



Analyzing focus in authentic data from an explicit task context

Kordula De Kuthy, Ramon Ziai, Detmar Meurers

January 14, 2016



Outline of the Talk

Motivation

Related Work

Switchboard Corpus
Lisa Annotation Scheme

Identifying Focus given Explicit Questions

Annotation Scheme
Difficult Cases

Identifying Focus given Implicit Questions

Annotation Guidelines
The Snowden Interview
The Stuttgart 21 Corpus

Conclusion



Motivation

- Why do we want to annotate corpora with information structure?
- Information structure notions such as *focus* and *topic* have been intensively investigated in the linguistic literature.
- The theories
 - mostly rely on intuitions about hand-crafted examples
 - are increasingly complex, without being fully operationalized
 - make contradicting empirical predictions (De Kuthy and Meurers 2012)
 → It is important to test these theories with authentic data.
- Insights gained from analyzing authentic data can help
 - empirically validate the key notions, and
 - extend the empirical relevance and reach of the theoretical proposals.



What is focus?

Aspects discussed in the literature:

- **answer to a question** (to an explicit or an implicit one)
- **informative** part of an utterance
- the part of an utterance that signals **alternatives**
- indicates **new** or **important** / **contrastive** information
- **asserted** / **at issue** material
- often signalled by **prosodic or syntactic prominence** (*language-dependent*)



Exploring these aspects for annotating focus

- **answer to a question:**
 - Analyze data in tasks where questions are explicitly given.
 - No explicit question given? Annotate the implicit question.
- **alternatives:**
 - Are alternatives present in the immediate context?
- **contrastive information:**
 - Explicitly mark preceding elements that are relevant for the contrast.
- **prosodic prominence:**
 - For spoken data, establish a connection between prosody and focus.
- **syntactic structure:**
 - What is the nature of the units that play a role in focus annotation?



Related work on Annotating Information Structure

1. **Switchboard:** a transcribed corpus of telephone conversations annotated for
 - syntax, disfluency, speech acts, animacy,
 - information status, coreference, prosody,
 - and a more recent annotation of focus (Calhoun et al. 2010).
2. **LISA** annotation scheme (Ritz et al. 2008)



Switchboard Corpus: The annotation scheme

- For the annotation of focus two new terms are introduced:
 - *Kontrast*: Coding of words for whether they have a salient alternative in the context (*kontrast*) or not (*background*).
 - *Trigger*: Links certain *kontrasts* to the word(s) motivating their marking.
- For *kontrast*, annotators identified words or NPs which were
 - “salient with an implication that this salience is in comparison or contrast to other related words or NPs explicitly or implicitly evoked in the context”
- For background and trigger, words of specific parts-of-speech were annotated (nouns, verbs, adj., adv., pronouns, demonstratives).



Types of Kontrast values

- **correction:** corrects or clarifies another word or NP
- **contrastive:** intended to contrast with another word mention in the context
- **subset:** highlights one member of a more general set that has been mentioned
- **adverbial:** a focus-sensitive adverb
- **answer:** fills an open proposition set up in the context such that it would make sense if only that word or phrase were spoken
- **other:** *kontrast*, but not one of the other types



Focus annotation in Switchboard: An example

Telephone conversation between two strangers about cutting the costs of the prison system:

A: [Have]_{background} [[private enter-enterprise]_{contrastive} [do]_{background} [it]_{background}

B: You mean the prisons

A: [They're already]_{background} [[talking]_{contrastive} about [it]_{background} [I mean]_{nonappl}
[they're talking about]_{background} , uh, [[having]_{contrastive} [it]_{background} , uh, as a
[you know]_{nonappl} [[business]_{contrastive} . Uh, to, uh, [you know]_{nonappl} , to so
[the]_{background} [you know]_{nonappl} the [[government]_{contrastive} [doesn't]_{background}
[have]_{background} to [[deal]_{contrastive} with [it]_{background} .



Focus annotation in Switchboard: Problems reported

- Annotation of *kontrast* was done by two annotators.
- Three main sources of disagreement between annotators:
 - (i) Varying scope of *kontrast*:
 - *Kontrast* was marked at the word or the NP level
 - Segmentation is one of the most challenging aspects of discourse annotation (Calhoun et al. 2010).
 - (ii) Disagreement between the subtypes of *kontrast*
 - (iii) Disagreement between *other kontrast* and *background*



Annotation of Focus with LISA

- At the SFB 632 Information Structure: creation of corpora for typologically diverse languages for IS-related research
- Tagsets and annotation guidelines were developed for several linguistic layers (Dipper et al. 2007).
- The LISA annotation scheme (Ritz et al. 2008) covers three dimensions of information structure:
 - information status (given, new, ...)
 - topic (aboutness topic, frame-setting topic)
 - focus (new information focus, contrastive focus)



Annotation of Focus with LISA: Guidelines

- Definition of focus:
 - That part of an expression which provides the most relevant information in a particular context
 - as opposed to the rest of information making up the background of the utterance.
- Focus on a subexpression indicates that it is selected from possible alternatives that are either implicit or given explicitly.
- Focus can extend over different domains in the utterance (affixes, words, constituents, clause) and can be discontinuous.



Annotation of Focus with LISA: Two types of new-information focus

- **solicited new-information focus:** part of a sentence carrying information explicitly requested by another discourse participant
- It is assumed that for each sentence a preceding explicit or implicit question exists.
- **unsolicited new-information focus (nf-unsol):** the part that supplies new or missing information according to implicit question



Annotation of Focus with LISA: Example (Dipper et al. 2007)

- (1) a. **[[Heute ist mir was ganz tolles passiert.]]_{nf-unsol}**
'Something totally fantastic has happened today to me.'
- b. **[[Da sollten nämlich, nämlich, der Thomas und der Ludwig,]]_{nf-unsol}**
die sollten [[Tomaten holen gehen]]_{nf-unsol}
'Thomas and Ludwig were supposed to go and get some tomatoes.'
- c. und **[[die Mama hat zuerst den Thomas]]_{nf-unsol}** losgeschickt
'And mum first sent Thomas '
- d. und der **[[ist dann los gegangen]]_{nf-unsol}** und **[[kam aber ohne Tomaten]]_{nf-unsol}** wieder
'and he went, but came back without tomatoes'



Annotation of Focus with LISA: an example (cont.)

- (2) a. und **[[dann sollte der der Ludwig los gehen]]_{nf-unsol}**
'and then Ludwig was supposed to go'
- b. und **[[dem ist dann genau dasselbe]]_{nf-unsol}** passiert
'and to him the same thing happened'
- c. **[[und dann sollte ich los gehen]]_{nf-unsol}**
'and then I was supposed to go'
- d. und ich **[[bin in die Stadt gegangen]]_{nf-unsol}** und **[[habe den richtigen Weg gefunden]]_{nf-unsol}** und **[[habe Tomaten mitgebracht]]_{nf-unsol}**
'and I went into town and found the right way and got some tomatoes'
- e. und **[[da hat sich die Mama sehr drüber gefreut.]]_{nf-unsol}**
'and mum was very happy about this.'



Annotation of Focus with LISA: Problems

- Problems reported by Ritz et al. (2008)
 - annotator disagreement about the size of the focus: phrasal head or whole phrases were marked as focus
 - "The definition of where foci start will need improvement."
- We observed:
 - The size of the focus in parallel example varies, in particular sometimes the main verb is part of the focus, sometimes not.
 - It is impossible to evaluate the size of the focus without knowing which implicit Question Under Discussion the annotators had in mind.
 - question under discussion should be made explicit in corpus



Our Perspective

- Make all aspects relevant to focus annotation fully explicit.
 - An annotation scheme should
 - explicitly annotate *QUDs*
 - mark *not-at-issue* content
 - mark *focus, contrast, topic*
 - make *form characteristics of potential focus units* explicit
 - Approach pursued:
 - Start by analyzing language in authentic task contexts including explicit questions: answers to reading comprehension questions
 - Then analyze increasingly less contextually constrained data: interviews and dialogues
- supports incremental theory validation and refinement



Focus with Explicit Questions

- Even with explicit questions, focus annotation in authentic data is not without challenges:
 - The size of individual foci has been identified as a source of disagreement (Ritz et al. 2008; Calhoun et al. 2010).
 - Non-well-formedness can make robust focus annotation difficult.
- To tackle these issues, we explored focus annotation in authentic reading comprehension data (Ziai and Meurers 2014).
- This data source is interesting because
 - the task makes explicit questions natural,
 - the context of the utterances is known and linguistically encoded and
 - it is known what content a felicitous answer should include.



Authentic Data: The CREG Corpus

- The data comes from CREG (Meurers et al. 2011; Ott et al. 2012) a German reading comprehension corpus we compiled.
- It was collected in the German programs of Kansas University and Ohio State University.
- CREG is a task-based corpus containing
 - 164 reading texts,
 - 1,517 reading comprehension questions,
 - 2,057 target answers provided by the teachers, and
 - 36,335 learner answers written by American learners of German.
- Each learner answer was rated by two annotators with respect to whether it answers the question correctly or not.



Corpus Specifics

- Focus annotation was performed on the CREG-1032 subset
 - balanced set: as many correct as incorrect answers
 - token length ≥ 5
 - both annotators agreed on meaning assessment
- The OSU subset was annotated for the current study since it contains longer answers and more answers per question.

	KU data set	OSU data set
target answers questions	136	87
student answers students	610	422
avg. token #	141	175
	9.71	15.00

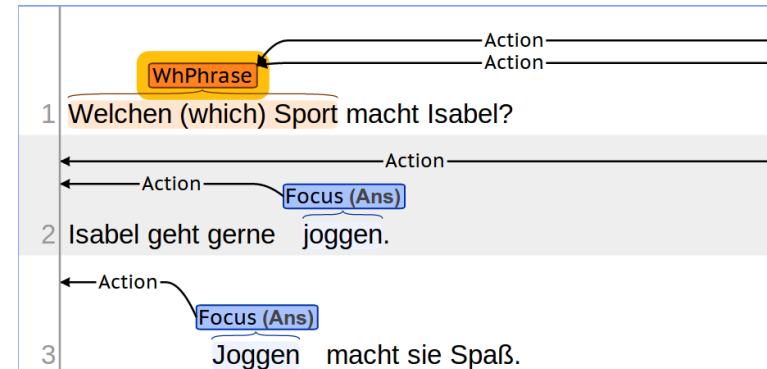


Annotating Focus with Explicit Questions

- Existing focus annotation schemes do not explicitly take questions into account (Calhoun et al. 2010; Dipper et al. 2007).
 - ⇒ We need to create a new annotation scheme.
- Our scheme consists of three types of categories:
 - Question Form** encodes the surface form of a question.
 - Focus** marks the focused words or phrases in an answer.
 - Answer Type** expresses the semantic category of the focus in relation to the question form



Example Annotation in *brat*



- Q: 'Which sport does Isabel do?'
 TA: 'She likes to go [jogging]_F.'
 SA: '[Jogging]_F is fun for her.'



Question Form

- WhPhrase:** Is applied to *wh*-questions and contains the whole phrase (usually identical to Vorfeld)
 - Example: *Why was Schorlemmer afraid in the beginning?*
- Yes/No:** For polarity questions, marks only the verb in the beginning of the question
 - Example: *Does one have to be a German citizen?*
- Alternative:** For alternative questions, all alternatives are marked.
 - Example: *Is he for or against the EU law?*
- Imperative:** For question imperatives, marks the imperative verb.
 - Example: *Give reasons for this other nickname.*
- NounPhrase:** For (rare) cases, where a question is abbreviated through a noun phrase
 - Example: *Place of residence?*



Focus

- A question opens up a specific alternative set (Rooth 1992)
 - Example: *Where was Mozart born?*
 - set of places
- Focus is the part of the answer that reduces the alternative set, inspired by Krifka (2007)
- Consequently, in deciding whether a word is in focus or not we test for semantic contribution:
 - If the word is substituted, does the meaning (and the reduction of the alternative set) of the expression change?



Focus

Q: Where does Heike live?

A: She lives [in Berlin.]_F

- Why? If *in* is replaced by e.g. *near*, the meaning changes.

Q: In what city does Heike live?

A: She lives in [Berlin]_F.

- Why? Alternative set is more restricted, *in* is obligatory.



Focus

- Multiple foci are possible
 - e.g., in double wh-questions: *Who bought which hat?*
- Only one kind of focus, contrastive focus is not distinguished separately.
- ‘QUD’ attribute encodes whether the question was answered directly.
 - Does the Question Under Discussion (Roberts 2012) differ from the explicit reading comprehension question?



Answer Type

- Expresses the semantic category of the focus in relation to the question form.
- Describes nature of the question-answer congruence by specifying the semantic class of the set of alternatives.
- Set of labels specific to data set strikes a balance between generality and specificity.



Answer Types with Example Answers

Category	Example (translated)
Time_Date	The movie starts <i>at 5:50</i>
Living_Being Thing	<i>The father of the child</i> padded through the dark outskirts. For the Spaniards <i>toilet and stove</i> are more important than the internet.
Abstr. _Entity Report	The applicant needs <i>a completed vocational training as a cook</i> . The speaker says “ <i>We ask all youths to have their passports ready.</i> ”
Reason	The maintenance of a raised garden bed is easier <i>because one does not need to stoop</i> .
Location	She is from <i>Berlin</i> .
Action	In the vegetable garden one needs to <i>hoe and water</i> .
Property	Reputation and money are <i>important</i> for Til.
Yes_No	<i>The mermaid does not marry the prince</i> .
Manner	The word is used <i>ironically</i> in this story.
Quantity/ Duration	The company seeks <i>75 employees</i> .
State	If he works hard now, <i>he won't have to work in the future</i> .

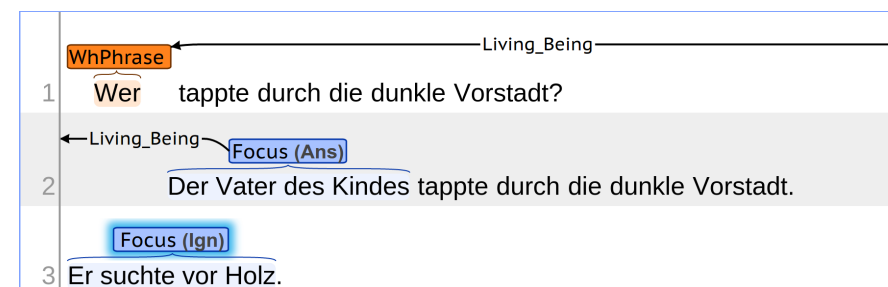


Indirect Answers

- So far, we assumed the explicit question to be the Question Under Discussion (Roberts 2012).
- Sometimes learners answer different questions which can be related to the explicit question (or not).
- In these cases, we annotate the new QUD and encode the nature of divergence from the old one as
 - 'Generalized',
 - 'Narrowed',
 - or 'Ignored'
 in the 'QUD' attribute.



Indirect Answers: Example



Q: 'Who padded through the dark outskirts?'

TA: '[The child's father]_F padded through the dark outskirts.'

SA: '[He searched for wood]_F.'

(QUD: 'Why did he pad through the dark outskirts?')



Annotation Results: Inter-annotator agreement

Type of distinction	Source	# tok.	%	κ
Binary (focus + background)	Student	6329	82.8	.65
	Target	6983	84.9	.69
Detailed (13 Answer Types + background)	Student	5198	72.6	.61
	Target	6839	76.5	.67

- Substantial agreement ($\kappa > .6$) throughout.
- Student answers harder to annotate than target answers.
- Answer Types as scaffold for focus work well.



Disagreement Example 1: Given Material

Q: Warum nennt der Autor Hamburg das
"Tor zur Welt der Wissenschaft"?
'Why does the author call Hamburg the
"gate to the world of science"?'

SA: [Hamburg hat viel renommierte Universitäten]_F (ann. 1)
Hamburg hat [viel renommierte Universitäten]_F (ann. 2)
'Hamburg has many renowned universities'

- Annotator 1 marks whole sentence
 - correct, alternative set is the set of reasons, not necessarily including *Hamburg*
- Annotator 2 excludes Given material
 - incorrect, because *Hamburg* is needed to distinguish between alternatives



Disagreement Example 2: Prepositions

Q: Wofür ist der Aufsichtsrat verantwortlich?

'What is the supervisory board responsible for?'

SA: Der Aufsichtsrat ist für [die Bestellung]_F verantwortlich. (ann. 1)
Der Aufsichtsrat ist [für die Bestellung]_F verantwortlich. (ann. 2)

'The supervisory board is responsible for the appointment.'

- Annotator 1 excludes preposition *für*
 - correct, because it cannot be exchanged and is needed only for well-formedness
- Annotator 2 marks whole preposition phrase
 - incorrect, *für* not needed to distinguish between alternatives



Focus given Explicit Questions: Conclusion & Outlook

- We showed that focus can successfully be annotated (with substantial agreement) in authentic data.
- We need a definition of focus which is semantically motivated
 - but precise enough to clearly select among possibilities in syntactic scope
- More annotation on more data is needed:
 - We plan to continue annotation efforts in our reading comprehension data.
 - We are also exploring crowdsourcing annotation, which helps to reduce the theory-dependence of information-structural notions.
- On that basis, we will try to automatically identify focus.
 - Manual annotation is semantically motivated but automatic classification needs to rely on surface properties.
 - Should lead to further insights on which surface properties reliably correlate with which semantic categories.



From Explicit to Implicit Questions

- Making use of a reading comprehension task providing explicit questions greatly facilitates
 - a) identifying the QUD and nature of alternatives, and
 - b) determining the location and extent of the focus in the language material.
- How do we proceed if we do not have explicit questions?
- We need a discourse model that
 - deals with explicit and implicit Questions under Discussion
 - defines what the relation between QUDs and answers is
 - and which QUDs can be inserted when and where.
- Adopt discourse model of Roberts (2012), Büring (2003) in which
 - discourse is organized along strategies of inquiry,
 - consisting of a series of questions and subquestions.



Questions under Discussion (Roberts 2012; Büring 2003)

- Informative discourse serves to eliminate uncertainty about the state of the world.
- With most of what we communicate, we strive to (partially) answer the Big Question *What is the way things are?*
- To that end, we devise a discourse strategy consisting of more specific questions.
- An (explicit or implicit) question is under discussion until it has been answered, i.e., *resolved*.



Relevance and not-at-issue content

- Only some of the material in a felicitous conversational move is relevant to the QUD (Simons et al. 2011):
 - An assertion is relevant to a QUD iff it contextually entails a partial or complete answer to the QUD.
 - Parts of an utterance that are not relevant to a QUD constitute *not-at-issue content*.
- Typical not-at-issue content (Potts 2005; Simons 2007):
 - supplemental expressions (appositives, non-restrictive modifiers, parentheticals)
 - expressives (epithets, ...)
 - evidentials (*know/think that*, ...)



An Annotation Approach for Focus and QUDs

(jointly developed with Arndt Riester)

- Goal: Turn a text into a discourse tree.
- The root node represents the most general QUD.
- Each node in the tree represents the current QUD at that position.
- A terminal nodes represent an answer to the specific QUD.



The Annotation Procedure

- Prepare text:
 - Read text carefully, making sure you understand it.
 - Split sentences into clauses.
 - Mark not-at-issue content.
- Build discourse tree:
 - Try to group as many clauses as possible under a common question.
 - Insert the implicit QUD above the clauses.
 - The clauses then represent a series of partial answers to the QUD.
 - In the tree, these partial answers are sibling nodes.
 - In every clause, mark those constituents that directly answer the current QUD as the focus of the clause.



Pilot Annotation with Authentic Data

- Idea: explore annotation scheme by stepwise relaxing the strict task structure we relied on in reading comprehension data
- (i) Interviews still offer some explicit questions as guidance:
 - Interview with Edward Snowden (English)
- (ii) Panel discussions are more free, but still evolve around a common topic.
 - Public mediation panel on building of a train station (German)



Authentic Data: The Snowden Interview Corpus

- We compiled a corpus from recordings and transcripts of an interview with Edward Snowden broadcasted on German TV.
- The corpus was annotated together with Arndt Riester.
- The annotation includes part-of-speech, information status (given/new) and prosodic prominence.
- Parts of the interview were annotated with QUDs and focus.



QUDs and Focus in the Snowden interview

Q0: *Who is Edward Snowden?*

Q1: *What did Edward Snowden do before he became a whistle-blower?*

A1: **[You]**_T **[were working until last summer for the NSA]**_F

Q1.1: *What happened there?*

A1.1: and [during this time]_{nai} **[you]**_T **[collected secretly thousands of confidential documents]**_F . [Sound](#)

Q2a: *What was the decisive moment*

Q2b: *or was there a long period of time*

Q2c: *or something happening* [Sound](#)

Q3: *why did you do this?* [Sound](#)



QUDs and Focus in the Snowden interview (cont.)

A2a: [I would say]_{nai} [sort of the breaking point]_T [is seeing the Director of National Intelligence, [James Clapper]_{nai} , directly lie under oath to Congress]_F . [Sound](#)

A3: [There's no saving an intelligence community that believes it can lie to the public and the legislators]_F [who need to be able to trust it and regulate its actions]_{nai} [Sound](#)

Q3.1: *What did Snowden realize?*

A3.1: [Seeing that really meant for me]_{nai} [there was no going back]_F [Sound](#)

A2.c/A3: [Beyond that,]_{nai} [it]_T was [the creeping realization that no one else was going to do this]_F . [Sound](#)



Authentic Data: The Stuttgart 21 corpus

- We compiled a corpus from recordings and transcripts of debates.
- Public mediation meetings between proponents and opponents of the Stuttgart 21 (S 21) construction project
 - Participants of the debate include local politicians, company representatives, and experts in various subjects.
 - The debate is led by a neutral mediator (Dr. Heiner Geißler).
 - The mediation took place in nine sessions, October to November 2010.
- Total length of the recordings: ≈ 50 hours
- To further pilot the annotation scheme and procedure, we annotated a small part of the corpus with QUDs and focus.



QUDs and Focus in Stuttgart 21: Example annotation

Q0: Können wir das mal abklären, was das für Gutachten sind?

can we that clarify what these for appraisals are

Q1: Also, was ist jetzt?

Ok, what is now?

Q1.1: What kind of proposals are there?

A1.1.a: Es gibt offenbar [ein neues überarbeitetes Konzept.]_F

there is apparently a newly corrected proposal

Q1.2: From when are the proposals?

A1.2.a: [Das erste]_T stammt [aus dem Jahre 2005.]_F.

The first one is from the year 2005.

Q2: Und jetzt, was gibt's neues?

And now, what is there new?

TG: **A1.1.b:** [Nein]_F, [es geht, es geht darum.]_{nai}

No, it is, it is about.

A1.2.b: [Außerdem]_{nai} stammt das erste [nicht aus 2005.]_{CF}

in addition comes the first not from 2005

A1.1.c: Und es gibt [eins mit Verstärkerfahrten]_F

and there is one with peak-time trains



QUDs and Focus in Stuttgart 21 (cont.)

NAI: Also, für die Zuschauer.

well for the audience

Q1.3: Which concepts are mentioned in these proposals?

Q1.3.1: What are Verstärkerfahrten?

A1.3.1: [Verstärkerfahrten, das]_T sind [Züge, die in Spitzenzeiten

peak-time trains that are trains, which in peak hours

zusätzlich eingesetzt werden.]_F

in addition installed are.

Q1.3.1.1: Why am I giving this explanation?

A1.3.1.1: [Das]_T [versteht kein Mensch sonst.]_F

that understands nobody otherwise

Q1.3.2: What do we need during peak times?

A1.3.2 [Nicht wahr,]_{nai} also, man braucht [halt]_{nai} [in Spitzenzeiten,

indeed ok one needs in peak times

im Berufsverkehr]_T braucht man [mehr Züge, als,

in the rush hour needs one more trains than

[sagen wir mal,]_{nai} vormittags um elf.]_F

lets say in the morning around 11



QUDs and Focus in Stuttgart 21 (cont.)

A1.1.c: [So,]_{nai} und [dafür]_T gibt es [offenbar]_{nai} neue, neue, [ein

so and therefore are there apparently new new a

neues Gutachten, oder eine neue Berechnung, oder, von demselben

new appraisal or a new calculation or of the same

Institut.]_F

institute

Q1.3.3: What is the institute called SMA?

A1.3.3: Und SMA, [nicht wahr]_{nai}, [das]_T ist ein ein [ein Institut aus

and SMA, indeed, that is an an an institute from

Zürich, [nicht wahr]_{nai}, das Gutachten angefertigt hat für den

Zurich, isn't it, which reports prepared has for the

Gesamtcomplex em S21 und drum herum.]_F

entire complex S21 and all around



Annotating focus: observation summary

- Adding implicit QUDs to the text helps
 - make the discourse structure of a text more transparent
 - identify at-issue and not-at-issue material
 - determine the size of the focus constituent.
- The identification of possible QUDs needs to be more constrained, integrating
 - top-down: Starting from the top QUD identify more specific QUDs.
 - bottom-up: Based on the utterance, determine potential QUDs.
- Overall pilot annotation results support the feasibility of annotation approach combining QUD and focus.
 - Plan to propose an annotation project based on the authentic interview and panel discussion data.



Conclusion

- Exploring how authentic data can be annotated for information structure, we want to make a contribution to an explicit and empirically adequate approach to analyzing focus.
- Our approach separates
 - a) identifying the QUD and nature of alternatives, and
 - b) determining the location and extent of the focus in the language material.

Making use of a reading comprehension task providing explicit questions greatly facilitates a).

- Our annotation pilots support that discourses can be structured in terms of QUDs, which then permit reproducible IS annotation.



Thank you!

Contact:

SFB 833

Project A4

Nauklerstr. 35, 72074 Tübingen

Phone: +49 7071 29-77167

{kdk,rziai,dm}@uni-tuebingen.de



References

Büring, D. (2003). On d-trees, beans, and b-accent. *Linguistics and Philosophy*, 26(5):511–545.

Calhoun, S., Carletta, J., Brenier, J., Mayo, N., Jurafsky, D., Steedman, M., and Beaver, D. (2010). The NXT-format switchboard corpus: A rich resource for investigating the syntax, semantics, pragmatics and prosody of dialogue. *Language Resources and Evaluation*, 44:387–419.

De Kuthy, K. and Meurers, D. (2012). Focus projection between theory and evidence. In Featherston, S. and Stolterfoth, B., editors, *Empirical Approaches to Linguistic Theory – Studies in Meaning and Structure*, volume 111 of *Studies in Generative Grammar*, pages 207–240. De Gruyter.

Dipper, S., Götze, M., and Skopeteas, S., editors (2007). *Information Structure in Cross-Linguistic Corpora: Annotation*



Guidelines for Phonology, Morphology, Syntax, Semantics and Information Structure, volume 7 of *Interdisciplinary Studies on Information Structure*. Universitätsverlag Potsdam, Potsdam, Germany.

Krifka, M. (2007). Basic notions of information structure. In Fery, C., Fanselow, G., and Krifka, M., editors, *The notions of information structure*, volume 6 of *Interdisciplinary Studies on Information Structure (ISIS)*, pages 13–55. Universitätsverlag Potsdam, Potsdam.

Meurers, D., Ziai, R., Ott, N., and Kopp, J. (2011). Evaluating answers to reading comprehension questions in context: Results for German and the role of information structure. In *Proceedings of the TextInfer 2011 Workshop on Textual Entailment*, pages 1–9, Edinburgh. ACL.

Ott, N., Ziai, R., and Meurers, D. (2012). Creation and analysis of a reading comprehension exercise corpus: Towards evaluating



meaning in context. In Schmidt, T. and Wörner, K., editors, *Multilingual Corpora and Multilingual Corpus Analysis*, Hamburg Studies in Multilingualism (HSM), pages 47–69. Benjamins, Amsterdam.

Potts, C. (2005). *The logic of conventional implicatures*. Oxford University Press, Oxford.

Ritz, J., Dipper, S., and Götze, M. (2008). Annotation of information structure: An evaluation across different types of texts. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*, pages 2137–2142, Marrakech, Morocco.

Roberts, C. (2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, 5(6):1–69.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, 1(1):75–116.



Simons, M. (2007). Observations on embedding verbs, evidentiality, and presupposition. *Lingua*, 117(6):1034–1056.

Simons, M., Tonhauser, J., Beaver, D., and Roberts, C. (2011). What projects and why. In *Proceedings of SALT*, volume 20, pages 309–327.

Ziai, R. and Meurers, D. (2014). Focus annotation in reading comprehension data. In *Proceedings of the 8th Linguistic Annotation Workshop (LAW VIII, 2014)*, pages 159–168, Dublin, Ireland. COLING, Association for Computational Linguistics.