

Interpretable Natural Language Processing

Fundamental studies for applied results



<https://agirussia.org>

Anton Kolonin
akolonin@aigents.com
Messengers: akolonin



The Plan

A bit of theory

Few cases

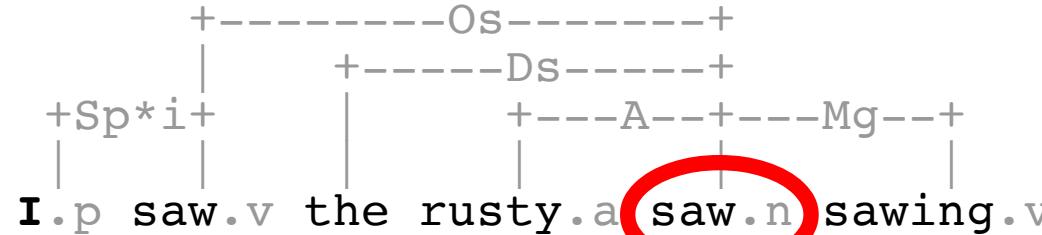
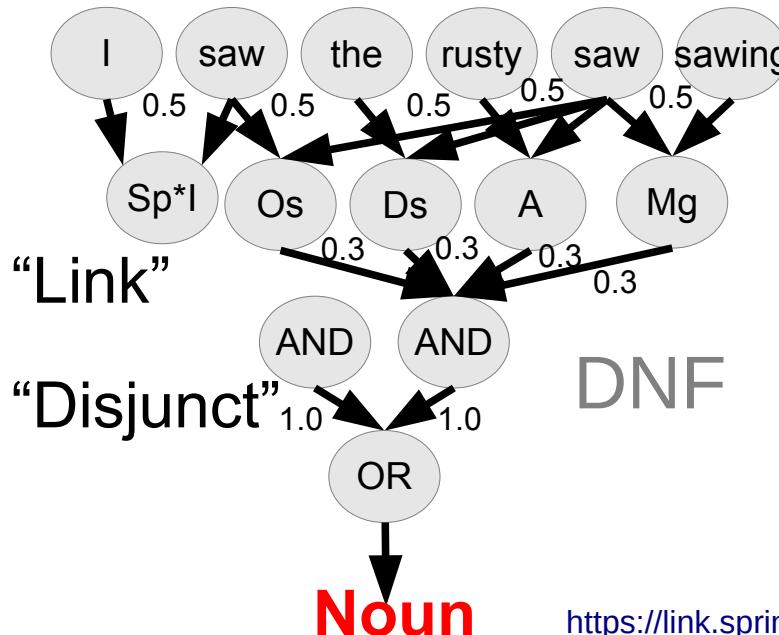
A bit more theory

A few more cases

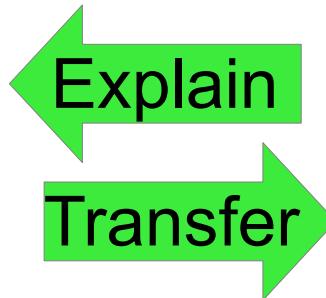
A bit of theory:
Fuzzy/probabilistic/soft-logic symbolic “models”

Bridging the Symbolic-Subsymbolic gap in NLP

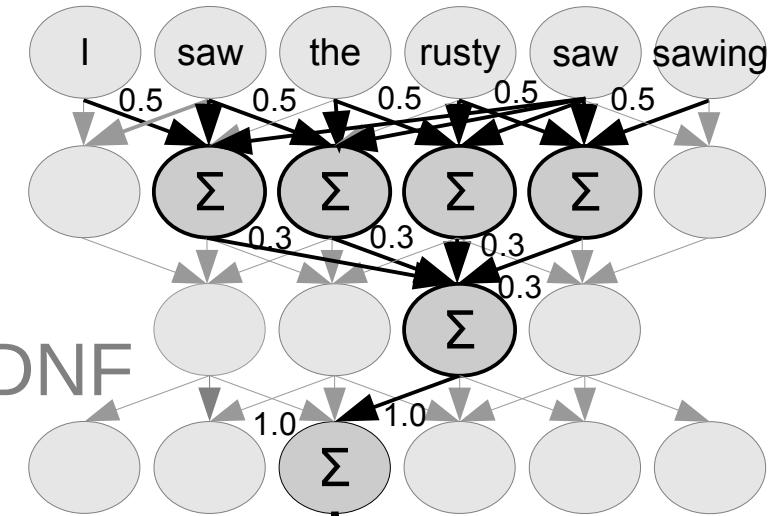
Formal
Link Grammar



Deep
Language Model



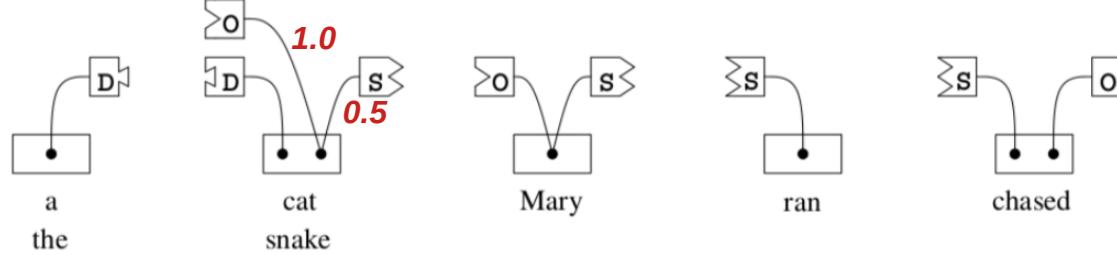
Soft-DNF



Noun

https://link.springer.com/chapter/10.1007/978-3-319-97676-1_11
https://doi.org/10.1007/978-3-030-27005-6_11
<https://github.com/singnet/language-learning/>

Link Grammar – Connectors/Costs, Disjuncts

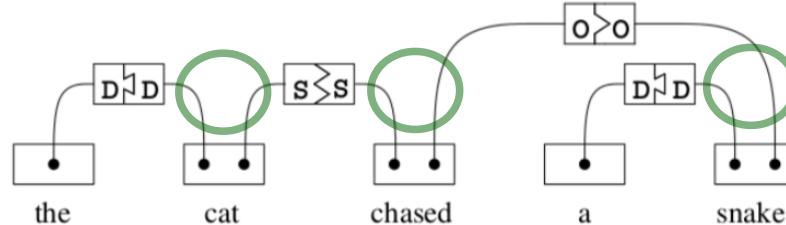


Connectors & Costs

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+;
```

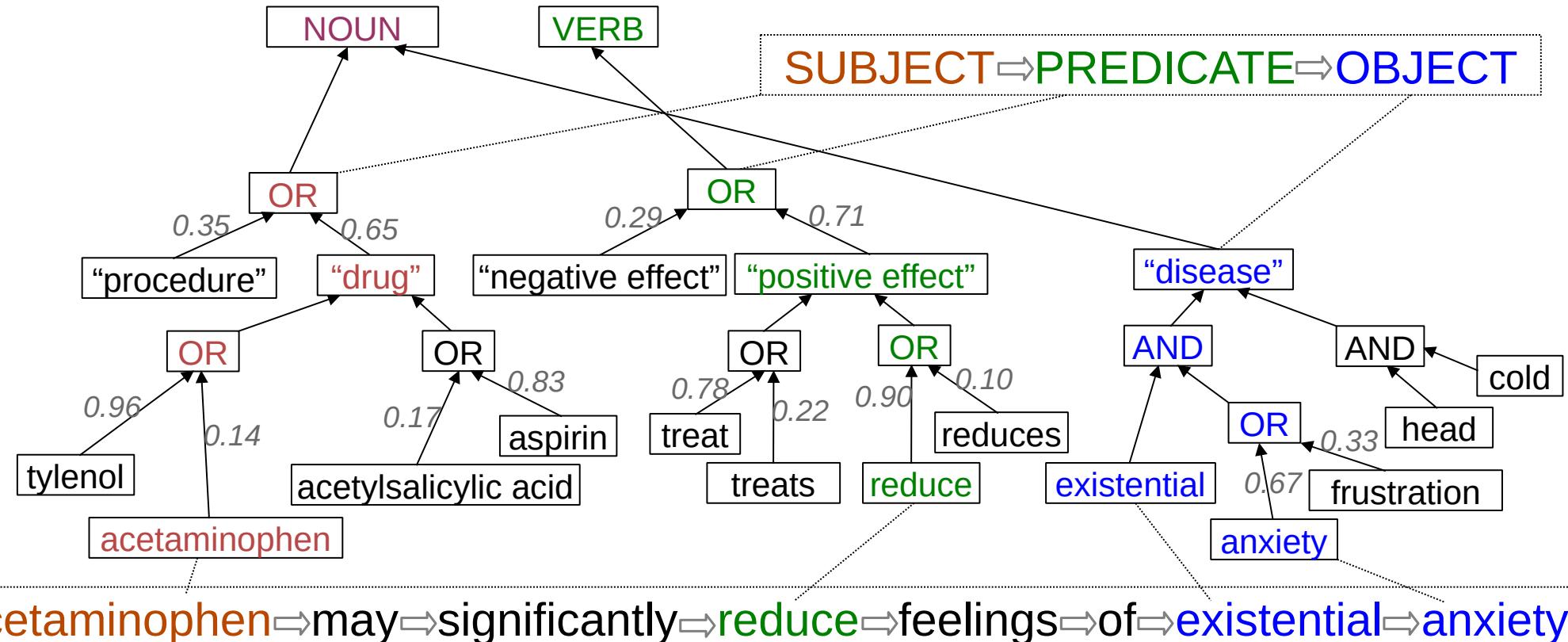
Note that although the symbols ‘‘&’’ and ‘‘or’’ are used to write down disjuncts, these are **not** Boolean operators, and do **not** form a Boolean algebra. They do form a non-symmetric compact closed monoidal algebra. The diagram below illustrates puzzle pieces, assembled to form a parse:



Disjuncts of Connectors

<https://arxiv.org/abs/1401.3372>
https://en.wikipedia.org/wiki/Link_grammar
<https://github.com/opencog/link-grammar>

NLP patterns (words, punctuation, phrases) for Interpretable Natural Language Processing (Aigents® “Deep Patterns”)



<https://ieeexplore.ieee.org/document/7361868>
<https://github.com/aigents/aigents-java>

Generalized Text Mining with Aigents® “Deep Patterns”

Classification

Category:

“Healthcare”

tylenol
acetaminophen
placebo

IS

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

HAS

“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.

<https://ieeexplore.ieee.org/document/7361868>
<https://github.com/aigents/aigents-java>

Few cases: Using symbolic “models”

Case 1: Property Attribution / Aigents® “Deep Patterns”

```
<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.

 Smallatraumatic soft tip. Ultra-thin 1 × 2 flat wire braid pattern”

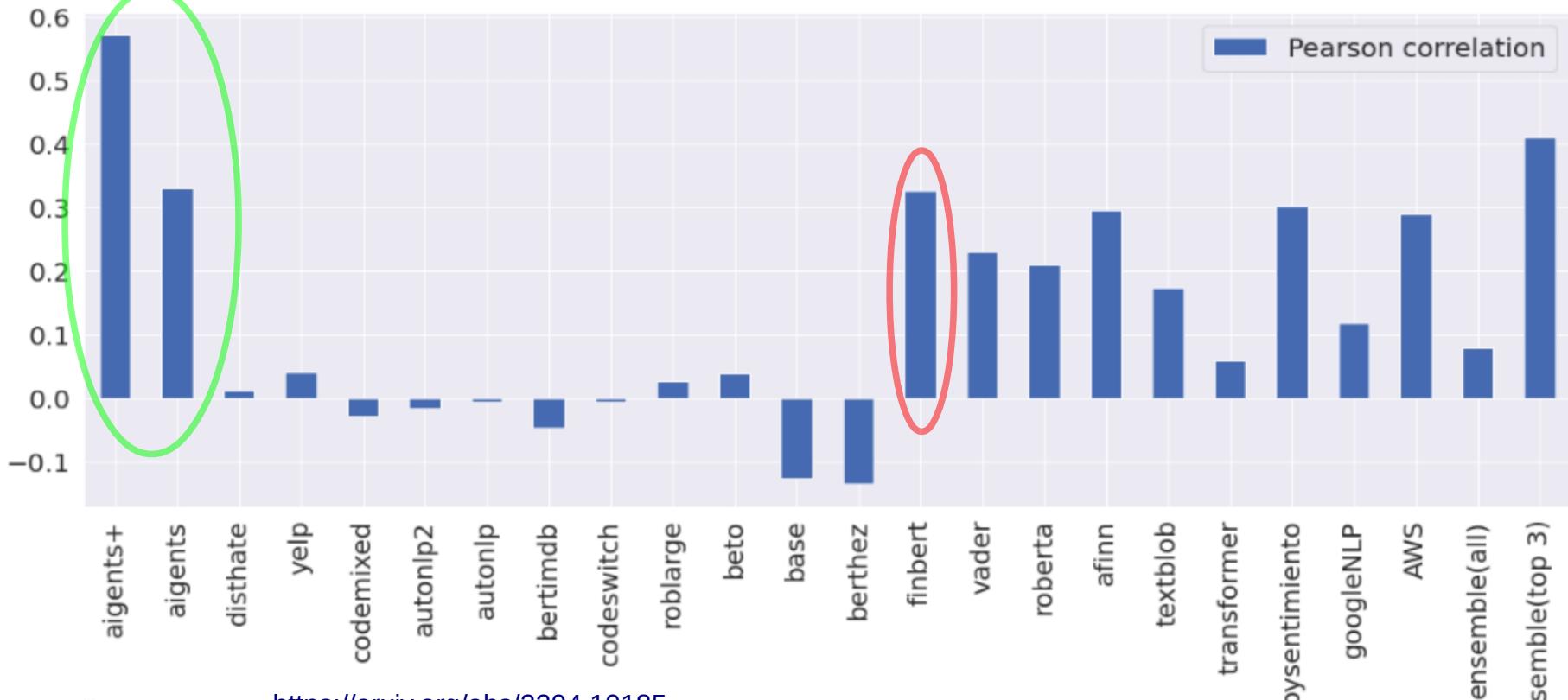
=

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

<https://ieeexplore.ieee.org/document/7361868>
<https://github.com/aigents/aigents-java>

Case 2: Interpretable Sentiment Analysis

Aigents® “interpretable” model vs. Bert fine-tuned on financial data
Average correlation across all models



Case 2.1: Aigents® News Feed Sentiment Analysis

The screenshot shows a news feed interface from aigents.com. At the top, there are navigation icons and a user profile for "Madmind Siberian". Below the header is a search bar labeled "Input search text" with a plus icon. The news feed lists several items:

- today**: putin says he has noted joe biden's sharp anti-Russian rhetoric (red bar, negative sentiment). 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 (yellow bar, neutral sentiment). <https://moderndiplomacy.eu/2020/09/12/kamala-harris-as-vice-president-attractive-for>
- yesterday**: **learn more about blockchain domains:** introduction video see you on thursday (green bar, positive sentiment). https://www.reddit.com/r/CryptoCurrency/comments/j6rjh/ama_with Brad_K
- yesterday**: The Russians may know more than we do : cia analyst says putin I doctors amid covid secrecy (pink bar, negative sentiment). https://www.reddit.com/r/politics/comments/j6t6yp/the_russians_may_know_more
- yesterday**: alexa-derived chatbot tech allegedly understands the british accent (green bar, positive sentiment). https://www.theregister.co.uk/2020/09/17/amazon_alex_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 (pink bar, negative sentiment). <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 (green bar, positive sentiment). 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 a news feed interface from aigents.com. At the top, there are navigation icons and a user profile for "Madmind Siberian". Below the header is a search bar labeled "Input new thing name or template" with a plus icon. The news feed lists several items:

- biden** (pink bar, negative sentiment)
- jinpingle** (green bar, positive sentiment)
- online social help** (yellow bar, neutral sentiment)
- personal artificial intelligence** (red bar, negative sentiment)
- putin** (pink bar, negative sentiment)
- sanders** (pink bar, negative sentiment)
- trump** (pink bar, negative sentiment)
- {agi hlai [strong ai] [strong artificial intelligence] [artificial general intelligence]}** (blue bar, neutral sentiment)
- {chatbot chat-bot [chat bot]}** (green bar, positive sentiment)
- {published putin vladimir}** (blue bar, neutral sentiment)
- {bitcoin btc btc/usd btcusd}** (yellow bar, neutral sentiment)
- {dollar usd usd/xau usdxau}** (yellow bar, neutral sentiment)
- {ethereum eth eth/usd ethusd}** (yellow bar, neutral sentiment)
- {gold xau xau/usd xauusd}** (yellow bar, neutral sentiment)

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>

Case 2.2: Aigents® Bot Sentimental Reactions in Chats

AigentsTest
15 members, 2 online

Life is wonderful 21:51 ✓

Aigents Bot Admin
Anton Kolonin
Life is wonderful
21:51

Death is not good 21:52 ✓

Aigents Bot Admin
Anton Kolonin
Death is not good
21:52

When I am in the bad mood, it sucks 21:53 ✓

Aigents Bot Admin
Anton Kolonin
When I am in the bad ...
21:53

Write a message... <https://t.me/AigentsBot>

Aigents
bot

5Ogm0fxNSEL7YoRYmvzmxo9U.jpg¹
title: biden to nominate ex-iowa gov
[https://www.reddit...](https://www.reddit.com/r/JoeBiden)

reddit
President-elect Joe Biden • r/JoeBiden
President-elect Joe Biden | We are the United States of America. There is not a single thing we cannot do. Are you with...

17:51

December 11

my telegram report, period 4 21:37 ✓

 report.html
40.8 KB
21:37

Forwarded from Anton Kolonin
Deaths sucks 21:51

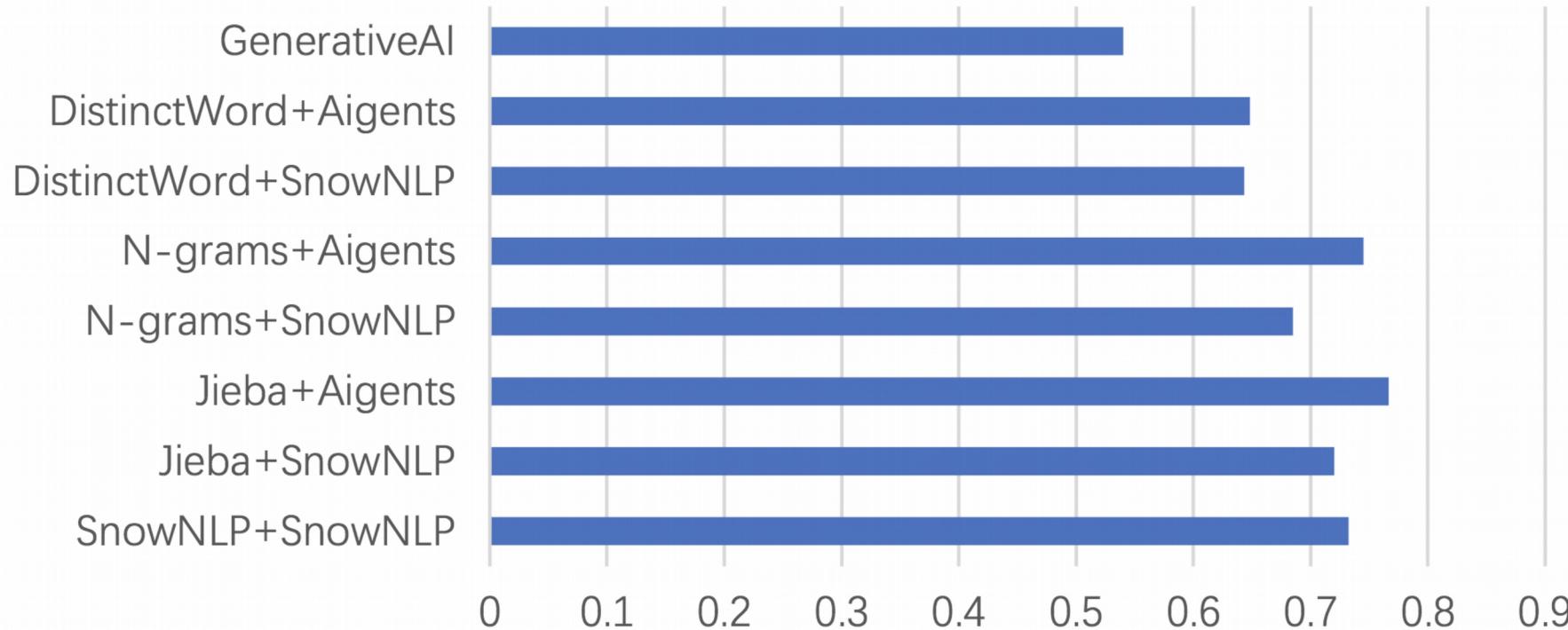
Forwarded from Anton Kolonin
When I am in the bad mood, it sucks 21:53

Write a message... <https://agents.medium.com/aigents-bot-for-telegram-groups-1dba32140047>

Case 2.3: Social Media Sentiment Analysis

H. Zhenghao, A. Kolonin: INTERPRETABLE SENTIMENT ANALYSIS AND TEXT SEGMENTATION FOR CHINESE LANGUAGE. 2024

Overall F1-score comparison



https://github.com/aigents/pygents/blob/main/pygents/aigents_api.py

https://github.com/aigents/pygents/blob/main/notebooks/hlp/sentiment/sentiment_test.ipynb

Case 3: Exploring Cognitive Distortions

COGNITIVE DISTORTION

Cognitive distortion refers to irrational and distorted thought patterns that skew one's perception of reality and contribute to negative emotions and behaviors.

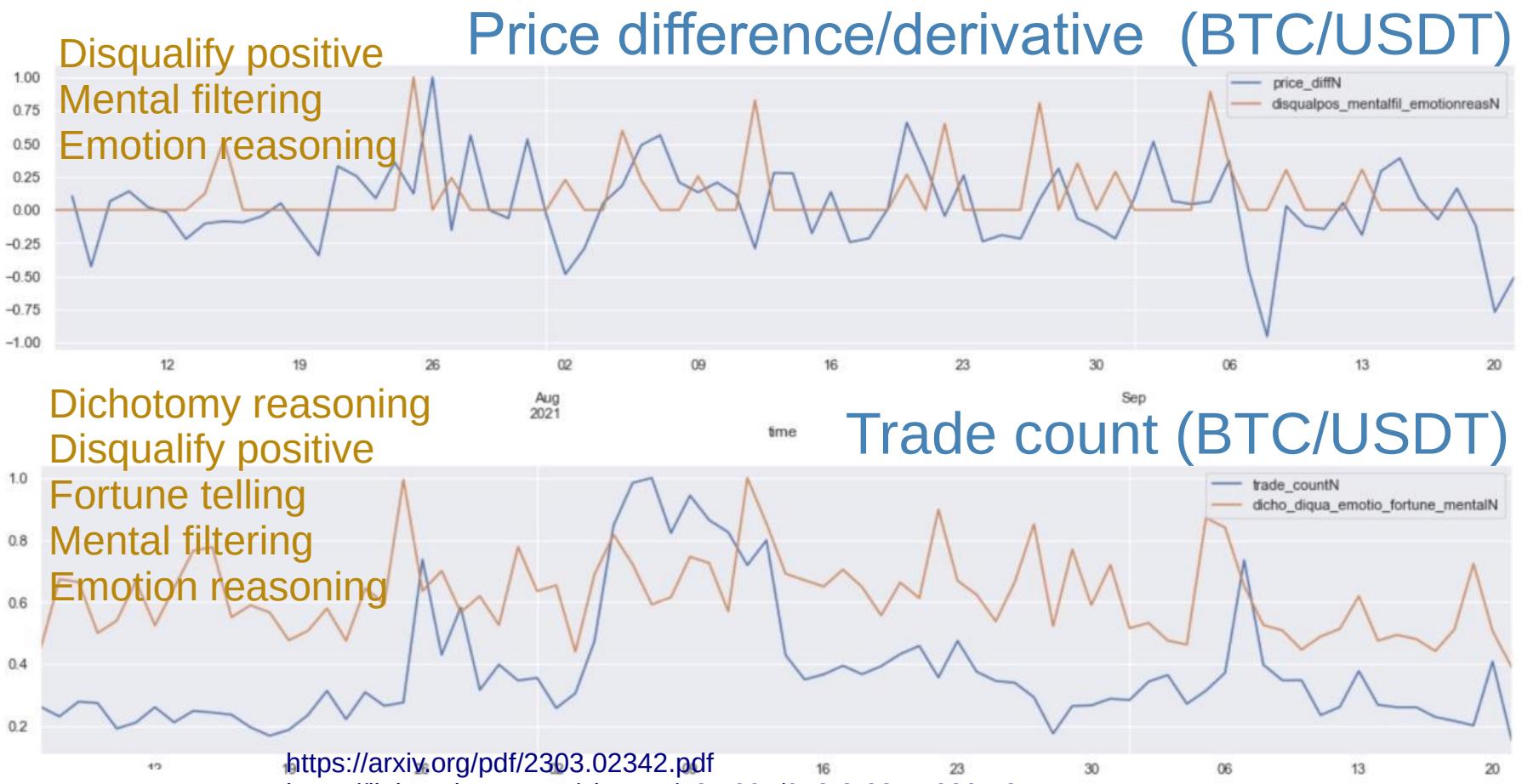
DEFINITION

Cognitive distortions are often deeply ingrained and can significantly impact an individual's mental well-being. Addressing cognitive distortions involves recognizing and challenging these distorted thoughts through strategies such as therapy, mindfulness, exercise, journaling, and social support.

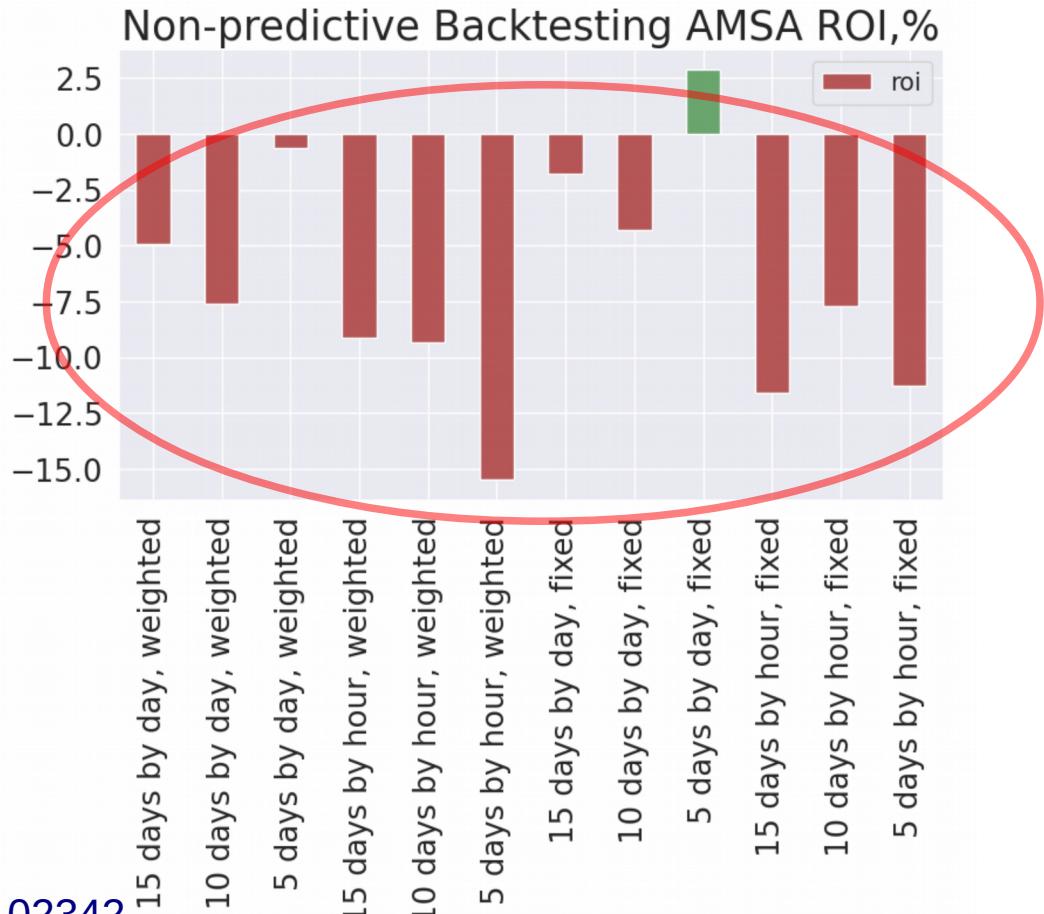
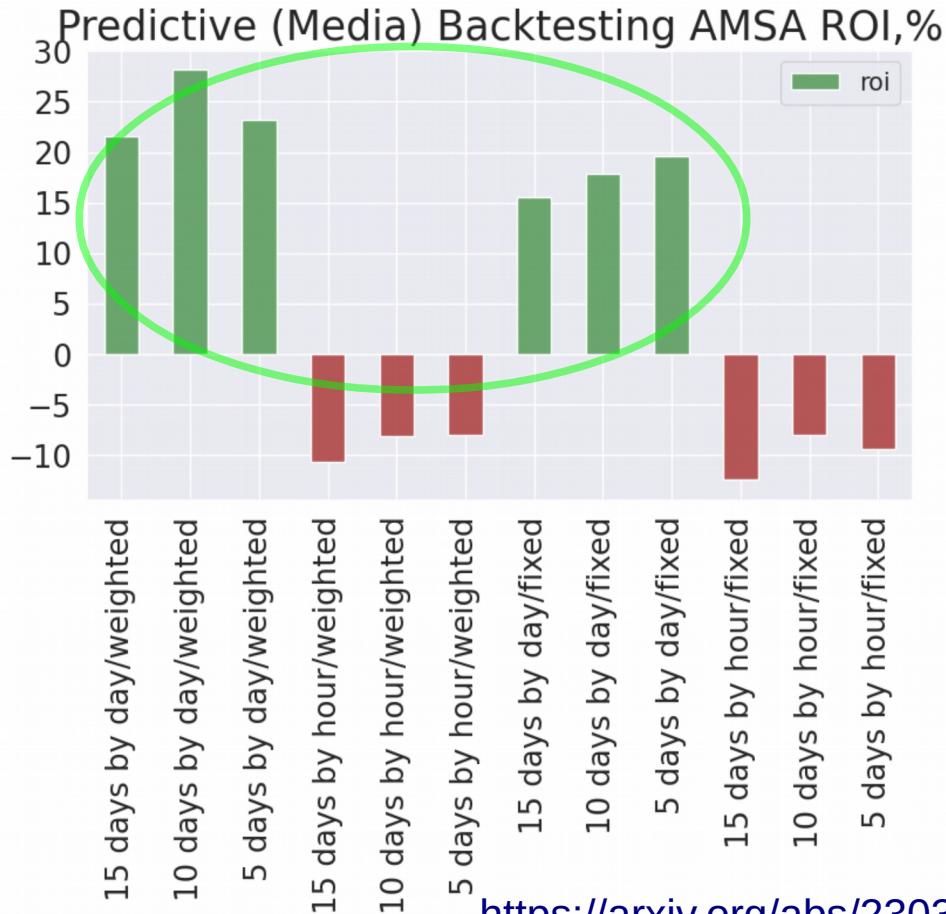
EXAMPLES

- All-or-Nothing Thinking
- Overgeneralization
- Filtering
- Catastrophizing
- Personalization
- Mind Reading
- Labeling
- Magnifying or Minimizing
- Emotional Reasoning
- Should Statements

Case 3.1: Cognitive Distortions on Media vs. Market



Case 3.1: Market Prediction on Cognitive Distortions



<https://arxiv.org/abs/2303.02342>

https://link.springer.com/chapter/10.1007/978-3-031-33469-6_19

Case 3.3: Cognitive Distortions for Psycho-Therapy

Cognitive Behavioral Therapy case: “Anxiety due to catastrophizing dismissal”

П: Здравствуйте! Как у вас дела?

К: Плохо. Выпустили лузера! (диагностируем #увольнение и #персонализация)

П: Как это случилось? Кто? (уточняем социальную ситуацию)

К: Да на работе. Уволили в общем. (диагностируем #увольнение)

П: Вас это пугает или вы обижены на них? (уточняем эмоцию - #тревога или #раздражение)

К: Конечно. Мне же ипотеку платить надо. (#тревога)

П: Вас уволили с работы и вы боитесь, что не сможете её оплатить? (#увольнение, #тревога)

К: Да. (подтверждаем диагноз в части социальную ситуацию и эмоции)

П: Что явилось причиной этой ситуации?

К: Мне просто не повезло - от компании ушел инвестор. (диагностируем #персонализация)

П: Вы в этом виноваты? (уточняем когнитивное искашение - #персонализация)

К: Нет. Просто я чувствую, что все пропало. (меняем диагноз на #катастрофизацию)

П: Вы боитесь, что после увольнения не будет новой работы? (#тревога, #увольнение, #катастрофизация)

К: Точно. Хотя я хороший специалист, на самом деле. (подтверждаем диагноз в части эмоции, социальной ситуации и когнитивного искашения с дальнейшим переходом к корректирующим воздействиям)

<https://www.sciencedirect.com/science/article/pii/S1877050922017458>

<https://arxiv.org/abs/2302.09377>

Case 3.3: Cognitive Distortions for Psycho-Therapy

A. Arinicheva, A. Kolonin: Diagnosis of Cognitive Distortions in Public, Group, and Personal Text Communications. 2024

Different thresholds

F_β score, $\beta=0.1$

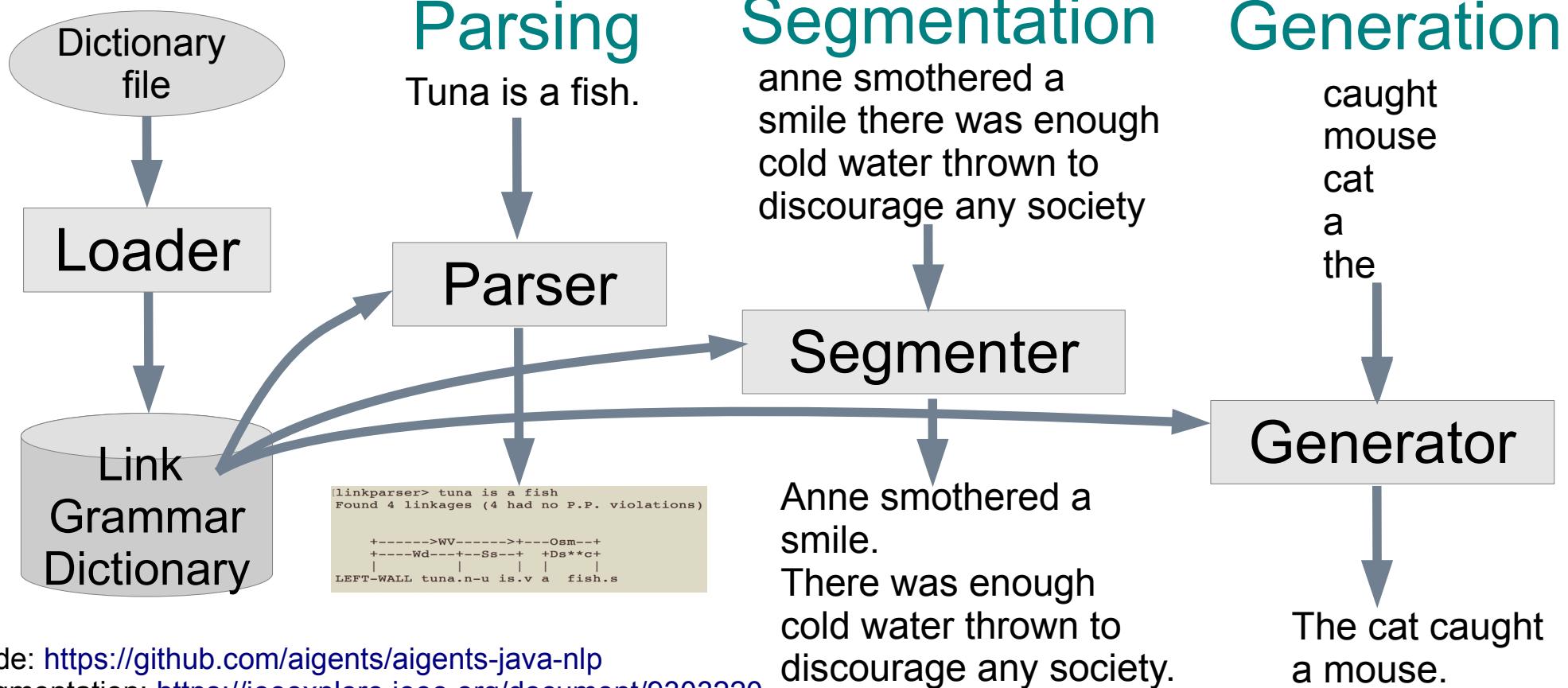
Contradictive statements	0.691	0.691	0.700	0.757	0.924	0.910	0.190	0.000	0.000
Positive statements	0.703	0.707	0.720	0.773	0.880	0.916	0.641	0.135	0.000
Negative statements	0.724	0.727	0.737	0.766	0.844	0.944	0.880	0.337	0.000
Rude statements	0.428	0.577	0.606	0.547	0.301	0.037	0.000	0.000	0.000
Mindreading	0.306	0.402	0.458	0.409	0.273	0.135	0.000	0.000	0.000
Dichotomous Reasoning	0.547	0.679	0.822	0.874	0.740	0.414	0.135	0.000	0.000
Emotional Reasoning	0.477	0.505	0.487	0.442	0.301	0.135	0.037	0.000	0.000
Should statements	0.892	0.947	0.970	0.979	0.385	0.072	0.000	0.000	0.000
Labeling and mislabeling	0.471	0.543	0.646	0.744	0.739	0.670	0.215	0.000	0.000
Personalizing	0.507	0.622	0.736	0.760	0.594	0.260	0.000	0.000	0.000
Magnification and Minimization	0.448	0.569	0.584	0.452	0.281	0.000	0.000	0.000	0.000
Overgeneralizing	0.528	0.633	0.728	0.670	0.524	0.370	0.135	0.000	0.000
Catastrophizing	0.190	0.190	0.163	0.105	0.105	0.037	0.000	0.000	0.000
Disqualifying the Positive	0.104	0.092	0.070	0.037	0.000	0.000	0.000	0.000	0.000
Fortune-telling	0.087	0.063	0.072	0.072	0.037	0.000	0.000	0.000	0.000
Mental Filtering	0.065	0.067	0.072	0.037	0.037	0.000	0.000	0.000	0.000
Average (all)	0.776	0.072	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Average (without Pos/Neg/Rud/Con)	0.496	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000



https://github.com/aigents/pygents/blob/main/pygents/aigents_api.py

https://github.com/aigents/pygents/blob/main/notebooks/nlp/sentiment/distortions_test.ipynb

Case 4: Link-Grammar based applications



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

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

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

V. Ramesh, A. Kolonin, 2020-2021

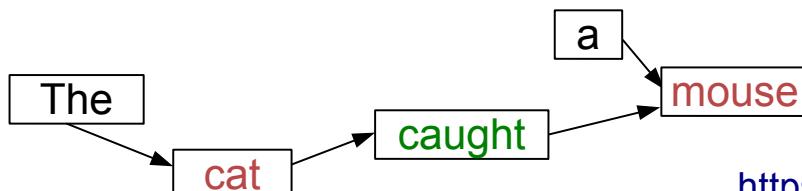
Case 4.2: Grammatical Language Generation

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



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

<https://arxiv.org/abs/2105.00830> V. Ramesh, A. Kolonin, 2021

Copyright © 2024 Anton Kolonin, Aigents®

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

Case 4.3: Question Answering with Link Grammar

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

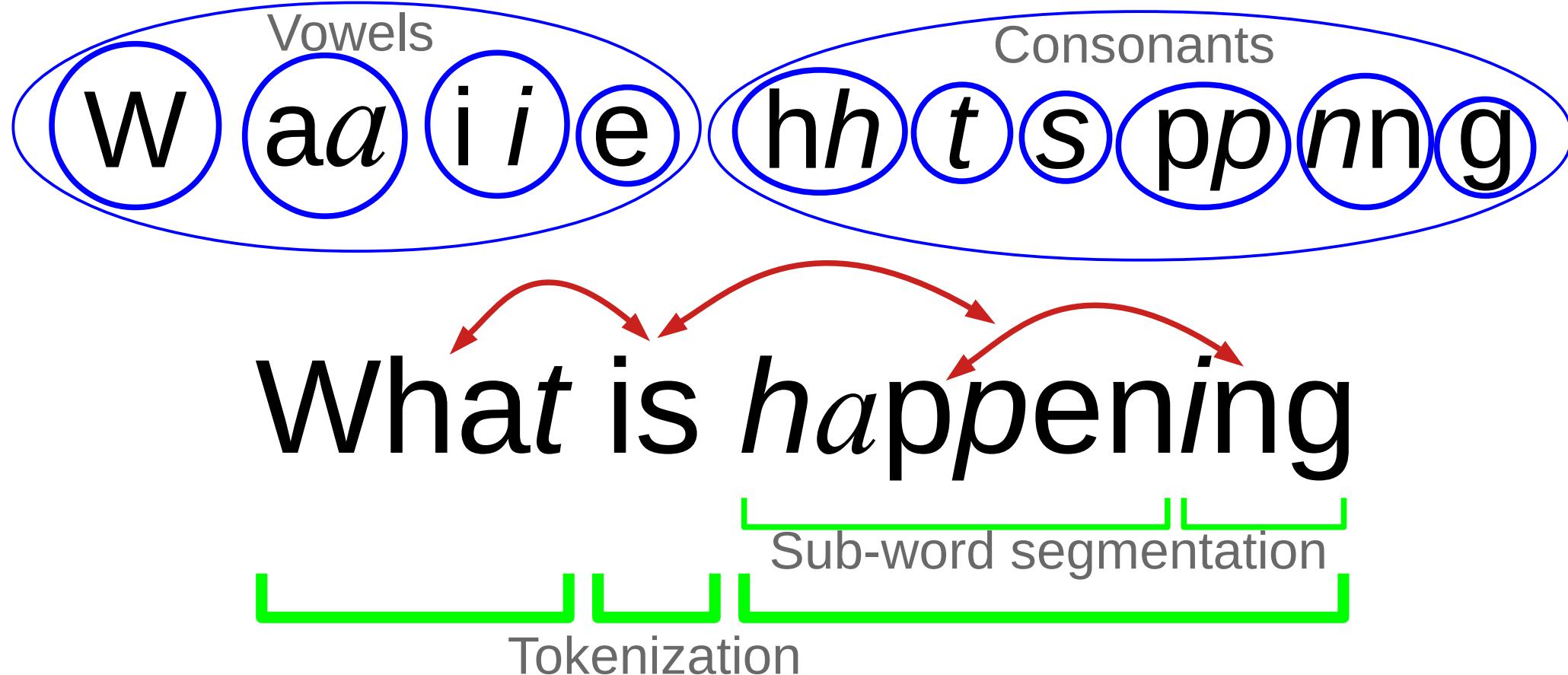
<https://www.youtube.com/watch?v=MKIOqO9FRq0>

https://link.springer.com/chapter/10.1007/978-3-030-93758-4_22 V. Ramesh, A. Kolonin, 2021

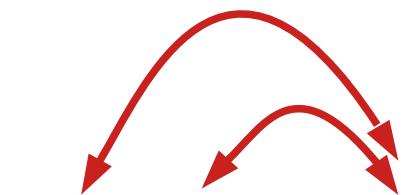
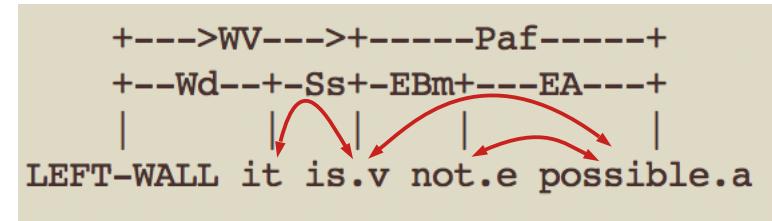
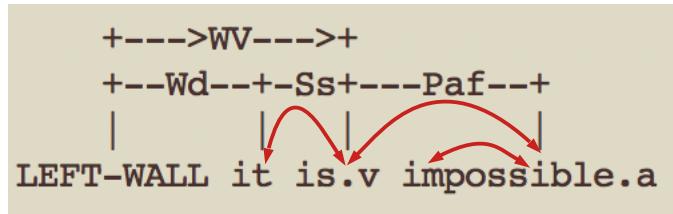
A bit more theory: Learning symbolic “models”

Learning Lexicon, Punctuation, Morphology and Grammar

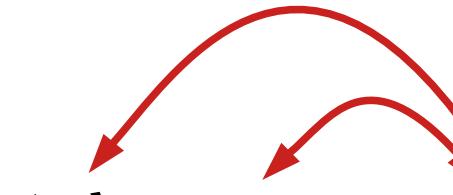
Clustering, Segmentation and Parsing



Blurring boundaries between grammatical parsing and morphological parsing causes tokenization ambiguity

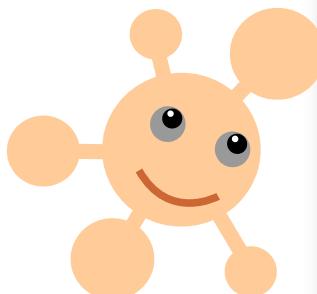


这不可能
(Google)



这 不 可 能
(Jieba)

Minimizing Uncertainty



New Tab x +

how are you

- how are you - Google Search
- how are
- how are you doing
- how are you answers
- How Are You Feeling - Song by TAYLOR DEE
- How Are You Today? - Song by Maple Leaf Learning
- how are you doing answer
- how are you synonyms
- how are you in spanish
- how are things going

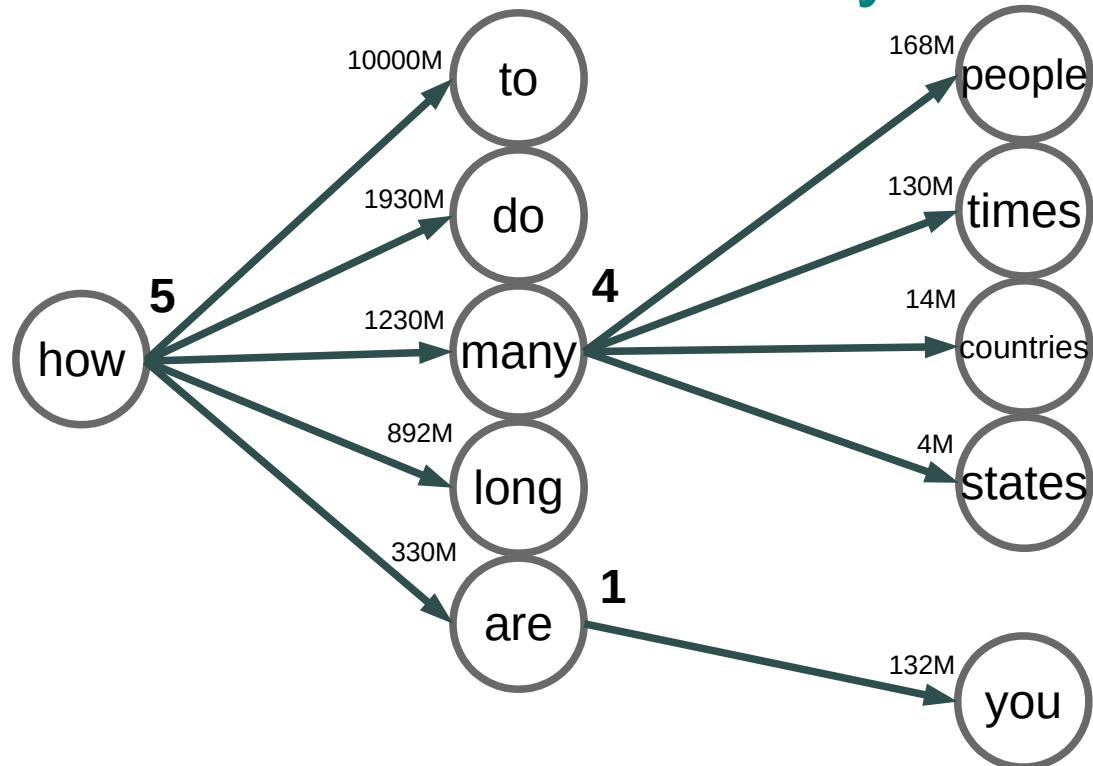


New Tab x +

how many

- how many - Google Search
- how many countries in the world
- how many weeks in a year
- how many states in usa
- how many continents
- how many people in the world
- how many words
- how many continents are there
- how many bones in human body
- how many episodes in house of dragons

Unsupervised Learning for Text Segmentation based on Probability and Uncertainty Measures



Metrics/Indicators:

Mutual Information¹
Conditional Probability^{1,2}
Transition Freedom^{2,3}

¹ <https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=6983&context=etd>

² <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2655800/>

³ Karl Friston. The free-energy principle: a unified brain theory? <https://www.nature.com/articles/nrn2787>

A few more cases: Learning and using symbolic “models”

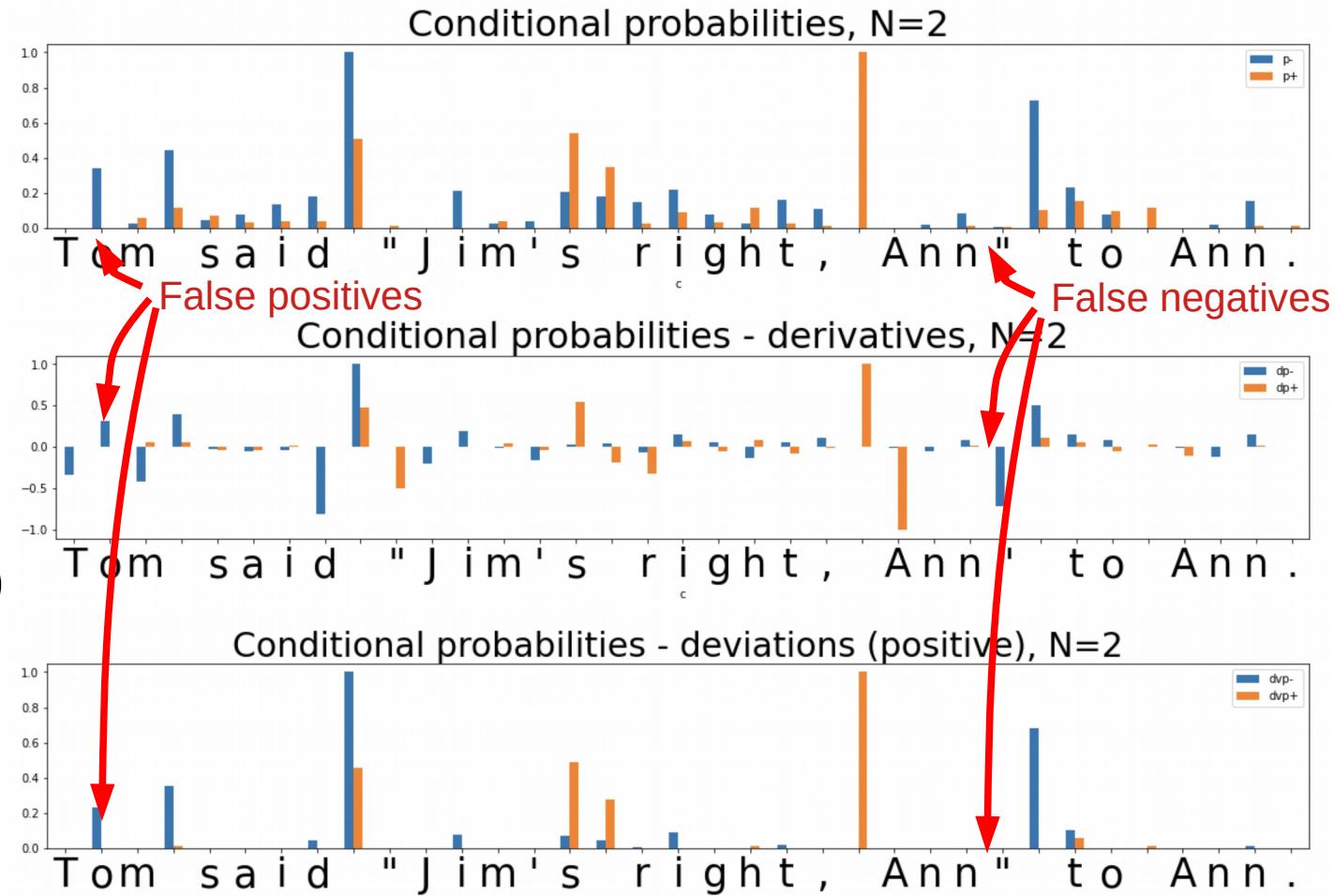
Case 5: Unsupervised Text Segmentation (Probability)

Metrics/Indicators:

Ngram (Character)
Conditional
Probability
(of Transition)

$P(\text{Ngram}_{n+1})/P(\text{Ngram}_n)$

$P("m")/P(m")$

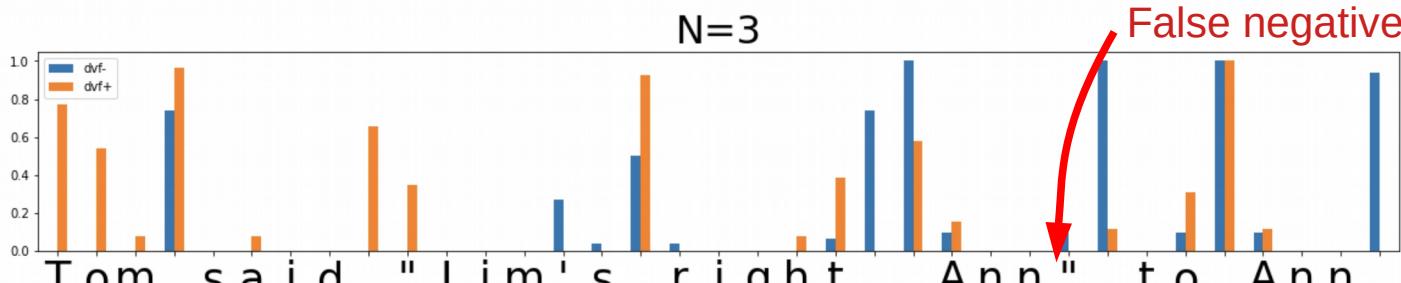
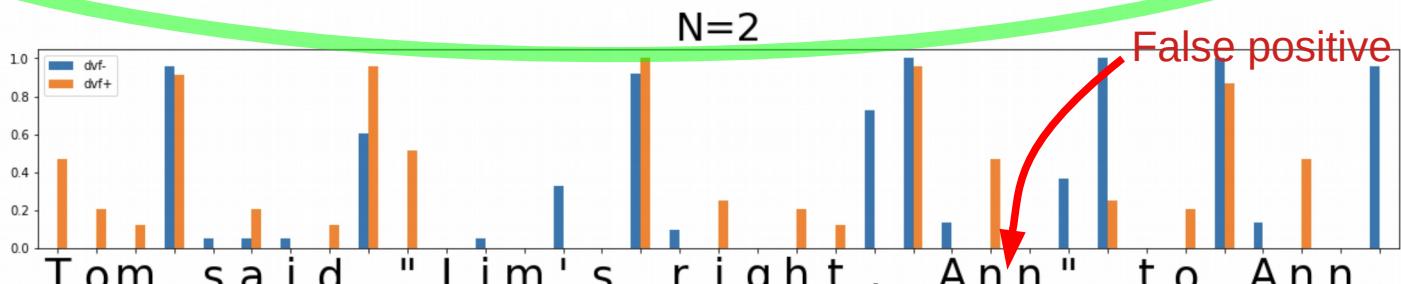
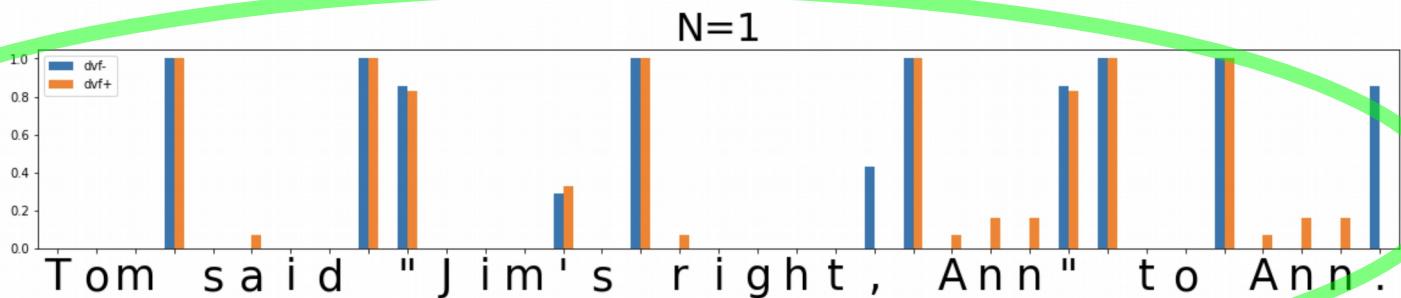


Case 5: Unsupervised Text Segmentation (Uncertainty)

Metrics/
Indicators:

Transition
Freedom
Deviation

(varying “N”
of N-gram)



Case 5: Unsupervised Text Segmentation (Hyper-parameters)

English

F1 - Brown ddf- & ddf+ filter=0 parameters=10967135

Hyper-Parameters:

Metric:
Transition
Freedom

[1]	0.5	0.75	0.82	0.79	0.79	0.81	0.89	0.89	0.89
[2]	0.46	0.54	0.62	0.67	0.85	0.92	0.81	0.71	0.37
[3]	0.56	0.67	0.72	0.73	0.69	0.61	0.46	0.36	0.19
[4]	0.54	0.68	0.7	0.6	0.43	0.3	0.19	0.15	0.1
[5]	0.51	0.55	0.52	0.38	0.25	0.16	0.11	0.1	0.08
[6]	0.48	0.46	0.38	0.25	0.17	0.12	0.1	0.08	0.07
[7]	0.42	0.34	0.24	0.15	0.11	0.1	0.08	0.08	0.07
[1, 2]	0.47	0.58	0.82	0.94	0.94	0.91	0.89	0.79	0.56
[2, 3]	0.51	0.62	0.74	0.79	0.83	0.81	0.66	0.46	0.24
[1, 2, 3]	0.5	0.69	0.79	0.87	0.91	0.89	0.78	0.58	0.25
[1, 2, 3, 4]	0.55	0.75	0.84	0.86	0.84	0.75	0.52	0.31	0.15
[4, 5, 6, 7]	0.56	0.6	0.51	0.33	0.2	0.14	0.1	0.08	0.07
[1, 2, 3, 4, 5]	0.56	0.78	0.86	0.84	0.74	0.53	0.31	0.17	0.1
[1, 2, 3, 4, 5, 6, 7]	0.59	0.78	0.82	0.69	0.49	0.26	0.15	0.09	0.07
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9

Threshold
for model
compression

F1 - Brown ddf- & ddf+ filter=0.0001 parameters=8643703

Combination
of Ngram N-s

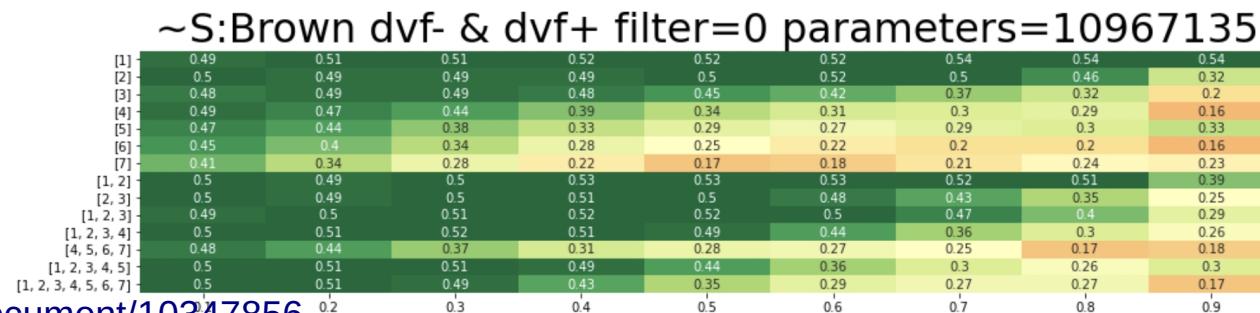
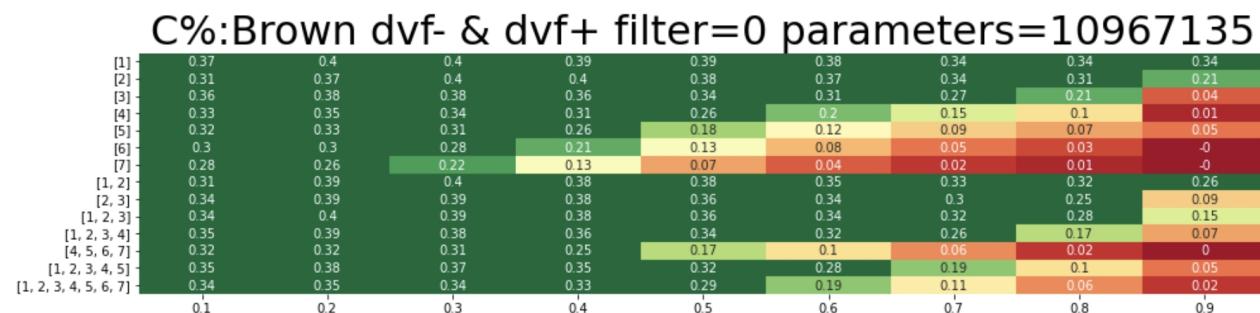
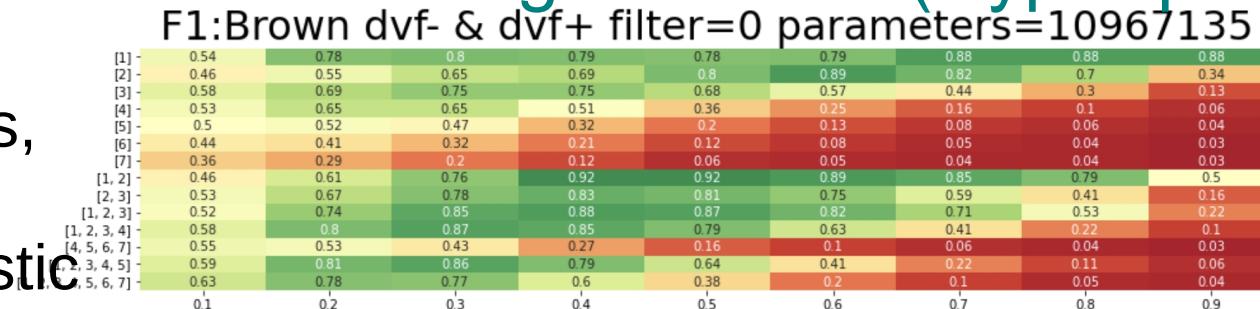
[1]	0.73	0.96	0.98	0.99	0.96	0.94	0.95	0.95	0.89
[2]	0.46	0.54	0.64	0.77	0.91	0.94	0.89	0.7	0.44
[3]	0.55	0.66	0.74	0.78	0.72	0.65	0.49	0.37	0.19
[4]	0.54	0.67	0.7	0.61	0.45	0.32	0.21	0.16	0.1
[5]	0.51	0.55	0.52	0.38	0.26	0.17	0.12	0.1	0.08
[6]	0.48	0.46	0.38	0.26	0.18	0.13	0.1	0.09	0.07
[7]	0.42	0.35	0.25	0.16	0.12	0.1	0.09	0.08	0.08
[1, 2]	0.51	0.64	0.82	0.96	0.96	0.96	0.9	0.88	0.68
[2, 3]	0.5	0.62	0.74	0.85	0.89	0.86	0.71	0.51	0.27
[1, 2, 3]	0.53	0.69	0.81	0.91	0.93	0.92	0.82	0.6	0.36
[1, 2, 3, 4]	0.55	0.75	0.86	0.88	0.88	0.81	0.57	0.33	0.17
[4, 5, 6, 7]	0.56	0.6	0.52	0.35	0.22	0.15	0.1	0.09	0.07
[1, 2, 3, 4, 5]	0.57	0.79	0.88	0.86	0.78	0.59	0.33	0.18	0.1
[1, 2, 3, 4, 5, 6, 7]	0.59	0.79	0.83	0.71	0.5	0.28	0.16	0.09	0.08
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9

Threshold for
segmentation

Case 5: Unsupervised Text Segmentation (Hyper-parameters)

English,
Brown corpus,
F1 vs.
culture-agnostic
metrics

Maximizing F1,
compression
factor (C%)
and
normalized
anti-entropy
(~S) in the
space of hyper-
parameters

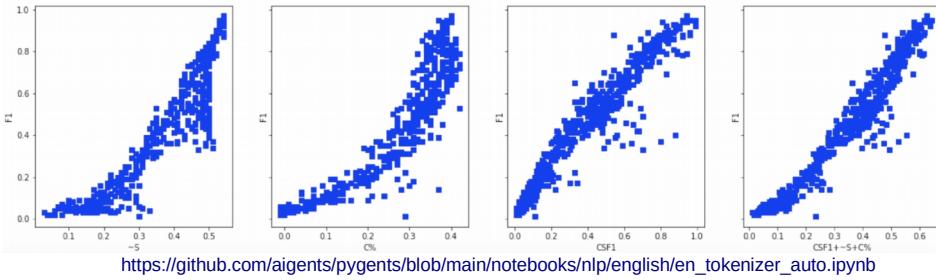


<https://ieeexplore.ieee.org/document/10347856>

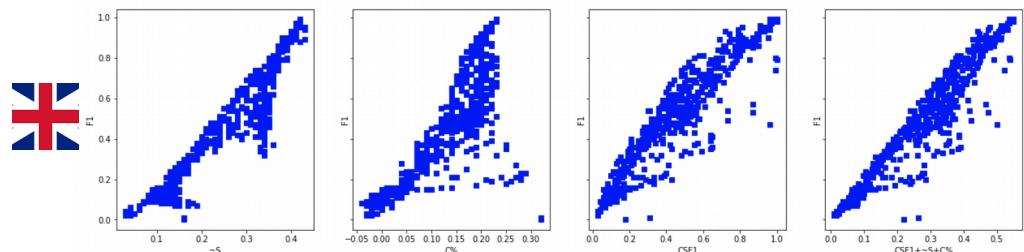
https://link.springer.com/chapter/10.1007/978-3-031-44865-2_1

Case 5: Unsupervised Text Segmentation (Different corpora)

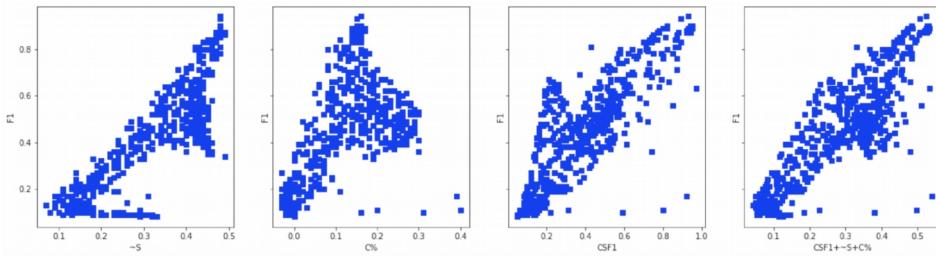
English, Train: Brown, Test: Brown 1000



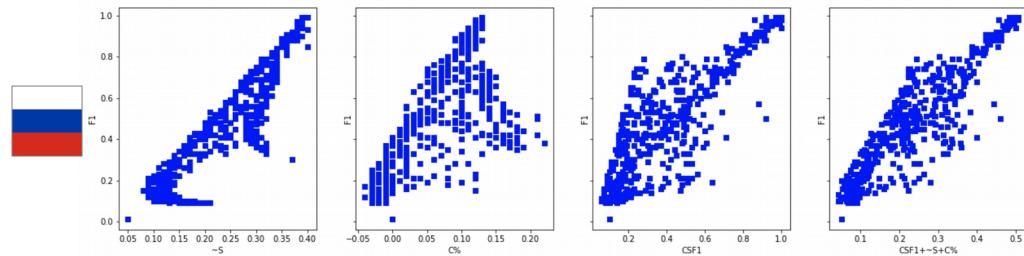
English, Train: Brown, Test: MagicData 100



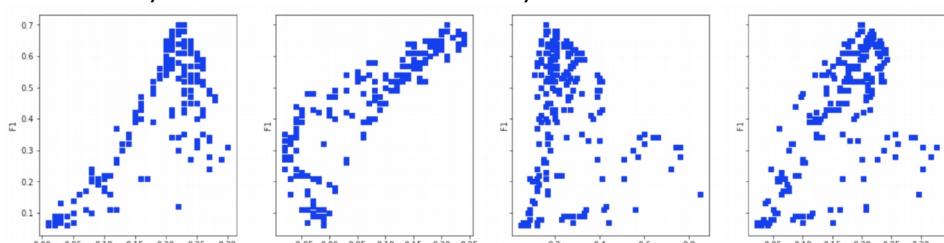
Russian, Train: RusAge, Test: RusAge 1000



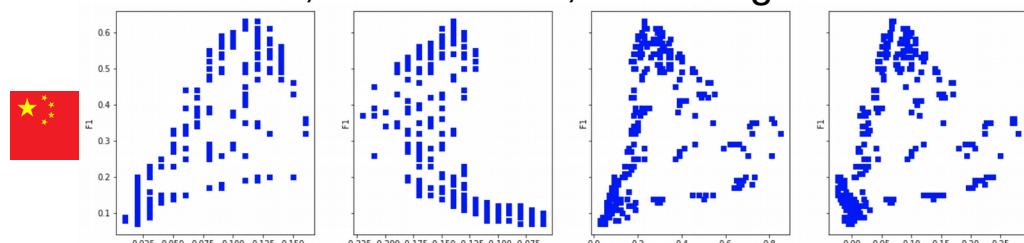
Russian, Train: Brown, Test: MagicData 100



Chinese, Train: CLUE News, Test: CLUE News 1000



Chinese, Train: Brown, Test: MagicData 100



Case 5: Unsupervised Text Segmentation ("Freedom"-based Tokenization against Lexicon-based one, referring to Rule-based)

Language	Tokenizer	Tokenization F_1	Lexicon Discovery Precision
English	Freedom-based	0.99	0.99 (vs. 1.0)
English	Lexicon-based*	0.99	-
Russian	Freedom-based	1.0	1.0 (vs. 1.0)
Russian	Lexicon-based*	0.94	-
Chinese	Freedom-based	0.71	0.92 (vs. 0.94)
Chinese	Lexicon-based*	0.83	-

*Lexicon-based Tokenization - greedy/beam search on word length (optimal) or frequency

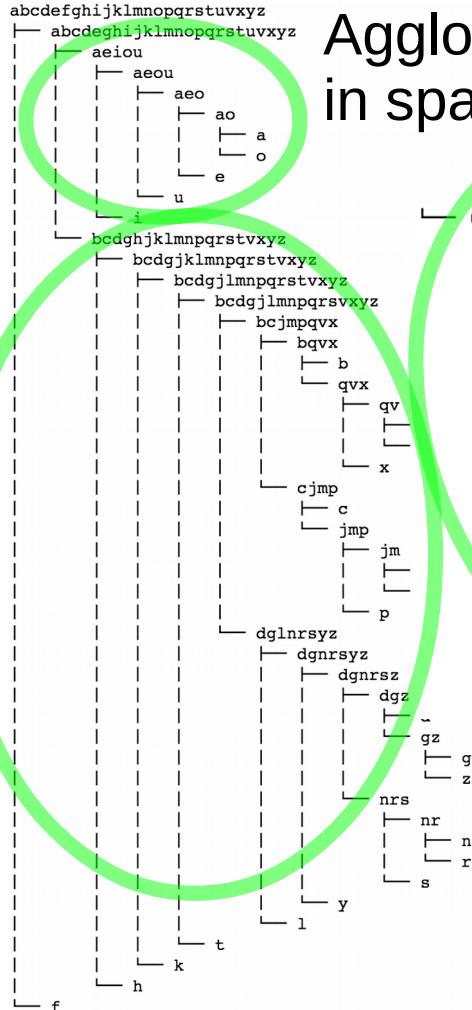
<https://aclanthology.org/2022.emnlp-main.239/>

<https://ieeexplore.ieee.org/document/10347856>

https://link.springer.com/chapter/10.1007/978-3-031-44865-2_1

Case 6: Unsupervised Character Category Learning

Agglomerative Clustering in space of Transitions

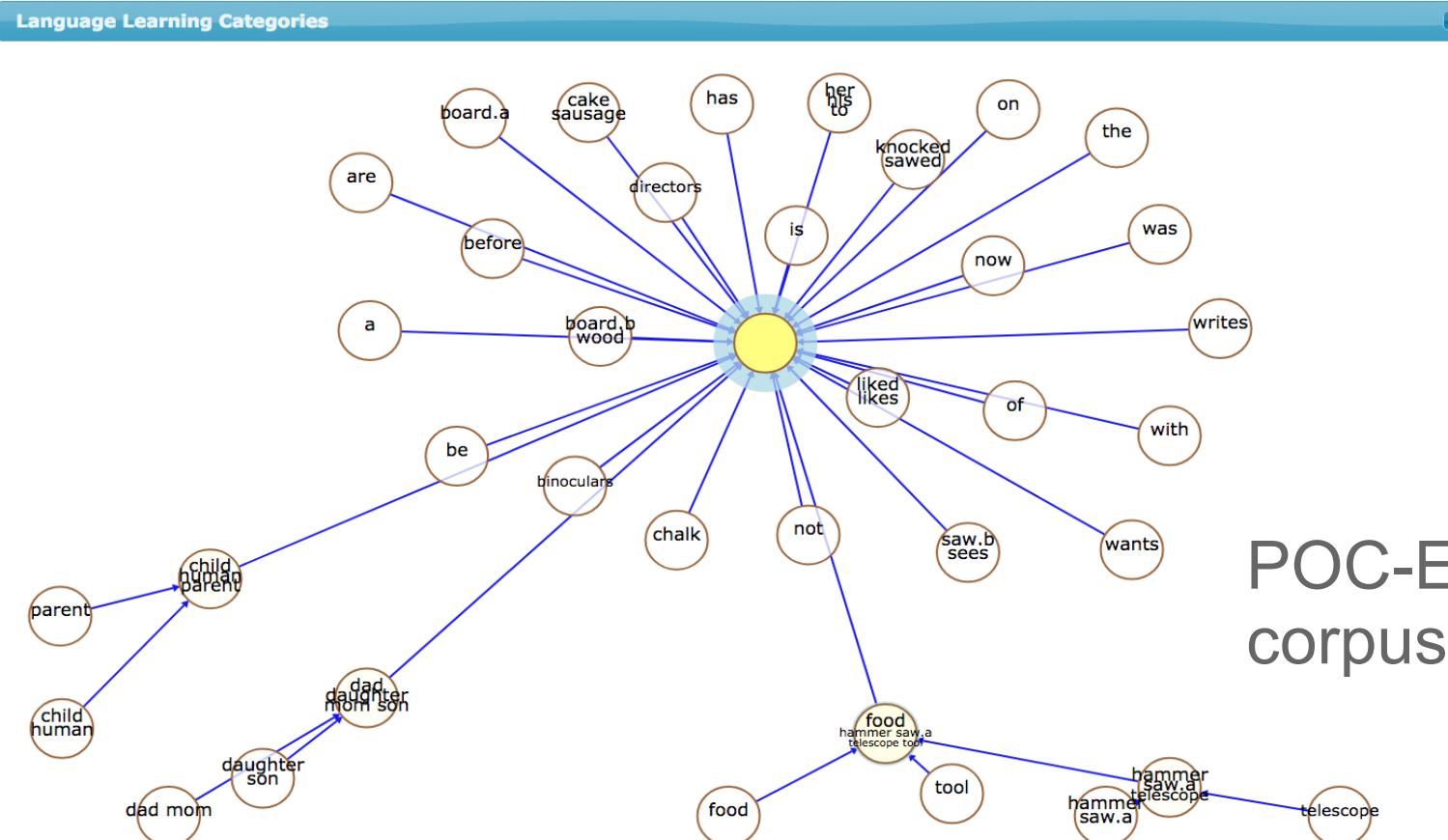


RusAge Test/Small,
Cosine Similarity

Copyright © 2024 Anton Kolonin, Agents®

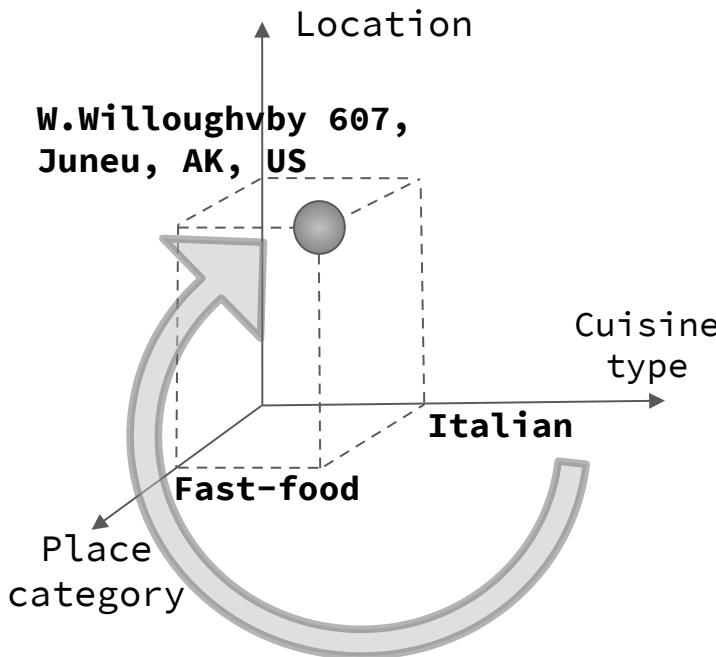
<https://aclanthology.org/2022.emnlp-main.239.pdf>

Case 7: Unsupervised Ontology Learning on Parses



POC-English corpus

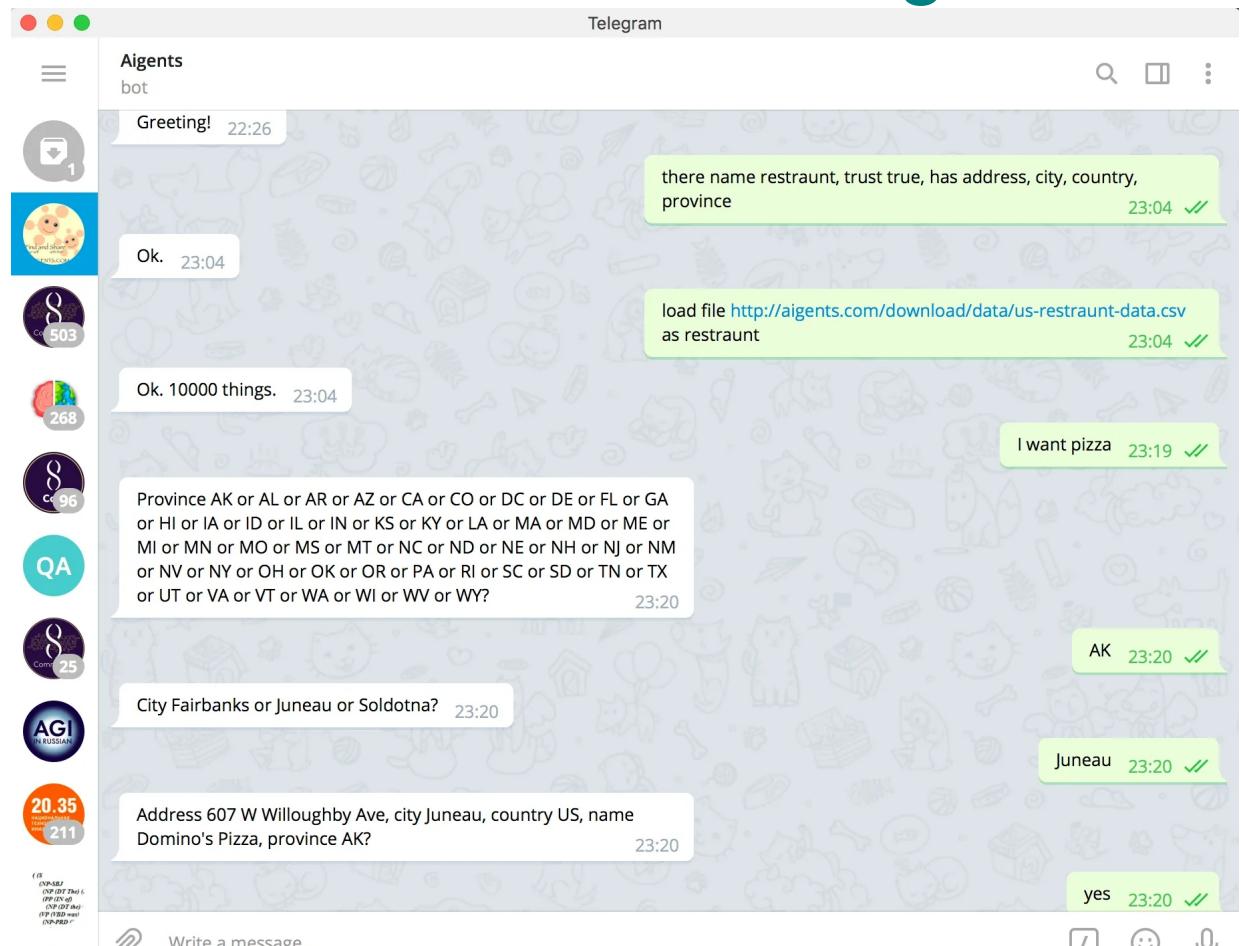
Case 8: Data-driven Conversational Intelligence



Dynamic data-driven dialog flow minimizing uncertainty
Just add RAG augmentation?

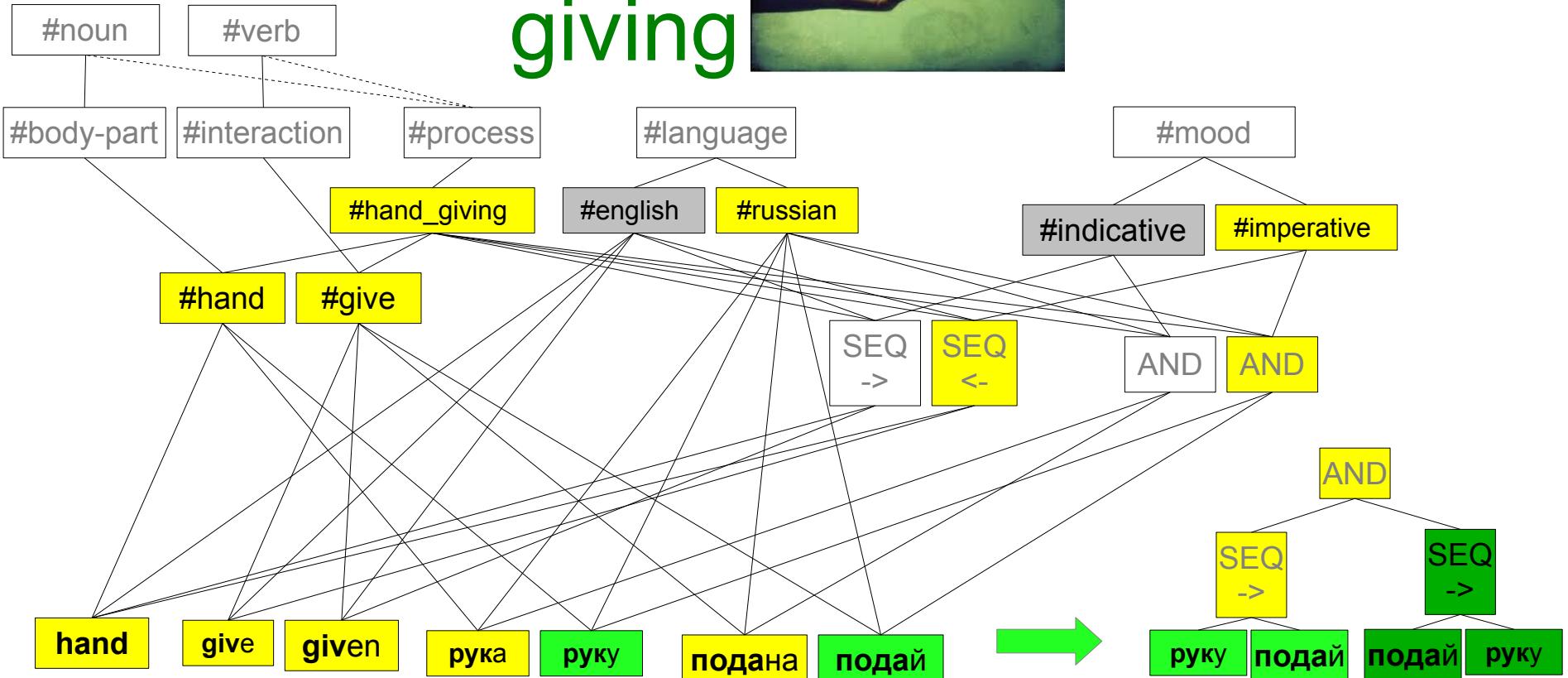
<https://t.me/AigentsBot>
<https://blog.singularitynet.io/chat-with-your-data-using-aigents-bots-99b76cae65f2>

Copyright © 2024 Anton Kolonin, Aigents®



Case 9: Graph-based contextual language generation

Hand giving



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

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

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

INLP Workshops
<https://agents.github.io/inlp/>

Extended presentation in Russian
<https://www.youtube.com/watch?v=RU2T0OFJJB4>

Thank You and Welcome!



<https://agirussia.org>

Anton Kolonin
akolonin@aigents.com
Messengers: akolonin

