# Information Structure and Anaphoric Links – A Case Study and Probe

Eva Hajičová, Anna Nedoluzhko
Charles University in Prague

- *“Semantics (as the study of meaning) is central to the study of communication. (p. ix) … The final category of meaning … is thematic meaning, or what is communicated by the way in which a speaker or writer organizes the message in terms of ordering, focus, and emphasis*.” (p.22) (Leech 1974)

- Communication – anaphoric links
- Semantics – information structure

# Topic-Focus Articulation

- Topic-focus articulation (information structure): reflects the cognitively based ‚given' – ‚new' strategy  BUT belongs to the systems of individual languages rather than to the domain of cognition
- the description of TFA is an integral part of  the representation of the (literal) meaning of the sentence (Sgall et al. 1986) = TFA is semantically relevant
- a declarative sentence asserts that its Focus holds about its Topic:

FOCUS (TOPIC)

# Primary distinction: contextual boundness

- The primary opposition for the description of TFA: the distinction of contextually bound (*cb*) and non-bound (*nb*) items.
- "Contextual boundness" should not be understood in a straightforward etymological way: an *nb* element may be 'known' in a cognitive sense (from the context or on the basis of background knowledge) but structured as non-bound, 'non-identifiable':
  - John called Mary a Republican.
  - Then she insulted HIM. X Then she INSULTED him.

# Contextual Boundness (TFA)

a specific TFA attribute in the Prague Dependency Treebank
values:

- **t** for a non-contrastive contextually bound node
- **c** for a contrastive contextually bound node
- **f** for a contextually non-bound node

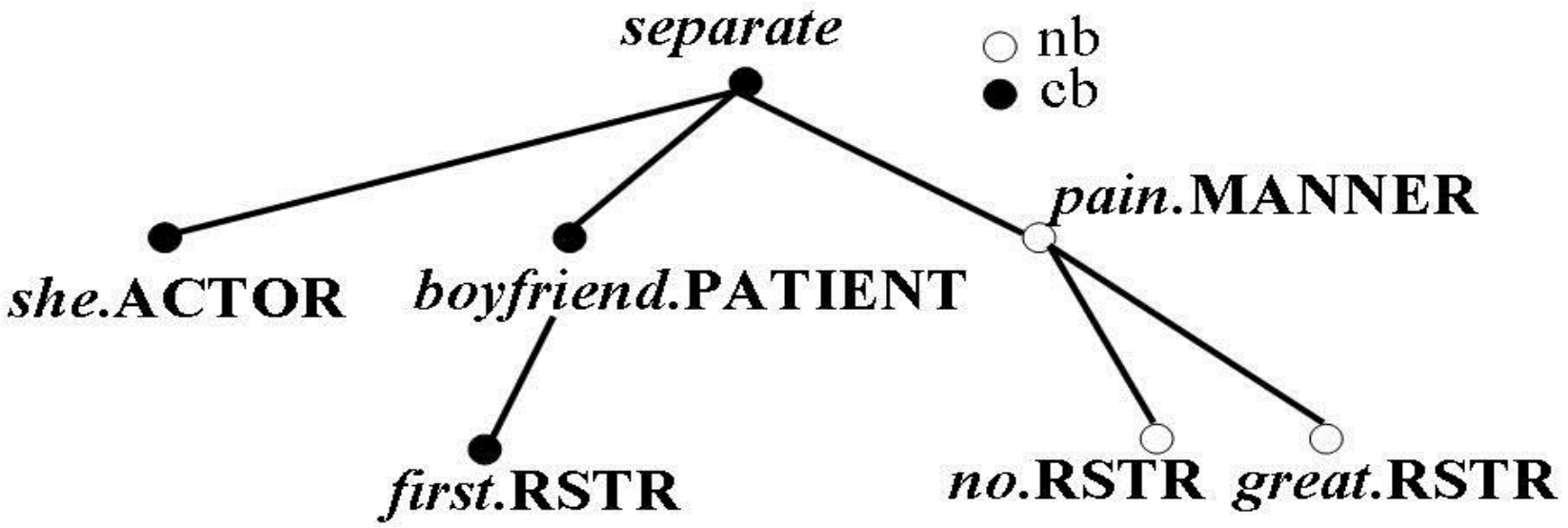
(Preceding context: Irena had two boyfriends and separated from both of them).

*She had separated from her first boyfriend with no great pain*.

The dichotomy of Topic and Focus on the basis of the values of this attribute:

TOPIC*: She had separated from her first boyfriend*

FOCUS: *with no great pain*


separate
nb
cb
she.ACTOR
boyfriend.PATIENT
pain.MANNER
first.RSTR
no.RSTR
great.RSTR

# Anaphoric links

- coreference

  *Irena was in love. She knew that.*

- bridging relations

  *I looked into the room. The ceiling was very high*.

- reference to a text segment

  *Mary suggested to visit John's mother. She also proposed to go swimming. Her last wish was to look at the city center. John denied it all.*

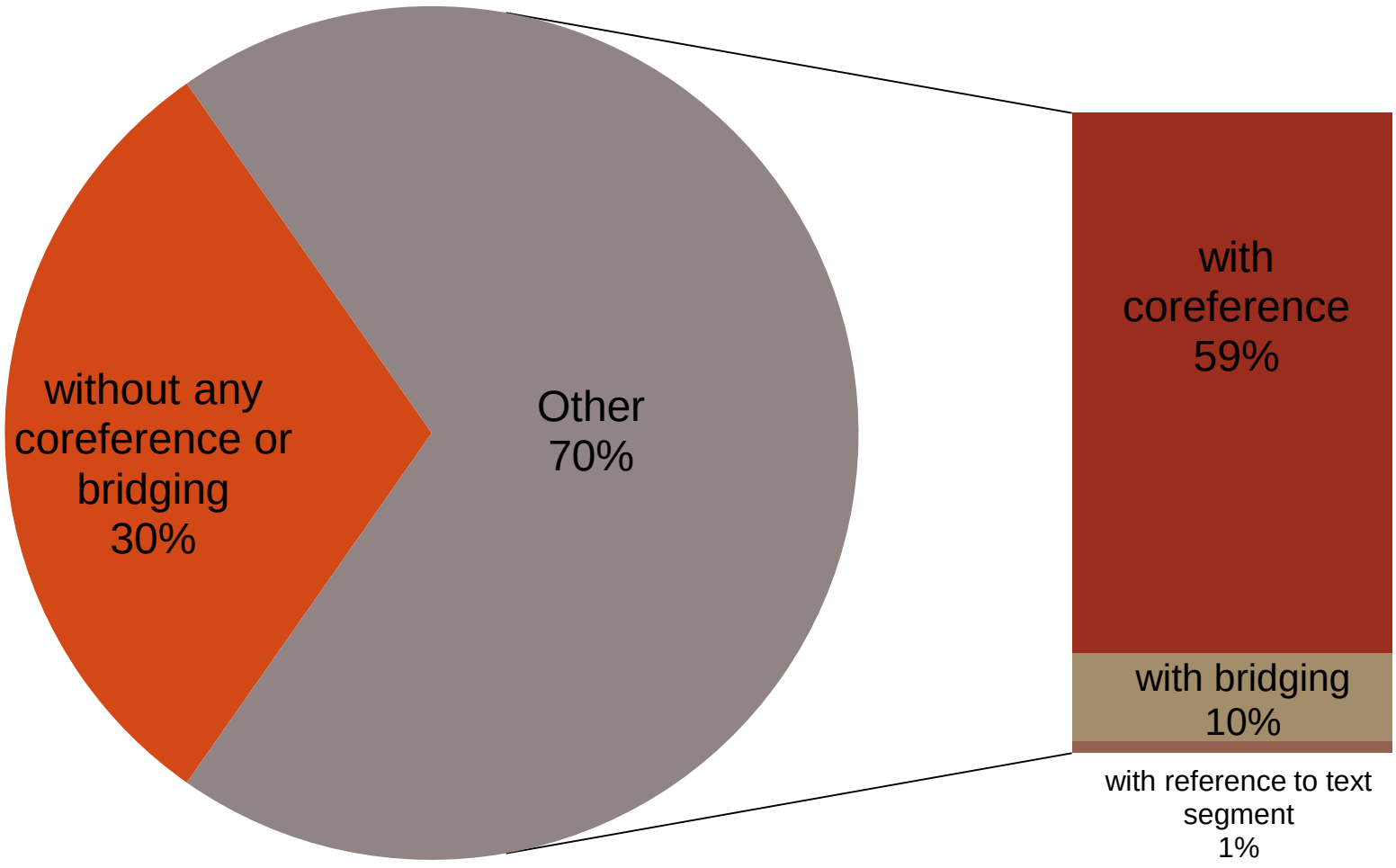
# Contextual Boundness and Anaphoric Links

*Irena had two boyfriends and separated from both of them*.

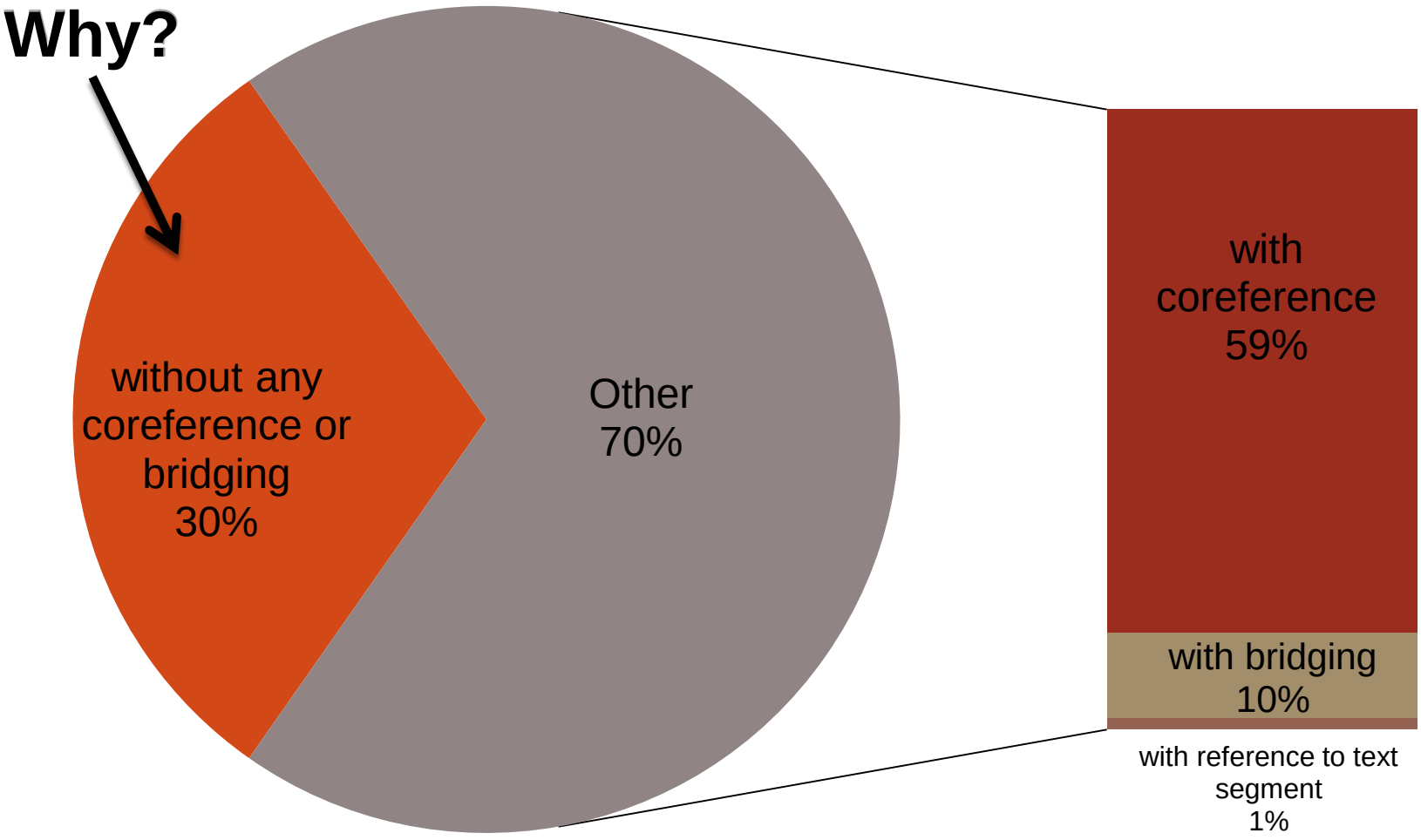
bridging relation:
set – element of the set

*She had separated from her first boyfriend*$_{cb}$ *with no great pain*.

# Contextually Bound Nouns



# Contextually Bound Nouns



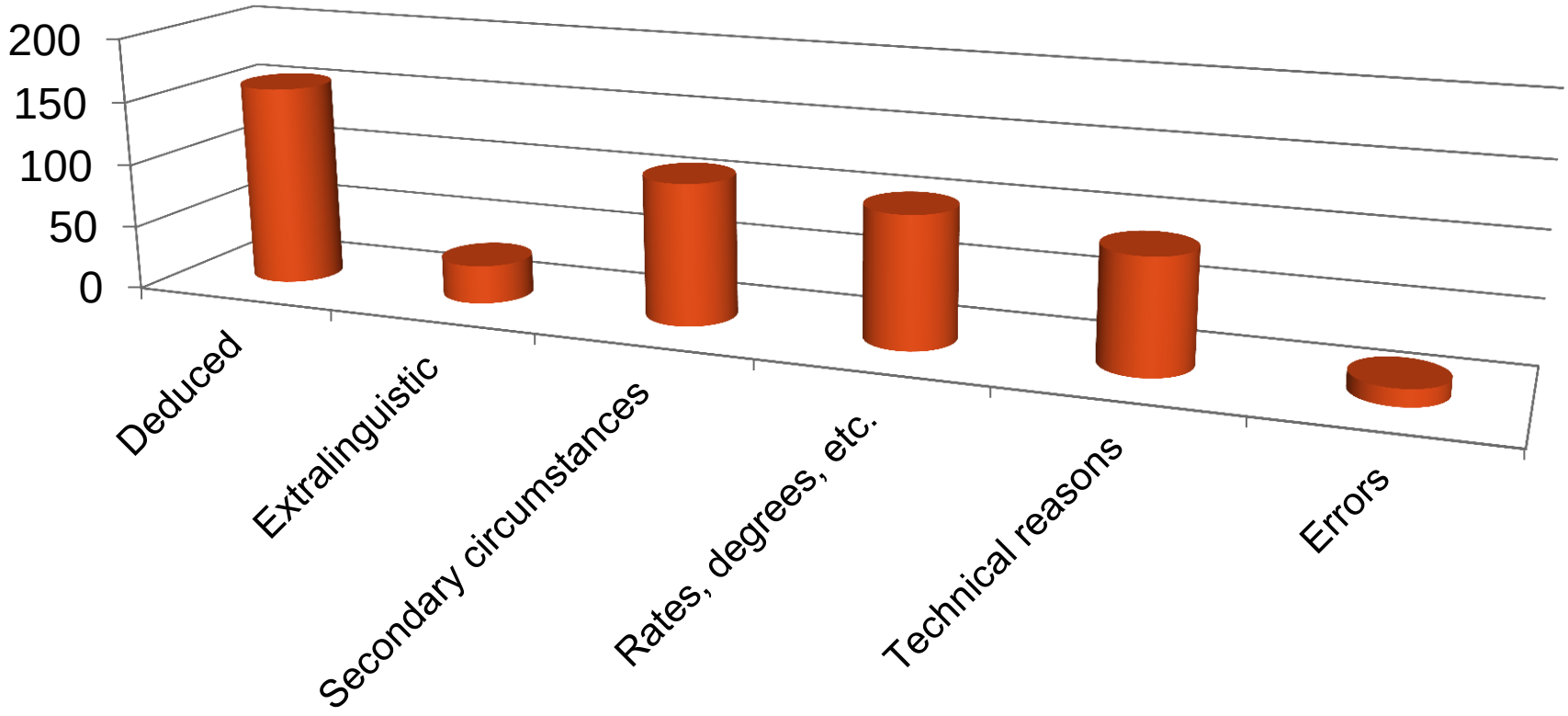
# Analyzed Data

- 500 cases from PDT
  - not elided, present on surface
  - study of possible reasons
  - distribution between reasons is not sharp

# Reasons of contextual boundness without anaphoric links

- Contexual boundness is deduced from some kind of semantic or pragmatic relation to the previous context
- Contextual boundness of a noun group has extralinguistic reasons - world knowledge
- Contextually bound expressions represent rates, degrees, scales, proportions, etc.
- Technical reasons
- Errors

# Reasons of contextual boundness without anaphoric links



# Reasons (1): Contexual boundness is deduced from some kind of semantic or pragmatic relation to the previous context

*According to this law, the insurance companies, which do not have the authorization, are not allowed to tender their products here.*

*Yet it seems that the best hope for alleviating the waves is the stormy weather that forecasters expect from today.*

# Reasons (1): Summary

- Contextual boundness is deduced from previous context from relations close to bridging relations (types *SUBSET*, *CONTRAST*, *ANAPH*, etc.); however, these relations have not been annotated in PDT because of their vague nature and high ambiguity in the given context;
- Contextual boundness is deduced from the relation of co-hyponymy in a broad sense;
- Contextual boundness is deduced from general topic of the text;
- Contextual boundness is given by anaphoric adjectives;
- Contextual boundness is part of formal characteristics (addresses, sport lists, etc.)

# Reasons (2) Contextual boundness of a noun group has extralinguistic reasons

*Complete information for small investors in Lidove noviny are at two sites.*

the reader has the paper in his hands when reading this sentence

*(The) double page was prepared by Jaromír Složil*.

the same, can be used with a definite article or without

# Reasons(3): A noun group refers to „scene“ settings

temporal setting

*In December last year, a detached workplace of the future institute appeared in the Musical Theatre in Berlin.*

local setting

*Another trend, which is even prescribed by the construction or reconstruction of administration buildings in some countries, is structured cabling*.

# Reasons (4): Contextually bound expressions represent rates, degrees, scales, proportions, etc.

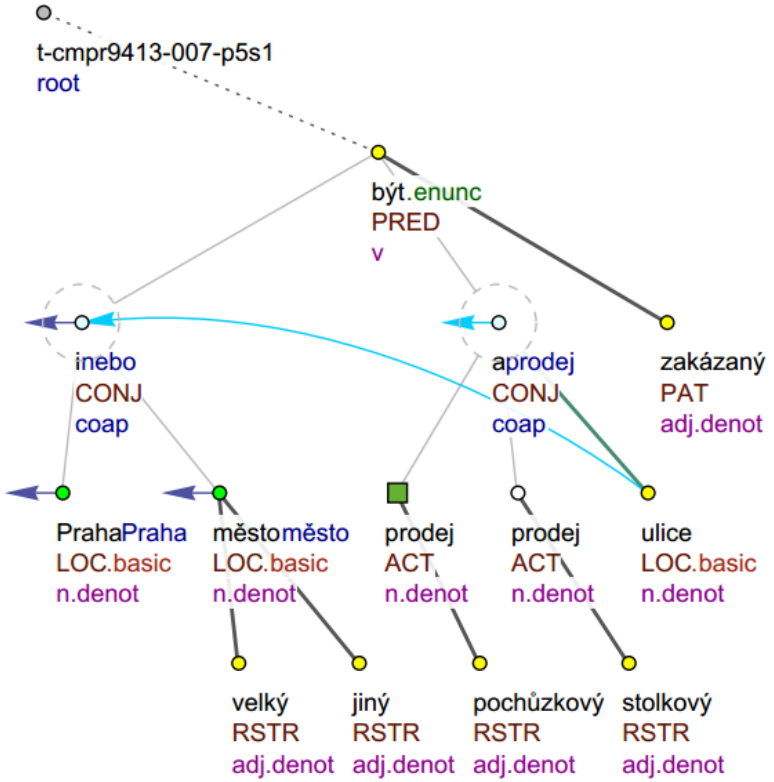
- *According to the Czech Statistical Office, the largest amount of production was achieved in wheat, which was estimated to 3,718 million tons.*
- *The Indonesia's minimum daily wage is one and a half dollar and sometimes workers must work during 10 to 12 hours.*
- *During the same period, the employment at the airport increased from 1,250 to 2,200 people*

# Reasons (5): Technical reasons

- reference – yes, link – no
- due to
  - conventions based on tectogrammatical structure of PDT trees (reconstruction of ellipses, expressions with some tectogrammatical functors)
  - multiword expressions (including constructions with functional words).

anaphoricity of unlinked expressions may be deduced

# Reasons (5): Technical reasons




*V Praze i v jiných velkých městech je pochůzkový [prodej] a stolkový prodej na ulicích zakázaný.*
*[In Prague and other big cities, walk [selling] and table selling on the streets is prohibited.]*

# Summary

- Largest group: (i) association given by semantic and pragmatic reasons
- Next largest group: (ii) world knowledge and broader extralinguistic context
- Technicalities: mainly in coordinate structures: reconstruction of deletions
- Lessons learned for future work: re-examination of annotation of coreferential links for groups (i) and (ii)

THANK YOU FOR YOUR ATTENTION

QUESTIONS?