Speaking Scripts: Sentence as a Set of Instructions

Kuznetsov readings:
Semantic representations in cognitive technologies
Moscow, September 2018

Ivan Rygaev
Laboratory of Computational Linguistics
Institute for Information Transmission Problems, Russian Academy of Sciences
irygaev@gmail.com

Classical Semantics

- Meaning of a sentence is a logical proposition
  - It has a truth-value
- Can be represented in different ways
  - Predicate logic expression
  - Semantic graph
  - etc.
- Static model
  - Better suitable for the representation of the knowledge in the mind rather than the semantics of the sentence

Formal Semantics

- Linguistic “wars” in 1940s
  - What should be studied – ordinary or formal language?
  - But both parties agreed: ordinary language is not logical
- Richard Montague
  - English as a Formal Language (1970)
  - Ordinary language can be seen/used as a formal one
- However
  - That does not mean that this is the purpose of language and that all the semantics reduces to it

An Example

- Sentences
  - *The writer wrote the novel and burned it*
  - *The writer burned the novel which he had written*
  - *The novel was written and burned by the writer*
  - *Who burned the novel is the writer who had written it*
- They have common propositional representation
  - writer(x), novel(y), write(x, y), burn(x, y)
- But they are not interchangeable in discourse
  - This representation cannot be used as interlingua
  - Hence it is not complete

Information Structure

- Information packaging into a message
  - Goal: communication (information transfer)
  - Presupposes dynamical process
- Usually considered as secondary
  - Belongs to pragmatics, not semantics
  - Modifies the “primary” syntactic structure, which exists independently

My Goal

- Rehabilitate information structure
  - It is between semantics and syntax (not outside)
  - Reflects the speaker intentions
  - Creates the syntactic structure from a semantic graph (not just modifies an existing one)
- Formula
  - Speaker knowledge + communicative intention + beliefs about the hearer knowledge -> a syntactic tree
  - The information structure forms the future syntactic tree on the semantic level

Information Categories

- Given/New
  - ‘New’ contains new information for the hearer
  - ‘Given’ contains the information which is already known
- Why is ‘Given’ included into the sentence?
  - To link new information to the existing one in the hearer’s mind
- *The cat is sleeping*
  - New: sleep(x)
  - Given: cat(x)
- The purpose of ‘Given’ is to find a mental file

Mental Database

- There are mental representations of the objects:
  - Which we perceive
  - Which we are told about
  - Which we infer
- The speaker does not have access to the hearer MDB
  - But he has reasonable believes about its content
  - He cannot directly activate a referent in the hearer MDB or take its ID to include into the sentence
  - He can provide only some descriptive info that the hearer can use as a pattern to search in her mental database (Given)

Hearer State Evaluation

- Goal:
  - To communicate that a certain person X arrived
- Strategy:
  - The person's name can be used (*John arrived*), but only if the hearer knows the name (according to the speaker)
  - Otherwise another expression must be used:
  - *That guy which we saw yesterday arrived*
  - Both sentences convey the same message
  - But for different hearers
  - The referring expression must be suitable for the hearer to identify the referent in context

Dynamic Semantics

- File Change Semantics (Heim 1982), DRT (Kamp 1981)
  - Definite descriptions serve to find existing mental referents
  - Indefinite descriptions – to create new ones
- *Peter built a house*
  - find x: named(x, 'Peter')
    create y: house(y)
    update: build(x, y)

x
named(x, 'Peter')
build(x, y)

y
house(y)
build(x, y)

Topic

- Several definite descriptions
  - *Peter built the house*
  - *The house was built by Peter*
- Common structure?
  - find x: named(x, ‘Peter’)
    find y: house(y)
    update: build(x, y)
- Topic referent is captured in the update instruction:
  - update x: build(x, y)
  - update y: build(x, y)

Discourse Representation Theory

- DRT already has this distinction (Van Der Sandt 1992)
  - Preliminary DRSs contain separate presupposition sections which are resolved and removed in Proper DRSs
  - These sections directly correspond to search instructions in my terminology
  - Resolving presuppositions as anaphora involves the search for their antecedents
  - It is Preliminary DRS which constitutes a genuine representation of a sentence
  - While Proper DRS is a natural device for a mental representation

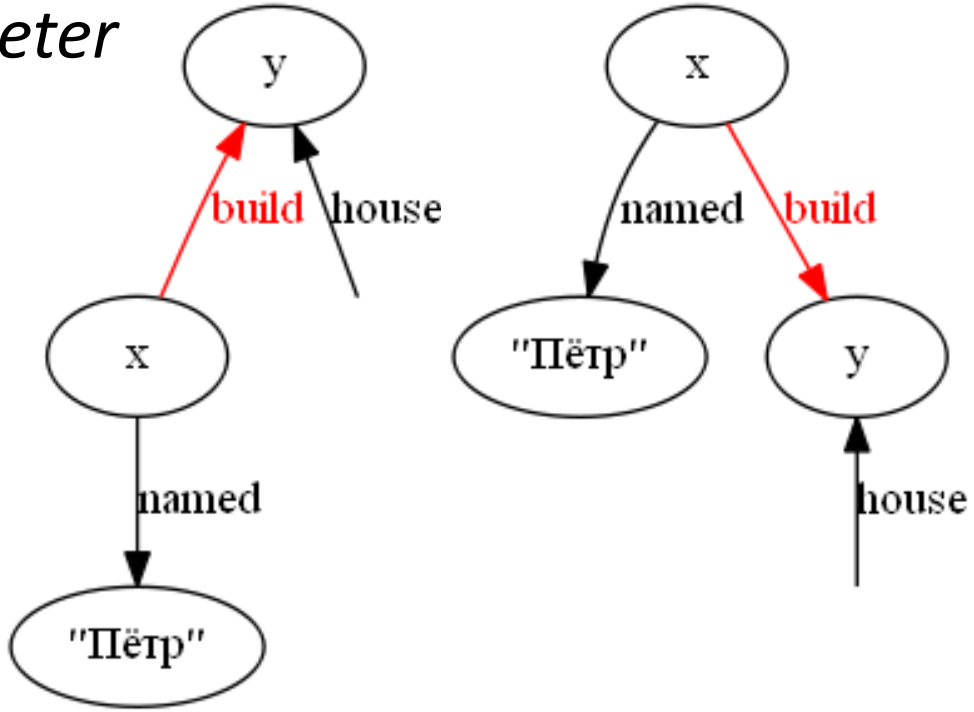
Semantic Graph

- Speaker knowledge:
  - named(x, ‘Peter), build(x, y), house(y)
- Semantic graph:

  - Nodes correspond to referents (variables)
  - Arcs correspond to predicates
- Not a tree-like structure
  - Topic defines the head
  - The tree appears automatically

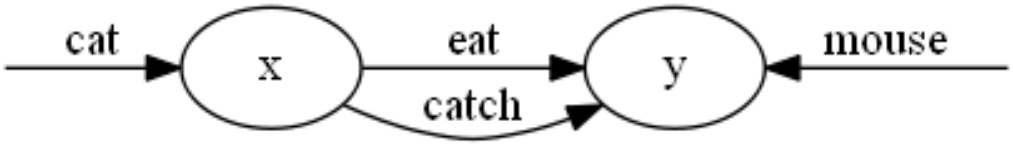
Building Trees

- The update instruction defines the head
  - Arrow direction shows semantic dependencies
  - Nodes position – information (and syntactic) dependencies
- *The house was built by Peter*
  - update y: build(x, y)
    find y: house(y)
    find x: named(x, ‘Peter’)
- *Peter built the house*
  - update x: build(x, y)
    find x: named(x, ‘Peter’)
    find y: house(y)



Loops

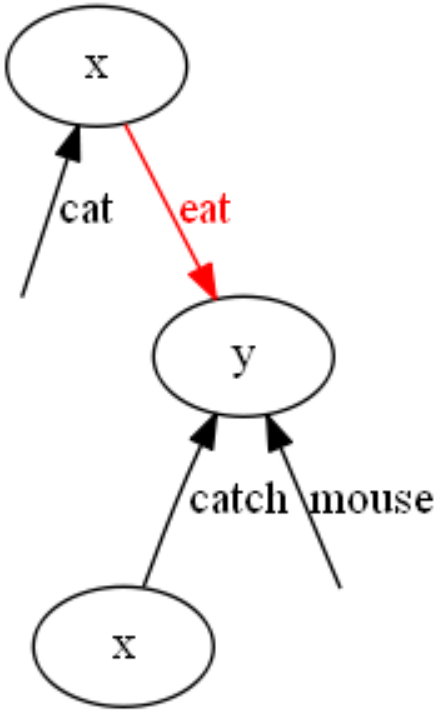
- More complex phrases:
  - *The cat is eating the mouse which it caught*
  - *The cat caught the mouse which it is eating*
- Speaker knowledge:
  - cat(x), eat(x, y), catch(x, y), mouse(y)
- Semantic graph:



  - Splitting into instructions helps to remove the loops (by duplication of the nodes)

Removing Loops

- *The cat is eating the mouse which it caught*
  - update x: eat(x, y)
  - find x: cat(x)
  - find y: mouse(y), catch(x, y)
- Constituents structure
  - It is arcs (predicates) which are linguistically realized
  - Empty nodes – pronouns
  - [[cat$_i$]] [eat [[mouse] [[PRO$_i$] catch]]]
  - [the [cat$_i$]] [is eating [the [mouse] [[it$_i$]caught]]]



Conclusions

- Semantic structures:
  - Representation of the sentence itself and mental representation it creates in the mind are different things
  - While the latter can be seen as a proposition
  - The former is a sequence of instructions
  - A text is a script to create data in the hearer MDB
- Instructions:
  - find – finds the existing mental file of the referent
  - create – creates a new mental file for the referent
  - update – updates the referent mental file

Conclusions

- Information structure:
  - Reflects the speaker intention to add certain information to the corresponding hearer's mental file
  - Plays a key role in the creation of the syntactic structure of the sentence
  - The speaker intention leads to a sequence of instructions
  - The sequence of instruction automatically builds a tree out of a semantic graph
  - Information categories Given/New and Topic/Comment are reflected in the instruction types

Thank you for you attention!
Questions?