Discrepancies between database- and pragmatically driven NLG: Insights from QUD-based annotations

Christoph Hesse Maurice Langner Anton Benz Ralf Klabunde

Special thanks to our annotators

Mariya Hristova Dorothea MacPhail Hathiya Muzni
Jeniffer Callou Johanna Wrede

Outline

1. Generation (NLG) of pragmatically rich texts
2. Question-Under-Discussion-based annotation
    1. The annotation process
    2. QUDs
    3. Focus/background
    4. Topic/comment
    5. At-issue/non-at-issue
3. Results

1 Generation (NLG) of pragmatically rich texts

Generation (NLG) of pragmatically rich texts

NLG from (1) database of technical vehicle specs, (2) QUD-trees of texts' discourse structure, annotated factual and evaluative assertions

Text plan:

- Typically based on rhetorical relations: (a) not constraint-based, and (b) descriptive rather than explanatory

Assumption: QUD-based approaches more suitable for theory-driven text plans

- Text spans and sentences are answers to implicit/explicit questions
- Discourse coherence is the systematic answering of discourse questions by partitioning chunks of information and sequencing

Discourse Q: Is this car good?
QUD_1: Is the car's engine good?
$QUD_{1.1}$: Is the engine's acceleration good?
$QUD_{1.2}$: Is the engine's max speed good?
$QUD_{1.3}$: Is the engine's performance good?
QUD_2: Is the car's suspension good?

Generation (NLG) of pragmatically rich texts (cont')

Theories of

- dialogue moves (Carlson, 1983; Ginzburg, 1996)
- contextual relevance (Roberts, 1996)
- information-structural concepts (e.g., focus/background distinction; Roberts, 1996; van Kuppevelt, 1995; von Stutterheim, 1997)
- temporal progression in narration (Klein & von Stutterheim, 1987)
- analysis of coherence relations and of subordination (Klein & von Stutterheim, 1987; van Kuppevelt, 1995)

Despite theoretic interest in QUDs, only 1x attempt at annotation guidelines and tools (De Kuthy, Reiter, Riester, 2018; Riester, Brunetti, De Kuthy, 2018; Riester, 2019) and no application in NLG

Generation (NLG) of pragmatically rich texts (cont')

A fully equipped Africa Twin with electronic suspension costs 18,665 euros. Including ancillary delivery costs and luggage system, it then joins the club of 20,000-euro touring enduros. First test rides show: The matter with the chassis is complicated. The new conventional suspension, which is fitted without exception in the basic Twin, and as standard in the Sports Adventure, does its job extremely well. The electronic suspension option available for the latter (1,600 euros extra) is even more sensitive, but the difference will not be noticeable for everyone. We bet: The top version will sell well.

objective facts from database

subjective expressive/evaluative content

2 QUD-based annotation

QUD-based annotation

- 30 German texts (driving reports)
- QUD-dependent information-structural distinctions:
  - focus/background
  - topic/comment
  - at-issue/not-at-issue information

2.1 The annotation process

The annotation process

- Divide text into sections
- Identify individual assertions
- Identify at-issue/not-at-issue information
- Divide assertions into text spans: focus/background, topic/comment
- Formulate QUDs fitting focus/background, topic/comment anlysis
- Indexing of constituents: split-focus/split-background & enumeration
- Bottom-up: left- and right-branching hierarchical QUD structures (representation of rhetorical relations, text sections, all the way up to discourse questions)

Resulting QUD trees have concrete, factual information (database queries) on leaf nodes, and level of abstraction to more evaluative/argumentative aspects increases towards the root node

2.2 QUDs

QUDs

- QUDs formulated with database queries in mind
- Applying this approach is easy when assertions are more fact-based, and difficult when they are subjective/evaluative/expressive content

Easy: Singleton factoids

QUD What's the car's acceleration? **A** from 0 to 100 km/h in 5.7 sec

Also easy: List enumeration (multiple facts)

QUD What models are available? **A** [list of *n* models]

Difficult: Subjective/evluative/expressive content

QUD How might changes to the interior affect the target audience? **A** ...

- Nuanced arguments, e.g. foregrounding/backgrounding

2.3 Focus/background

Focus/background

- Focus & background are propositional attributes
- Focus = new part of a proposition (answer to QUD)
- Background = old part of a proposition (presupposed in/part of QUD)
- Foci map to syntactic constituents, the focus domains
- Focus domain contains focus exponent = most salient element of this focus domain
- Two functions of focus: (1) new information, (2) contrast information
- Given/new distinction not always applicable

Focus/background (cont')

Split-focus: multiple focus constituents actually one A to QUD

QUD What about the power unit?

A [In der praktischen Außenhaut]$_{F_1}$ des 3,60 m kurzen Fünftürers, [war der Antrieb erstmal kaum zu erkennen.]$_{F_1}$
[In the practical outer skin]$_{F_1}$ of the 3.6 meter short five-door car, [the engine was hardly noticeable at first.]$_{F_1}$

- Focus F_1 split
- 3.6 m, five-door segment does not answer the QUD
- Surprisal (hardly noticeable) from counterfactual intuitions: small car (*3.6 m short* not *long*, five doors also uncommon with small cars, *practical* code for small too?) $>$ no room for big power unit
- QUD What about the power unit? must be part of a larger argument structure which captures this surprisal/counterfactual

Focus/background (cont')

Coordination: multiple distinct foci partial/cumulative A to QUD

QUD How is the Renault Captur?

A Der Renault Captur $[\text{wächst}]_{F_1}$ und $[\text{verändert seinen Charakter.}]_{F_2}$
The Renault Captur $[\text{grows}]_{F_1}$ and $[\text{changes its character.}]_{F_2}$

- 2 foci because 2 different aspects
- 2 coordinated main clauses, each with its own focus, but answering single QUD
- Focus as A to QUD not a 1-to-1 relationship (1x focus $\neq$ 1x QUD)

Focus/background (cont')

Ellipsis: verb + preposition pied-piping foci

QUD How have the aesthetics changed, compared to the old Captur?

A [das sieht scharf trainiert (aus)]$_{\mathrm{F}_1}$und [(das sieht) angriffslustig aus]$_{\mathrm{F}_2}$
[that looks sharply trained]$_{\mathrm{F}_1}$ and [(that looks) ready to attack]$_{\mathrm{F}_2}$

- German verb to look/appear has 2 versions: prefixed *aussehen* and non-prefixed *sehen* . . . *aus*, which can do pied-piping
- Notice how verb ellipsis works: verb stem *sieht* in the first clause, preposition *aus* in the second clause

2.4 Topic/comment

Topic/comment

- Topic = discourse referent of the comment
- Comment = a statement about the topic
- Topic often introduced via indefinite means and later on picked up anaphorically
- Topic structures might involve contrastive topics (comparison of discourse topics)
- Topics are contained in phrasing of QUDs
- Topics map to database variables
- For the genre of driving reports, topics are vehicles and vehicle parts
- Indexing for coreference (metonymy can pose problems)

2.5 At-issue/non-at-issue

At-issue/non-at-issue

- At-issue content = all information relevant given context (topic/background)
- Not-at-issue content = additional/optional information, not speaking to the QUD (not discourse-irrelevant in general)
- Not-at-issue content is a seprate assertion, potentially with a QUD subordinated to QUD of at-issue content
- This sub-QUD can then have its own independent focus structure
- Degrees in at-issueness > embedding structures in QUD trees
- Examples of non-at-issue:
  - Evaluative adverbs (e.g., *surprisingly*, *remarkably*)
  - Embedding verbs (e.g., $[\text{They say that}]_{\text{NAI}}\,[[\text{the car}]_{\text{BG}}\,[\text{is overpriced}]_{\text{F}}]$)

At-issue/non-at-issue (cont')

Licencing conditions of evaluative adverbs

Surprisingly, the new Kawasaki consumes 4 liters less gas despite its 5% weight increase in comparison to the previous generation of this model.

- Technical understanding of the connection between weight and gas consumption necessary to decode the licencing conditions of *surprisingly*
- This sort of information cannot be queried from database but must be inferred from domain knowledge
- and it must inform the phrasing of QUDs and QUD tree structures

- Database contains marks 1.0 to 6.0 for each evaluation criterium
- Annotators need to take special care in identfying not-at-issue content (e.g., using tests proposed by Simons, Tonhauser, Beaver, Roberts, 2010) so that it can be properly mapped to these marks

3 Results

Results

- γ coefficient (Mathet et al., 2015) rather than Kirppendorff's α because α sensitive to length of text segements while γ is not
- $\gamma = 1$ perfect annotator agreement, $\gamma < 0$ worse than random

Category/Tag	γ
Not-at-issue	-0.075
Focus	0.115
Background	0.08
Auxiliary/misc	0.32
Overall	0.13

- Annotators disagree on information-structural labels
- Low agreement because of differences in tree structure rather than disagreemen about which facts are relevant, what is focus, what background, and what at-issue because that is very high
- Since phrasing of QUDs heavily influences tree structure, structural differences do not necessarily mean annotators differ in their text interpretation

Results

- Challenge not *which* facts to realized, but *how* to realize them so they support a specific argument
- Not all QUDs are fact-oriented, but can also be evaluative/subjective/expressive
- QUD structure not simply derivable from focus structure or at-issueness
- Number of levels in hierarchical QUD trees highly depend on complexity of argument
  - Technical facts are evaluated positively/negatively
  - Depending on their evaluation, facts play different roles as evidence in authors' argumentation
  - Argumentation often anticipates counterfactual of implicit assumptions and causal logic, which must be spelled out in QUDs
- A QUD analysis of these pragmatically rich texts is not reducible to QUDs translating to database queries of technical facts and marks

Conclusion for NLG

- Linguistic analysis relevant and important for NLG
- Without good annotation poor results with ML of discourse structures
- Linguistic annotation is non-trivial
- Extend small corpus with additional ADAC data for unsupervised learning

Thank you!

/christoph-hesse/question-under-discussion
/MMLangner/QUDA

Special thanks to our annotators
Mariya Hristova Dorothea MacPhail Hathiya Muzni
Jeniffer Callou Johanna Wrede