

# Markup of human-bot dialogues: known problems and new challenges

**Valeria Generalova**

*Russian State University for the Humanities*

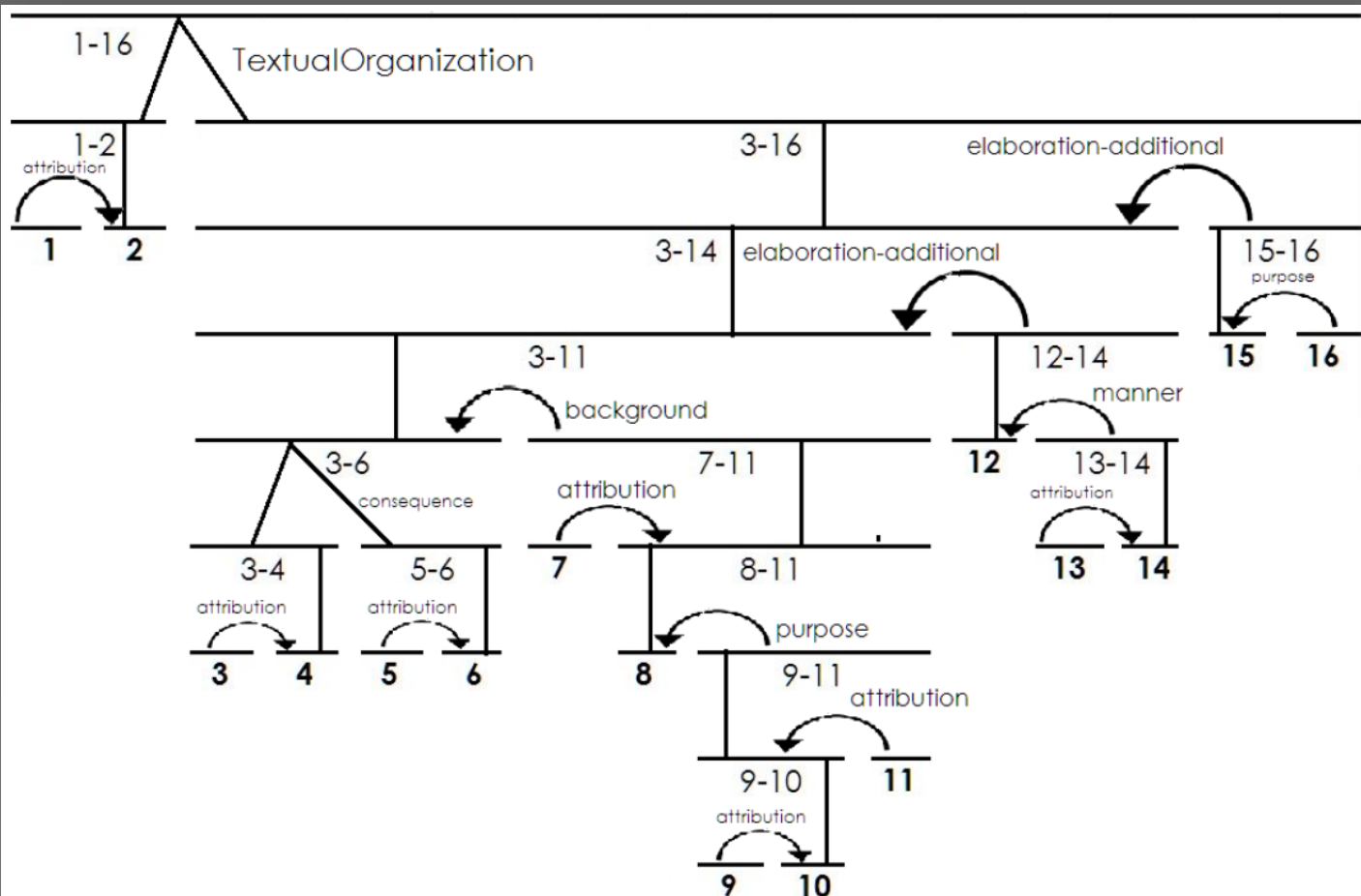
**61st Student's Conference for Linguistics**

**24 May 2017, Zurich**

# Dataset ≠ collection of texts ≠ corpus

- Linguistic corpora have linguistic annotation
- The more annotation, perhaps, the better the corpus
- Annotation has to be relevant for the users of the corpus
  - rhyme types in poetry corpora
  - prosody in spoken corpora
  - pragmatics in dialogue corpora

# Narratives and dialogues are not alike



```
<diaml xmlns:
"http://www.iso.org/diaml/">
<dialogueAct xml:id="e1" target="Is there an
earlier connection?"
sender="#c" addressee="#a"
communicativeFunction="prop.Question"
dimension="task"/>
<dialogueAct xml:id="e2" target="No there isn't"
sender="#a" addressee="#c"
communicativeFunction="answer"
dimension="task"
functionalDependence="#e1"/>
<dialogueAct xml:id="e3" target="I'm sorry"
sender="#a" addressee="#c"
communicativeFunction="apology"
dimension="social obligations"/>
</diaml>
```

# What is the problem with human-bot dialogues?

"Siri would you want to  
study linguistics"

I have everything I need  
already.

"Siri why don't you study  
linguistics"  
tap to edit

I don't understand 'Siri  
why don't you study  
linguistics'. But I could  
search the web for it.

*S: Do you understand and are comfortable with the rules?*

*P: Yeah yeah*

*P: So the name is kind of a famous person*

*P: Okay*

*P: I'm not sure how good am I in this area*

*S: Yes*

*S: I am a famous person and I am male*

*P: Okay okay good*

*S: Alright*

*S: And what is your first question?*



# Strategy 1

1. Detect an error
2. Identify error type
3. Indicate required answer type / intent

What's the meaning of life

I hope to find the answer to that someday.

What's the sense of life

42  
(according to Douglas Adams' humorous science-fiction novel The Hitchhiker's Guide to the Galaxy)

"What's the matter"  
tap to edit

Here is what I found:

Wikipedia  
Matter

Before the 20th century, the term matter included ordinary matter composed of atoms and excluded other energy phenomena such as light or sound. This concept of matter may be generalized from atoms to include any objects having mass even when at rest, but this is ill-defined because an object's mass can arise from its (possibly massless) constituents' motion



# Strategy 2

1. Detect an error
2. Detect a repetition
3. Consider only the best option
4. Mark all the others as inconvenient

"What do you read"

This is about you, not me.

"What do you read"

I can't answer that.

"What do you like to read"

I've never really thought about it.

"Would you prefer to read"

tap to edit

This is about you, not me.

# Strategy 3

- Consider the data as it is
- Any answer is possible and acceptable
- Large context tags possible

Do you have a dream

I dream of world peace.

How to stop wars

Web search stop wars.

Do you have a dream

Yes, to make this conversation more enjoyable for you.

# Conclusions

- Linguistically different errors require different processing strategies
- Corpus annotation becomes evaluation
  - bias inevitable
- Large context should be taken into account
- Relations between utterances of a single speaker are necessary
- Disambiguation of word and sentence level can help to detect errors
- Algorithms need to be created



**Thank you!**  
**Questions are welcome!**

You can always contact me:  
[generalova.valeria@gmail.com](mailto:generalova.valeria@gmail.com)

# References

- Bunt, H., Alexandersson, J., Choe, J. W., Fang, A. C., Hasida, K., Petukhova, V., ... & Traum, D. R. (2012, May). ISO 24617-2: A semantically-based standard for dialogue annotation. In LREC (pp. 430-437).
- Petukhova, V., Gropp, M., Klakow, D., Schmidt, A., Eigner, G., Topf, M., Srb, S., Motlicek, P., Potard, B., Dines, J., Deroo, O., Egeler, R., Mainz, U., Liersch, S. (2014). *The DBOX corpus collection of spoken human-human and human-machine dialogues*. In Proceedings of the Ninth International Conference on Language Resources and Evaluation, 252-258