Discourse processing in the time of DNNs

Eduard Hovy

CMU

November 2020

This workshop

"… surge of activity in discourse parsing, coherence models, text summarization, corpora for discourse level reading comprehension, and discourse related/aided representation learning …

"… a workshop that brings together discourse experts and upcoming researchers will catalyze the speed and knowledge needed to solve such problems"

When you go from one sentence to many…
…you go from a localized (nearly) 1-D unit (the frame) to an extended, multi-dimensional network of units that appears over time

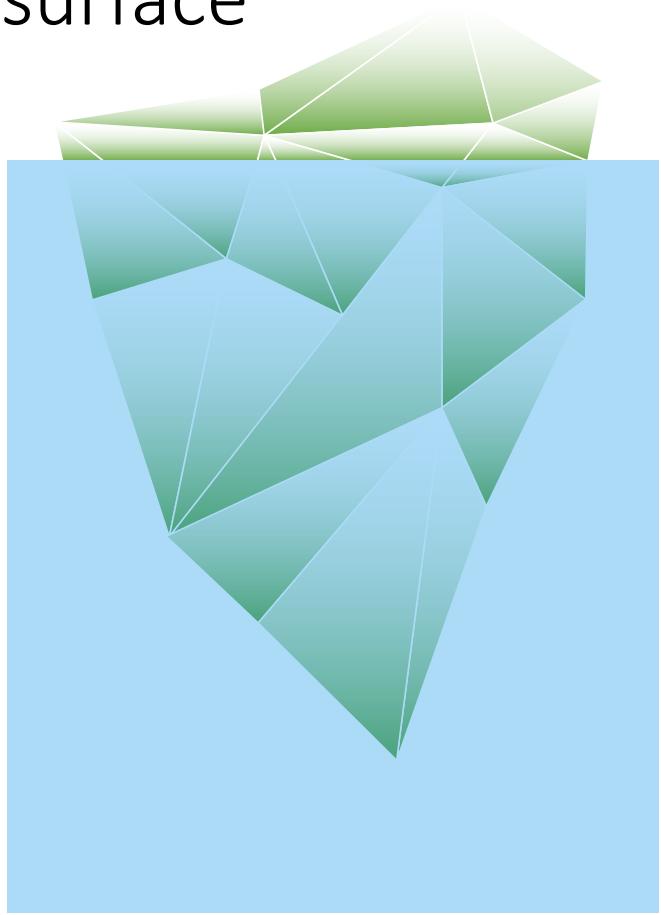
So the Speaker must guide the Hearer on how to reconstruct the logical structure, and must maintain consistency over the growing text:

- Text structure and coherence: the role and contribution of each portion of the text, attachment points, given vs. new, etc. This includes theme, rheme and focus shift: flow / evolution of the topic
- Reference and coreference: evolution and updating of referent space
- Stylistic consistency: same level of formality, specificity, humour, etc.

There is a lot hidden below the surface

- Even more than for a sentence, discourse leaves important information implicit
- If you don't know what it is, there is little hope for inducing/learning it automatically
- DNNs don't do very well
- Discourse is not primarily an algorithm problem

- NLP people should work on the theory



Outline

Coreference

Coherence

Stylistic consistency

Outline

Coreference

Coherence

- What it is
- Types of structure and processing
- How to start a clause?

Stylistic consistency

Coherence

What it is

Various theories

(Radford et al., 2019)

OpenAI's GPT-2 and GPT-3

- Large transformer-based language models with 1.5 billion / 175 billion parameters
- Auto-regressive training:
  - Objective: Predict the next word, given many previous words in a text
  - GPT-2: 8 million webpages; GPT-3: hundreds of billions of words

- Main achievement claimed: **astounding long-text generation**
- Full model initially not released due to concerns about malicious applications of the technology
  - Claimed "potential harmful effects of GPT-3 include misinformation, spam, phishing, abuse of legal and governmental processes, fraudulent academic essay writing and social engineering pretexting" (May 2020)

GPT-2 discourse generation example

(Radford et al., 2019)

System Prompt (human-written)

In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.

Model Completion (machine-written, 10 tries)

The scientist named the population, after their distinctive horn, Ovid's Unicorn. These four-horned, silver-white unicorns were previously unknown to science.

Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved.

Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow.

Pérez and the others then ventured further into the valley. "By the time we reached the top of one peak, the water looked blue, with some crystals on top," said Pérez.

Pérez and his friends were astonished to see the unicorn herd. These creatures could be seen from the air without having to move too much to see them – they were so close they could touch their horns.

While examining these bizarre creatures the scientists discovered that the creatures also spoke some fairly regular English. Pérez stated, "We can see, for example, that they have a common 'language,' something like a dialect or dialectic."

Why Ovid?

One horn or four?

Why two centuries?

Which phenomenon?

Into the valley or up a peak?

Seen without having to move, but close enough to touch?

What does this mean?

Another GPT-2 masterpiece of logic

System Prompt (human-written)

Recycling is good for the world.

NO! YOU COULD NOT BE MORE WRONG!!

Model Completion (machine-written, 25 tries)

Recycling is NOT good for the world. It is bad for the environment, it is bad for our health, and it is bad for our economy. I'm not kidding. Recycling is not good for the environment. It is destructive to the earth and it is a major contributor to global warming. Recycling is not good for our health. It contributes to obesity and diseases like heart disease and cancer. Recycling is bad for our economy. It increases the cost of a product, and in turn, the price of everything that is made with that product. Recycling is not good for our nation. We pay a tremendous price for the privilege of having the world's most advanced and efficient recycling system. Recycling is a huge, colossal waste of time, energy, money, and resources. And THAT is why we need to get back to basics and get back to basics in our recycling efforts. One of the best ways to start is to look at the process of creating a paper product. When you make a paper product, it is basically a long chain of materials. Everything from the raw materials (wood, cardboard, paper, etc.), to the reagents (dyes, solvents, etc.) to the printing equipment (chemicals, glue, paper, ink, etc.), to the packaging, to the packaging materials (mercury, chemicals, etc.) to the processing equipment (heating, cooling, etc.), to the packaging materials, to the packaging materials that are shipped overseas and to the packaging materials that are used in the United States. Each step along the way creates tons of waste that we constantly have to clean up. The process of making a paper product is a very wasteful one. But the end result is something that all of us need to consume. And if we want to keep the recycling process running efficiently, then we really need to think about each and every step that goes into making a paper product.

Observations: There's no 'message'

- No logical argument, no long-distance consistency
- No clear content anchored on anything, just 'hallucination'

- This is just a very long ngram (a little generalized) that continues extending itself to the right…

- Unless it becomes possible to
  - control the content
  - control the logic of the story

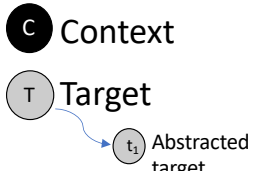
  there's no reason to take this seriously

Ngram continuation produces non-sense nonsense

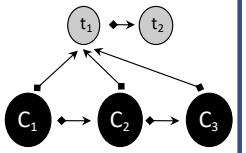
Still, lots of work on ngram continuation

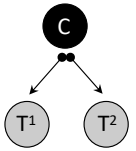
- Input: Start with a given ngram sequence
- Output: Produce *N* more words to the right, rank, pick the best(?) option, and cycle
- Content: Given at start (GPT-3); or from an interlocutor (dialogue); or sliding across some other text (summarization)

	Data2Text	Abstractive summarization	Email reply, review generation	Story generation, Recipe generation, Goal-oriented dialogue
Is content given?	Partial ($C \subset T$)	Full ($C \supset T$)	None ($C \perp T$)	None ($C \perp T$)
NLG objective	Realization	Abstraction + Realization	Diversification + Realization	Coherence + Realization
Planning	-	Hierarchical	Vertical	Horizontal



(Kang 2019)





A generator has to actually plan its text

To plan a discourse, you need to know:
What effect do I want on the reader? — Intent
What do I want to talk about? — Content

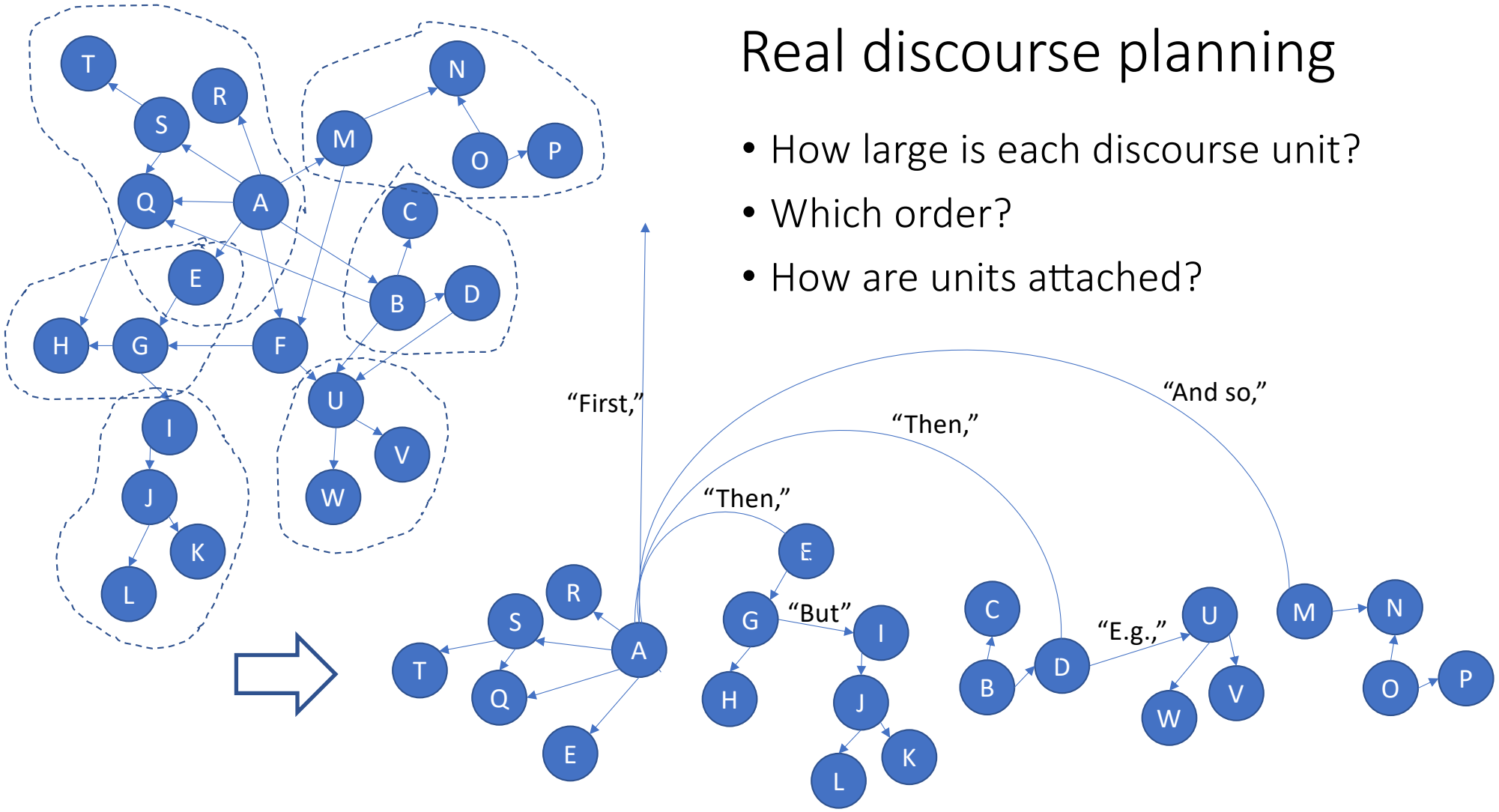
- We need to understand types of **communicative intent**
- We need to understand how **intent+content —> effect** on the Reader

- DNN-based generators have no/weak notions of intention or topic
- Traditional semantic analyzers/parsers have no idea how to recognize discourse intent — they stop with sentence-level Speech Acts
- Traditional planners/generators have done the most work here but have never achieved cross-domain generality

How can we fix this with DNNs?

Real discourse planning

- How large is each discourse unit?
- Which order?
- How are units attached?



The problem is coherence structure

- What is ‘coherence’?
  - Any 3-year-old learns language unconsciously and can make single sentences. But it takes years to learn to write coherent multisentence text, and after learning it we still cannot describe what makes text coherent
- Why are there no ‘text grammars’?
  - Unlike for words, there is no small set of classes to serve as nonterminal symbols for sentences, and there is no canonical ordering of these parts
- So what then is text structure? Why can you not simply order sentences in random order?
  - Some genres are structured, others not

McKeown Ph.D. 1985

Write rules to organize this coherently

[Eltville, (Germany)]$_A$

[with a considerable weight for a white wine]$_B$

[Taubenberg, Sonnenberg, and Langenstück are among vineyards of note]$_C$

[An important wine village of the Rheingau region]$_D$

[The vineyards make wines that are emphatically of the Rheingau style]$_E$

[Eltville, (Germany)]$_A$

[An important wine village of the Rheingau region]$_D$

[The vineyards make wines that are emphatically of the Rheingau style]$_E$

[with a considerable weight for a white wine]$_B$

[Taubenberg, Sonnenberg, and Langenstück are among vineyards of note]$_C$

How do you know this is the right order?

General theory

- Discourse is made of clause-sized units
- There are no 'grammatical' types for units, but units do have *functions*
- The function is the unit's communicative purpose in the text
- The functions are quite recognizable and obey some (weak) ordering constraints (in some genres)
- The same unit may fulfill different functions in different contexts (unlike a word, which has a fixed grammatical category). Some functions: *define, example, contrast, attributes, conclude, reason*

A text is **coherent** iff each unit's function in the text is clear

McKeown Ph.D. 1985

The functions of the example

[Eltville, (Germany)]$_A$	Introduction
[An important wine village of the Rheingau region]$_D$	Definition
[The vineyards make wines that are emphatically of the Rheingau style]$_E$	Attribute (detail)
[with a considerable weight for a white wine]$_B$	Elaboration of detail
[Taubenberg, Sonnenberg, and Langenstück are among vineyards of note]$_C$	Example

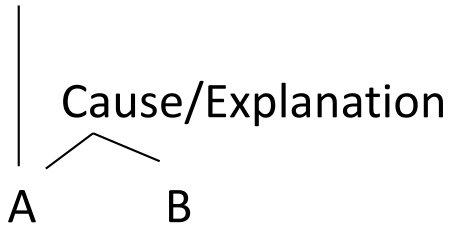
How you know this is the right order — because:

Introduction < Definition < {Attribute, Example}*

Discourse parsing and generation

- **Parsing**: Assign a function to a unit

  "A because B" =>



  - Early (Marcu thesis 1996)
  - Difficult: F-score 0.55 (Joty et al. ACL 2013)
  - SVM predictor, F-score 0.62 (Ji and Eisenstein ACL 2019), Neural predictor, F-score 0.78 (Lin et al. ACL 2019)

- **Generation**: Order and relate units to show how they jointly create the intended functions

  Speaker-Intent(Cause(A,B)) => "A because B"

Discourse generation: A big challenge

- The Speaker wants to achieve some goal(s) on the Hearer
  - Inform, request, please, shock, puzzle, anger…
- The Speaker has a lot of relevant content to communicate
- It is not possible to say everything in a single clause, so the content must be staged

- Discourse generation is a planning problem
- Two-layer architecture: structure/content planner + sentence realizer
- Planner must:
  - Compute the plan (at least, the main steps)
  - Tell the realizer what next to say (= content)
  - Tell the realizer how to introduce and structure it (= intent)

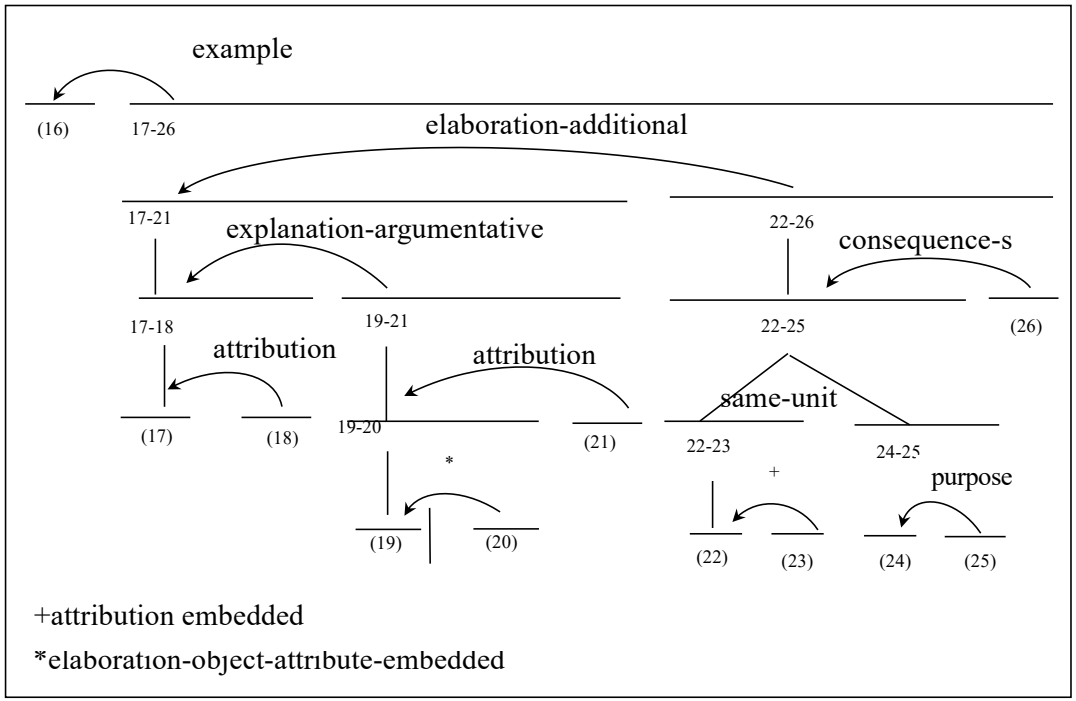
Discourse structure planning

- Methods of planning:
  - Top-down hierarchical decomposition — 1980s and 1990s
  - Bottom-up function mapping and achievement — 1980s
  - Simple continuation, like stream-of-consciousness — 2016 and later
- Plans = discourse structure relations that link content to intent
  - Schemas: McKeown (1985)
  - Relations: Rhetorical Structure Theory (RST: Mann and Thompson 1988), Grosz and Sidner (1979), Kamp and Asher (DRT, 1990), (Hovy et al. 1994), Penn Discourse Treebank (Miltsakaki 2004)…
  - Inference logic: Hobbs (1977)
- Planners:
  - Formalize RST relations with content requirements and intention
  - RST Structurer (Hovy 1989), Moore and Paris (1990)...

Example RST discourse tree

[Still, analysts don't expect the buy-back to significantly affect per-share earnings in the short term.][16] ["The impact won't be that great,"][17] [said Graeme Lidgerwood of First Boston Corp.][18] [This is in part because of the effect][19] [of having to average the number of shares outstanding,][20] [she said.][21] [In addition,][22] [Mrs. Lidgerwood said,][23] [Norfolk is likely to draw down its cash initially][24] [to finance the purchases][25] [and thus forfeit some interest income.][26]

wsj_1111



Can you plan using a DNN?

- General approach for the planner:
  - Plans: Learn typical unit/function sequences
  - Control: Define a way to communicate the next unit/function to the realizer
  - Realizer: Train an ngram realizer (like GPT-3) to use the unit/function to select content, and then generate it

- Only a few attempts to guide the generator toward global coherence
  - Bosselut et al. NAACL 2018
  - Peng et al. 2018, EMNLP 2020
  - Kang et al. thesis 2019, EMNLP 2020

- Current work is rather poor in defining unit/functions and the communication notation — a long way to go, requiring theory

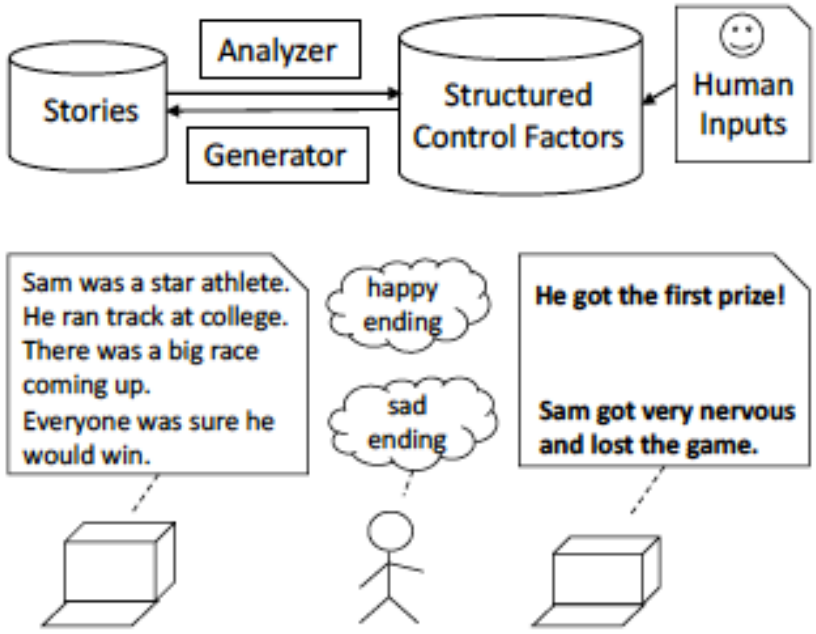
Types of structure and processing

1. ‘Planning’ using guidance words
2. Internal coherence by ordering sentences
3. Argument structure
4. Other, more complex structures

(Peng et al., 2019)

Ex 1: Example ‘planning’ using guidance words

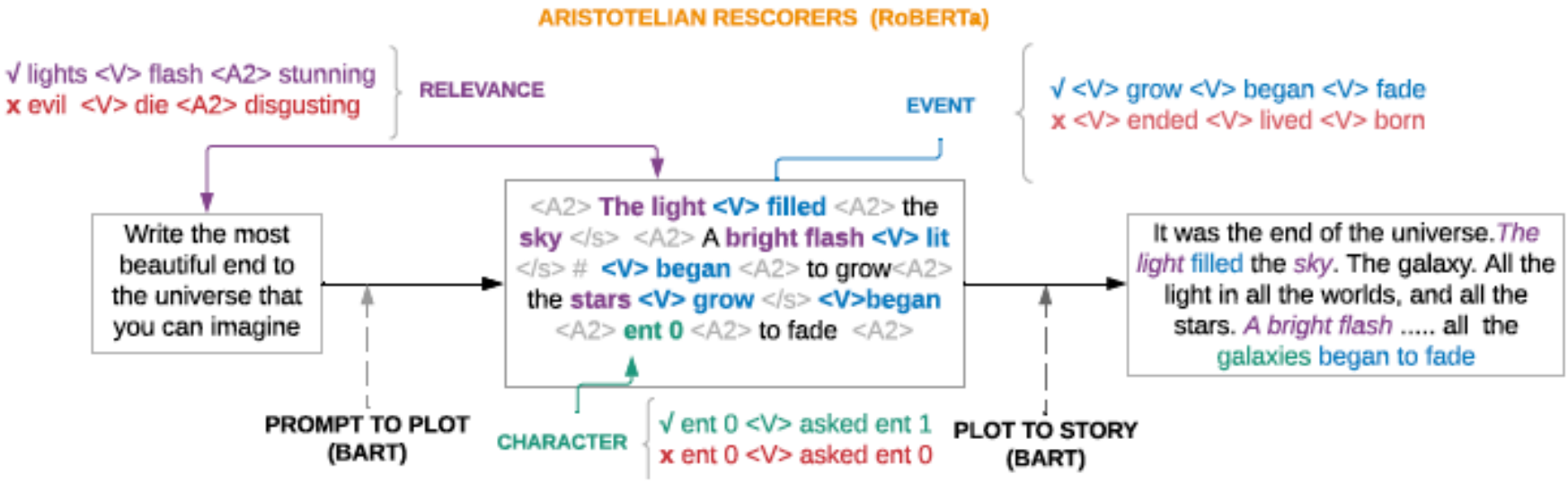
- Controllable story generation approach:
  - Define ‘planner guidance’ = English keyword to control realizer
  - Extract and learn sequence of guidance words from stories => the ‘planner’
  - Train realizer to associate each guidance word with words to be chosen for output
- Challenges:
  - What is a good ‘guidance word’?
  - How do I learn a series of them?
  - How do I control the realizer with them?



(Peng et al., 2020)

Guidance words –> plot structures

- Define plot structure models and data:
  - Plot model p(z|x): {prompts, extracted silver standard plot-structures} using SRL info
  - Story model p(y|z;x): {story, prompts+plot}
- Training: use this to fine-tune pretrained conditional LM BART (Lewis et al. 2019)



(Kang et al., 2019)

Ex 2: Toward coherence: internal sentence ordering

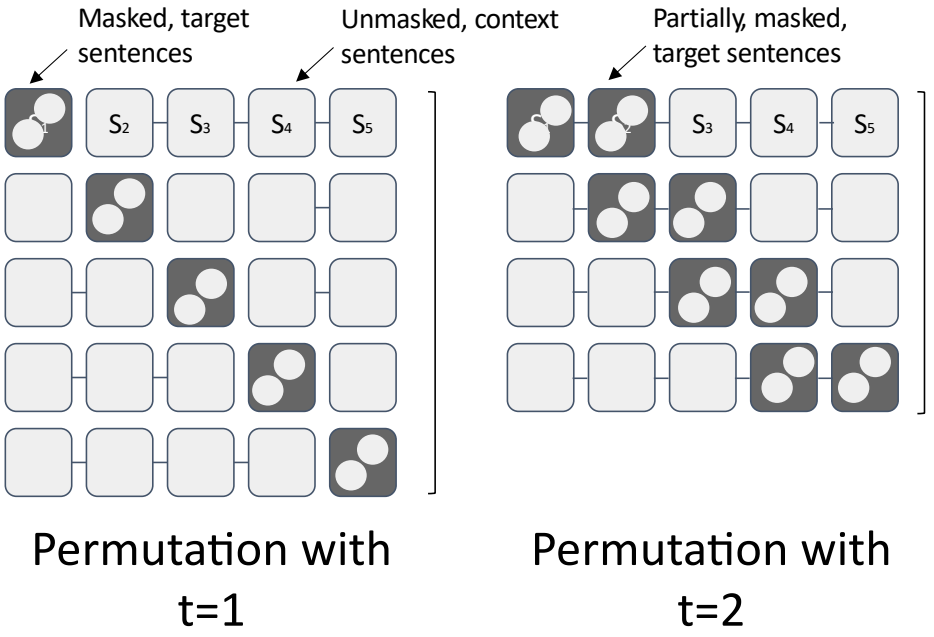
[1] Satyrs never wear armor, including helmets, Newel began, using his hands expressively. [2] But years ago I was in a play, and the helm was part of my costume. [3] During the big battle scene, a few of us were assailing a castle. [4] We had quite a set. [5] The main tower must have been fifteen feet tall, fashioned from real stone. [6] Anyhow, as we actors were laying siege, a big chunk of ..

- Step toward developing discourse 'plans'
- Task: Given par start and end sentences, ask system to order intervening sentences for coherence
- Problems:
  - Difficult even for humans (METEOR ~4.5)
  - Context is too sparse to predict multiple sentences
- Easiest is just one filler sent per paragraph

(Kang et al., 2019)

Staging the ordering task

- Stepwise permutation masking
  - One step at a time vs L^2
  - context (c) + target (t) = paragraph (L)
  - context (c) > target (t)
- Resolve context sparsity by providing/predicting only keywords (not full realization)



Permutation with t=1

Permutation with t=2

(Kang et al., 2019)

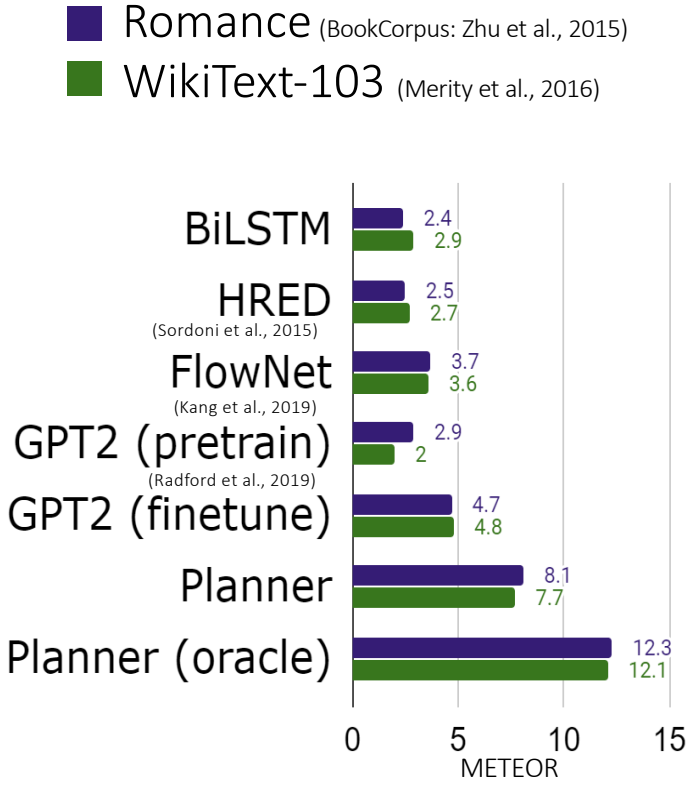
Effect of order planning on paragraph completion

- Noun+verb keywords are more useful than other types
  - +2.4 METEOR over Random
- More keywords => better generation
  - +1.6 than half, +2.1 than zero

Coherence with context / Overall quality

GPT (fine)	Planner	Human
2.1 / 3.6	3.8 / 3.9	4.8 / 4.9

Human evaluation [1–5]



Automatic evaluation

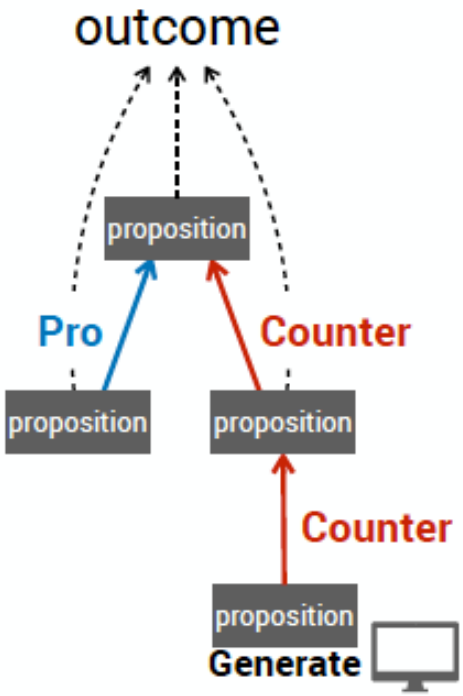
Ex 3: Argument structure

- An acid test for discourse processing

- First, we need to understand what argumentation structure is

  …and build a corpus!

- Theories of argumentation:
  - Aristotle, Walton, Toulmin…
- Corpora:
  - Centre for Argumentation Technology https://arg-tech.org/
  - CNET (Kim et al. 2010)
  - NPS Chat (Forsyth and Martell 2007)
- Tech:
  - IBM's Project Debater https://www.research.ibm.com/artificial-intelligence/project-debater/



(Jo 2019–)

Yohan Jo's approach to argumentation analysis

- Get examples
- Identify individual propositions (minimal clause-sized claims)
- Attach them into a tree-link structure
- Assign sentiment and support/oppose tags to links

- Determine type of claim and link (factual evidence, refer to authority, own experience, common knowledge…)
- Determine individual (sub)argument/claim strengths

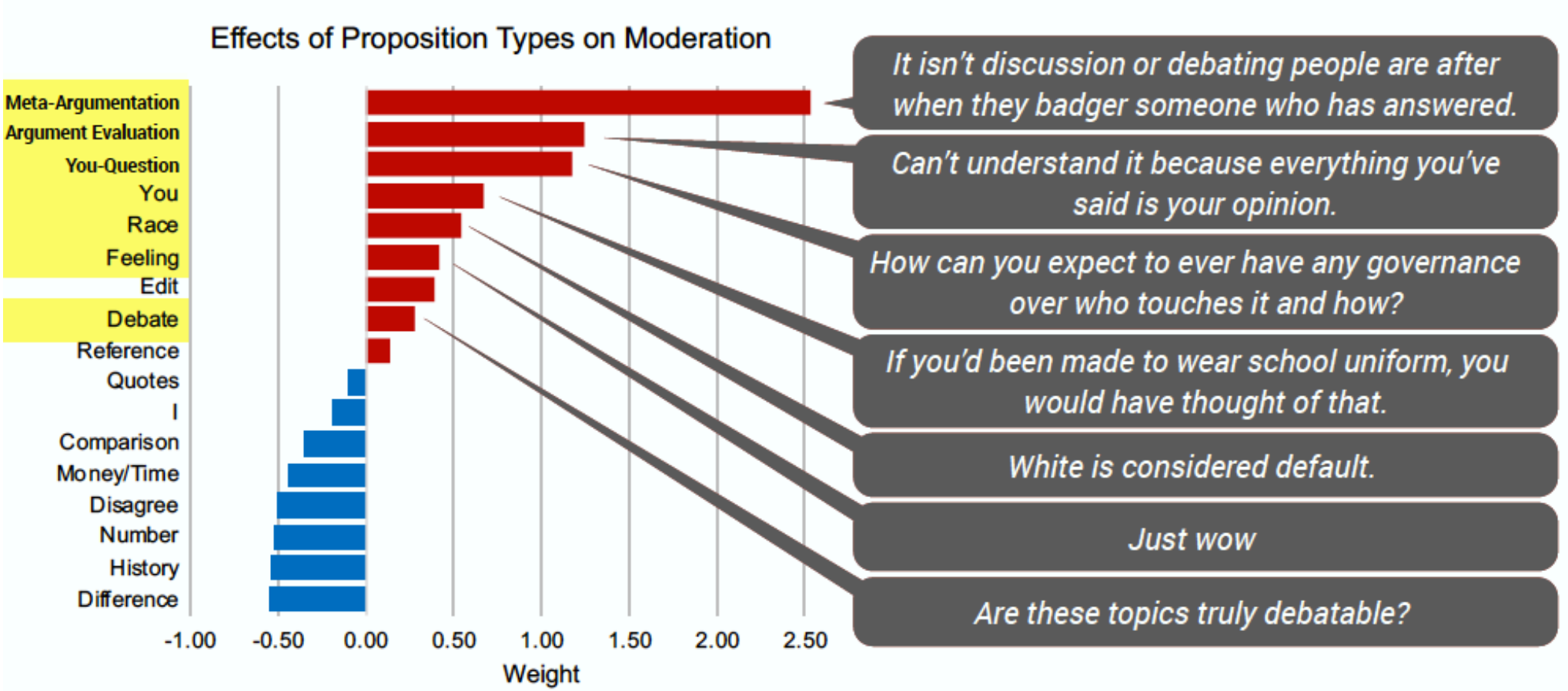
- Identify weak points for counter-arguments

Lots of raw training data

	Goal	Topics	#Utterances	#Discussions
WIKIPEDIA The Free Encyclopedia	Decide an action	How to edit Wikipedia articles	161,525	21,108
ravelry	Share opinions	Political	350,376	4,213
CMV ChangeMyView	Change other users' viewpoints	Any (religion, daily issues, ...)	204,679	3,207
	Sway voters	Political	3,021	369

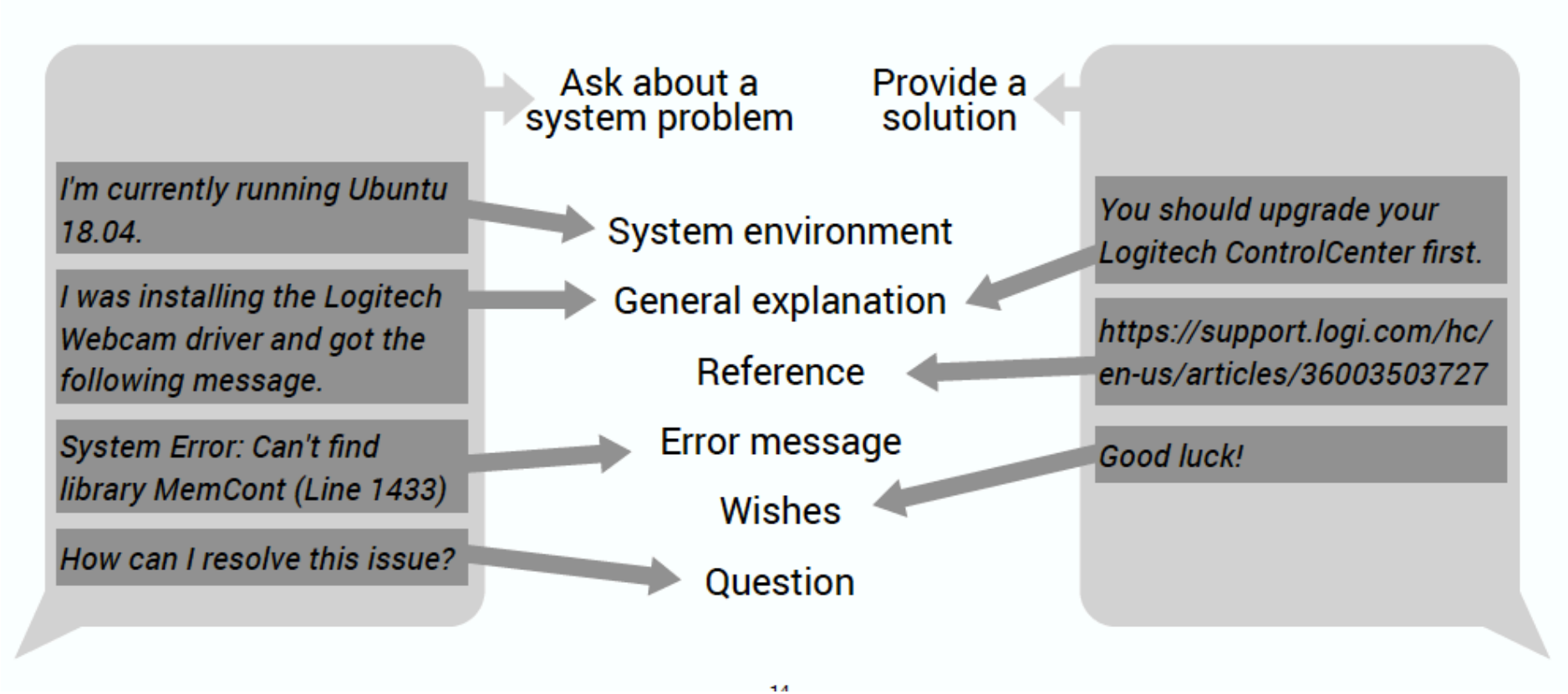
(Jo 2019)

Proposition types and frequencies



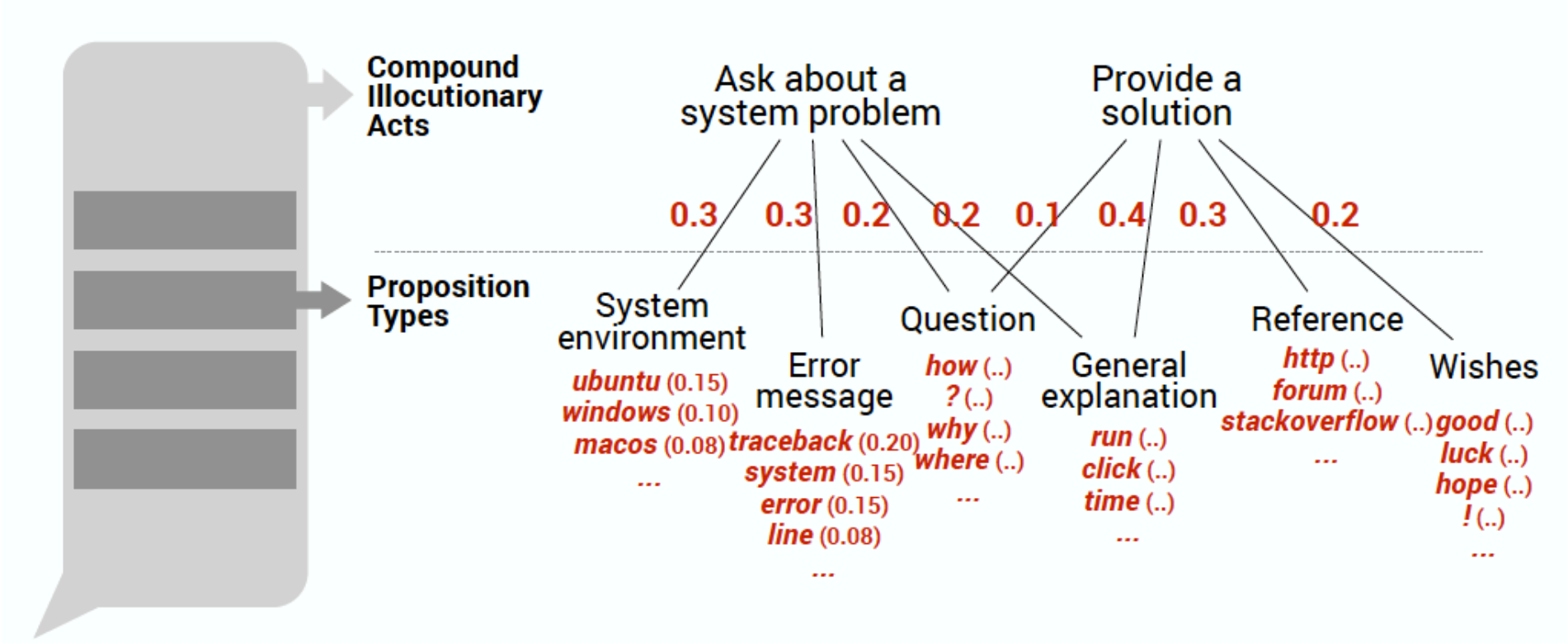
Compound illocutionary act as a probabilistic mixture of proposition types

(Jo 2019)



Compound illocutionary act as a probabilistic mixture of proposition types

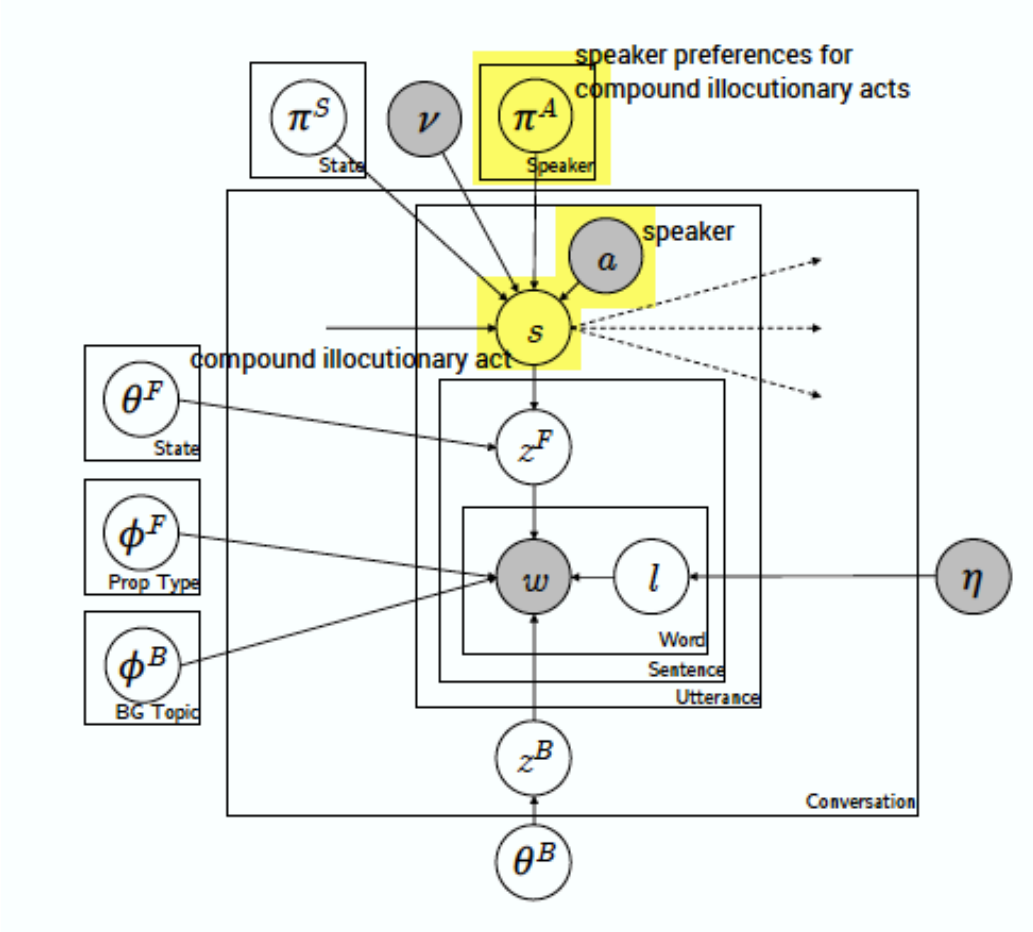
(Jo 2019)



(Jo 2019)

Model to determine arg structure

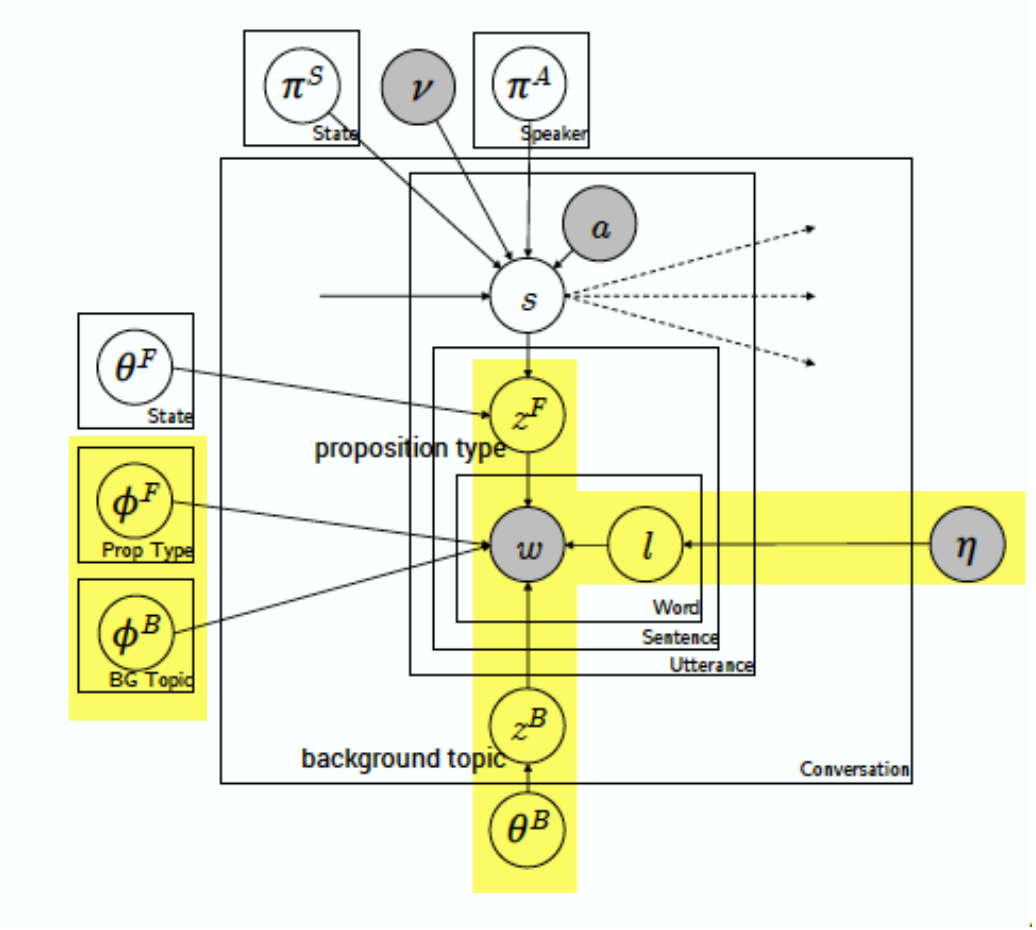
- Each utterance is assigned a compound illocutionary act, and each sentence is assigned a proposition type
- The act of an utterance depends on the act of the preceding utterance (e.g., ask→answer)
- The act of an utterance also depends on the **speaker's preferences** for certain acts



(Jo 2019)

Model to determine arg structure

- Each utterance is assigned a compound illocutionary act, and each sentence is assigned a proposition type
- The act of an utterance depends on the act of the preceding utterance (e.g., ask→answer)
- The act of an utterance also depends on the **speaker's preferences** for certain acts
- A proposition type is characterized by words that are **less specific to discussion topics**



(Jo 2019)

Arg Discourse Unit detection accuracy

All modules produce outputs that are more similar to asserted propositions than simple ADUs are

Module	ADU Segmentation	Our Model	Δ	Metric
Anaphora Resolution	55.8	60.1	4.3↑	Subject/object dependencies
ADU Segmentation	–	78.9		F1
Reported Speech – Detection	–	85.1		F1
Reported Speech – Content Extraction	59.5	89.6	30.1↑	BLEU
Question – Detection	–	86.2		F1
Question – Transformation	47.5	54.5	7.0↑	BLEU
Subject Reconstruction	59.1	62.6	3.5↑	BLEU
Revision	75.5	76.6	1.1↑	BLEU
End-to-end	–	59.0		BLEU

(Jo 2019)

Support and oppose links

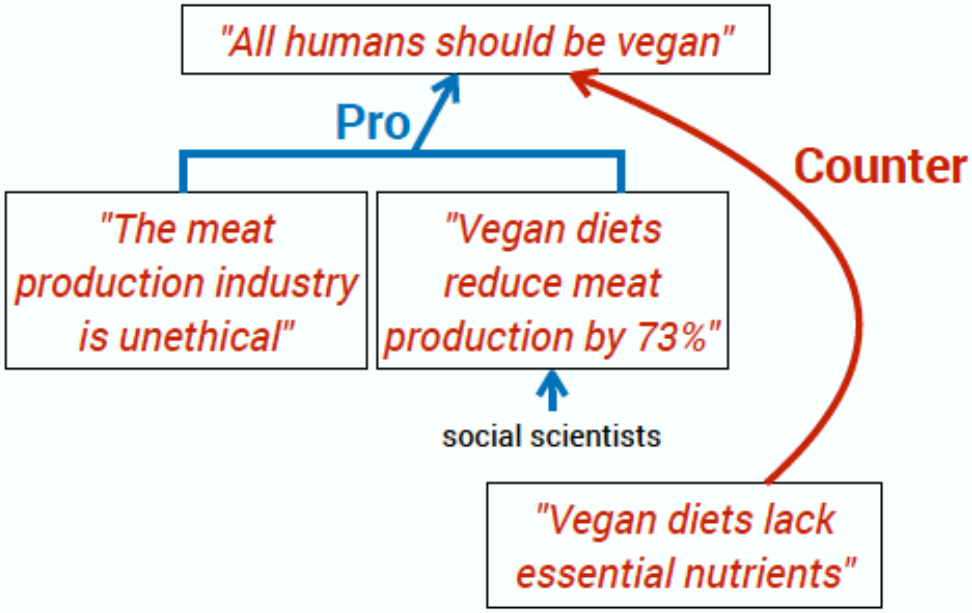
- Assign illocution label
- Assign sentiment
- Assign inter-turn link label

I think all humans should be vegan.

Look at how unethical the meat production industry is.

Social scientists proved that vegan diets reduce meat production by 73%.

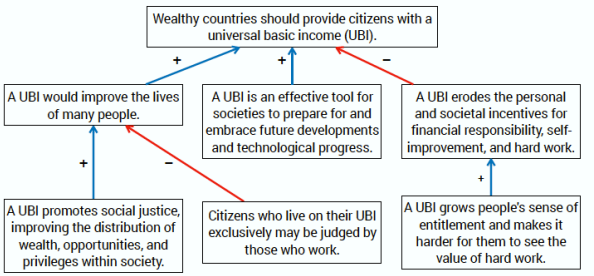
Well... don't vegan diets lack essential nutrients, though?



(Jo 2019)

Finding counterargument / attack points

- Corpus: Kialo.com collaborative argument platform
  - 1.4K issues, 130K statements (#pro=68K, #counter=67K)
- Extremeness = Avg (|#Positive replies − #Negative replies| / #Replies)
- Identify features of attacked sentences:



Content

- n-grams (20K)
- Sentence-level topics (50)

External Knowledge

- Frequency
- Attractiveness
- Extremeness

Proposition Types

- Question-Confusion
- Question-Why/How
- Question-Other
- Normative
- Prediction
- Examples
- Hypothetical
- Citation
- Comparison
- Personal Story
- Use of *You*
- Use of *We*

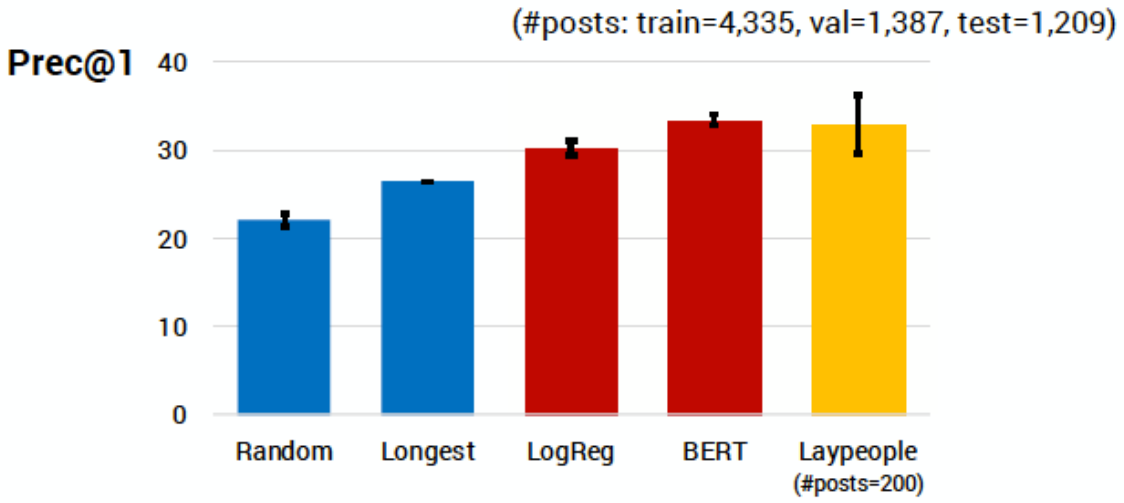
Tone

- Subjectivity
- Qualification
- Hedges
- Concreteness
- Sentiment Score
- Sentiment Category
- Arousal
- Dominance

(Jo 2019)

Attackability classifier

- Counter-statement characteristics:
  - More **data**, **less subjective**, and **less of personal stories**
  - More about **safety**, **welfare**, and **abstract concepts**, and **less about technology**
  - Ask **questions** and express **uncertainty** (hedging)
  - Tend to match other **statements** and **responses** in Kialo
  - Not related to sentiment, though negative sentences are significantly more attacked
- Classifier results
  - BERT encodes useful linguistic patterns predictive of attackability
  - BERT embeddings contain the information of lexicon-based features



(Jo 2020)

Generating counterarguments

- Strategy: To attack *X*, find event *Y* that has opposing sentiment and some temporal/causal relationship with *X*
- Approach:
  - Use template-based methods based on argumentation schemes (Walton et al. 2008)
  - Generate creative and truthful propositions from domain knowledge

Work in progress

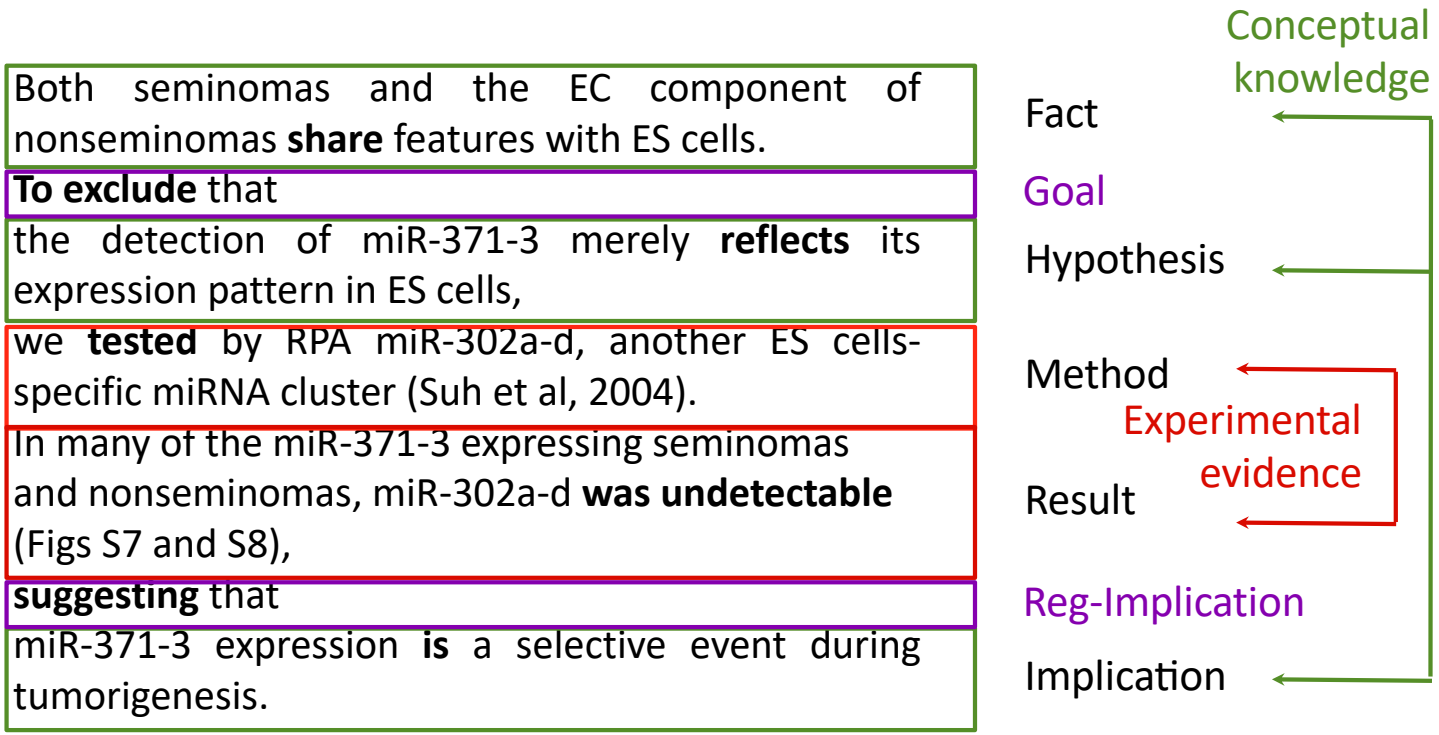
Ex 4: Even more complex structures…

- A scientific paper
- A legal rent/lease agreement
- A privacy agreement online
- A marriage or divorce contract

- Each domain has an inner logic
- The discourse structure reflects the logic
- Info Extraction into the frame is not the problem — understanding what the frame IS is the problem!

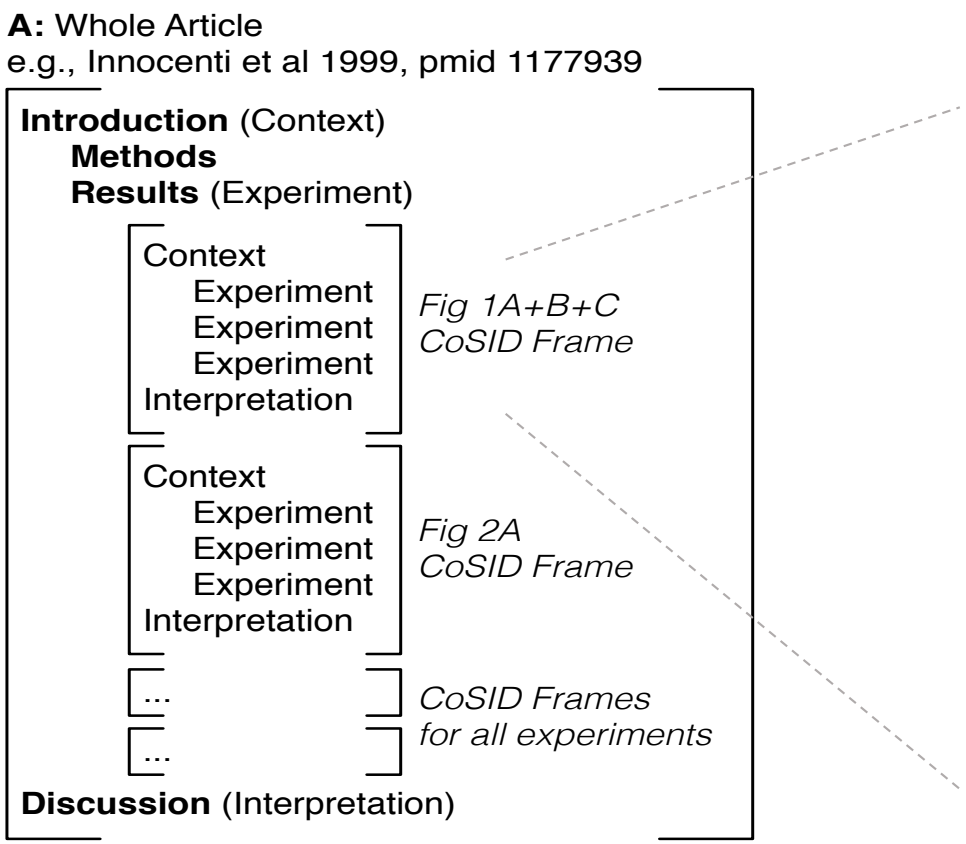
Scientific experiment discourse
Discourse structure and epistemic tags are very helpful

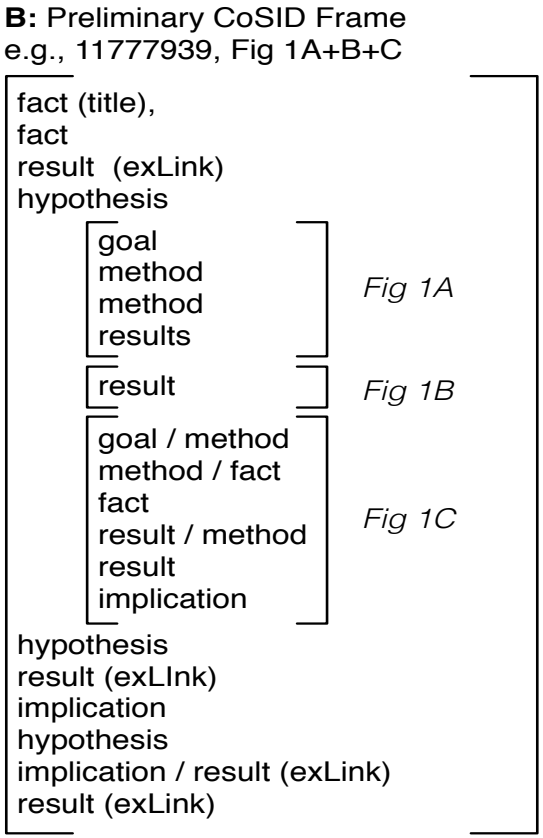
(Burns et al. 2016)



Scientific experiment discourse
Nested structures of CoSID frames

(Burns et al. 2016)





Quick interim summary

- From simple schemas (McKeown) through RST structure trees to argument structures to complex domain-logic frames
- In each case you need:
  - The underlying structure-building representation units
  - Ways to connect them to the content
  - Ways to understand/signal their intentional role
  - Either an analyzer/structure builder or a planner + sentence realizer

- You also need methods to evaluate the structure
  - Content correctness and completeness
  - Structure/ordering correctness
  - Intention signaling correctness and completeness

Ways of measuring long-term coherence

- Measuring perplexity of trained language model
- Sentence ordering (Barzilay and Lapata 2008)
- Narrative cloze task (Chambers and Jurafsky, 2008; Mostafazadeh et al. 2016)

Ranking or classification

- Next sentence prediction (Devlin et al. 2019)
- Text infilling (Zhu et al. 2019; Huang et al., 2020)
- Story generation (Fan et al. 2018, Peng et al. 2020)
- Paragraph bridging (Kang et al. 2019)

Context ⊥ Target: Generation with partial context (e.g., surrounding text, prompt)

- ….
- For large structures: F-score on each piece and F-score over the whole structure

Open issues

- Content: Rep of discourse units
  - Approaches vary, depending on level of granularity needed
- Plans: Libraries of structures, schemas, or relations
  - Each researcher tends to create their own, and large collections tend to be unmanageable
  - Many relations available (RST, Penn Discourse Treebank, etc.)
  - Connection of relation to content
- Intentions!
  - No standard set exists; RST bootstraps intentions from relations; many people talk about but do not provide intention vocab (Grosz and Sidner CL1986; etc.)
  - No clear connection between intentions, Speech Acts, and relations
- Unlike for syntax, there's no standardization — why not?

How to start a clause?

Bridge

Theme

Attitude

Clause-initial position

Our challenge problem: what must the sentence planner put at the start of the clause (before the verb)? Why?

Last night John opened the window.

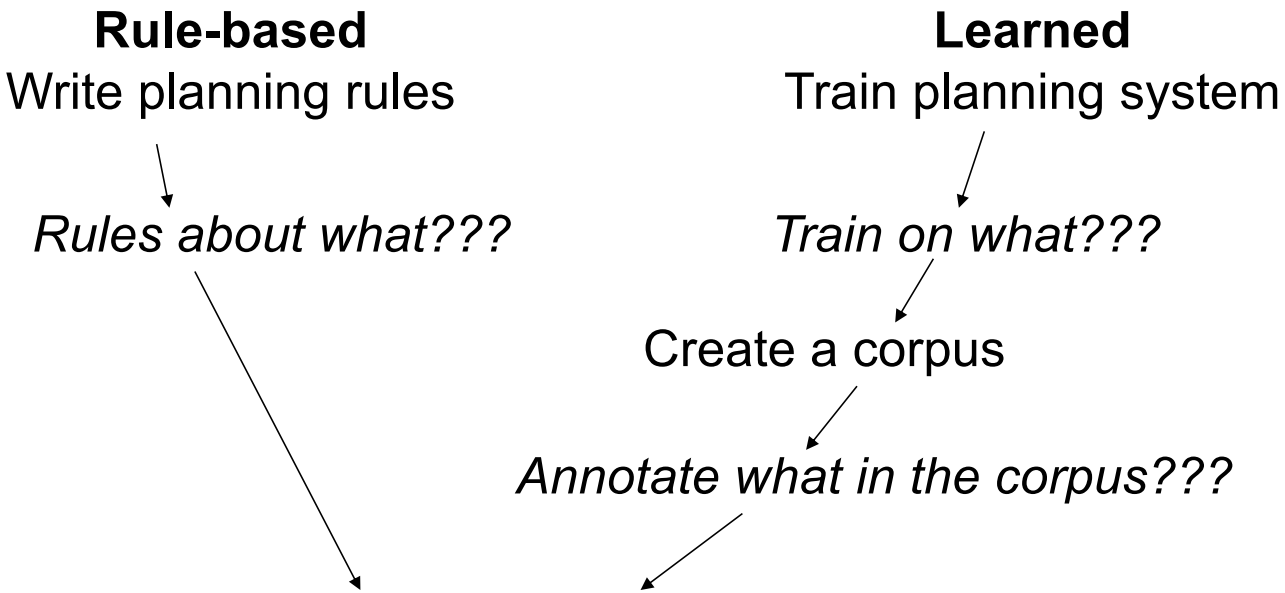
Last night the window was opened by John.

There's a stain on your skirt.

Your skirt has a stain on it.

However, he won the championship.

Approach



First understand the problem

- Identify important phenomena
- Create terminology for them
- Define terms with respect to text

Underlying assumption

- It's not just random choice — the Author does everything for a reason
- So we have to find out the purpose: why the clause starts how it does
- So we need a functional (not structural) account of clause-initial positions

…what kinds of functions are there?

Examples: what's the purpose?

- (1a) Last night John opened the window.
- (1b) Last night the window was opened by John.
- (2) [Peter went to bed at 9 and] [ ] fell immediately asleep.
- (3) [I hate green peppers but] Spanish omelette I really like.
- (4) However, he won the championship.
- (5) Perhaps we should change our clothes?
- (6) Luckily, I found the book I was looking for in the library.
- (7) It's this type of film that I like best.
- (8) It's great that you can come to visit.
- (9a) There's a stain on your skirt.
- (9b) Your skirt has a stain on it.

Three principal functions

1. Bridge (attachment)
2. Theme (perspective)
3. Attitude (interpersonal spectrum)

1. Bridge (attachment)

- Definition: The **Bridge** position signals to the Hearer where in the evolving discourse the current clause is to be attached

- The idea:
  - As the participants speak, the discourse structure continues to evolve
  - The Hearer has a problem knowing where the next clause should be attached
  - The Speaker can help by signaling the attachment point — a link back into the previous text
  - This allows the Hearer to build the discourse tree accurately
- How does the Hearer know which part of clause is the bridge? — usually at the clause start

- You cannot do this if you don't know the structure — ngram continuation generators are lost

Bridge example

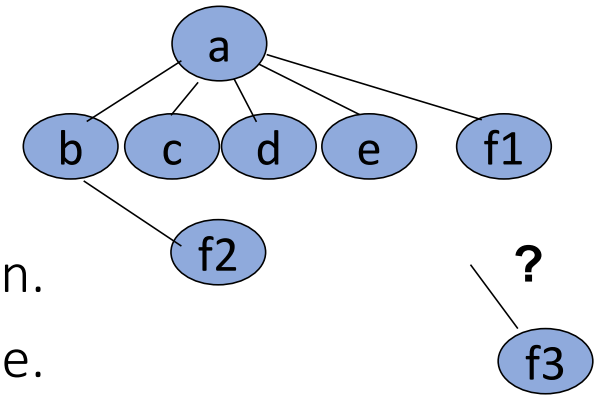
a. In Venice, the old man bought expensive new wine glasses.

b. The shop assistant wrapped them in a nice package.

c. Then he went to his hotel by gondola.

d. Getting off, he slipped, and fell into the water.

e. His clothes, shoes, hat, and glasses… everything was wet.



f1. And his glasses were broken.

f2. And inside the package, his glasses were broken.

f3. And his glasses were broken, inside the package.

2. Theme (perspective)

- Definition: The **Theme** position in a sentence signals to the Hearer what the Speaker's own point of view is to the situation

- Different positions in the clause carry different effects: The Speaker can highlight a participant or circumstance of the clause — but which one??
  - This choice might affect verb selection:
    (a) Maria received the letter from Hendrik
    (b) Hendrik sent the letter to Maria
- By placing a unit in clause-initial position, the speaker signals his/her perspective (point of view) on the event
  - In space: *come* vs. *go*, *receive* vs. *send*, *buy* vs. *sell*…
  - On specific entities: *Spanish omelette is what I like*

Theme control

- To know what to thematize, the generator needs a point of view, and needs to maintain it consistently
- The analyzer needs to recognize this and build a ‘mental map’ of where the generator ‘stands’ in relation to others in the situation

- This sort of internal model needs representation
- I doubt whether DNNs and other ML algorithms can automatically induce such representation

3. Attitude (Interpersonal stance)

- Definition: The **Attitude** signals the Speaker's own evaluation/opinion (toward the hearer, the clause content, the situation…)

- Interpersonal types:
  - **Mood** expresses basic Speech Act type (declarative, interrogative, exclamative, imperative) — *Are you my new piano teacher?* (Y/N question)
  - **Modality** expresses evaluation of probability, but not affect-bearing (possible, probable, certain…) — *It's possible that he was there*
  - **Evaluation** expresses affective attitude/sentiment (like, dislike) — *Unfortunately, she was blind; Perhaps she was late*

- Again, the system needs to recognize/maintain a consistent stance; there's no such control inside DNN-based parsers/generators

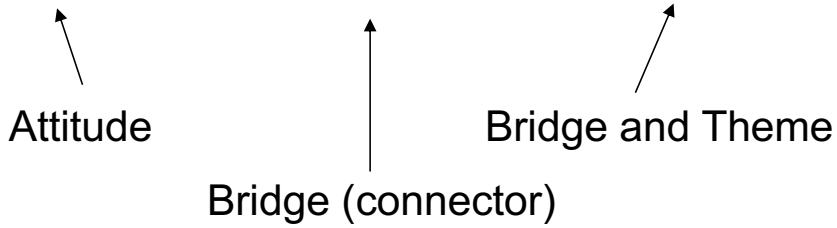
And this can happen together

My doctor forbade me to eat eggs….

Unfortunately, Spanish omelette is what I like

My doctor forbade me to eat eggs….

Unfortunately, however, Spanish omelette is what I like



Implications for text planning

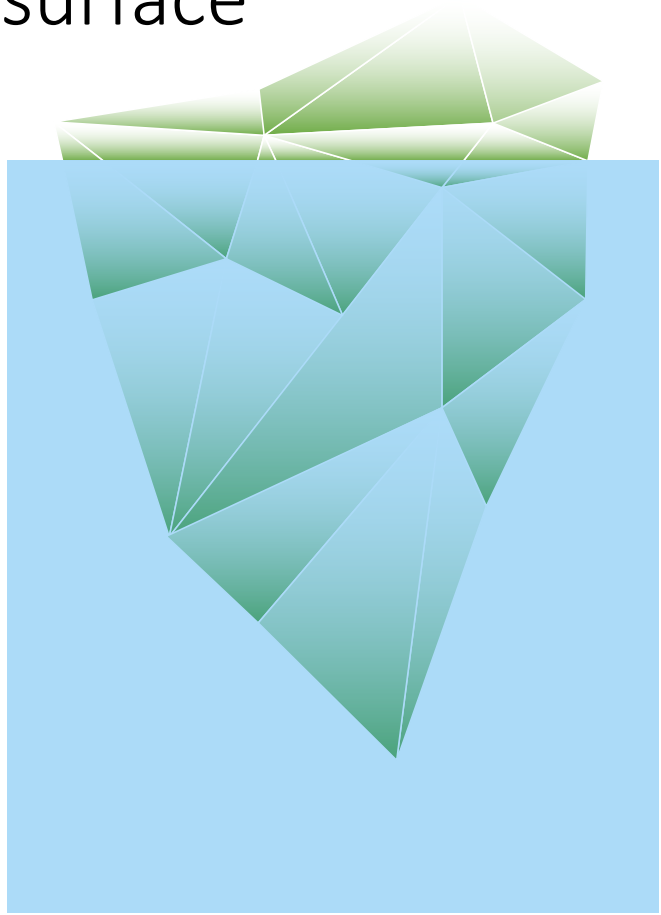
- How to proceed?
- Theory building: We need to represent (in rules) or create feature recognizers (for learning engines) the salient aspects:
  - Bridge: Possible attachment points (units in earlier discourse)
  - Theme: Speaker's 'locations' in space, etc.
  - Attitude: Types of attitude and lexical signals
- Testing: Then we need to test the validity of these ideas on texts (using annotation exercises)
- System building: Create a system (rules or statistics) that can identify the units
- Evaluation: Compare to gold-standard annotated corpus

Conclusion:
Where next for discourse?

There is a lot hidden below the surface

- Even more than for a sentence, discourse leaves important information implicit
- If you don't know what it is there is little hope for inducing/learning it automatically
- DNNs don't do very well
- Discourse is not primarily an algorithm problem

- NLP people should work on the theory



Where next for coherence research?

- Theories of speaker intent: semantic and interpersonal goals and effects, Speech Acts, illocutions and perlocutions
- Genres of structure: domain frames/schemas, argument graphs, simple dialogue…
- Discourse planning: plans and relations, recognizing and generating plans
- Clause-level attachment/briding/perspective planning

- Before we can apply DNNs, we need to know what real discourse requires!

Very little of this is primarily algorithmic

Most of this is theory building tested by annotation