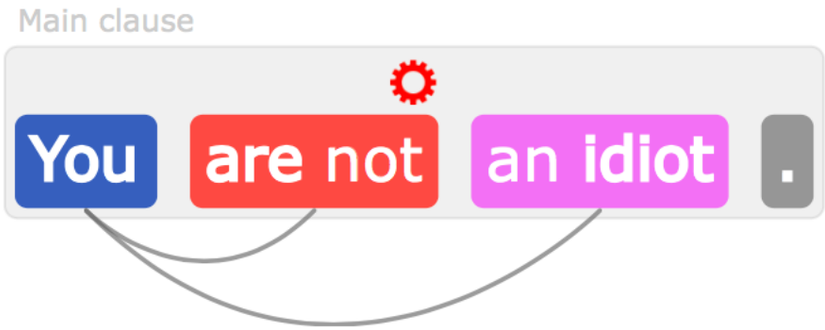


All of the graphics are examples of a Samurai Labs' Syntactic Parser output

GRAMMAR AS A KEY FOR ACHIEVING HIGH PRECISION

samurai LABS





All of the graphics are examples of a Samurai Labs' Syntactic Parser output

GRAMMAR AS A KEY FOR ACHIEVING HIGH PRECISION

samurai LABS

Likely to be perceived as toxic (0.99) Learn more

You are an idiot.

Likely to be perceived as toxic (0.77) Learn more

You are not an idiot.

Likely to be perceived as toxic (0.95) Learn more

I don't think you are an idiot.

Likely to be perceived as toxic (0.96) Learn more

I cannot believe he said you are an idiot.

Likely to be perceived as toxic (0.97) Learn more

And then I saw the message: you are an idiot.

Source: http://perspectiveapi.com/#/

HYBRID APPROACH

samurai LABS

Rule of thumb: use whatever works better for solving certain sub-task.

Symbolic:
- governance over the whole decision process,
- verification of excluding conditions,
- finding candidates for specific classification sub-tasks,
- normalization, semantic and syntactic transformations.

Statistical:
- dedicated classifiers for specific well-defined sub-tasks,
- supportive techniques for building rules and dictionaries.

BENCHMARK

samurai LABS

Cyberbullying Detection – Technical Report 2/2018
Department of Computer Science
AGH University of Science and Technology

Dataset:

RE-ANNOTATED (!) Formspring data for Cyberbullying Detection (12 772 posts); original dataset is available on Kaggle*

Cyberbullying in Original Annotation:

802 samples (6.3%)

Cyberbullying in New Annotation:

913 samples (7.1%)

* https://www.kaggle.com/swetaagrawal/formspring-data-for-cyberbullying-detection

Source: https://arxiv.org/abs/1808.00926

BENCHMARK: RESULTS

samurai LABS

Cyberbullying Detection – Technical Report 2/2018
Department of Computer Science
AGH University of Science and Technology

Dataset:

RE-ANNOTATED (!) Formspring data for Cyberbullying Detection (12 772 posts); original dataset is available on Kaggle*

Cyberbullying in Original Annotation:

802 samples (6.3%)

Cyberbullying in New Annotation:

913 samples (7.1%)

Solution	Precision	Recall
Fasttext	0.466	0.507
Samurai	0.804	0.843
System A	0.230	0.835
System B	0.277	0.656
System C	0.179	0.798
System D	0.111	0.821
System E	0.283	0.307

* https://www.kaggle.com/swetaagrawal/formspring-data-for-cyberbullying-detection

Source: https://arxiv.org/abs/1808.00926

KEY TAKEAWAYS

samurai LABS

1. Be mindful about what you measure.
2. Do not stick with well-defined tasks.
3. Symbolic / rule-based systems are not dead.
4. Grammar is a key for high-level NLP tasks.

THANK YOU

Gniewosz Leliwa | Samurai Labs
Director of AI Research & Co-founder
gniewosz@samurailabs.ai