

# Interplay of Coreference and Discourse

Research and Annotations

*Anna Nedoluzhko*  
*Charles University, Prague*

# Outline

- Corpora Annotated with Coreference and Discourse Relations
- Coreference to Discourse Transitions
  - structural
  - diachronic
- Examples of Coreference ↔ Discourse Interplay (not only) from Prague Dependency Treebank
- Other Dimensions (TFA, Saliency)
- What Does It Mean? (= Conclusions)

# Coreference + Discourse

Cohesion  
Coherence  
(Halliday & Hasan  
1976 ...)

coreference

discourse

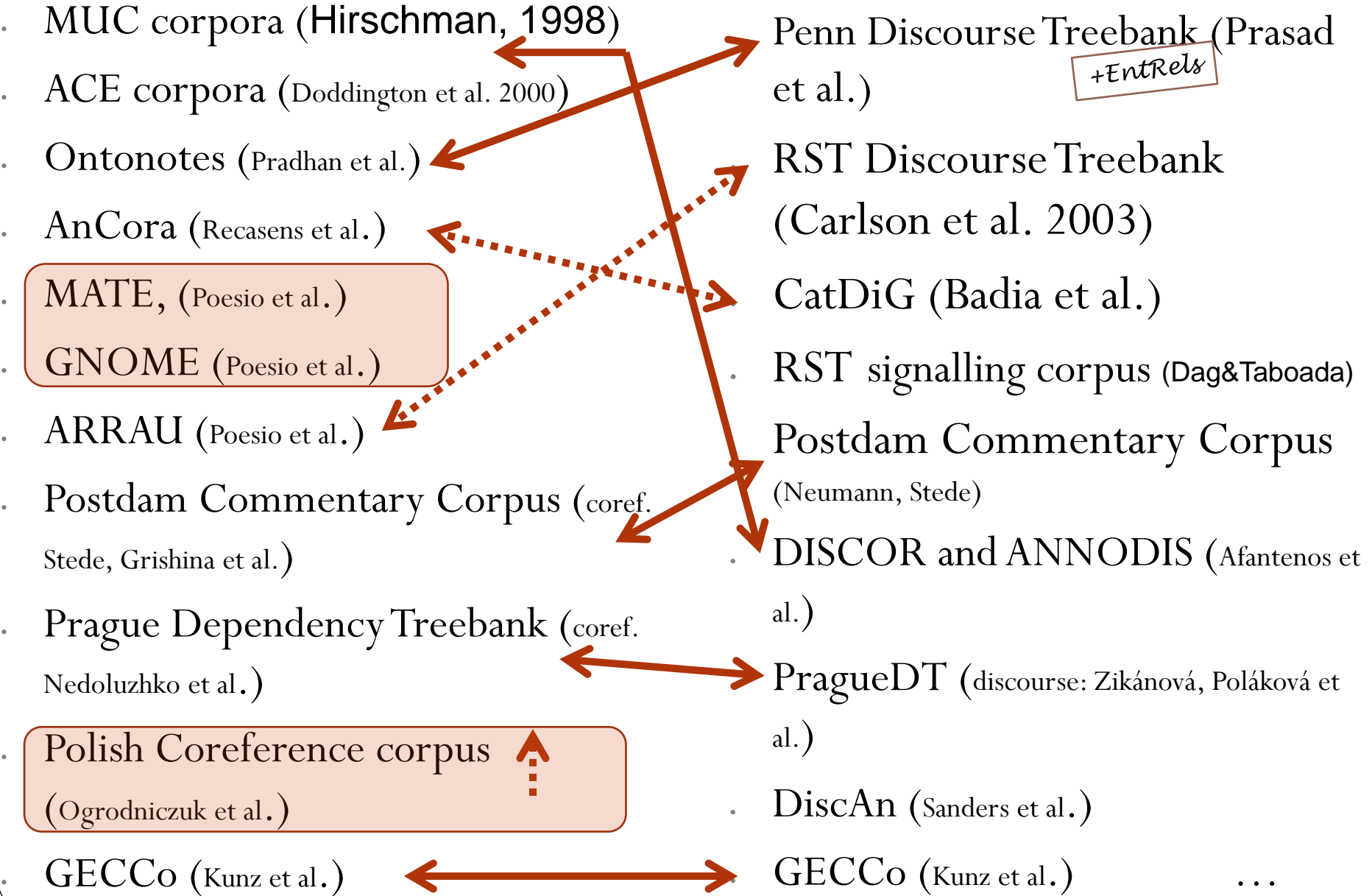


Maybe together again?  
(many dreams, incl.

Poesio et al. 2016)

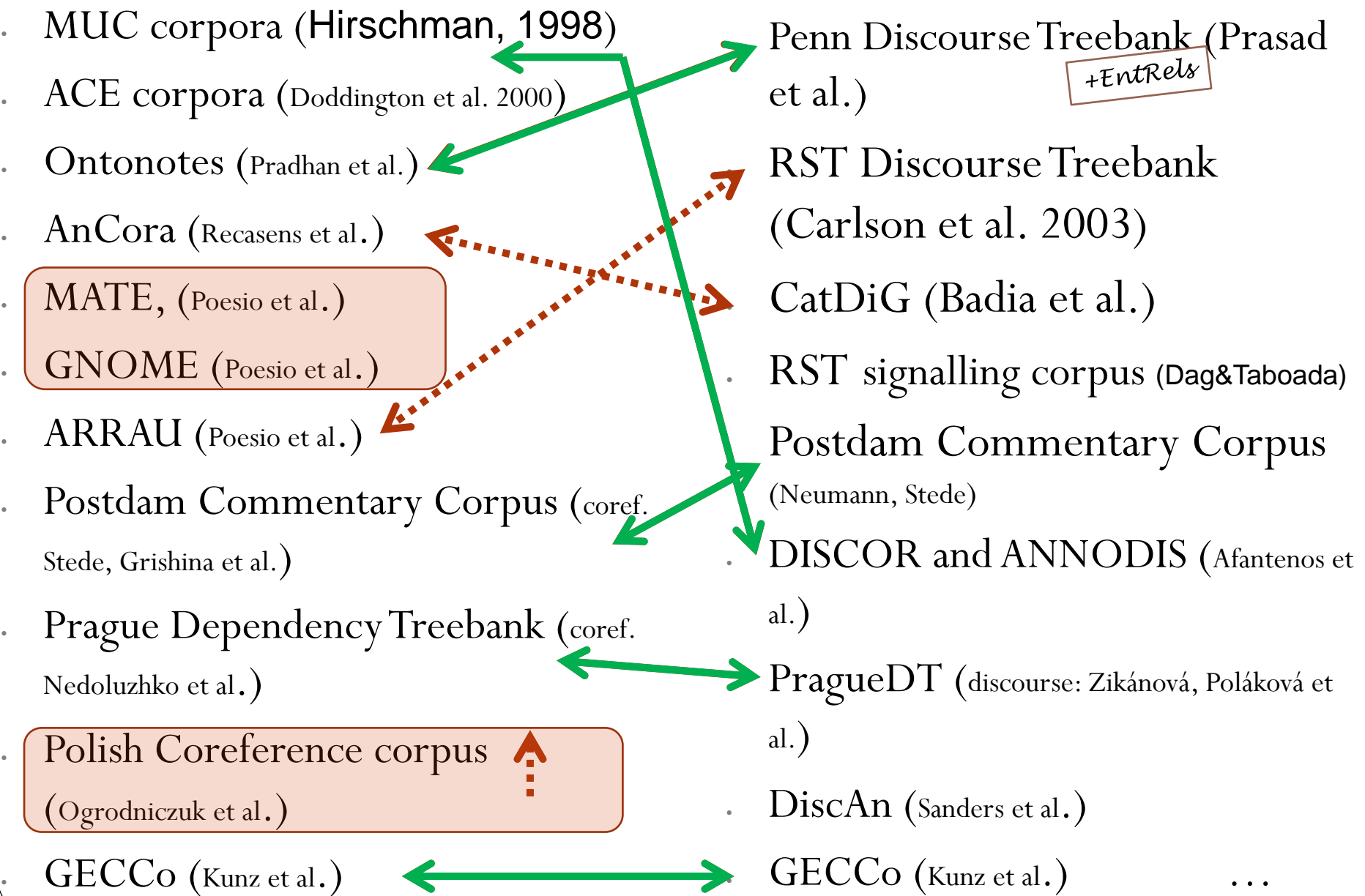
Coreference annotation

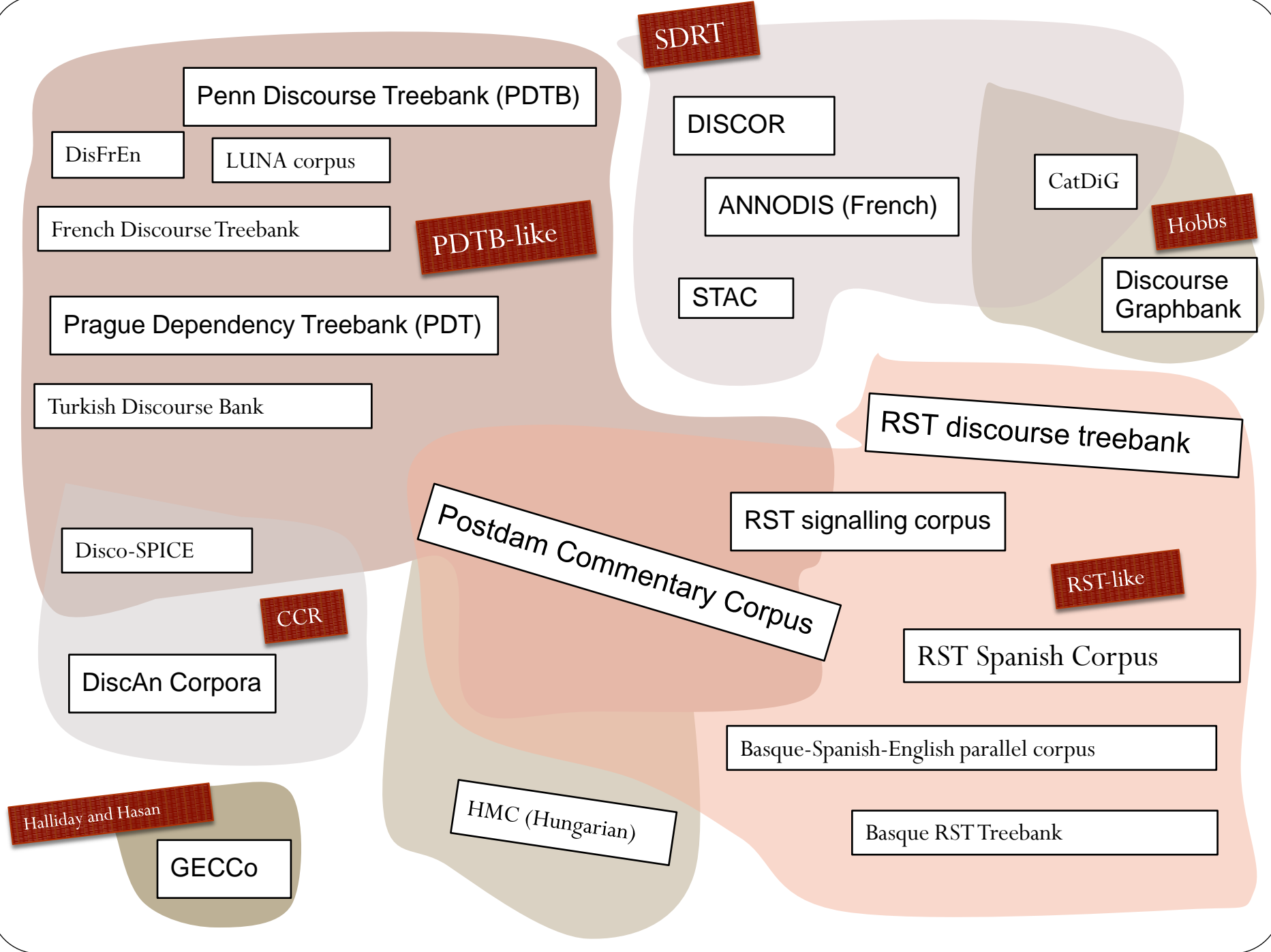
Discourse annotation



# Coreference annotation

# Discourse annotation





Penn Discourse Treebank (PDTB)

DisFrEn

LUNA corpus

French Discourse Treebank

PDTB-like

Prague Dependency Treebank (PDT)

Turkish Discourse Bank

Disco-SPICE

CCR

DiscAn Corpora

Halliday and Hasan

GECCo

Postdam Commentary Corpus

HMC (Hungarian)

SDRT

DISCOR

ANNODIS (French)

STAC

Hobbs

CatDiG

Discourse Graphbank

RST discourse treebank

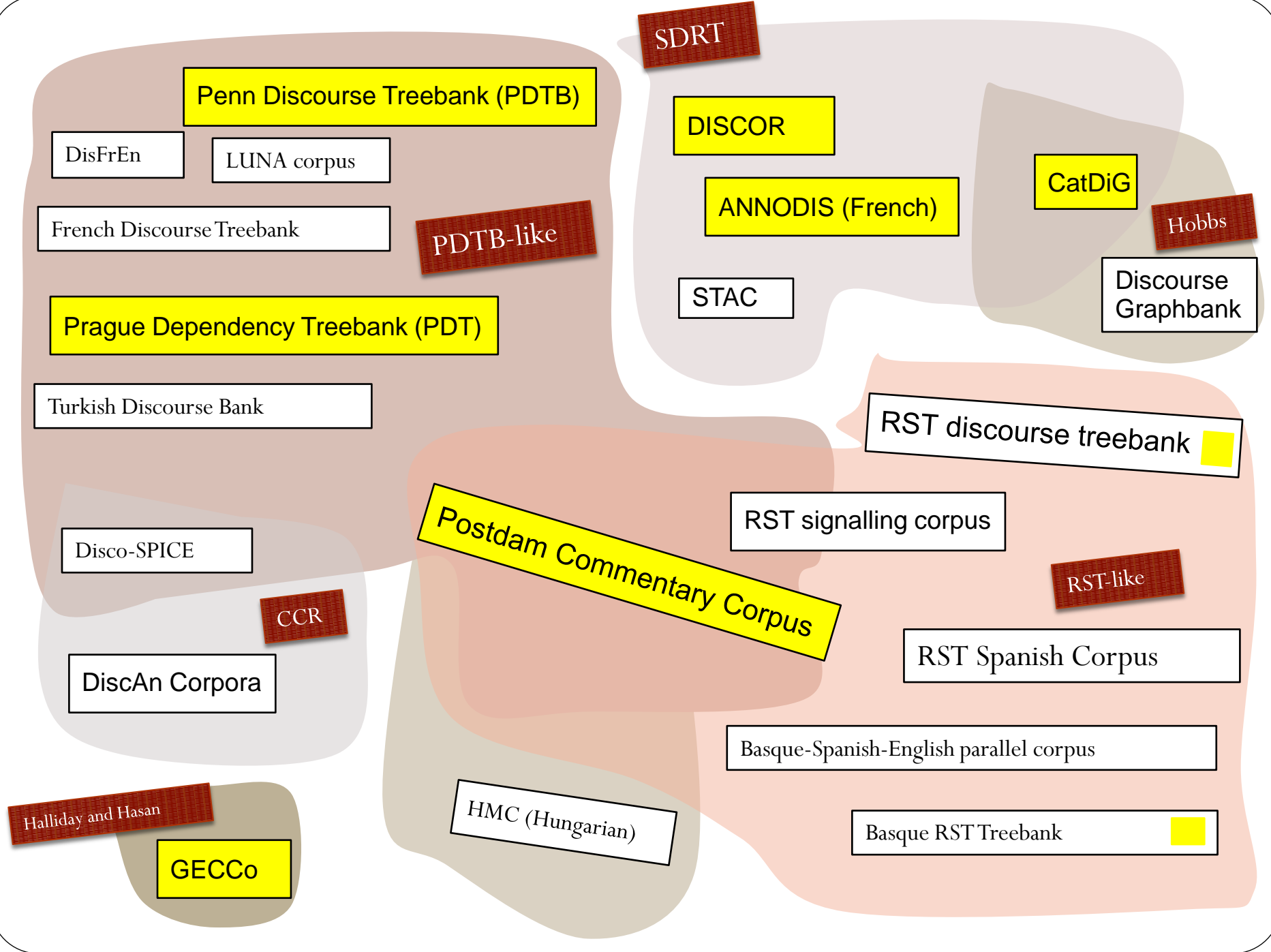
RST signalling corpus

RST-like

RST Spanish Corpus

Basque-Spanish-English parallel corpus

Basque RST Treebank



Penn Discourse Treebank (PDTB)

DisFrEn

LUNA corpus

French Discourse Treebank

PDTB-like

Prague Dependency Treebank (PDT)

Turkish Discourse Bank

Disco-SPICE

CCR

DiscAn Corpora

Halliday and Hasan

GECCo

SDRT

DISCOR

ANNODIS (French)

STAC

CatDiG

Hobbs

Discourse Graphbank

Postdam Commentary Corpus

RST discourse treebank

RST signalling corpus

RST-like

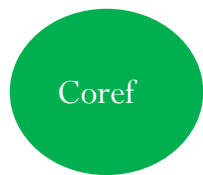
RST Spanish Corpus

Basque-Spanish-English parallel corpus

HMC (Hungarian)

Basque RST Treebank

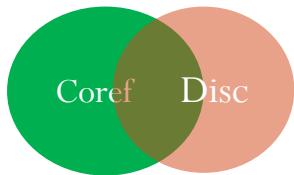
# Coreference + Discourse Relations



entity-based of discourse-based coherence

(co-existing)

- coreference  $\leftrightarrow$  logico-semantic relations
- bridging relations  $\leftrightarrow$  logico-semantic relations

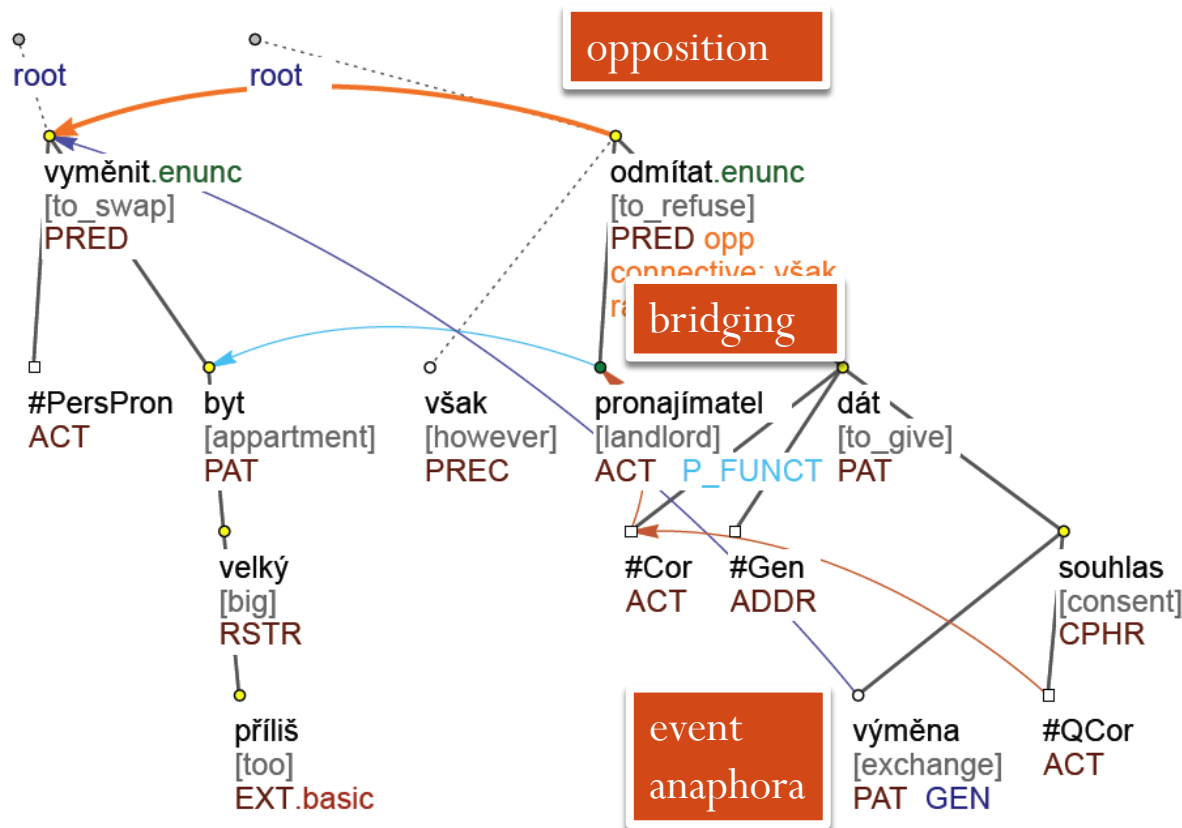


(overlapping)

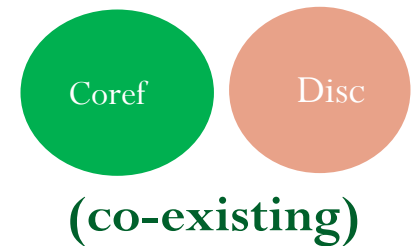
- event anaphora (discourse deixis, abstract coreference...)
- anaphoric connectives (en:*that's why, because of that*, de:*deswegen, demzufolge*, cz:*kvůli tomu*)
- contrast (cz:*zatímco* [ $\approx$  while])
- ?attribution (other anaphoric rules)



# Bridging, Coreference and Discourse



from PDT



*Chtěl bych vyměnit příliš velký byt. Pronajímátele však odmítá dát k výměně souhlas.*

≈ lit. en. *I'd like to swap an apartment that is too big. **However**, the landlord refused to give consent to **the exchange**.*

# Coreference and Discourse: co-existing

from PDT

*Další možností, jak snížit cenu bytů na volném trhu, je posílit nabídku o byty z **trhu regulovaného**. Tento trh s byty je naprosto zdeformovaný.*

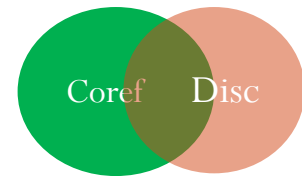
*≈ en. Another possibility to reduce the housing costs on the market, is to increase the supply by apartments from **the regulated market**. This housing market is totally distorted.*

(implicit) opposition



coreference

# Coreference → Discourse: Event anaphora



*Ted* arrived late. *He* wanted to irritate Mary.

.....



*Ted arrived late. This* irritated Mary.



*Ted arrived late. His late arrival* irritated Mary.



*Ted arrived late. Mary was really irritated that he was late.*



*Ted arrived late. It was impossible for him to come on time.*

elaboration

explanation

specification

explication

equivalence

reformulation

exemplification

generalization

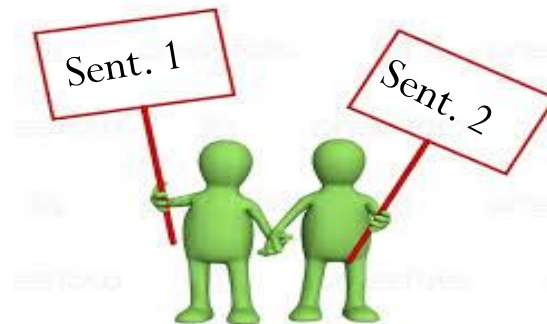
# Event anaphora + discourse

from PDT

- (1) *Loňské léto bylo v Uherském Hradišti dost bouřlivé [...].*
- (2) ***Juran musel budovat tým za pochodu v rozeběhnuté soutěži...***
- (3) *Já to považuji za výhodu.*
- (4) *Když přijde trenér do hotového mužstva, musí vycházet z hráčů, které má, a tomu přizpůsobit hru.*

- (1) *Last summer in Uherske Hradiste was quite stormy.*
- (2) ***Juran had to build a team during the competition.***
- (3) *I consider **this** to be an advantage.*
- (4) *When a new trainer comes to an already completed team, he must base the strategy on the players he has, and to adapt the game to them.*

explication



coreference

# Anaphoric Connectives

coreference

Ted was a great cook. Mary loved him for *this*.



Ted was a great cook. *This* is why Mary loved him.



Ted was a great cook. *Therefore* she loved him.

*that's why*



Ted was a great cook. *So* she loved him.

*thus*

discourse

German *trotz dem* → *trotzdem*, Czech *kvůli tomu, přestože*

# Conclusion

- coreference-discourse interplay is richer in discourse
- discourse perspective is ‘deeper’
- different weight of coreference relations for textual coherence
- ‘important’ coreference is not accidental (thematic progressions, information structure, ...)
- there are standard transfers from coreference to discourse (*for this – therefore, event anaphora – coreference between events*)
- other topics (salience, information structure, ellipses, attribution)

# References

- S.Botley, et al. (eds.) 1999. Corpus-based and Computational Approaches to Discourse Anaphora
  - Laurence Danlos, Bertrand Gaiffe. Event coreference and discourse relations. L. Kulda. Language, Music and Cognition, Kluwer Academic Publishers, 2004.
  - M.A.K. Halliday & Ruqaiya Hasan. Cohesion in English. 1976
  - E. Hovy, M. Marcus, M. Palmer, L. Ramshaw, and R. Weischedel. 2006. Ontonotes: the 90% solution. In NAACL-HLT.
  - Lapshinova-Koltunski, E. and Kunz, K. (2014). Annotating cohesion for multilingual analysis. LREC.
  - Louis Annie, Aravind Joshi, Rashmi Prasad, Ani Nenkova. (2010) Using entity features to classify implicit discourse relations.
  - M. Poesio et al. Anaphora Resolution. Algorithms, Resources, and Applications. 2016.
  - Brian Reese, Julie Hunter, Nicholas Asher, Pascal Denis and Jason Baldridge. Reference Manual for the Analysis and Annotation of Rhetorical Structure. 2007 (DISCOR)
  - B. Webber, A. Joshi, M. Stone, A. Knott. Anaphora and Discourse structure. ACL, 1994
- ... and many others

# Acknowledgements

We acknowledge the support from the Grant Agency of the Czech Republic (grant 16-05394S). This work has been using language resources developed and stored and distributed by the LINDAT/CLARIN project of the Ministry of Education, Youth and Sports of the Czech Republic (project\_LM2015071).



# Thank you!



coreference

discourse