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, Salience)
- What Does It Mean? (= Conclusions)

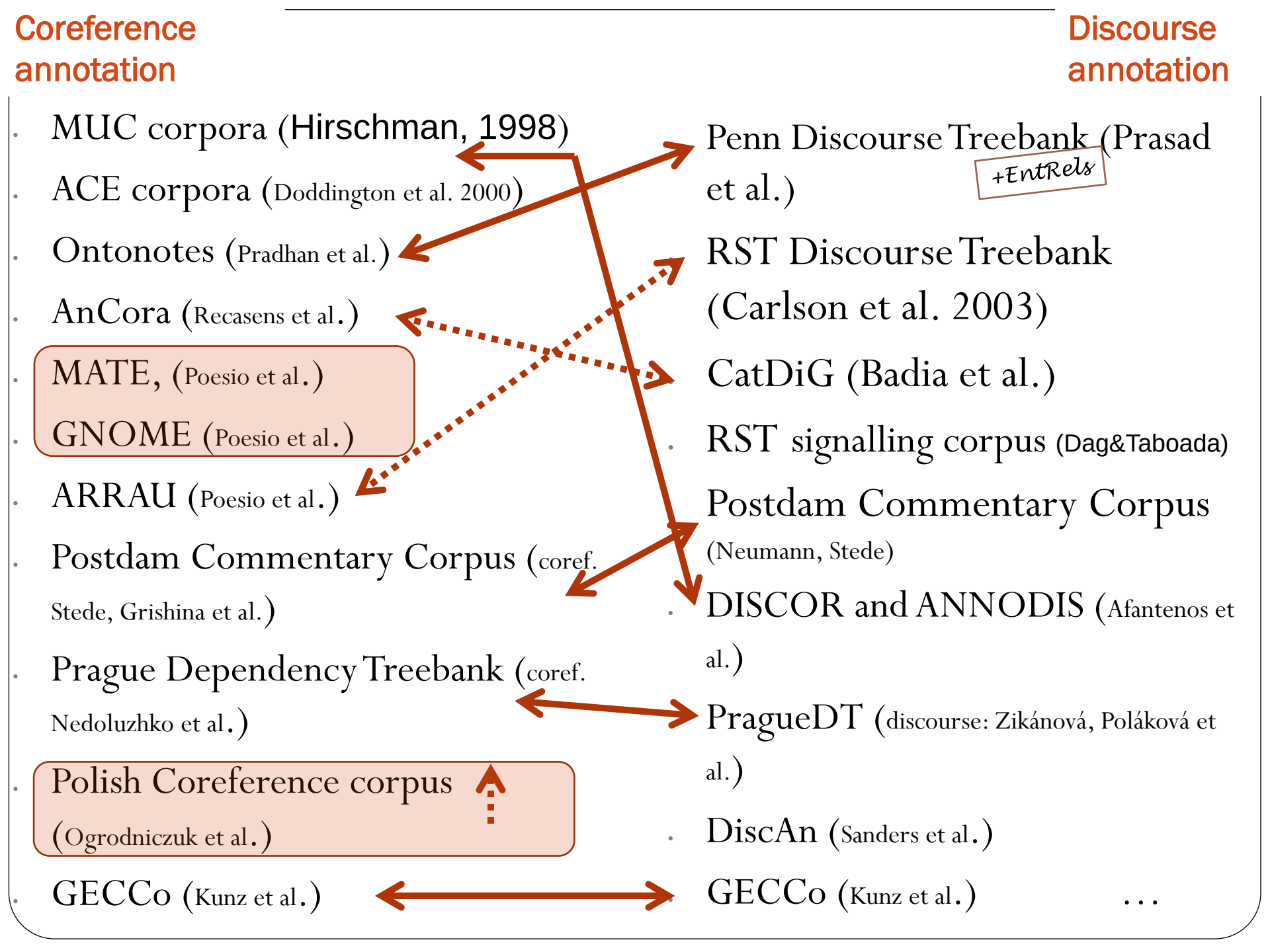
Coreference + Discourse

Cohesion
Coherence
(Halliday&Hasan 1976 …)

coreference

discourse

Maybe together again?
(many dreams, incl.
Poesio et al. 2016)


Coreference annotation
Discourse annotation
MUC corpora (Hirschman, 1998)
ACE corpora (Doddington et al. 2000)
Ontonotes (Pradhan et al.)
AnCora (Recasens et al.)
MATE, (Poesio et al.)
GNOME (Poesio et al.)
ARRAU (Poesio et al.)
Postdam Commentary Corpus (coref. Stede, Grishina et al.)
Prague Dependency Treebank (coref. Nedoluzhko et al.)
Polish Coreference corpus (Ogrodniczuk et al.)
GECCo (Kunz et al.)
Penn Discourse Treebank (Prasad et al.)
+EntRels
RST Discourse Treebank (Carlson et al. 2003)
CatDiG (Badia et al.)
RST signalling corpus (Dag&Taboada)
Postdam Commentary Corpus (Neumann, Stede)
DISCOR and ANNODIS (Afantenos et al.)
PragueDT (discourse: Zikánová, Poláková et al.)
DiscAn (Sanders et al.)
GECCo (Kunz et al.)
...

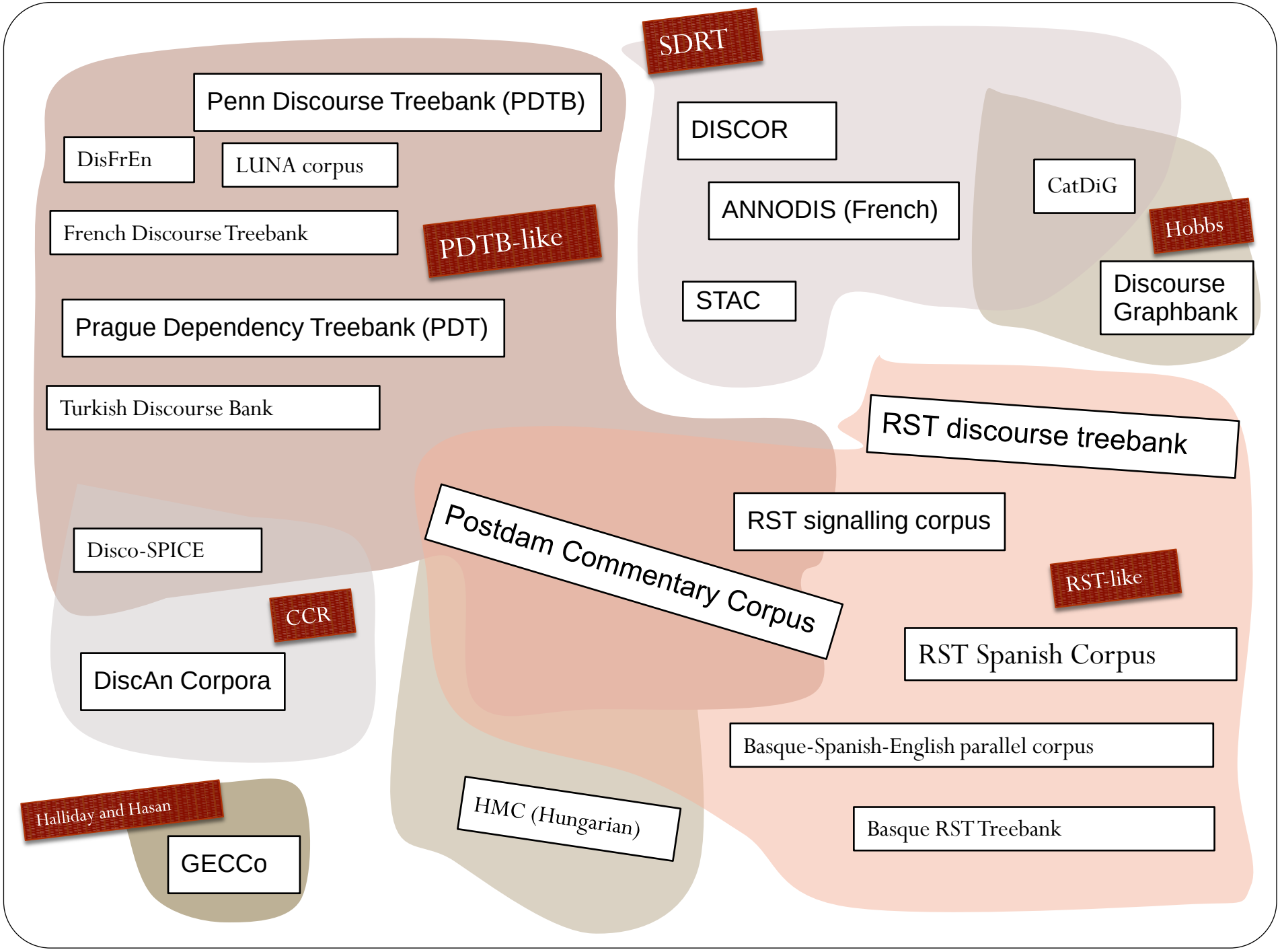
Coreference annotation

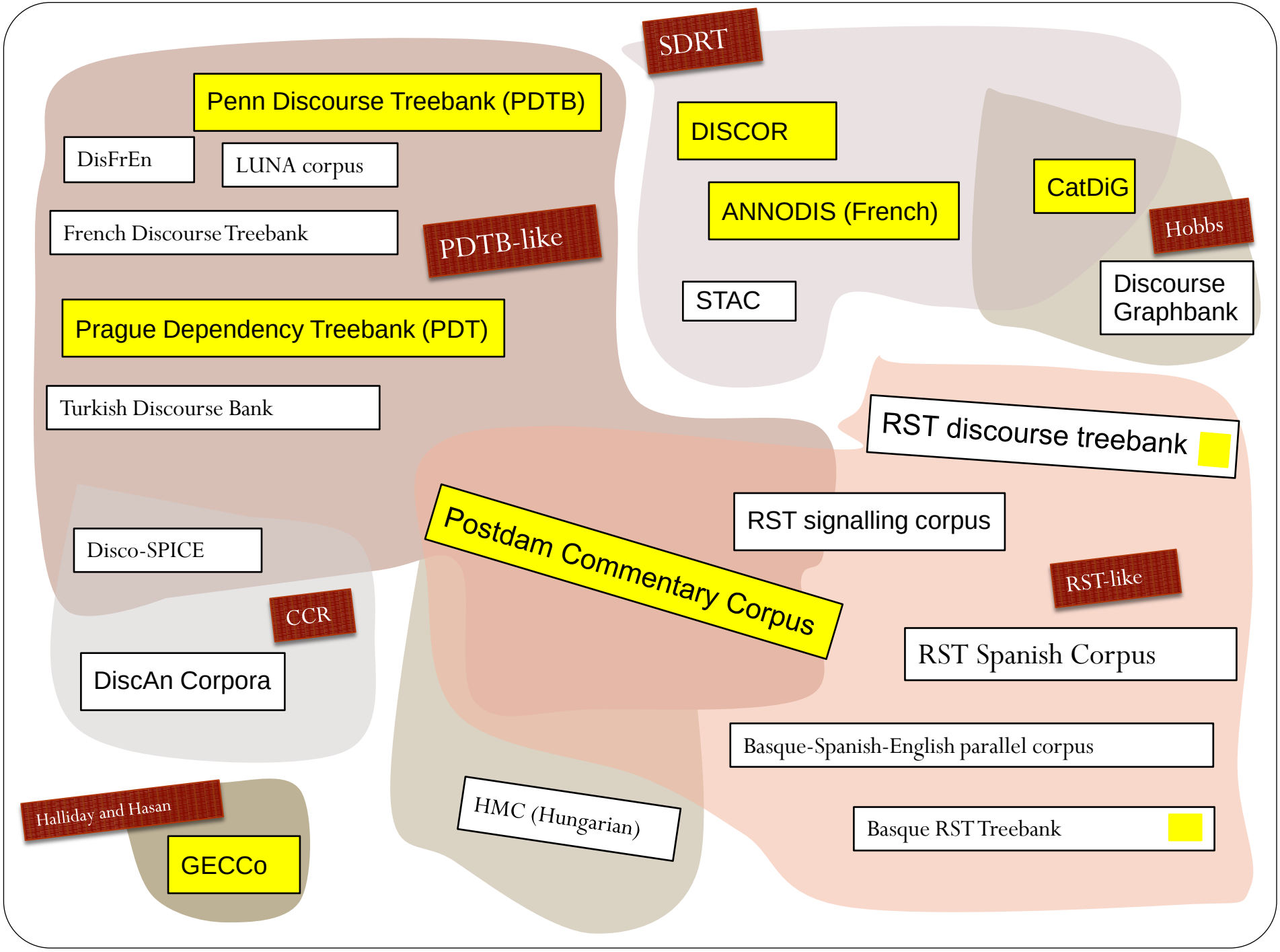
- MUC corpora (Hirschman, 1998)
- ACE corpora (Doddington et al. 2000)
- Ontonotes (Pradhan et al.)
- AnCora (Recasens et al.)
- MATE, (Poesio et al.)
- GNOME (Poesio et al.)
- ARRAU (Poesio et al.)
- Postdam Commentary Corpus (coref. Stede, Grishina et al.)
- Prague Dependency Treebank (coref. Nedoluzhko et al.)
- Polish Coreference corpus (Ogrodniczuk et al.)
- GECCo (Kunz et al.)

Discourse annotation

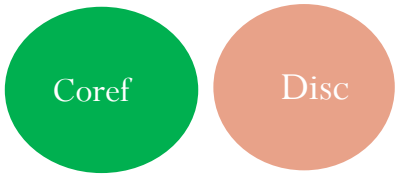
- Penn Discourse Treebank (Prasad et al.) +EntRels
- RST Discourse Treebank (Carlson et al. 2003)
- CatDiG (Badia et al.)
- RST signalling corpus (Dag&Taboada)
- Postdam Commentary Corpus (Neumann, Stede)
- DISCOR and ANNODIS (Afantenos et al.)
- PragueDT (discourse: Zikánová, Poláková et al.)
- DiscAn (Sanders et al.)
- GECCo (Kunz et al.)

...


SDRT
Penn Discourse Treebank (PDTB)
DISCOR
DisFrEn
LUNA corpus
CatDiG
ANNODIS (French)
Hobbs
French Discourse Treebank
PDTB-like
Discourse Graphbank
STAC
Prague Dependency Treebank (PDT)
Turkish Discourse Bank
RST discourse treebank
RST signalling corpus
Disco-SPICE
Postdam Commentary Corpus
RST-like
CCR
RST Spanish Corpus
DiscAn Corpora
Basque-Spanish-English parallel corpus
Halliday and Hasan
HMC (Hungarian)
Basque RST Treebank
GECCo


SDRT
Penn Discourse Treebank (PDTB)
DISCOR
DisFrEn
LUNA corpus
CatDiG
ANNODIS (French)
Hobbs
French Discourse Treebank
PDTB-like
Discourse Graphbank
STAC
Prague Dependency Treebank (PDT)
Turkish Discourse Bank
RST discourse treebank
Postdam Commentary Corpus
RST signalling corpus
Disco-SPICE
RST-like
CCR
RST Spanish Corpus
DiscAn Corpora
Basque-Spanish-English parallel corpus
HMC (Hungarian)
Halliday and Hasan
Basque RST Treebank
GECCo

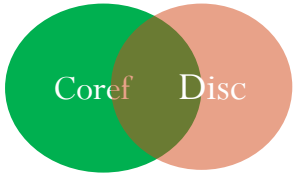
Coreference + Discourse Relations



(co-existing)

entity-based of discourse-based coherence

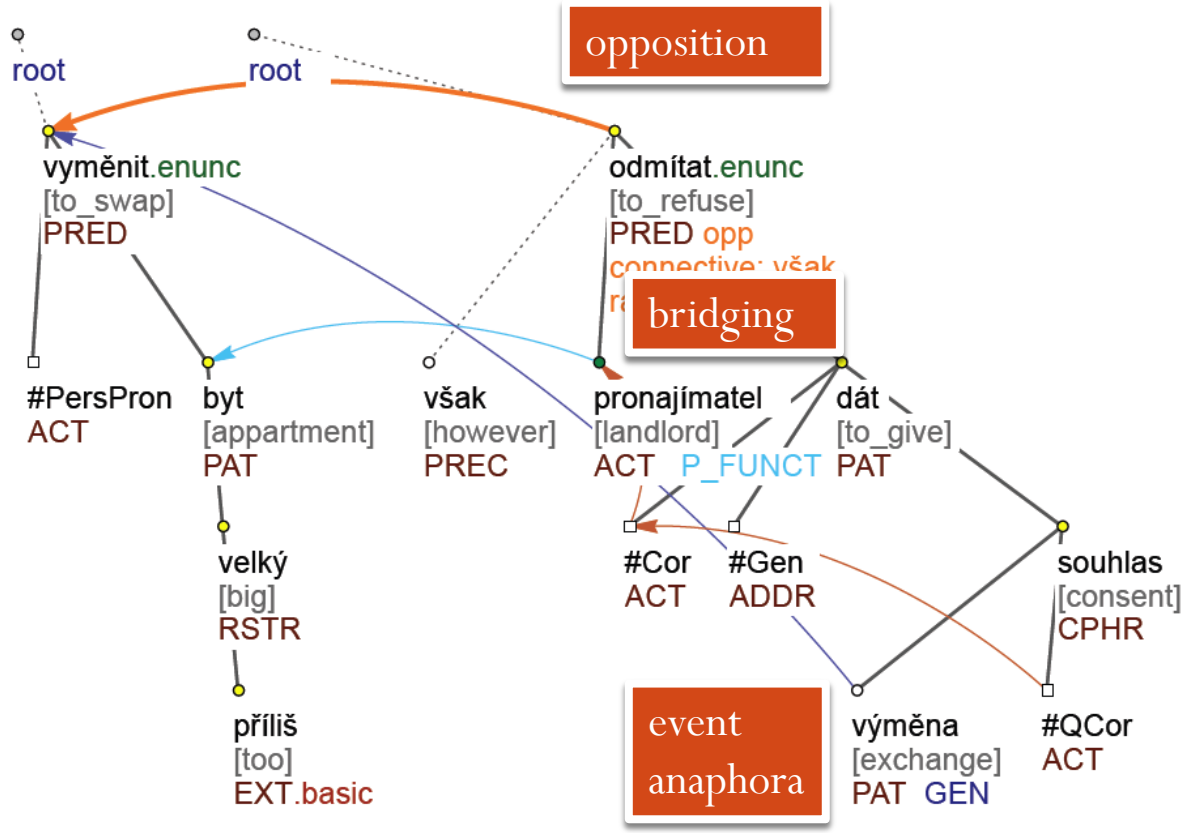
- coreference ↔ logico-semantic relations
- bridging relations ↔ logico-semantic relations



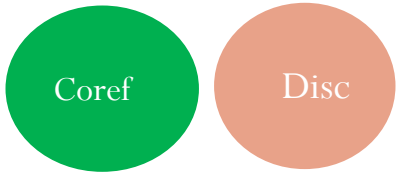
(overlapping)

- event anaphora (discourse deixis, abstract coreference…)
- anaphoric connectives (en:***that****'s why*, *because of* ***that***, de:***des****wegen*, ***dem****zufolge*, cz:*kvůli* ***tomu***)
- contrast (cz:*zatímco* [≈ while])
- ?attribution (other anaphoric rules)

Bridging, Coreference and Discourse



from PDT



(co-existing)

Chtěl bych vyměnit příliš velký byt. Pronajímatel 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

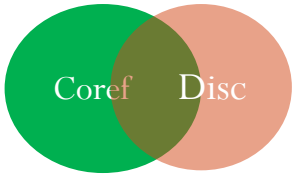
Další možností, jak snížit cenu bytů na volném trhu, je posílit nabídku o byty z ***trhu regulovaného***. ***Tento trh s byty*** *je naprosto zdeformovaný.*

≈ 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

explication

specification

equivalence

reformulation

examplification

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****.*

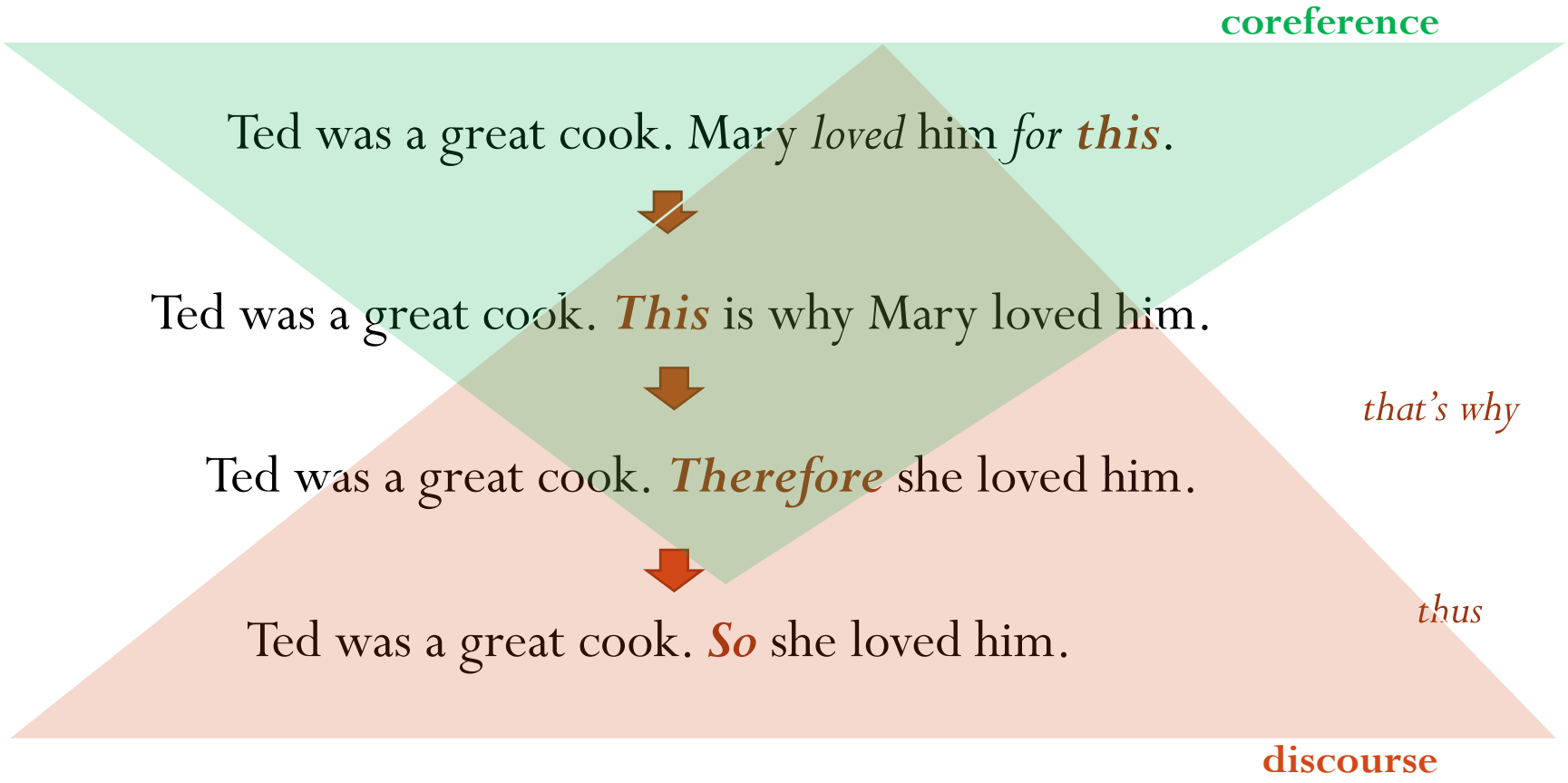
(3) I consider ***this*** *to be an advantage.*

(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



German *trotz dem* → *trotzdem*, Czech *kvůli* ***tomu***, *přes**to**že*

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 (saliences, 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 multillingual 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