

*Discourse*

**Annotation**

*tomorrow* 😊

**Theories and ~~Tools~~**

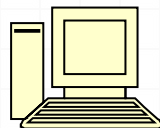
**Corpora**

*Anja Nedoluzhko*



# What today's lecture is **NOT** about

- 0 computational processing of discourse, discourse parsing of any type
- 0 event annotation (Martha Palmer, James Pustejovsky et al.)
- 0 various forms of connectives (DRDs) in different languages (other classes)
- 0 just a little about spoken corpora



*weil = because*  
*da = because*  
*wo = where*  
*wenn = when*  
*wie = how*  
*dass = that*  
*obwohl =*  
*although*  
*der/die/das =*  
*who*





# What today's lecture **IS** about

- I. Very generally about discourse-related topics
- II. Discourse theories (H&H, RST, SDRT, PDTB, DGB, CCR) and corpora annotated with discourse relations



Overview of discourse theories will be made from the point of view of relations and conceptions (not connectives, they will be addressed within other courses)

- III. Comparison of theories and approaches

# I. General words about discourse-related topics

- 0 large/small discourse units,
- 0 coreference,
- 0 cohesion/coherence,
- 0 spoken/written,
- 0 elementary discourse units,
- 0 requirements to discourse structure

## Processing Large Discourse Units

- Logical document structure

*a hierarchy of segments of the document, each of which corresponds to a visually distinguished semantic component of the document (Summers 1998)*



**BUNDESGERICHTSHOF**  
IM NAMEN DES VOLKES

**URTEIL**

VI ZR 101/06

Verkündet am:  
27. März 2007  
Holmes,  
Justizangestellte  
als Urkundsbeamtin  
der Geschäftsstelle


in dem Rechtsstreit

Nachschlagewerk: ja  
BGHZ: nein  
BGHR: ja

BGB § 823 Abs. 1, G, § 1004, StGB § 185, TMG § 10

Ein Unterlassungsanspruch wegen eines in ein Meinungsforum im Internet eingestellten ehrverletzenden Beitrags kann auch dann gegen den Betreiber des Forums gegeben sein, wenn dem Verletzten die Identität des Autors bekannt ist.

## Discourse Processing



**BUNDESGERICHTSHOF**  
IM NAMEN DES VOLKES

**URTEIL**

VI ZR 101/06

Verkündet am:  
27. März 2007  
Holmes,  
Justizangestellte  
als Urkundsbeamtin  
der Geschäftsstelle

in dem Rechtsstreit

Nachschlagewerk: ja  
BGHZ: nein  
BGHR: ja

BGB § 823 Abs. 1, G, § 1004, StGB § 185, TMG § 10

Ein Unterlassungsanspruch wegen eines in ein Meinungsforum im Internet eingestellten ehrverletzenden Beitrags kann auch dann gegen den Betreiber des Forums gegeben sein, wenn dem Verletzten die Identität des Autors bekannt ist.

## Processing Large Discourse Units

- Logical document structure

*a hierarchy of segments of the document, each of which corresponds to a visually distinguished semantic component of the document (Summers 1998)*

## Discourse Processing

### Analysis of DRD-related Contrasts in Spoken Czech, English and German

Kerstin Anna Kunz\*, Ekaterina Lapshinova-Koltunski\*\*, Anna Nedoluzhko\*\*\*

\*University of Heidelberg, \*\*Saarland University, \*\*\*Charles University in Prague  
kerstin.kunz@ucl.uni-heidelberg.de, e.lapshinova@mx.uni-saarland.de, nedoluzko@ufal.mff.cuni.cz

The aim of the present paper is to analyse contrasts in Czech, English and German in terms of discourse-relational devices (DRDs). The novelty of our approach lies in the nature of the resources we are using. Advantage is taken of existing resources, which are, however, annotated on the basis of two different frameworks. We use an interoperable scheme unifying DRDs of both frameworks in more abstract categories and considering only those phenomena that have a direct match in the three languages.

Our aim is two-fold: On the one hand, we intend to compare similarities and differences between spoken and written discourse regarding the use of DRDs and test the interoperable scheme that was originally designed for the analysis of written discourse, as described in Lapshinova et al. (2015). This will help us to check if this scheme is sufficient, or if additional classes of discourse-relational devices (DRDs) have to be added. On the other hand, we are interested in the contrasts existing between typologically close (English vs. German) and typologically more distant (Czech vs. German/English) languages.

Our assumption is that the conditions of spoken language production influence the creation of discourse-relational devices. This has already been postulated for lexicogrammar in various works (e.g., Miller, J. & R. Weinert (2009) and shown by corpus-based works comparing English and German (Amoia et al. 2012, for coreference, and Kunz & Lapshinova 2014, for several cohesive types). For instance, the number of semantically vague devices should be relatively high, because of mutual presence of the speech participants, immediacy and spontaneity of the communication. In addition, the number of devices used should be higher, due to reduced short-term memory capacity and revisions. Although the conditions of spoken language production should find their reflex in all three languages under investigation, we expect some features of spoken discourse to be language-specific.

Quantitative contrastive analyses on the level of discourse require annotated corpora involving time-consuming compilation and annotation, especially in a multilingual context. Therefore, we have decided to take advantage of the existing resources reflecting systemic peculiarities and realisational options of the languages under analysis. We use Czech, English and German data, annotated on the basis of two different frameworks: Functional Generative Description, as described in Sgall

interviews available in both corpora. Although not being clearly representative spoken data (and not capturing prosodic information), these will allow us to get first results on the differences not only between languages, but also between different genres (written and spoken), and thus provide insights for future analysis.

The scheme used for the analysis includes relations of contingency, contrast, expansion and temporal relations. In the Czech data, these categories are further classified into subcategories, e.g., purpose, explication, semantic and pragmatic reason-result and condition for contingency relations. The scheme for German and English contains annotations of the general categories only. However, it also integrates modal adverbs (such as *well, certainly or of course*) which although not connecting two propositions directly, play an important role for cohesion in spoken language.

In our presentation, we provide more information on our hypotheses, the resources and the scheme applied, as well as the results of our analysis.

### References

- Amoia, M., K. Kunz and E. Lapshinova-Koltunski. 2012. Coreference in Spoken vs. Written Text: a Corpus-based Analysis. Proceedings of the IREC2012, Istanbul.
- Halliday, M.A.K. and Ruqaiya Hasan. 1976. Cohesion in English. Longman, London, New York.
- Lapshinova, E., A. Nedoluzhko, and K. Kunz. 2015. Across languages and genres: Creating a universal annotation scheme for textual registers. In Ines Rehbein and Heike Zinsmeister, editors, Proceedings of the Workshop on Linguistic Annotations, NAACL-2015, Denver, USA.
- Kunz, K. and E. Lapshinova-Koltunski. 2015. Cross-linguistic analysis of discourse variation across registers. Ajmer, K. and Hassegard, H. (eds.). Nordic Journal of English Studies, 14(1), pp. 258-288.
- Miller, J. & R. Weinert. 2009. Spontaneous Spoken Language: Syntax and Discourse. Oxford: OUP.
- Sgall, P., E. Hajičová, and J. Panevová. 1986. The meaning of the sentence in its semantic and pragmatic aspects. Reidel, Dordrecht.

## Processing Large Discourse Units

### • Logical document structure

*a hierarchy of segments of the document, each of which corresponds to a visually distinguished semantic component of the document (Summers 1998)*

### • Content zones

*a continuous portion of a text document that fulfills a functional role for the text as a whole, contributing to the overall message or purpose, as it is characteristic for the genre of the text (defined for each sentence)*

Teufel and Moens (2002) - aimed at recovering content zones from conference papers in CL.

- Aim, research goal
- Textual: statements about section structure
- Own: description of the authors' work (methodology, results, discussion)
- Background: generally accepted scientific background
- Contrast: comparison with other work
- Basis: statements of agreement with other work
- Other: description of other researchers' work

# Discourse Processing

## Synthesis of pyrazole and pyrimidine Troeger's base-analogues

Rodrigo Abonia, Andrea Albernez, Hector Larrabondo, Jairo Quiroga, Braulio Isuasty, Henry Isuasty, Angelina Hormaza Adolfo Sanchez, and Manuel Noguera

1  
PERKIN

Troeger's base analogues bearing fused pyrazolic or pyrimidinic rings were prepared in acceptable to good yields through the reaction of 3-alkyl-5-amino-arylpyrazoles and 6-aminopyrimidin-4(3H)-ones with formaldehyde under mild conditions (i.e. in ethanol at 50°C in the presence of catalytic amounts of acetic acid). Two key intermediates were isolated from the reaction mixtures, which helped us to suggest a sequence of steps for the formation of the Troeger's bases obtained. The structures of the products were assigned by <sup>1</sup>H and <sup>13</sup>C NMR, mass spectra and elemental analysis and confirmed by X-ray diffraction for one of the obtained compounds.

### Introduction

Although the first Troeger's base 1 was obtained more than a century ago from the reaction of p-toluidine and formaldehyde [1], recently the study of these compounds has gained importance due to their potential applications. They possess a relatively rigid chiral structure which makes them suitable for the development of possible synthetic enzyme and artificial receptor systems [2], chelating and biomimetic systems [3] and transition metal complexes for regio- and stereoselective catalytic reactions [4]. For these reasons, numerous Troeger's base derivatives have been prepared bearing different types of substituents and structures (i.e. 2-5 Scheme 1), with the purpose of increasing their potential applications [2,3,5].



Scheme 1 The original Troeger's base 1 and some interesting derivatives and analogues.

However, some of the above methodologies possess tedious work-up procedures or include relatively strong reaction conditions, such as treatment of the starting materials for several hours with an ethanolic solution of conc. hydrochloric acid or TFA solution, with poor to moderate yields, as is the case for analogues 4 and 5 [5].

Considering these potential applications, we now report a simple synthetic method for the preparation of 5,12-dialkyl-3,10-diaryl-1,3,4,8,10,11-hexaazetetracyclo[6.6.1.0 2.6 .0 9.13]penta-deca-2(6),4,9(13),11-tetraene 8a-e and 4,12-dimethoxy-1,3,5,9,11,13-hexaazetetracyclo[7.7.1.0 2.7.0 10.15 .1 hepta-deca-2(7),3,10(15),11-tetraene-6m14-diones 10a,b based on the reaction of 3-alkyl-5-amino-1-arylpyrazoles 6 and 6-aminopyrimidin-4(3H)-ones 9 with formaldehyde in ethanol and catalytic amounts of acetic acid. Compounds 8 and 10 are new Troeger's base analogues bearing heterocyclic rings instead of the usual phenyl rings in their aromatic parts.

### Results and discussion

In an attempt to prepare the benzotriazolyl derivative 7a, which could be used as an intermediate in the synthesis of new hydroquinolines of interest, [6], a mixture of 5-amino-3-methyl-1-phenylpyrazole 6a, formaldehyde and benzotriazole in 10 ml of ethanol, with catalytic amounts of acetic acid, was heated at 50°C for 5 minutes. A solid precipitated from the solution while it was still hot. However, no consumption of benzotriazole was observed at TLC.

The reaction conditions were modified and the same product was obtained when the reaction was carried out without using benzotriazole, as shown in Scheme 12. On the basis of NMR and mass spectra and X-ray crystallographic analysis we established that the structure is 5,12-dialkyl-3,10-diaryl-1,3,4,8,10,11-hexaazetetracyclo[6.6.1.0 2.6 .0 9.13]penta-deca-2(6),4,9(13),11-tetraene 8, a new pentacyclic Troeger's base analogue.

Co\_Gro    Other    Aim    Gap/Weak    Own\_Mthd    Own\_Res    Own\_Conc

## Processing Large Discourse Units

- Logical document structure

*a hierarchy of segments of the document, each of which corresponds to a visually distinguished semantic component of the document (Summers 1998)*

- Content zones

*a continuous portion of a text document that fulfills a functional role for the text as a whole, contributing to the overall message or purpose, as it is characteristic for the genre of the text (defined for each sentence)*

- Topic-based segmentation

*a sequence of non-overlapping text segments that completely covers the text. Each unit consists of one or more sentences that address a common topic.*

WordNet

... is largely a matter of intuition

... is a costly procedure → no large human annotated corpora available

- connectives,
- word repetition,
- bridging relations, related words, pronouns, ellipsis, lexical chains

## Discourse Processing

[1.7] A man named Lionel Gaedi went to the Port-au-Prince morgue in search of his brother, Josef, but was unable to find his body among the piles of corpses that had been left there. [1.8] "I don't see him—it's a catastrophe," Gaedi said. [1.9] "God gives, God takes." [1.10] Chris Rolling, an American missionary and aid worker, tried to extricate a girl named Jacqueline from a collapsed school using nothing more than a hammer. [1.11] He urged her to be calm and pray, and as night fell he promised that he would return with help. [1.12] When he came back the next morning, Jacqueline was dead. [1.13] "The bodies stopped bothering me after a while, but I think what I will always carry with me is the conversation I had with Jacqueline before I left her," Rolling wrote afterward on his blog. [1.14] "How could I leave someone who was dying, trapped in a building! ... [1.15] She seemed so brave when I left! [1.16] I told her I was going to get help, but I didn't tell her I would be gone until morning. [1.17] I think this is going to trouble me for a long time." [1.18] Dozens of readers wrote to comfort Rolling with the view that his story was evidence of divine wisdom and mercy.

© Manfred Stede / NAACL Tutorial 2013



# Processing Large Discourse Units

- Logical document structure

*a hierarchy of segments of the document, each of which corresponds to a visually distinguished semantic component of the document (Summers 1998)*

- Content zones

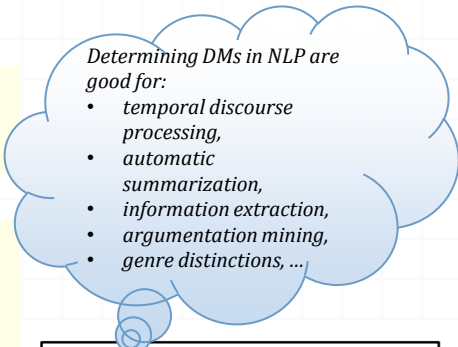
*a continuous portion of a text document that fulfills a functional role for the text as a whole, contributing to the overall message or purpose, as it is characteristic for the genre of the text (defined for each sentence)*

- Topic-based segmentation

*a sequence of non-overlapping text segments that completely covers the text. Each unit consists of one or more sentences that address a common topic*

- Discourse modes (C. Smith, 2003)

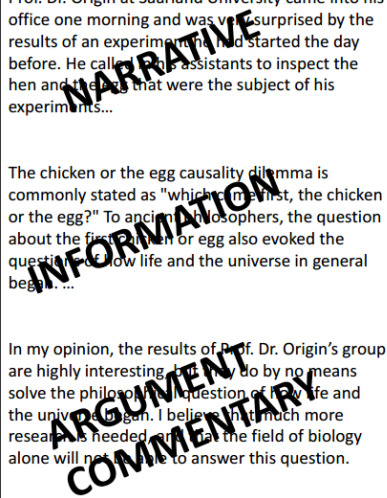
*narrative, information, argument (commentary), report, description*



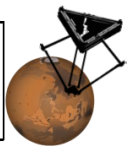
Prof. Dr. Origin at Saarland University came into his office one morning and was very surprised by the results of an experiment he had started the day before. He called his assistants to inspect the hen and the eggs that were the subject of his experiments...

The chicken or the egg causality dilemma is commonly stated as "which came first, the chicken or the egg?" To ancient philosophers, the question about the first chicken or egg also evoked the question of how life and the universe in general began...

In my opinion, the results of Prof. Dr. Origin's group are highly interesting, but they do by no means solve the philosophical question of how life and the universe began. I believe that much more research is needed and that the field of biology alone will not be able to answer this question.



On Monday, NASA **announced** that signs of liquid water **have been found** on Mars. The Mars Reconnaissance Orbiter spacecraft **found** evidence of the liquid on the Martian surface, in long dark spots on the Red Planet thought to be formed because of water flow.

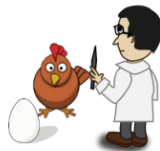


**REPORT**  
STATE, EVENT  
temporal progression  
related to speech time.

The sand-hills here run down to the sea, and end in two spits of rock jutting out opposite each other, till you lose sight of them in the water. One is called the North Spit, and one the South.



**DESCRIPTION**  
STATE, on-going EVENT  
metaphorical progression  
through scene



one text  
≈ one genre



one passage  
≈ one discourse  
mode



Situation entity types (A. Friedrich)

- state
- event
- generic sentence
- generalizing sentence
- general stative
- abstract entity
- report
- speech act

## Processing Large Discourse Units

- Logical document structure

*a hierarchy of segments of the document, each of which corresponds to a visually distinguished semantic component of the document (Summers 1998)*

- Content zones

*a continuous portion of a text document that fulfills a functional role for the text as a whole, contributing to the overall message or purpose, as it is characteristic for the genre of the text (defined for each sentence)*

- Topic-based segmentation

*a sequence of non-overlapping text segments that completely covers the text. Each unit consists of one or more sentences that address a common topic*

- Discourse modes (C. Smith, 2003)

*narrative, information, argument (commentary), report, description*

## Coreference

### grammatical coreference

- mostly possible to identify the antecedent on the basis of grammatical rules of the given languages
- within one sentence

## Discourse Processing

- \* arguments in constructions with verbs of control  
(John wants to [#Cor.ACT] kiss Mary.)
- \* reflexive pronouns (John shaved himself)
- \* relative pronouns (John, who came late, apologized.)
- \* coreference with verbal modifications that have dual dependency (John saw Mary [#Cor.ACT] stand on the windowsill and cry.)
- \* reciprocity (John and Mary kissed [#Rcp.PAT].)

## Processing Large Discourse Units

- Logical document structure

*a hierarchy of segments of the document, each of which corresponds to a visually distinguished semantic component of the document (Summers 1998)*

- Content zones

*a continuous portion of a text document that fulfills a functional role for the text as a whole, contributing to the overall message or purpose, as it is characteristic for the genre of the text (defined for each sentence)*

- Topic-based segmentation

*a sequence of non-overlapping text segments that completely covers the text. Each unit consists of one or more sentences that address a common topic*

- Discourse modes (C. Smith, 2003)

*narrative, information, argument (commentary), report, description*

## Coreference

grammatical coreference

textual coreference

- not restricted to grammatical means alone, context
- different means (pronominalisation, grammatical agreement, repetitions, synonyms, paraphrasing, hyponyms/hyperonyms, etc.)
- often occurs between entities in different sentences

## Discourse Processing

- \* personal and possessive pronouns (*Jonh left Mary. He wanted to see his mother*),
- \* demonstrative pronouns *ten, ta, to* (*It means that he doesn't really love Mary.*)
- \* with textual ellipsis (zeros) (*Více si Ø vážil své matky.*)
- \* nouns (*John asked his mother to advise him how he should behave with Mary, but mother ignored her son's wish.*)
- \* local adverbs (*John asked mother to come to Mary's place with him but she decided not to go there.*)
- \* some adjectives (*At last, Mary came to Prague herself and found the Prague atmosphere quite casual.*)
- \* reference to events (*Mary suggested Jonh to go to the theater, but Jonh ignored her wish.*)
- \* If antecedent is a whole segment of (previous) text larger than one sentence (phrase) — special type of textual coreference segm(ent) without explicitly marked antecedent: (*The next day Mary suggested to visit his mother. Then she proposed to go swimming. Her last wish was just to look at the city center. Jonh denied all of it.*)

## Processing Large Discourse Units

- Logical document structure

*a hierarchy of segments of the document, each of which corresponds to a visually distinguished semantic component of the document (Summers 1998)*

- Content zones

*a continuous portion of a text document that fulfills a functional role for the text as a whole, contributing to the overall message or purpose, as it is characteristic for the genre of the text (defined for each sentence)*

- Topic-based segmentation

*a sequence of non-overlapping text segments that completely covers the text. Each unit consists of one or more sentences that address a common topic*

- Discourse modes (C. Smith, 2003)

*narrative, information, argument (commentary), report, description*



## Coreference resolution

## Processing Small Discourse Units

coherence relations

connectives, cue phrases

# COHESION

In a cohesion analysis the connectivity of the discourse is primarily tied to the explicit marking of semantic relations. These explicit cues make a text a text.

BUT: Cohesion is necessary but not sufficient condition for the creation of text (Halliday and Hasan, 1976).

*Texture* (quality that makes a particular set of words and sentences a text)



Figure from Taboada (2015)

Halliday and Hasan (1976)

# COHERENCE

Cohesive elements like connectives are viewed as important but not necessary features of discourse: they are linguistic markers expressing the underlying conceptual relations that are of a cognitive nature (Sanders, 1992). \*

Coherence and cohesive relations are the threads that make up a text.

Discourse is coherent if

all of its **pieces belong together**



coherence  
(relational coherence)

and they have some **common thread**



cohesion  
(entity-based coherence)

after Taboada (2015)

# Requirements to a theory of discourse structure



*Ted Sanders (1992)*

A satisfying theory of discourse structure should meet

DESCRIPTIVE  
ADEQUACY

A theory discourse structure makes it possible to describe the structure of all kinds of natural texts.

PSYCHOLOGICAL  
PLAUSIBILITY

A theory of discourse structure should at least generate plausible hypotheses on the role of discourse structure in the construction of the cognitive representation.

*(it should make sense) ☺*

# Elementary Discourse Units (EDUs)

## Sentences

*Give me your essay, please. Then you may leave the room.*

*Give me your essay, please. The best one.*

*When you leave the room, give me your essay.*

## Clauses

*John attended the lecture despite his illness.*

nominalizations

*Exhausted, I called for a taxi cab.*

“free adjuncts”

restrictive and non-restrictive relative clauses

*John, although he was ill, attended the lecture.*

embeddings

*The red car that's parking in front of you belongs to me.*

*The red car, which was brought here by my dad, belongs to me.*

# Elementary Discourse Unit (EDU)

Abstract objects  
(Asher, 1993)

events, states, conditions and dialogue acts, that are typically expressed as sentences, but they can also be smaller or larger units (clauses, paragraphs, dialogue segments)

**BUT!**  
*Vary from researcher to researcher, depending on the level of granularity needed*

Pragmatic elementary units (Polanyi et al. 2004)

- meaning,
- discourse function,
- independent continuation

A span of text, usually a clause, but in general ranging from minimally a (nominalization) NP to maximally a sentence. It denotes a single event or type of event, serving as a complete, distinct unit of information that the surrounding discourse may connect to. An EDU may be structurally embedded in another.



# Elementary Discourse Unit (EDU)



Nicolas Asher

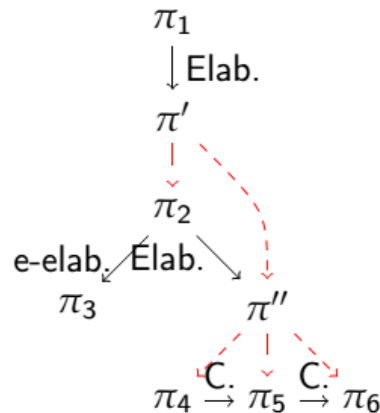
basic intuition – they also exist underneath the clausal level, and they can either be titles, oppositions, relative clauses, some NP (e.g. *principles*), etc.

[*Interprovincial Pipe Line Co. said*] $\pi_1$   
[*it will delay a proposed two-step, 830 million dollar* [US\$705.6 million] $\pi_3$   
*expansion of its system*] $\pi_2$  [*because Canada's output of crude oil is shrinking.*] $\pi_4$

opposition

title, NP

[Principes de la sélection naturelle.] $\pi_1$  [La théorie de la sélection naturelle [telle qu'elle a été initialement décrite par Charles Darwin,] $\pi_3$  repose sur trois principes :] $\pi_2$  [1. le principe de variation] $\pi_4$  [2. le principe d'adaptation] $\pi_5$  [3. le principe d'hérédité] $\pi_6$



# Written vs. Spoken



- 0 **Written text** - of a single author
- 0 The author is organizing his thoughts together, make them coherent as much as he can

**Definitely different!**

- a number of discourse structure annotated corpora for dialogues (Asher's STAC, Sidarenka (chats), L. Degand's LOCAS-F, Italian?)
- annotating discourse on spoken texts requires first of all annotation of DMs.

- 0 **Conversation** as a game of message exchange involving a kind of signaling game, a play with reactions:
  - 0 X plays  $\varphi \rightarrow$  Y decodes a message  $\rightarrow$  Y decides what signal to send in return  $\rightarrow$  X decodes a message.

# Outline

I. Very generally about discourse-related topics



II. Discourse theories (H&H, RST, SDRT, PDTB, DGB, CCR) and corpora annotated with discourse relations

III. Comparison of theories and approaches

Penn Discourse Treebank (PDTB)

French Discourse Treebank

PDTB-like

Prague Dependency Treebank (PDT)

Turkish Discourse Bank

Postdam Commentary Corpus

RST signalling corpus

Halliday and Hasan

GECCo

DiscAn Corpora

Sanders et al.

Discourse Graphbank

Hobbs

DISCOR

SDRT

ANNODIS (French)

STAC

LOCUS-F

RST corpus

Discourse dependency treebank

RST-like

RST Spanish Corpus

Basque-Spanish-English parallel corpus (RST)

Basque RST Treebank

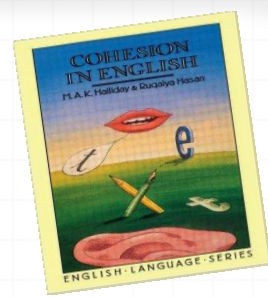
## THEORIES & CORPORA

- 0 Halliday and Hasan – Cohesion in English
- 0 RST – Rhetorical Structure Theory
- 0 SDRT - Segmented Discourse Representation Theory
- 0 PDTB – Penn Discourse Treebank
- 0 Discourse Graphbank
- 0 CCR – Cognitive approach to coherence relations



# M.A.K.Halliday and Ruqaiya Hasan

## *Cohesion in English*



- 0 We can distinguish what is text and what is not → text has structure
- 0 But the unity of a text is a unity of a different kind
- 0 Text is realized by sentences but it does not have the same structural integration among its parts as we find among the parts of a sentence or clause
- 0 Investigate the resources that English has to create text
- 0 **Cohesion** – a set of explicit cues – is what makes a text a text

- 0 **Reference:** identity between entities

*John loves Mary. However he is afraid to kiss her.*

- 0 **Substitution:** similarity between different instantiated entities of the same type

*These biscuits are stale. Get some fresh ones.*

- 0 **Ellipsis**

*Would you like to hear another verse? I know twelve  $\emptyset$  more.*

- 0 **Conjunction:** logico-semantic relations between propositions (e.g. addition, contrast, cause)

*John love Mary. However he is afraid to kiss her.*

- 0 **Lexical cohesion** (similarity between entities of the same type based on sense relations (e.g. hyperonymy, part-whole relations))

*Why does this little boy wriggle all the time? Girls don't wriggle.*



# German-English Contrasts in Cohesion (GECCo)



- 0 based on the definition of cohesion and cohesive devices in English by (Halliday & Hasan, 1976)
- 0 elaborated for a contrastive analysis of two languages
- 0 comparable and parallel texts in English and German (ca. 80,000 sentences)
- 0 various registers, including written and spoken dimensions





# German-English Contrasts in Cohesion (GECCo)



reference	personal head, modifier, <i>it/es</i> -endo- and -exophoric, demonstrative head, modifier, local, temporal, comparative particular and general
conjunctive relations	additive, adversative, causal, temporal, modal
substitution	nominal, verbal, clausal
ellipsis	nominal, verbal, clausal
lexical cohesion	general nouns, repetition, synonymy, antonymy, hyperonymy, hyponym, meronymy
chains	Nr of chains, chain length, Nr of antecedents

## Outline

# THEORIES & CORPORA

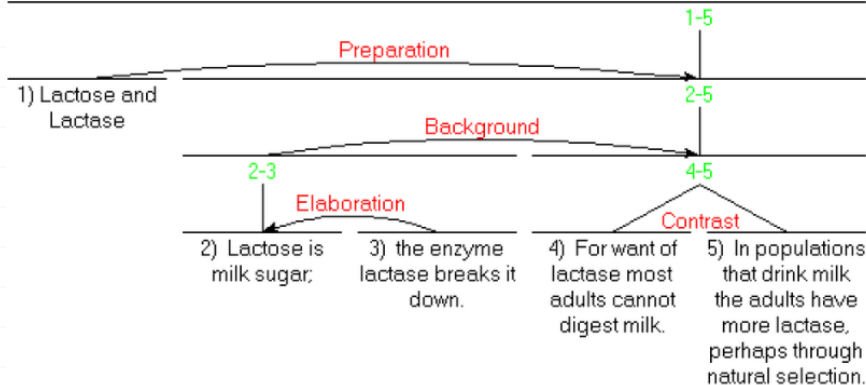
- 0 Halliday and Hasan – Cohesion in English ✓
- 0 RST – Rhetorical Structure Theory
- 0 SDRT - Segmented Discourse Representation Theory
- 0 PDTB – Penn Discourse Treebank
- 0 Discourse Graphbank
- 0 CCR – Cognitive approach to coherence relations

# Rhetorical Structure Theory

Mann, W. C. and S. A. Thompson (1988). **Rhetorical structure theory. Toward a functional theory of text organization.** Text 8(3), 243-281.



Taboada, Maite and William C. Mann. (2006). **Rhetorical Structure Theory: Looking back and moving ahead.** Discourse Studies, 8 (3), 423-459.



Scientific American, October 1972.

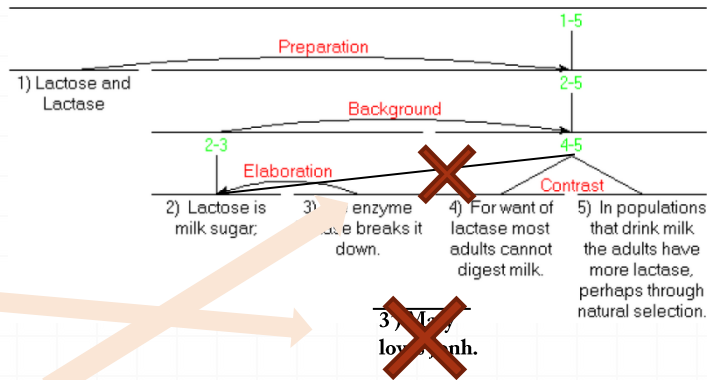
<http://www.sfu.ca/rst/>  
<http://www.sfu.ca/rst/05bibliographies/>

# Rhetorical Structure Theory

- Empirical perspective, comes from corpus analysis
- RST is about how text works (primarily written → spoken)

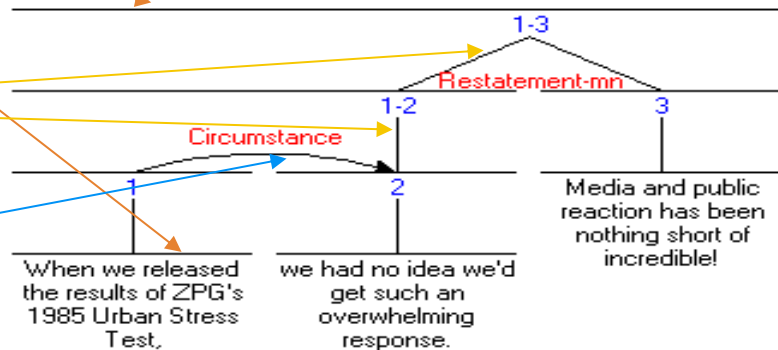
*RST principles*

- ✓ Coherent texts consist of minimal units, which are linked to each other, recursively, through rhetorical relations (coherence or discourse relations )
- ✓ Coherent texts do not show gaps or non-sequiturs.
- ✓ The resulting structure is a complete tree: no cross-dependencies.
  - completeness (one schema application contains the entire text)
  - connectedness (each span, except for the span that contains the entire text, is either a minimal unit or a constituent of another schema application)
  - uniqueness (each schema application contains a different set of text spans)
  - adjacency (the spans of each schema application constitute one contiguous text span)



# RST: Graphical Representation

- 0 A **horizontal line** covers a span of text (possibly made up of further spans)
- 0 A **vertical line** signals the nucleus or nuclei
- 0 A **curve** represents a relation, and the direction of the arrow, the direction of satellite towards nucleus



# Relations in RST

- RST describes coherence relations between discourse segments : RST analysis wants to answer the question how coherence in text is achieved
- Definition of the relations are based on functional and semantic criteria, not on morphological or syntactic signals
- no reliable or unambiguous signal for any of the relation
- Different lists of the relations exist:
  - Mann and Thomson – 24 relations (Cause, Contrast, Elaboration, Restatement, Evidence, Conditions, Antithesis...)
  - *Later* - List, Means, Preparation, Unconditional, Unless, two Restatements (nuclear and multinuclear), Joint (the declared absence of a relation)



# Relation names (Mann-Thompson 1988)

---

**Circumstance**

**Solutionhood**

**Elaboration**

**Background**

**Enablement and Motivation**

**Enablement**

**Motivation**

**Evidence and Justify**

**Evidence**

**Justify**

**Relations of Cause**

**Volitional Cause**

**Non-Volitional Cause**

**Volitional Result**

**Non-Volitional Result**

**Purpose**

**Antithesis and Concession**

**Antithesis**

**Concession**

**Condition and Otherwise**

**Condition**

**Otherwise**

**Interpretation and Evaluation**

**Interpretation**

**Evaluation**

**Restatement and Summary**

**Restatement**

**Summary**

**Other Relations**

**Sequence**

**Contrast**

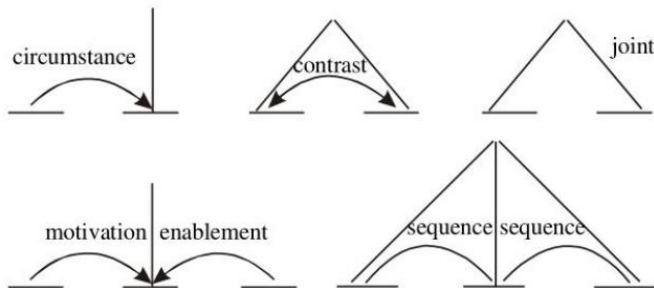
---

Other classifications are possible, both longer and shorter lists have been proposed.

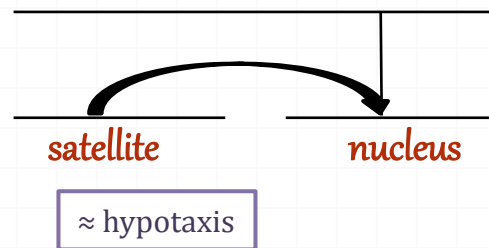
# Relations in RST - nuclearity

Coherence relations between discourse segments

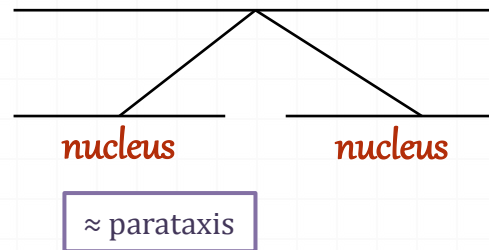
- asymmetric (“mononuclear”)
  - one nucleus, one satellite
- symmetric (“multinuclear”)
  - multiple nuclei



other schemas



mononuclear relation

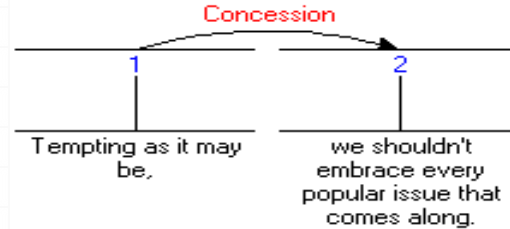


multinuclear relation



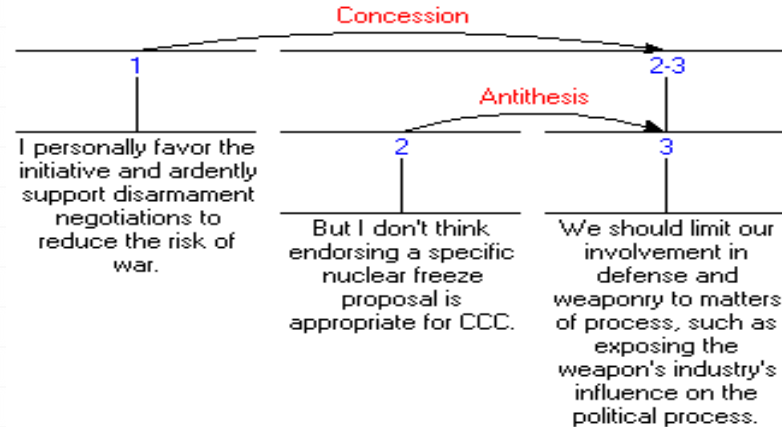
# Hypotactic (subordinate)

## 0 Sub-sentential Concession relation



## 0 Concession across sentences

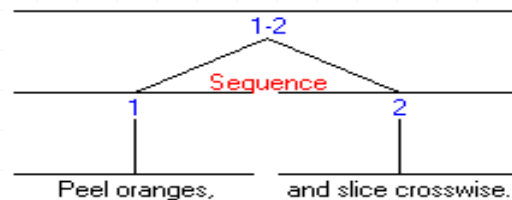
- 0 Nucleus (spans 2-3) made up of two spans in an Antithesis relation



# Paratactic (coordinate)

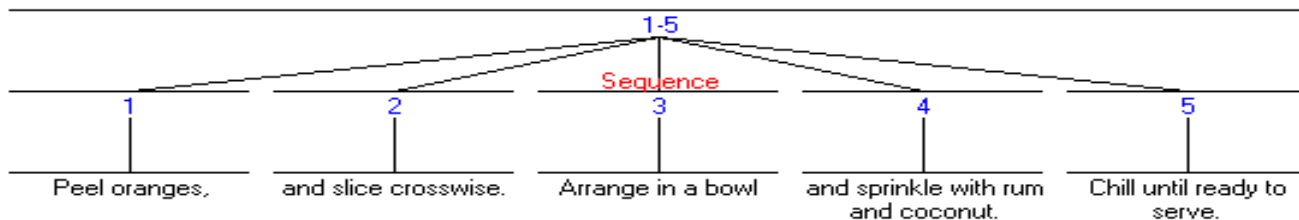
- At the sub-sentential level (traditional coordinated clauses)

*Peel oranges, and slice crosswise.*



- But also across sentences

*1. Peel oranges, 2. and slice crosswise. 3. Arrange in a bowl 4. and sprinkle with rum and coconut. 5. Chill until ready to serve.*



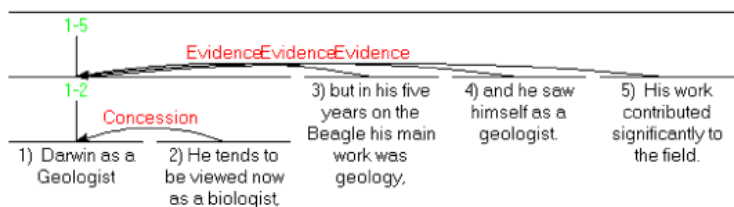
# Relations in RST - Constraints

## Example: Evidence

### A relation consists of:

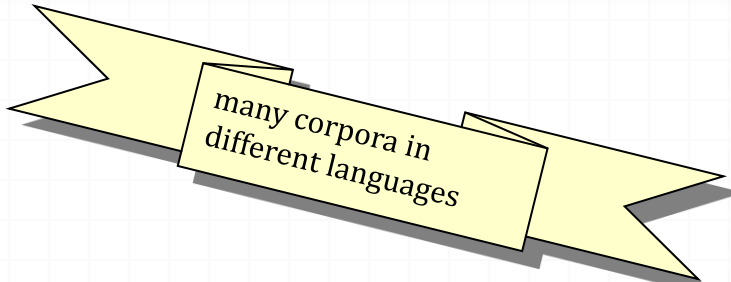
1. Constraints on the Nucleus,
2. Constraints on the Satellite,
3. Constraints on the combination of Nucleus and Satellite,
4. The Effect.

1. Constraints on the Nucleus  
The reader may not believe N to a degree satisfactory to the writer
2. Constraints on the Satellite  
The reader believes S or will find it credible
3. Constraints on the combination of N+S  
The reader's comprehending S increases their belief of N
4. Effect (the intention of the writer)  
The reader's belief of N is increased



*1) Darwin as a Geologist 2) He tends to be viewed now as a biologist, 3) but in his five years on the Beagle his main work was geology, 4) and he saw himself as a geologist. 5) His work contributed significantly to the field.*

# RST corpora



many corpora in  
different languages

- 0 RST Discourse Treebank (Carlson et al. 2003)
- 0 RST Signalling Corpus (Das and Taboada 2015)
- 0 Potsdam Commentary Corpus (Stede and Neumann 2014)
  - 0 220 German newspaper commentaries annotated with different types of linguistic information, including RST
- 0 Discourse Relations Reference Corpus
  - 0 [http://www.sfu.ca/rst/06tools/discourse\\_relations\\_corpus.html](http://www.sfu.ca/rst/06tools/discourse_relations_corpus.html)
  - 0 texts from RST web site +annotated Wall Street Journal articles from the RST Discourse Treebank +review texts from the SFU Review Corpus
- 0 GUM - The Georgetown University Multilayer Corpus
  - 0 POS, lemmas, syntax, constituent and dependency syntax, Information status (given, accessible, new)
  - 0 Entity and coreference annotation
- 0 Spanish RST Discourse Treebank (da Cunha, Iria, Juan Manuel Torres-Moreno and Gerardo Sierra 2011))
  - 0 <http://corpus.iingen.unam.mx/rst/>
- 0 Basque RST Discourse Treebank (Iruskieta et al.)
  - 0 <http://ixa2.si.ehu.es/diskurtsoa/en/>
- 0 Multiling RST Treebank (English, Spanish and Basque)
  - 0 <http://ixa2.si.ehu.es/rst>

# RST Discourse Treebank

L. Carlson et al. 2003

Carlson, Lynn; Daniel Marcu and Mary Ellen Okurowski. **RST Discourse Treebank**  
LDC2002T07. Web Download. Philadelphia: Linguistic Data Consortium, 2002.

- 385 WSJ articles from Penn Treebank, representing over 176,000 words of text. ~14% were double-tagged.
- Document length: 31 to 2124 words; average of 458.14 words
- Average # EDUs per document: 56.59.
- Average # words per EDU: 8.1.
- Nature of articles: general news, financial, business, cultural reviews, editorials
- Intended users: developers of automatic text processing systems

# RST Discourse Treebank

Taggers	Units	Spans	Nuclearity	Relations	# docs	Avg # EDUs
A, E	0.98	0.90	0.84	0.76	7	57.7
A, B	1.00	0.93	0.88	0.79	5	58.2
A, C	0.95	0.84	0.78	0.68	4	116.5
A, F	0.95	0.78	0.69	0.69	4	26.5
A, D	1.00	0.87	0.80	0.72	4	23.3

IAA

## Relations

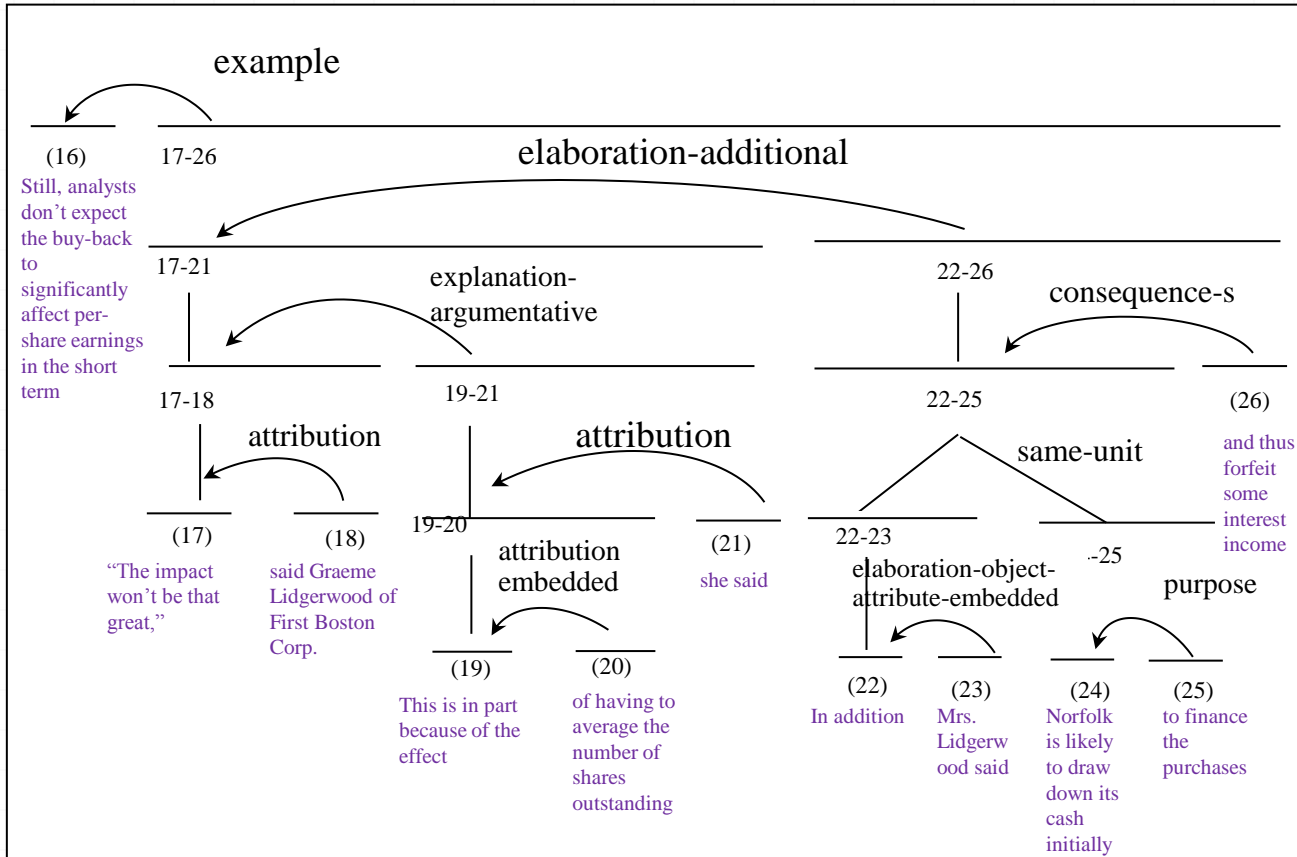
78 relations in 16 groups:

- **Attribution**: attribution, attribution-negative
- **Background**: background, circumstance
- **Cause**: cause, result, consequence
- **Comparison**: comparison, preference, analogy, proportion
- **Condition**: condition, hypothetical, contingency, otherwise
- **Contrast**: contrast, concession, antithesis
- **Elaboration**: elaboration-additional, elaboration-general-specific
- **Enablement**: purpose, enablement
- **Evaluation**: evaluation, interpretation, conclusion, comment
- **Explanation**: evidence, explanation-argumentative, reason
- **Joint**: list, disjunction
- **Manner-Means**: manner, means
- **Topic-Comment**: problem-solution, question-answer, topic-comment..
- **Summary**: summary, restatement
- **Temporal**: temporal-before, temporal-after, temporal-same-time..
- **Topic Change**: topic-shift, topic-drift

Carlson, Lynn; Daniel Marcu and Mary Ellen Okurowski. **RST Discourse Treebank** LDC2002T07. Web Download. Philadelphia: Linguistic Data Consortium, 2002.

# RST Discourse Treebank

wsj\_1111



[Still, analysts don't expect the buy-back to significantly affect per-share earnings in the short term.]<sup>16</sup> ["The impact won't be that great,"]<sup>17</sup> [said Graeme Lidgerwood of First Boston Corp.]<sup>18</sup> [This is in part because of the effect]<sup>19</sup> [of having to average the number of shares outstanding,]<sup>20</sup> [she said.]<sup>21</sup> [In addition,]<sup>22</sup> [Mrs. Lidgerwood said,]<sup>23</sup> [Norfolk is likely to draw down its cash initially]<sup>24</sup> [to finance the purchases]<sup>25</sup> [and thus forfeit some interest income.]<sup>26</sup>

# RST Signalling Corpus

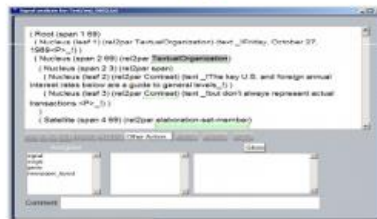
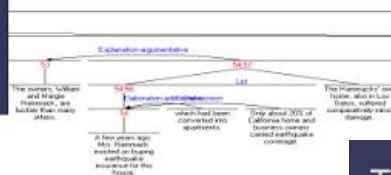
Das – Taboada, 2015

Das, Debopam and Maite Taboada (2015) **RST Signalling Corpus**. LDC.

Hypothesis: There are no (or very few) ‘implicit’ relations.

## Analysis process

- Load relations into UAM CorpusTool
- Annotation procedure
  - Examine relation
  - Find signals
  - Annotate signal information



## Taxonomy of signals (8 groups, 39 total)

- |                     |  |
|---------------------|--|
| 1. Discourse marker | • Conjunction, adverbial, prepositional phrase   |
| 2. Reference        | • Personal, demonstrative, comparative           |
| 3. Lexical          | • Indicative phrase/word                         |
| 4. Semantic         | • Synonym, antonym, hyponym, lexical chain       |
| 5. Morphological    | • Tense  |
| 6. Syntactic        | • Non-finite/relative clause, parallel structure |
| 7. Graphical        | • Colon, dash, bullet                            |
| 8. Genre            | • Attribution, pyramid scheme                    |

```
8 </segment>
9 <segment id="1" start="318" end="335" features="signal:single;dm:and" state="active"/>
10 <segment id="2" start="318" end="335" features="signal:single;morphological;tense" state="active"
11 =present - past"/>
12 <segment id="8" start="318" end="335" features="signal:single;semantic;lexical_chain" state="act"
13 =today - last week"/>
14 <segment id="11" start="471" end="493" features="signal:single;genre:inverted_pyramid_scheme" st
15 />
16 <segment id="12" start="471" end="493" features="signal:combined:(semantic+syntactic):(anonymous)
17 state="active" comment="few issues - Crawford"/>
```



# RST - Present and Future

- 0 Annotated corpora in different languages, maintained web

## RST workshops:

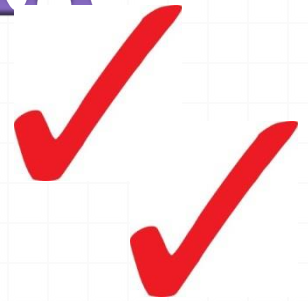
- 0 2007 - Collocated with the *16th Intercâmbio de Pesquisas em Linguística Aplicada (InPLA)*, Pontifícia Universidade Católica de São Paulo, Brazil.
- 0 2009 - Named "Brazilian RST Meeting", collocated with *The 7th Brazilian Symposium in Information and Human Language Technology (STIL)*, in São Carlos, Brazil.
- 0 2011 - 3rd Workshop, "RST and Discourse Studies" was held with *The 8th Brazilian Symposium in Information and Human Language Technology (STIL)*, in Cuiabá, Brazil.
  - 0 [http://corpus.iingen.unam.mx/RST\\_2011](http://corpus.iingen.unam.mx/RST_2011)
- 0 2013 - The 4th workshop "RST and Discourse Studies" was also held within *STIL (2013)* in Fortaleza, Brazil.
  - 0 <http://encontrorst2013.wix.com/encontro-rst-2013>
- 0 2015 - The 5th workshop "RST and Discourse Studies", collocated with *The 31st Conference of the Spanish Association for Natural Language Processing*. Alicante, Spain.
  - 0 <https://sites.google.com/site/workshoprst2015/>

# RST done!



# THEORIES & CORPORA

- 0 Halliday and Hasan – Cohesion in English
- 0 RST – Rhetorical Structure Theory
- 0 Discourse Graphbank
- 0 SDRT - Segmented Discourse Representation Theory
- 0 PDTB – Penn Discourse Treebank
- 0 CCR – Cognitive approach to coherence relations





# Jerry R. Hobbs' theory

Hobbs, Jerry R. 1985. **On the coherence and structure of discourse**. Technical Report 85-37, Center for the Study of Language and Information (CSLI), Stanford, CA.

Hobbs, Jerry R. 1985. **Literature and Cognition**., CSLI, 1990



Coherence relations, e.g. *contrast*, *elaboration*, *parallel*...

Theory of discourse coherence  
Theory of discourse interpretation

- ... hold between segments of a discourse
- ... are defined in terms of propositions that can be inferred from the assertions of discourse segments (in terms of formal logic)
  - The assertion of a clausal discourse segment is, roughly, what is predicated by the main verb

# Jerry R. Hobbs' theory

## Hobbs' relations

Occasion Relation

Cause

Evaluation Relation

Ground-Figure

----- Explanation Relations

Expansion Relations

----- Parallel

----- Elaboration

----- Exemplification

----- Contrast

----- Violated Expectation

### Parallel:

Infer  $p(a_1, a_2, \dots)$  from the assertion of  $S_0$  and  $p(b_1, b_2, \dots)$  from the assertion of  $S_1$ , where  $a_i$  and  $b_i$  are similar, for all  $i$ .

$$\frac{\frac{S_0}{\frac{p}{a_1} \frac{a_2}}{\frac{p}{b_1} \frac{b_2}}{S_1}}{\text{Set stack } A \text{ empty and set link variable } P \text{ to } T.}$$

**Elaboration:** Infer the same proposition  $P$  from the assertions of  $S_0$  and  $S_1$ .

*Go down First Street. Just follow First Street three blocks to A Street.*

go(Agent: you, Goal: x, Path: First St., Measure: y) for some x and y.

go (Agent: you, Goal: A St., Path: First St., Measure: 3 blks)

If we assume that x is A Street and y is 3 blocks, then the two are identical and serve as the proposition  $P$  in the definition.



# Discourse Graphbank

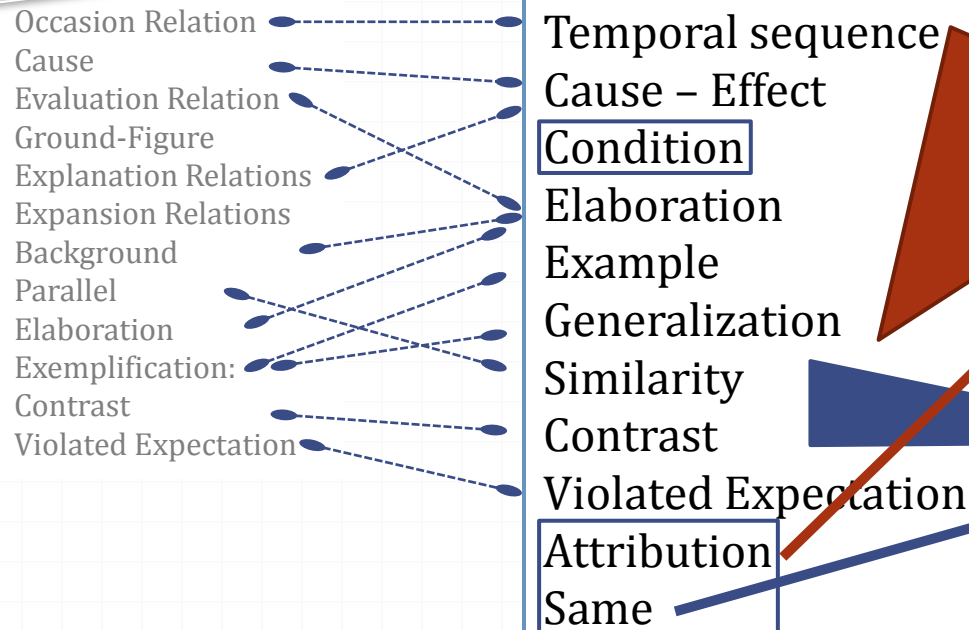
Wolf, F. and Gibson, E.: 2005, **Representing discourse coherence: A corpus based study**, Computational Linguistics 31(2), 249–287.

- 0 motivated by Hobbs(1985) and Kehler (2002)
- 0 vs. RST: polemics to tree structures
- 0 135 texts annotated with coherence relations (WSJ and AP Newswire)
- 0 DUs – mainly clauses
  - 0 delimiting discourse segments (*because, and, for example*, periods, semicolons, commas, etc.)
  - 0 treat attributions (*John said that . . .*) as discourse segments
- 0 Discourse Segment Groupings
  - 0 groups: e.g. attributed to the same source, topically related, might also be subgroups, consisting of several discourse segments each (making a partially hierarchical structure for the text)
  - 0 groups were allowed to partially overlapped, but this was not used by annotators

# Discourse Graphbank

## RELATIONS

Hobbs' relations:



asymmetrical (directed)

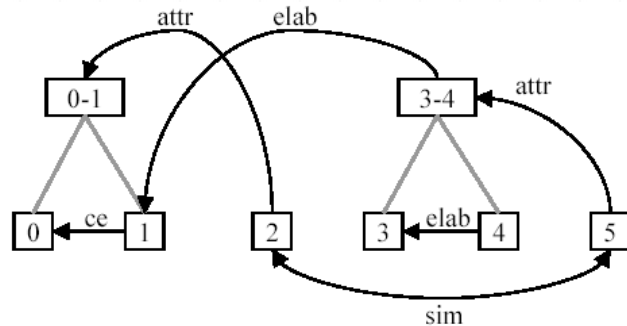
*following (Mann and Thompson  
1988; Marcu 2000)*

symmetrical (undirected)

# Discourse Graphbank

## REPRESENTATION

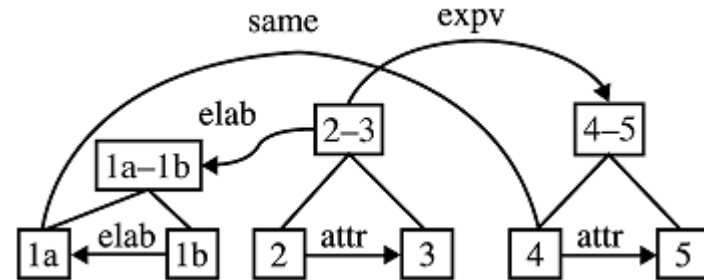
discourse structure is complex and requires a representation in terms of chain graphs



(ce=Cause-Effect; attr=Attribution; elab=Elaboration; sim=Similarity.)

0. Farm prices in October edged up 0.7% from September
1. as raw milk prices continued their rise,
2. the Agriculture Department said.
3. Milk sold to the nation's dairy plants and dealers averaged \$14.50 for each hundred pounds,
4. up 50 cents from September and up \$1.50 from October 1988,
5. the department said.

1. <sub>1a</sub>[ Mr. Baker's assistant for inter-American affairs, ] <sub>1b</sub>[ Bernard Aronson, ]
2. while maintaining
3. that the Sandinistas had also broken the cease-fire,
4. acknowledged:
5. "It's never very clear who starts what."



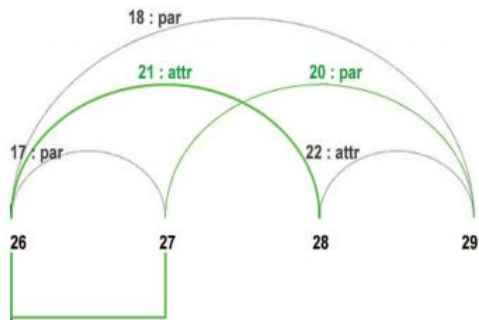


# DGB vs RST (*or* graphs vs. trees)

Markus Egg and Gisela Redeker, **How Complex is Discourse Structure?** LREC 2010

## DGB:

- crossed dependencies
- 41.22% of the segments have multiple parents (W&G 2005)



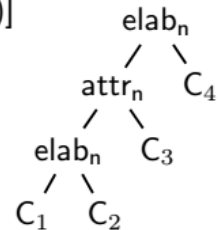
## Egg and Redeker:

distinguish the complexity inherent in the data and the one arising from specific design choices in W&G's annotation (on 14 texts from DGB)

Result: Many non-tree-like dependencies may be converted to tree-like ones.

RST-based analysis of (4)

(5) [= (3)]






(4) ( $C_1$ ) "He was a very aggressive firefighter. ( $C_2$ ) He loved the work he was in," ( $C_3$ ) said acting Fire Chief Larry Garcia. ( $C_4$ ) "He couldn't be bested in terms of his willingness and his ability to do something to help you survive." (ap-890101-0003)

# Discourse Graphbank done!



# THEORIES & CORPORA

- 0 Halliday and Hasan – Cohesion in English 
- 0 RST – Rhetorical Structure Theor 
- 0 Discourse Graphbank 
- 0 PDTB – Penn Discourse Treebank
- 0 SDRT - Segmented Discourse Representation Theory
- 0 CCR – Cognitive approach to coherence relations

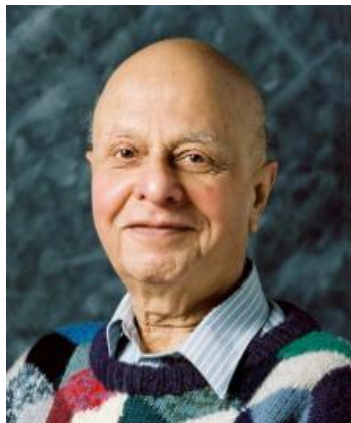
# Penn Discourse Treebank

<http://www.seas.upenn.edu/~pdtb/>

- 0 Authors, background, goals
- 0 Versions
- 0 Annotation principles
- 0 Related corpora

performed by Lucie Poláková

Prof. Aravind Joshi  
University of Pennsylvania



# PDTB – Authors

Prof. Bonnie Webber  
University of Edinburgh



Dr. Rashmi Prasad  
University of Wisconsin-  
Milwaukee  
(formerly UPenn)

UPenn  
Collaborators  
(among others):

Alan Lee



Eleni Miltsakaki



Nikhil Dinesh



# PDTB – Background

- 0 Premise: The meaning and coherence of a discourse results partly from how its constituents relate to each other.
- 0 DISCOURSE RELATIONS: Semantic, “informational”, relations between **abstract objects** (AOs) mentioned in discourse.
- 0 Abstract objects – events, states, propositions (Asher 1993)

# PDTB – Background

- 0 **Low-level:** “Shallow” or “local” discourse analysis: Lack of agreement on high-level discourse representation structures (trees, graphs...)
- 0 **Theory-neutral:** Allows corpus to be usable with different frameworks; allows for data-driven “emergent” theory of discourse structure.
- 0 **Lexically grounded:** relations anchored by lexical items
- 0 **Stand-off** representation: can be easily merged with other annotations

performed by Lucie Poláková

# PDTB – Goals

To annotate a large-scale corpus of discourse relations to:

- 0 Extend the scope of discourse-level NLP research and resulting applications;
- 0 Facilitate cross-linguistic empirical research on discourse relations.



# Penn Discourse Treebank - History

- 0 PDTB 1.0 – 2004 (Miltsakaki et al., 2004, LREC Portugal)
- 0 PDTB 2.0 – 2008 (Prasad et al., 2008, LREC Morocco)
- 0 PDTB 3.0 – work in progress, two NSF grants 2014-2017 (UPenn, University Wisconsin-Milwaukee)

# Annotation principles

What is annotated in PDTB?

- 0 Discourse relations – introduced by:
  - 0 Discourse connectives “but, then, for example, although...”
  - 0 Alternative lexicalizations of the connectives “the reason is”
  - 0 Implicit relations (no connective present)
- 0 Senses (semantics) of the relations – contrast, condition, cause...
- 0 Attribution “he said, I believe...”

# Annotation principles

- 0 A discourse connective is a discourse-level predicate taking two (and only two) text units (abstract objects) as its arguments:
- 0 Explicit connectives:

**Arg1**

**Relation**

**Arg2**

**She hasn't played any music since the earthquake hit.**

# Annotation principles

0 Implicit relations:

**Some have raised their cash positions to record levels.**

[Implicit = because] High cash positions help buffer a fund when the market falls.

0 AltLex:

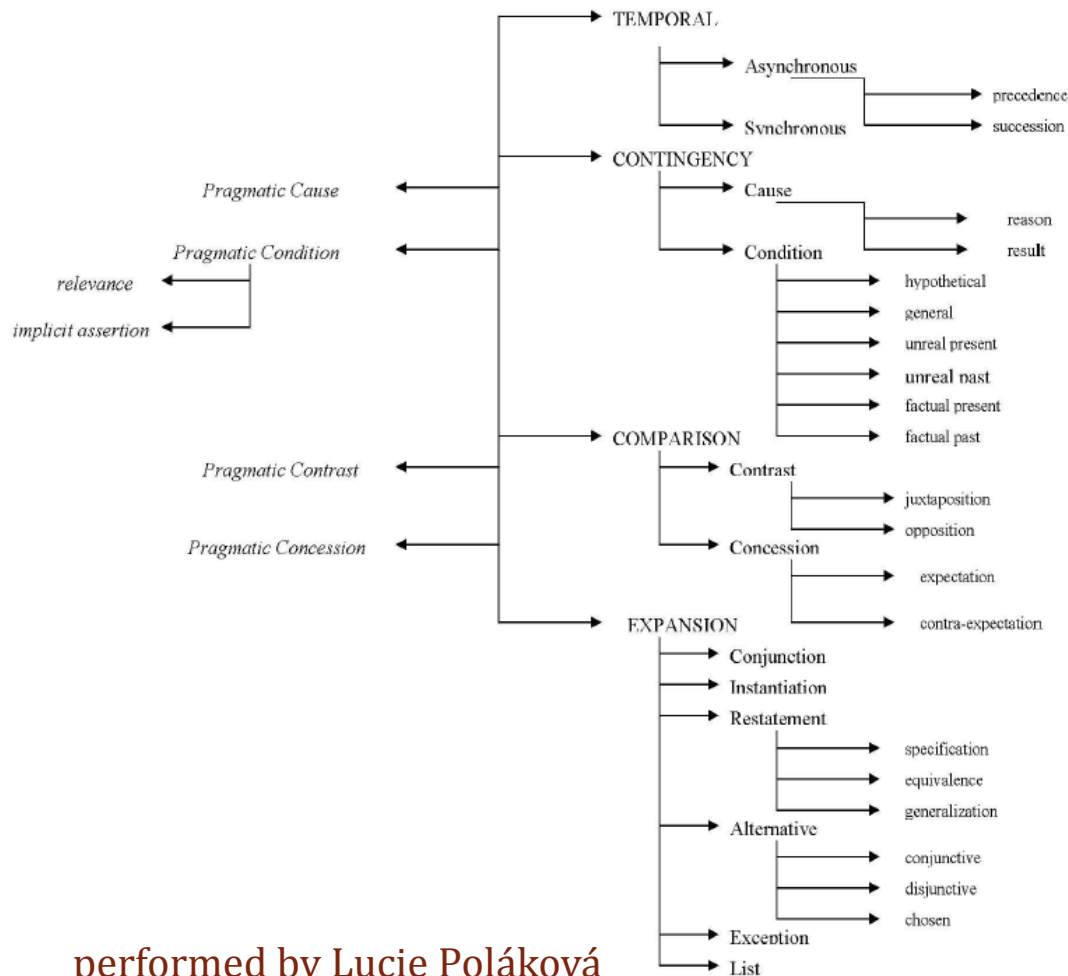
**A few years ago, the company offered two round-trip tickets on Trans World Airlines to buyers of its Riviera luxury car.** The promotion helped Riviera sales exceed the division's forecast by more than 10%, Buick said at the time.

# Senses (PDTB 2.0)

0 Three-level  
hierarchy:

4 classes, 15 types, 23  
subtypes

**Substantially revised  
for the 3.0 version!!**



performed by Lucie Poláková

# Attribution

= ascription of texts contents to agents that expressed them

- 0 NOT CONSIDERED A DISCOURSE RELATION in the PDTB
- 0 annotated for every relation and for each of its two arguments
- 0 4 features of attribution recognized:

**Source** – writer, other person, arbitrary, inherited (for arguments)

**Type** – verbs of communication (**say, claim, explain...**), propositional attitude (**think, suppose...**), factive (**regret, remember...**), control verbs (**persuade, promise, intend...**)

**Scopal polarity** – the attribution verb reverses the polarity of the argument (**deny, not think...**)

**Determinacy** – indeterminate if the attribution can be cancelled (**it cannot be said...**)

# Updates for PDTB 3.0

What is newly annotated in PDTB 3.0?

## 0 Intra-sentential relations!

0 free adjuncts (-ing + ed forms) with no explicit connectives

*Exhausted, I called for a taxi cab.*

= As a result of being exhausted...

0 VP coordinations

0 Subordinators (also prepositions): “by, despite, because of, instead of, in order to, to...”

0 Punctuation marks: colon, semicolon, dash

0 Implicit relations across paragraph boundaries

0 Extending some connectives: then – but then

performed by Lucie Poláková

# New sense hierarchy

- 0 Only **two levels** (subtype level cancelled, some fine distinctions not made anymore)
- 0 Level three only preserved via **directionality** of the relations
- 0 Pragmatic domain: **epistemic and speech act** readings
- 0 **New senses** introduced: purpose, similarity, negative condition...
- 0 Some senses redefined (conjunction etc.)



# Corpora inspired by the PDTB annotation

- 0 The BioDiscourse Relation Bank (BioDRB, Prasad et al., 2011) – English
- 0 Hindi Discourse Relation Bank (HDRB, Kolachina et al., 2012, Oza et al. 2009)
- 0 The Leeds Arabic Discourse Treebank (Al-Saif and Markert, 2010)
- 0 PDTB-style annotation of Chinese (Zhou and Xue, 2012)
- 0 Turkish Discourse Bank (Zeyrek et al., 2010)
- 0 Prague Dependency Treebank 3.0 (Bejček et al., 2013) – Czech
- 0 LUNA: PDTB-style annotation of Italian spoken dialogs (Tonelli et al., 2010)
- 0 Potsdam Commentary Corpus (Stede 2004, Stede and Neumann 2014) – German
- 0 French Discourse Treebank (Danlos et al. 2012)
- 0 Tüba-D/Z Treebank (Gastel et al 2011, Versley and Gastel 2012) – German (specific connectives, partly implicit relations)

# Literature

Most important:

- 0 The PDTB research group: *The PDTB 2.0 Annotation manual*, 2007.
- 0 Prasad et al. 2008: *The Penn Discourse Treebank 2.0*, LREC, Morocco.
- 0 R. Prasad, B. Webber and A. Joshi. 2014. *Reflections on the Penn Discourse TreeBank, Comparable Corpora and Complementary Annotation*. In: *Computational Linguistics* 40:921–950.

All related literature available at:

<http://www.seas.upenn.edu/~pdtb/bibliography-year.shtml>

performed by Lucie Poláková

# PDTB done!



# THEORIES & CORPORA

- 0 Halliday and Hasan – Cohesion in English ✓
- 0 RST – Rhetorical Structure Theor ✓
- 0 Discourse Graphbank ✓
- 0 PDTB – Penn Discourse Treebank ✓
- 0 SDRT - Segmented Discourse Representation Theory
- 0 CCR – Cognitive approach to coherence relations

# Segmented Discourse



## Representation Theory (SDRT)

Asher, N. (1993). Reference to Abstract Objects in Discourse. Dordrecht

Asher, N. and A. Lascarides (2003). **Logics of conversation**. Cambridge University Press

- 0 Primarily based on Discourse Representation Theory (DRT: Kamp, 1981; Kamp & Reyle, 1993)
- 0 Comes from **formal semantics**, initially from formal sentence semantic and then systematically extended to discourse
  - 0 interested in 'semantic scopes' – some sort of bijection between one structure to another between the relations
- 0 **Recursive Structure (SDRS)** - constituted by discourse units linked by discourse relations where a discourse unit is either an **elementary discourse unit (EDU)** or a **complex discourse unit (CDU)**
  - 0 DUs can be embedded one into another (EDUs)
  - 0 CDUs may not partially overlap

# Segmented Discourse Representation Theory (SDRT)

## A discourse structure in SDRT:

**A** - is a set of labels (units)

**Last** - is a label A (intuitively last clause)

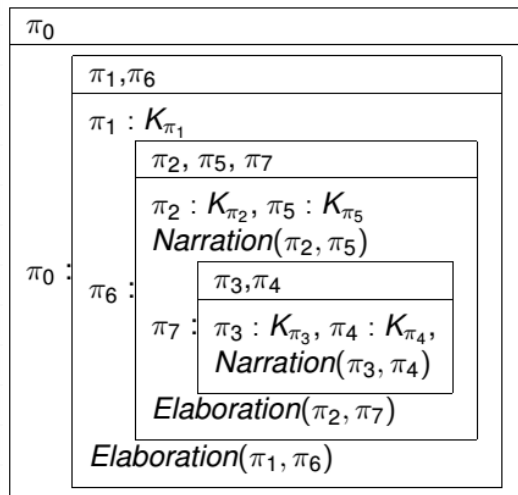
**F** - is a formula which assigns each member of A a member of a formula of the SDRS language

$$SDRS = \langle A, F, Last \rangle$$

- $\pi_1$  John had a great evening last night.
- $\pi_2$  He had a great meal.
- $\pi_3$  He ate salmon.
- $\pi_4$  He devoured lots of cheese.
- $\pi_5$  He then won a dancing competition.

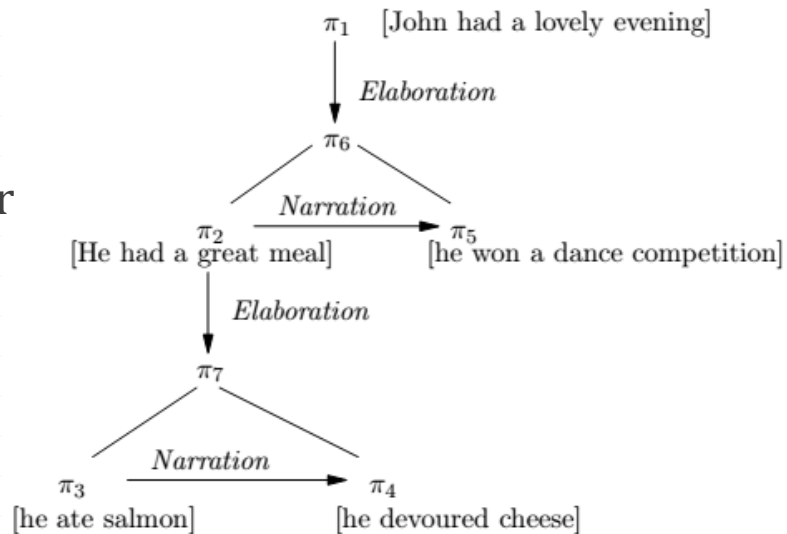
$\langle A, \mathcal{F}, LAST \rangle$ , where:

- $A = \{\pi_0, \pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6, \pi_7\}$
- $\mathcal{F}(\pi_1) = K_{\pi_1}, \mathcal{F}(\pi_2) = K_{\pi_2}, \mathcal{F}(\pi_3) = K_{\pi_3},$   
 $\mathcal{F}(\pi_4) = K_{\pi_4}, \mathcal{F}(\pi_5) = K_{\pi_5},$   
 $\mathcal{F}(\pi_0) = Elaboration(\pi_1, \pi_6)$
- $\mathcal{F}(\pi_6) = Narration(\pi_2, \pi_5) \wedge Elaboration(\pi_2, \pi_7)$
- $\mathcal{F}(\pi_7) = Narration(\pi_3, \pi_4)$
- $LAST = \pi_5$



# Segmented Discourse Representation Theory (SDRT)

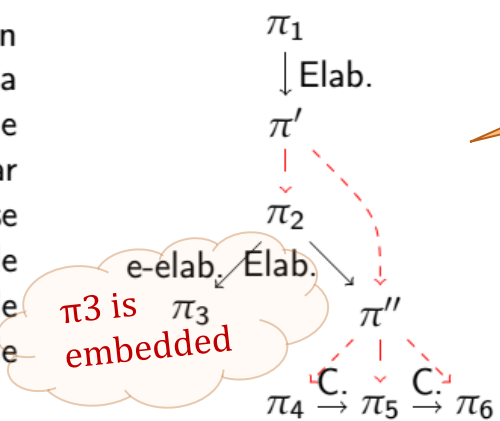
- 0 **Graph representation** - An SDRS can be seen as a directed acyclic graph where each DU is a vertex
  - 0 **Directed labelled edges** for rhetorical relations
  - 0 **Directed unlabelled edges** link CDUs to their content.
- 0 **Two kind of relations:**
  - 0 **Coordinating**, e.g. Result, Narration (drawn horizontally)
  - 0 **Subordinating**, e.g. Elaboration, Explanation, Contrast (drawn vertically)



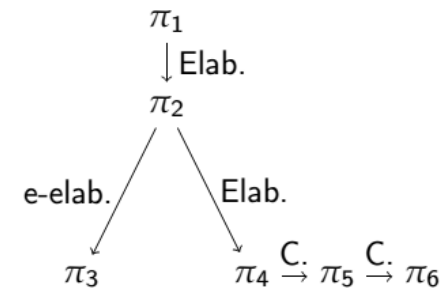
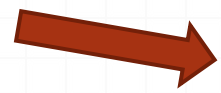
# Segmented Discourse Representation Theory (SDRT)

[Principes de la sélection naturelle.] $\pi_1$  [La théorie de la sélection naturelle [telle qu'elle a été initialement décrite par Charles Darwin,] $\pi_3$  repose sur trois principes :] $\pi_2$  [1. le principe de variation] $\pi_4$  [2. le principe d'adaptation] $\pi_5$  [3. le principe d'hérédité] $\pi_6$

*Principles of natural selection. The theory of natural selection, such as it was initially described by Charles Darwin, is based on three principles: 1. Principle of variation 2. Principle of adaptation 3. Principle of heredity*



computationally is very hard to find CDUs



Scopes here can be described as :  $C.(\pi_4, \pi_5) \wedge C.(\pi_5, \pi_6) \wedge Elab.(\pi_3, \pi_2) \wedge e - Elab.(\pi_3, \pi_2) \wedge Elab(\pi_3, [\pi_4, \pi_5, \pi_6]) \wedge Elab(\pi_1, [\pi_3, \pi_4, \pi_5, \pi_6])$ .



# SDRT - Right frontier constraint

0 **Right frontier constraint (RFC)** - A discourse constituent must be attached on the right frontier of the ongoing discourse. Polanyi (1985), Webber (1988) Asher (1993), Asher - Lascardes (2003)

0 Given a tree or a graph, an attachment cannot jump to a constituent on the left of the current one.

**!!Claimed to be intuitive!!**

John had a lovely evening

*Elaboration*

He had a great meal

*Narration*

He won a dancing competition

*Elaboration*

He ate salmon

*Narration*

He devoured cheese

It was a beautiful pink.

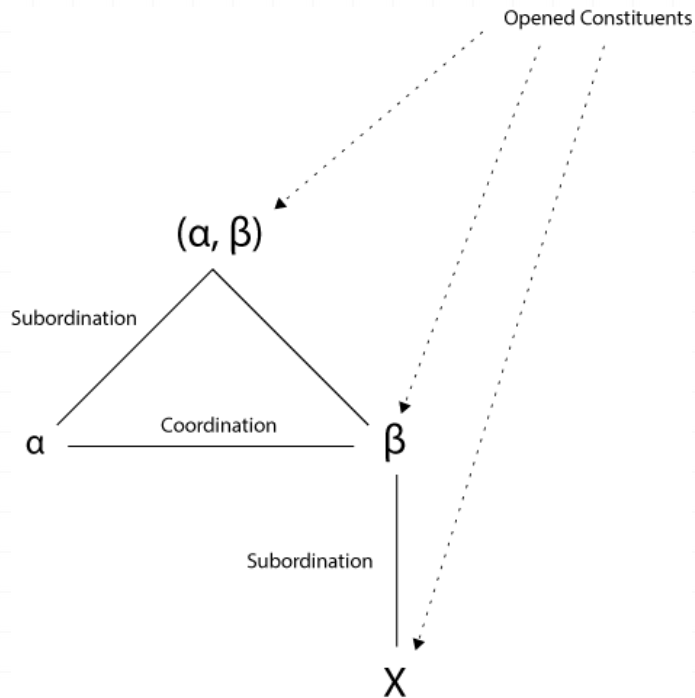


**Let  $\Delta$  be a discourse structure with  $\alpha$  the current (= last) constituent. A new constituent  $\beta$  can be attached by a DR to  $\gamma$  in  $\Delta$  only if:**

1.  $\gamma = \alpha$ , or
2.  $\alpha$  is subordinated to  $\gamma$ .

- a. John had a great evening last night.
- b. He had a great meal.
- c. He ate salmon.
- d. He devoured lots of cheese.
- e. He then won a dancing competition.
- f. ~~It (the salmon) was a beautiful pink.~~

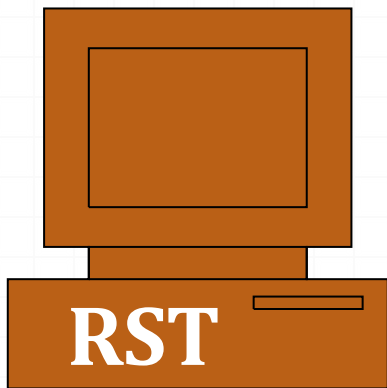
# SDRT - Right frontier constraint



- RFC requires that new discourse constituents cannot be integrated anywhere in the tree representing preceding discourse but only in a well-defined area situated at the Right Frontier of the tree.
- The constituents situated in this area are said to be *open* for attachment while all the others are said to be *closed* (Polanyi, 1988; Asher & Lascarides, 2003).

## Trees

tries to capture the intentions of the author as they are judged by the human annotator



long-distance attachments are not possible

## Graph

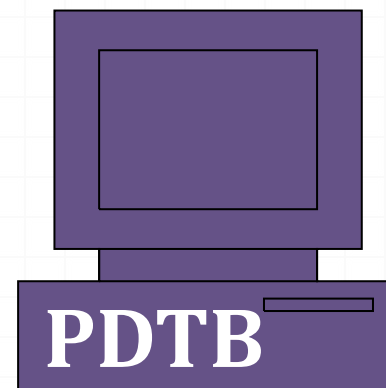
based on formal semantics



long-distance attachments are possible

## No assumption on text structure

based on lexically grounded relations



long-distance attachments are possible

# SDRT corpora

## 0 DISCOR

Brian Reese, Julie Hunter, Nicholas Asher, Pascal Denis and Jason Baldridge. Reference Manual for the Analysis and Annotation of Rhetorical Structure. 2007

## 0 ANNODIS

Afantenos S. D., Asher N., Benamara F., Bras M., Fabre C., Ho-Dac L.-M., Le Draoulec A. Muller P., Péry-Woodley M.-P., Prévot L., Rebeyrolle J., Tanguy L., Vergez-Couret M., Vieu L. (2012). An empirical resource for discovering cognitive principles of discourse organization: the ANNODIS corpus. *LREC 2012*, Istanbul, Turkey, July 2012.

## 0 GEOPO

Lydia-Mai Ho-Dac : La position initiale dans l'organisation du discours : une exploration en corpus, thèse de doctorat, Université de Toulouse-Le Mirail, novembre 2007

## 0 STAC

Stergos Afantenos Eric Kow Nicholas Asher Jérémy Perret. Discourse parsing for multi-party chat dialogues. Proceedings of EMNLP-2015, pp. 928–937, Lisbon.

# DISCOR (Discourse Structure and Coreference Resolution)

- 0 Goal: to test hypotheses about the interaction between discourse structure and the resolution of anaphoric links.
- 0 Data: MUC and ACE corpora (because already have coreference)
- 0 14 relations  
(cf. 78 in RST corpus)

Coordinating Relations		Subordinating Relations	
<i>Veridical</i>	<i>Nonveridical</i>	<i>Veridical</i>	<i>Nonveridical</i>
Continuation	Consequence	Background	Attribution
Narration	Alternation	Elaboration	
Result		Explanation	
Contrast		Commentary	
Parallel		Source	
Precondition			

# DISCOR (Discourse Structure and Coreference Resolution)

## EDU segmentation

- 0 ca. clauses, but not if embedded, e.g. do not segment wh-clause in  
Privately held Arnold, which had about \$750 million in billings and \$90.7 million in revenue last year, handles advertising for such major corporations as McDonald's Corp., ...
- 0 non-restrictive relative clauses introduce EDUs, unless doing so results in a discontinuous EDU
- 0 Many, but certainly not all, cases of syntactic subordination introduce a new elementary discourse unit. Complements of verbs of communication, for example, introduce EDUs (*say, note, announce, etc.*).
- 0 appositions are EDUs



# ANNODIS - (ANNOtation DIScursive)

- 0 project 2007-2010 – texts annotation and parsing
- 0 people: Nicholas Asher, Farah Benamara, Philippe Muller, Laure Vieu, Stergos Afantenos, etc.
- 0 an annotated corpus of French written texts (news, wikipedia, linguistic papers, reports) for the study of discourse organization – 86 stories, 687,000 words,
- 0 3 years of annotation
- 0 3188 Elementary Discourse Units (EDU) and 1395 Complex Discourse Units (CDU) linked by 3355 rhetorical relations (e.g. contrast, elaboration, result, attribution, etc.)
- 0 text annotated with various discourse phenomena
  - 0 bottom-up approach: applying a compositional and logical model of discourse organization (SDRT)
  - 0 top-down approach: starts from the text as a whole and focuses on the identification of configurations of cues signalling higher-level text segments, in an attempt to address the interplay of continuity and discontinuity within discourse (annotation of Enumerative Structures and Topical Chains)
- 0 annotation tool Glozz (will be shown by the STAC corpus)

# ANNODIS - (ANNOtation DIScursive)

alternation  
attribution  
background  
comment  
continuation  
contrast

entity-elaboration (e-elab)

elaboration  
explanation

flashback  
frame  
goal  
meta-relation

narration  
parallel  
result

temploc

## Rhetorical Relations

French

temporal?  
DISCOR

Coordinating Relations		Subordinating Relations	
<i>Veridical</i>	<i>Nonveridical</i>	<i>Veridical</i>	<i>Nonveridical</i>
Continuation	Consequence	Background	Attribution
Narration	Alternation	Elaboration	
Result		Explanation	
Contrast		Commentary	
Parallel		Source	
Precondition			

Muller P., Vergez M., Prevot L., Asher N., Benamara F., Bras M., Le Draoulec A. & Vieu L. (2012). Manuel d'annotation en relations de discours du projet Annodis, Carnets de grammaire 21, CLLE-ERSS.

- 17 relations
- were chosen because they are more or less common to all the theories of discourse, or correspond to well-defined subgroups in fine-grained theories
- The intermediate level of granularity was chosen as a compromise between informativeness and reliability of the annotation process. It corresponds to the level chosen in the PDTB, and a coarse-grained RST.



# SDRT corpora - STAC

0 running project: 2011-2016

0 ca 1100 negotiation dialogues, short texts (in English!) – much larger corpus than ANNODIS:

0 refined annotation tool Glozz

0 inter-annotator agreement better than for ANNODIS (basically because turns are annotated as EDU without further division)

The screenshot shows the Glozz 1.1.0-beta interface. The main window displays a dialogue with the following turns:

- 192 : amycharl : anyone want a sheep
- 193 : amycharl : [unclear]
- 194 : IG : sry
- 195 : sabercat : for what?
- 196 : amycharl : wheat preferably
- 197 : sabercat : dont have that :D
- 198 : amycharl : anything else is fine

Annotations are shown as colored boxes and lines connecting them. A red circle highlights a specific annotation point labeled "CDU". The right sidebar shows the "Relations" panel with a list of relation types: Anaphora, Question-answer, Result, Comment, Continuation, Conditional, Explanation, and Elaboration. Below this is a table with columns for "Sort/Type", "Sort/Date", "Show sel.", and "Visible".

Sort/Type	Sort/Date	Show sel.	Visible
u_Turn(1,37)	ID=1		
u_Turn(38,56)	ID=3		
u_Turn(57,71)	ID=5		
u_Turn(72,98)	ID=7		
u_Turn(99,132)	ID=9		

rennoc1

Points: 2

Soldiers: 0  
Resources: 5  
Dev. Cards: 0

Roads: 12  
Stimts: 3  
Cities: 4

Tomm  
Points: 4

L. Road

Soldiers: 1

Roads: 8

Game

\* It's rennoc1's turn to roll. Rolled a 11.  
\* rennoc1 gets 1 wheat. Tomm gets 1 wheat.

History

rennoc1: you know if you have more than 4, I think you can trade them into the bank.  
Dave: yeah but it's not ideal  
Tomm: That's true... 4 -> 1  
Tomm: Well, I might do, but depends on my roll, I'm afraid  
Tomm: Dave: Wheat for a clay?  
Dave: sure, can you do 2 for 2, or do you just want 1 for 1  
Tomm: just 1-4-1 I'm afraid  
Dave: fair enough  
Tomm: Oh... now I get wheat(!)  
Dave: heh

Chat



Road: Cost: 1 1

Settlement: Cost: 1 1 1 1

Game Options

Dave

Points: 2

Soldiers: 1  
Resources: 5  
Dev. Cards: 1

Roads: 11  
Stimts: 3  
Cities: 4

Sit Here

## STAC annotation

# STAC

## Two-level annotation:

segmentation of dialog turns into discourse units

labelling with domain-related speech acts  
(negotiation moves)

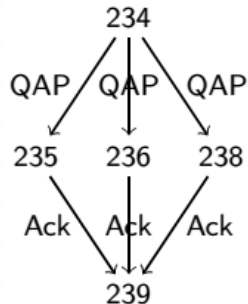
relational rhetorical annotation familiar from  
ANNODIS but with relations for dialogue.

- non-treelike structures certainly exist
- long distance crossing dependencies

reflected in parsers output

- domain level acts
  - offer : *I'll give you 2 clay for a rock*
  - counteroffer : *How about 2 clay for a wheat ?*
  - accept : *OK, it's a deal.*
  - refusal : *I don't think so.*
  - has-resource : *I have wheat*
  - strategic comment : *joel fancies a bit of your clay*
  - other (non relevant for negotiation)
- relational rhetorical annotation

234	gotwood4sheep	anyone got wheat for a sheep ?
235	inca	sorry, not me
236	CheshireCatGrin	nope. you seem to have lots of sheep !
237	gotwood4sheep	yup baaa
238	dmm	i think i'd rather hang on to my wheat i'm afraid
239	gotwood4sheep	kk i'll take my chances then...



# References

- 0 Asher, Nicholas (1993). *Reference to Abstract Objects in Discourse*. Dordrecht: Kluwer.
- 0 Asher, Nicholas and Lascarides, Alex (2003). *Logics of Conversation*. Cambridge: Cambridge U.P.
- 0 Asher, Nicholas . Discourse Parsing for Text and Dialogue [2013]. MELODI. IRIT, CNRS
- 0 Asher, Nicholas (with Antoine Venant, Philippe Muller, Pascal Denis, Stergos Afantenos, Farah Benamara). Annotating and learning models of discourse structure [2015]
- 0 Asher, Nicholas and Lascarides, Alex (1997). Bridging. *Journal of Semantics*
- 0 Cornish, Francis (1999). *Anaphora, Discourse and Understanding. Evidence from English and French*. Oxford: Oxford U.P.
- 0 Stergos D. Afantenos and Nicholas Asher [2010] Testing SDRT's Right Frontier . Proceedings of Coling 2010, pages 1–9,
- 0 Jacques Jayez. [2008]. The Right Frontier Constraint. ENS-LSH and L 2 C2
- 0 Lascarides, A. and N. Asher [2007] Segmented Discourse Representation Theory: Dynamic Semantics with Discourse Structure, in H. Bunt and R. Muskens (eds.) *Computing Meaning: Volume 3*, pp87--124, Springer.
- 0 Lascarides, A. Semantics and Pragmatics of NLP Segmented Discourse Representation Theory. SPNLP. (slides)
- 0 Nicholas Asher, Laure Vieu. Subordinating and coordinating discourse relations. *Lingua* 115 (2005), 591–610.
- 0 Philippe Muller, S. Afantenos, Pascal D., N, Asher. Constrained decoding for text-level discourse parsing [2012], Proceedings of COLING 2012: Technical Papers, pages 1883–1900,
- 0 Polanyi, Livia (1985). A theory of discourse structure and discourse coherence. *Proceedings of the 21st Meeting of the Chicago Linguistics Society*.
- 0 Brian Reese, Julie Hunter, Nicholas Asher, Pascal Denis and Jason Baldridge. [2007] DISCOR: Reference Manual for the Analysis and Annotation of Rhetorical Structure (Version 1.0)
- 0 Webber, Bonnie (1988). Discourse deixis and discourse processing. Technical Report MS-CIS-86-74, Department of Computer and Information Science, University of Pennsylvania.

# SDRT done!



# THEORIES & CORPORA

- 0 Halliday and Hasan – Cohesion in English ✓
- 0 RST – Rhetorical Structure Theor ✓
- 0 Discourse Graphbank ✓
- 0 PDTB – Penn Discourse Treebank ✓
- 0 SDRT - Segmented Discourse Representation Theory ✓
- 0 CCR – Cognitive approach to coherence relations ✓

# CCR (Cognitive approach to coherence relations)

Sanders, T. J. M., W. P. M. Spooren, and L. G. M. Noordman (1992). **Toward a taxonomy of coherence relations**. *Discourse Processes* 15, 1–35.



coherence relations are considered as cognitive entities (coherence relations and their linguistic marking affect the cognitive representation of discourse, e.g. text understanding, they are psychological entities rather than merely an analytic tool)

Proved by series of psycholinguistic experiments (Spooren, 1989; Haberlandt, 1982; Sanders 1986, e.g. linguistic marking appears to lead to faster processing of the following discourse segment).

# CCR (Sanders et al.)

## Investigating coherence relations

### relational account

- 0 Focus on the meaning of the relation and not on the meaning of each specific segment
- 0 Classification in terms of cognitive primitives

### language-based account

- 0 Focus not so much on relations themselves but rather on a study of linguistic devices that are used to signal relations explicitly



# CCR (Sanders et al.)

## Four cognitive primitives

*Although he worked hard, he failed the exam.*

discourse segment **S<sub>1</sub>**

expresses  
↓

proposition **P**

discourse segment **S<sub>2</sub>**

expresses  
↓

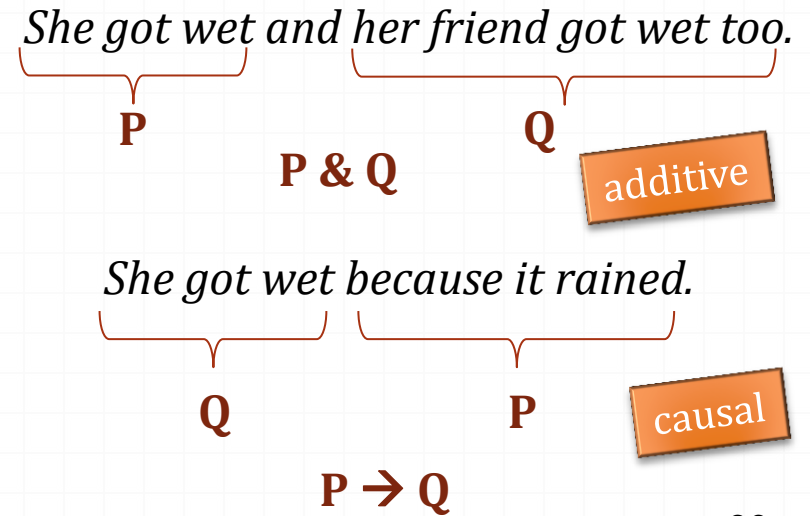
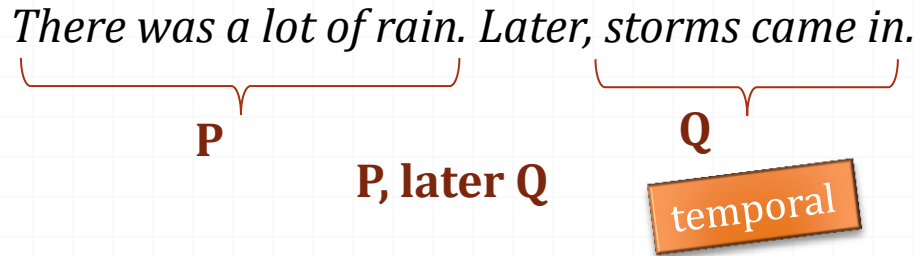
proposition **Q**

# CCR (Sanders et al.)

## Four cognitive primitives

0 basic operation (Additive – Temporal – Causal / Conditional)

originally (Sanders et al. 1992) only additive (weakly connected) vs. causal (strongly connected) were distinguished



# CCR (Sanders et al.)

## Four cognitive primitives

- 0 basic operation (Additive – Temporal – Causal / Conditional)
- 0 source of coherence (semantic/ pragmatic)

semantic (objective)

*She got wet because it rained.*



A relation is objective (semantic) when both segments happen in the real world and are thus facts. The speaker's opinion is not reflected in the relation.

*She is not at home, because her car is not there*



pragmatic (subjective)

A relation is subjective (pragmatic) if one or both segments express an opinion, argument, claim or conclusion of the speaker.

# CCR (Sanders et al.)

## Four cognitive primitives

- 0 basic operation (Additive – Temporal – Causal / Conditional)
- 0 source of coherence (semantic/ pragmatic)
- 0 order of segments (basic/non-basic)

*Although he worked hard, he failed the exam.*

P Q

A relation with a basic order has an antecedent, followed by a consequent. The antecedent is the cause of the argument, the consequent is the consequence or the claim.

basic

*She got wet because it rained.*

Q P

non-basic

A relation with a non-basic order has a consequent which precedes the antecedent. The cause or the claim thus precedes the cause of the argument.

# CCR (Sanders et al.)

## Four cognitive primitives

- 0 basic operation (Additive – Temporal – Causal / Conditional)
- 0 source of coherence (semantic/pragmatic)
- 0 order of segments (basic/non-basic)
- 0 polarity (positive/negative)

*She got wet because it rained.*

$S_1=Q$

$S_2=P$

positive

A relation is positive if the two discourse segments  $S_1$  and  $S_2$  function in the basic operation as antecedent (P) and consequent (Q) respectively.

*Although he worked hard, he failed the exam.*

$\text{not-}S_1=P$

$\text{not-}S_2=Q$

negative

A relation is negative if not  $S_1$  and  $S_2$ , but their negative counterparts, not- $S_1$  and not- $S_2$ , function in the basic operation.

# DiscAn corpora



## Towards a discourse annotation system for Dutch language corpora

- 0 DiscAn project – integrating existing corpora of Dutch discourse phenomena in the CLARIN infrastructure
  - 0 set of corpus analyses has been standardized (both in terms of raw data -- the texts -- and analyses) and opened up for further scientific research
- 0 text have been annotated for discourse phenomena during last ca 15 years at several universities in the Netherlands and Belgium
- 0 format: Excel tables, doc files, SPSS files etc. → brat?
- 0 visualization ANNIS

# DiscAn subcorpora

- 0 Degand 2.2: compiled for a study of the causal connectives "aangezien", "want" and "omdat" in Dutch news, 143 cases from a Dutch newspaper (NRC Handelsblad from 1994)

Degand, L. (2001). Form and function of causation: A theoretical and empirical investigation of causal constructions in Dutch. Leuven: Peeters.

- 0 PanderMaatSanders: causal connectives "daardoor", "daarom" and "dus" in Dutch news
- 0 Persoon Corpus: causal connectives "want" and "omdat" in Dutch spontaneous conversations, from the Corpus of Spoken Dutch

Persoon, I., Sanders, T., Quené, H. & A. Verhagen (2010). Een coördinerende omdat-constructie in gesproken Nederlands? Tekstlinguïstische en prosodische aspecten. Nederlandse Taalkunde, 15, 259-282.

- 0 SandersSpooren Corpus: causal connectives "want" and "omdat" in several types of discourse

# DiscAn subcorpora

Discourse phenomena	Author	Cases
Causal connectives	Bekker (2006)	500 explicit ( <i>doordat, want, dus, daarom, nadat, voordat</i> ) / 200 implicit
Causal connectives	Degand (2001)	150 ( <i>want, aangezien, omdat</i> ) from newspapers
Coherence relations	Den Ouden (2004)	70 (causal implicit, non-causal)
Connectives	Evers-Vermeul (2005)	600 historical data / 4400 from Childes
Causal connectives	Pander Maat & Degand (2001)	150 ( <i>dus, daarom</i> ) from newspaper corpora
Coherence relations	Pander Maat & Den Ouden (2011)	795 implicit and explicit relations from a self-assembled corpus of 40 press releases
Causal connectives	Pander Maat & Sanders (2000)	150 ( <i>dus, daarom, daardoor</i> ) from a newspaper-corpus (Volkskrant)
Causal connectives	Persoon (2010)	105 ( <i>omdat, want</i> ) from CGN
Causal connectives	Pit (2003)	200 ( <i>aangezien, omdat, doordat, want</i> ) newspaper / 100 ( <i>omdat, doordat, want</i> ) narrative; from newspaper (Volkskrant) and fictional books
Causal connectives	Sanders & Spooren (2009)	100 newspaper (Volkskrant) / 275 from CGN / 80 from Chat ( <i>want, omdat</i> )
Coherence relations	Sanders & van Wijk (1996)	100 childrens explanatory texts; ca. 1500 coherence relations
Coherence relations	Spooren & Sanders (2008)	1100 coherence relations (children elicit responses)
Causal connectives	Spooren et al. (2010)	275 ( <i>want, omdat</i> ) spoken, from CGN; 100 ( <i>want, omdat</i> ) written
Causal connectives	Stukker (2005)	300 ( <i>daardoor, daarom, dus</i> ) newspaper / 300 historical data ( <i>daarom, dus</i> )
Coherence relations	Vis (2011)	135 texts; 643 subjective relations
Connectives	Van Veen (2011)	1951 <i>waarom-</i> ( <i>why-</i> ) questions <b>96</b> their answers (Childes)



# DiscAn Annotation

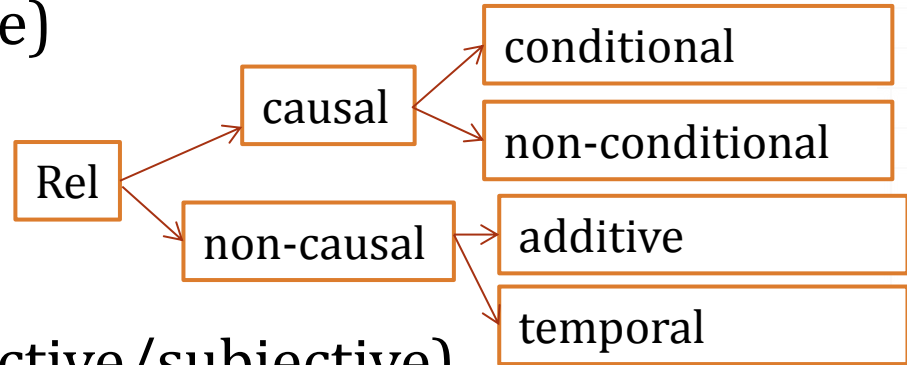
Four annotated categories:

0 Polarity (negative/positive)

0 Basic operation

0 source of coherence (objective/subjective)

0 order



BUT! Temporal and non-causal negative relations do not differ in source of coherence, because they are an objective representation of reality by nature.

BUT! Order for additive relations is not marked, because they are symmetric

Temporal relations can have a chronological order (basic order) and non-chronological (non-basic). Two events can also happen simultaneously. In this case, order is not annotated.

# References

- 0 Degand, L. (2001). Form and function of causation: A theoretical and empirical investigation of causal constructions in Dutch. Leuven: Peeters.
- 0 Persoon, I., Sanders, T., Quené, H. & A. Verhagen (2010). Een coördinerende omdat-constructie in gesproken Nederlands? Tekstlinguïstische en prosodische aspecten. Nederlandse Taalkunde, 15, 259-282.
- 0 Sanders, T.J.M. & Knott, A. (1998). The classification of coherence relations and their linguistic markers: An exploration of two languages. Journal of pragmatics, 30, (pp. 135-175) (41 p.).
- 0 Sanders, T. J. M., W. P. M. Spooren, and L. G. M. Noordman (1992). Toward a taxonomy of coherence relations. Discourse Processes 15, 1-35.
- 0 Ted Sanders. On the Cognitive and Text-Analytic Status of Coherence Relations.
- 0 Sanders, T.J.M., Vis, K. & Broeder, D. (2012). Project notes on the Dutch project DiscAn. Eighth Joint ACL - ISO Workshop on Interoperable Semantic Annotation Pisa.
- 0 Marianne Vergez-Couret, Laurent Prévot et Myriam Bras, « How Different Information Sources Interact in the Interpretation of Interleaved Discourse: The Case of Two-Step Enumerative Structures », Discours [En ligne], 11 | 2012, mis en ligne le 23 décembre 2012, consulté le 16 décembre 2015. URL : <http://discours.revues.org/8743> ; DOI : 10.4000/discours.8743

# CCR done!



# THEORIES & CORPORA

- 0 Halliday and Hasan – Cohesion in English
- 0 RST – Rhetorical Structure Theory
- 0 SDRT - Segmented Discourse Representation Theory
- 0 PDTB – Penn Discourse Treebank and PDTB-like
- 0 Discourse Graphbank
- 0 CCR – Cognitive approach to coherence relations



# So what?...



- 0 How to compare the theories / approaches / structures / corpora ...?
- 0 Can we translate from one corpus to another, thus extending the range of data available for performing automated tasks ?
- 0 Once you choose one of these theories you get the corpus that looks completely different than another corpus in another theory. (RST corpus looks completely different than PDTB)



## Comparative attempts

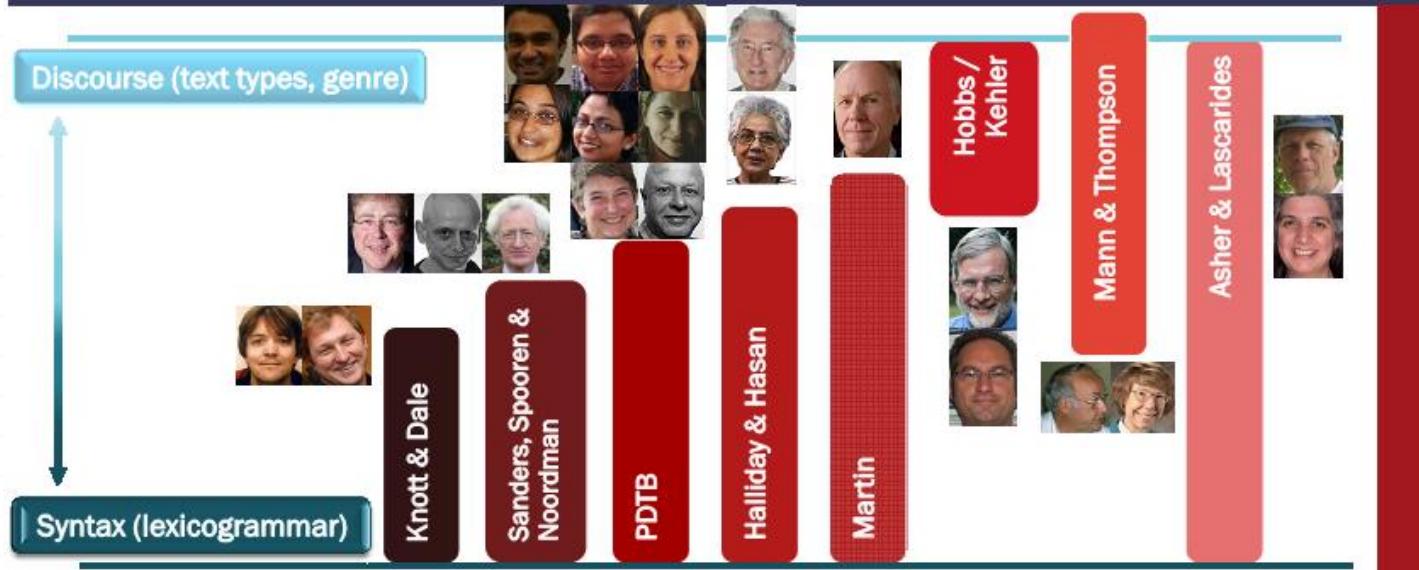
- 0 Taboada, M. (2015) Rhetorical relations are relations of coherence: What discourse coherence means, and how we can find it. Conference of the COST Action TextLink: Structuring Discourse in Multilingual Europe . Louvain-la-Neuve, Belgium. January 2015. (Plenary presentation)
- 0 Rehbein, Ines, Merel Scholman and Vera Demberg. Annotating discourse relations in spoken language: A comparison of the PDTB and CCR frameworks. DiSpoL meeting of TextLink, Saarbrucken, September, 2015.
- 0 Anja Nedoluzhko, Ekaterina Lapshinova, Kerstin Kunz: PDT vs. GECCo (2015: LAW-IX NAACL, TextLink meetings in Louvain-la-Neuve and Saarbrucken, LREC subm.)
- 0 Venant –Asher – Muller – Pascal – Afantenos: Expressivity and comparison of models of discourse structures, Sigdial 2013.

## Creating common standard


- 0 ISO standard for the annotation of semantic relations in discourse
- 0 applying Sanders' dimensions (April 2015, Fribourg)

Taboada, M. (2015) **Rhetorical relations are relations of coherence: What discourse coherence means, and how we can find it.** Conference of the COST Action TextLink: Structuring Discourse in Multilingual Europe . Louvain-la-Neuve, Belgium. January 2015.

## The space of coherence relations



# Comparative attempts

- 0 Taboada, M. (2015) Rhetorical relations are relations of coherence: What discourse coherence means, and how we can find it. Conference of the COST Action TextLink: Structuring Discourse in Multilingual Europe . Louvain-la-Neuve, Belgium. January 2015. (Plenary presentation) 
- 0 Rehbein, Ines, Merel Scholman and Vera Demberg. Annotating discourse relations in spoken language: A comparison of the PDTB and CCR frameworks. DiSpoL meeting of TextLink, Saarbrucken, September, 2015.
- 0 Anja Nedoluzhko, Ekaterina Lapshinova, Kerstin Kunz: PDT vs. GECCo (2015: LAW-IX NAACL, TextLink meetings in Louvain-la-Neuve and Saarbrucken, LREC subm.)
- 0 Venant –Asher – Muller – Pascal – Afantenos: Expressivity and comparison of models of discourse structures, Sigdial 2013.

# Creating common standard

- 0 ISO standard for the annotation of semantic relations in discourse
- 0 applying Sanders' dimensions (April 2015, Fribourg)



precedence - succession	reason - result	confrontation	conjunction
synchronous	pragmatic reason – result	opposition	instantiation
	purpose	pragmatic contrast	specification
	explication	restrictive opposition + exception	equivalence
	condition	concession	generalization
	pragmatic condition	correction (replacement)	conjunctive alternative
		gradation	disjunctive alternative
<b>TEMPORAL</b>	<b>CONTINGENCY</b>	<b>COMPARISON (CONTRAST)</b>	<b>EXPANSION</b>

discourse markers  
(attitude markers, modal particles) -  
in PDT not considered  
as connectives

TEMPORAL	CAUSAL	ADVERSATIVE	ADDITIVE	MODAL
temporal relation between events	relation of causality/dependence between	relation of contrast/alternative, for two events which are not true at the same time	relation of addition, for two events that are true/not true at the same time	relation between events connected by an evaluation of the speaker
<i>after, afterwards, at the same time..</i>	<i>because, therefore, that's why...</i>	<i>yet, although, by contrast...</i>	<i>and, in addition...</i>	<i>well, sure, of course, surely, eventually...</i>
<i>nachdem, danach, gleichzeitig...</i>	<i>weil, deshalb, aus diesem Grund...</i>	<i>doch, obwohl, im Gegensatz dazu...</i>	<i>und, außerdem...</i>	<i>klar, sicher, allerdings, jedenfalls, eigentlich, wohl...</i>

# Anja Nedoluzhko, Ekaterina Lapshinova, Kerstin Kunz: PDT vs. GECCo (2015)

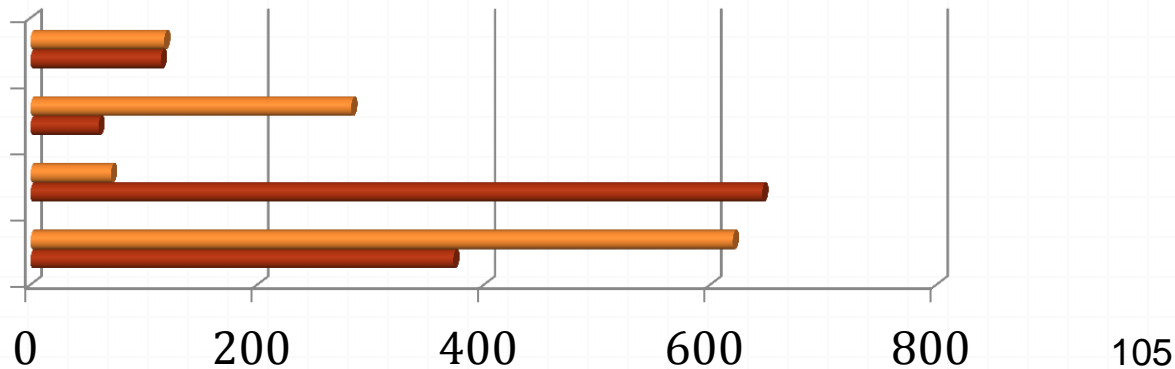


Prague Dependency Treebank (Prague)

German-English Contrasts in Cohesion (Saarland Uni)

■ PDT ■ GECCo

DSDs  
ellipsis  
bridging/lexical cohesion  
coreference



## Comparative attempts

- 0 Taboada, M. (2015) Rhetorical relations are relations of coherence: What discourse coherence means, and how we can find it. Conference of the COST Action TextLink: Structuring Discourse in Multilingual Europe . Louvain-la-Neuve, Belgium. January 2015. (Plenary presentation) ✓
- 0 Rehbein, Ines, Merel Scholman and Vera Demberg. Annotating discourse relations in spoken language: A comparison of the PDTB and CCR frameworks. DiSpoL meeting of TextLink, Saarbrucken, September, 2015.
- 0 Anja Nedoluzhko, Ekaterina Lapshinova, Kerstin Kunz: PDT vs. GECCo (2015: LAW-IX NAACL, TextLink meetings in Louvain-la-Neuve and Saarbrucken, LREC subm.) ✓
- 0 Venant –Asher – Muller – Pascal – Afantenos: Expressivity and comparison of models of discourse structures, Sigdial 2013.

## Creating common standard

- 0 ISO standard for the annotation of semantic relations in discourse
- 0 applying Sanders' dimensions (April 2015, Fribourg)

Ted Sanders, Let's try to make annotation systems communicate – **towards a systematic approach of coherence relations** . Fribourg, 2015.

## 0 Use CCR as an intermediate language

Basic Operation	Source of Coherence	Order	Polarity	Class	Relation
Causal	Semantic	Basic	Positive	1.	Cause–consequence
Causal	Semantic	Basic	Negative	2.	Contrastive cause–consequence
Causal	Semantic	Nonbasic	Positive	3.	Consequence–cause
Causal	Semantic	Nonbasic	Negative	4.	Contrastive consequence–cause
Causal	Pragmatic	Basic	Positive	5a.	Argument–claim
				5b.	Instrument–goal
				5c.	Condition–consequence
Causal	Pragmatic	Basic	Negative	6.	Contrastive argument–claim
Causal	Pragmatic	Nonbasic	Positive	7a.	Claim–argument
				7b.	Goal–instrument
				7c.	Consequence–condition
Causal	Pragmatic	Nonbasic	Negative	8.	Contrastive claim–argument
Additive	Semantic	—	Positive	9.	List
Additive	Semantic	—	Negative	10a.	Exception
				10b.	Opposition
Additive	Pragmatic	—	Positive	11.	Enumeration
Additive	Pragmatic	—	Negative	12.	Concession

from Sanders et al. 1992

### CCR – RST mapping

Basic op.	Source of coh.	Order	Polarity	CCR Relation	Additional criteria	RST Relation
Causal	Objective	Basic	Positive	Cause-consequence	+volitional	Volitional cause/result
					-volitional	Non-volitional cause/result
				Condition-consequence		Condition
Causal	Objective	Basic	Negative	Contrastive cause-consequence		Contrast
Causal	Objective	Non-basic	Positive	Consequence-cause	+volitional	Volitional cause/result
					-volitional	Non-volitional cause/result
				Consequence-condition		Condition
Causal	Objective	Non-basic	Negative	Contrastive consequence-cause		Contrast
Causal	Subjective	Basic	Positive	Argument-claim	+evaluation	Evaluation
					-evaluation	Interpretation
				Condition-claim		Condition
Causal	Subjective	Basic	Negative	Contrastive argument-claim		Anti-thesis
Causal	Subjective	Non-basic	Positive	Claim-argument	Content claim	Evidence
						Justify
				Claim-condition		Motivation
				Contrastive claim-argument		Condition
Causal	Subjective	Non-basic	Negative	Contrastive claim-argument		Anti-thesis
Additive	Objective	-	Positive	List	+temp order	Sequence
					-temp order -specification	Joint
					-temp order	
					+specification	Elaboration
						Restatement
						Summary
						Circumstance
						Background
Additive	Objective	-	Negative	Opposition		Contrast
				Exception		Contrast
Additive	Subjective	-	Positive	Enumeration		Presentational sequence
Additive	Subjective	-	Negative	Concession		Concession

from Sanders et al. 2015

Ted Sanders, Let's try to make annotation systems communicate – **towards a systematic approach of coherence relations** . Fribourg, 2015.

*Example 1:*

The door slammed because there is strong wind outside.

- RST: Non-Volitional cause / Explanation ?
- PDTB: (CONTINGENCY.Cause.) reason ?
- CCR: positive, causal, objective, non-basic (Consequence-cause)

*Example 3:*

John is tall but Fred is small.

- RST: Contrast ?
- PDTB: (COMPARISON.Contrast.)opposition ?
- CCR: Negative, additive, objective (opposition)

*Example 2:*

Max is a very good skier, because he won the competition twice last year.

- RST: Evidence ?
- PDTB:(CONTINGENCY.) Pragmatic cause / justification ?
- CCR: Positive, causal, subjective, non-basic (Claim-argument)

*Example 4:*

Jane married Mark even though she does not love him.

- RST: Antithesis ?
- PDTB: (COMPARISON.Concession.) contra-expectation ?
- CCR: negative, causal, objective, non-basic (Contrastive consequence-cause)

Rehbein, Ines, Merel Scholman and Vera Demberg. **Annotating discourse relations in spoken language: A comparison of the PDTB and CCR frameworks.** DiSpoL meeting of TextLink, Saarbrücken, September, 2015.

- 0 **idea:** Use CCR (Cognitive approach to Coherence Relations, Sanders, Spooren & Noordman 1992) as an intermediate language between different frameworks
- 0 **question:** To what extent can PDTB relations be analysed consistently using CCR dimensions?
- 0 2 annotators analyzed 1197 relations independently using PDTB 3.0 and CCR, respectively
- 0 Annotations mapped onto each other to investigate consistency of relation meanings across theories

Rehbein, Ines, Merel Scholman and Vera Demberg. **Annotating discourse relations in spoken language: A comparison of the PDTB and CCR frameworks.** DiSpoL meeting of TextLink, Saarbrücken, September, 2015.

- 0 Overall, 69% of the PDTB relations were consistently categorized as belonging to the target CCR class.
- 0 Analysis of random sample of 50 disagreements: 48% of disagreements due to differences between the theories.
- 0 Other disagreements due to difference in segmentation or interpretation of relation (14%) and to annotation errors (38%)

	<i>Polarity</i>	pos	pos	neg	neg	pos		
	<i>Basic op.</i>	temp	caus	caus	add	add		
	<i>S. of coh.</i>	uspec	uspec	uspec	uspec	uspec		
	<i>Order</i>	uspec	uspec	uspec	NA	NA	nra	count
Temp.	Asynchronous	75	3	0	11	11	0	36
Cont.	Cause	2	87	0	1	6	4	223
	Cause_belief	0	86	0	0	10	5	21
Comp.	Concession	0	0	57	37	6	0	54
	Contrast	1	4	14	75	4	2	161
Expan.	Conjunction	12	9	0	10	65	2	490
	Equivalence	0	26	0	4	47	23	47
	Instantiation	0	19	0	3	71	6	31
	Specification	1	32	0	5	55	7	109

Examples of differences between theories

- ▶ The connective 'but' indicates a negative relation in CCR, but not necessarily in PDTB.

(1) *She's by a Northern-based sire. [implicit but] I think he's dead now perhaps.*

- ▶ Argumentative relations classified as causal in CCR, but additive 'Expansion' in PDTB.

(2) *I used the weight room facility for exercising. [impl. because] I exercise from physiotherapy that I had to do.*

## Comparative attempts

- 0 Taboada, M. (2015) Rhetorical relations are relations of coherence: What discourse coherence means, and how we can find it. Conference of the COST Action TextLink: Structuring Discourse in Multilingual Europe . Louvain-la-Neuve, Belgium. January 2015. (Plenary presentation) ✓
- 0 Rehbein, Ines, Merel Scholman and Vera Demberg. Annotating discourse relations in spoken language: A comparison of the PDTB and CCR frameworks. DiSpoL meeting of TextLink, Saarbrücken, September, 2015. ✓
- 0 Anja Nedoluzhko, Ekaterina Lapshinova, Kerstin Kunz: PDT vs. GECCo (2015: LAW-IX NAACL, TextLink meetings in Louvain-la-Neuve and Saarbrücken, LREC subm.) ✓
- 0 Venant –Asher – Muller – Pascal – Afantenos: Expressivity and comparison of models of discourse structures, Sigdial 2013.

## Creating common standard

- 0 ISO standard for the annotation of semantic relations in discourse
- 0 applying Sanders' dimensions (April 2015, Fribourg) ✓



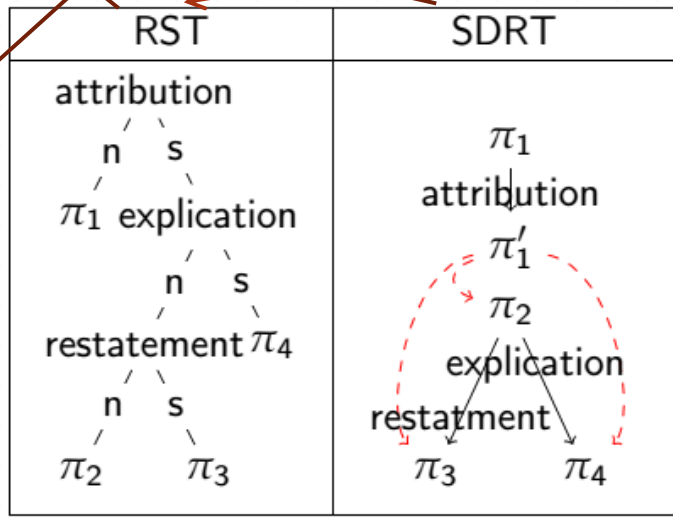
# RST vs. SDRT

Venant –Asher – Muller – Pascal – Afantenos: **Expressivity and comparison of models of discourse structures**, Sigdial 2013.

[Interprovincial Pipe Line Co. said] $\pi_1$  [it will delay a proposed two-step, 830 million dollar [(US\$705.6 million)] $\pi_3$  expansion of its system] $\pi_2$  [because Canada's output of crude oil is shrinking.] $\pi_4$

interested in relations

interested in scopes



Mixed Nuclearity Principle

- NS relations only transmit nucleus argument to a parent relation.
- $Restatement(\pi_2, \pi_3) \boxtimes Explanation(\pi_2, \pi_4) \boxtimes Attribution(\pi_1, \pi_2)$

immediately

- $Attribution(\pi_1, [\pi_2, \pi_3, \pi_4])$   
 $Restatement(\pi_2, \pi_3)$   
 $Explication(\pi_2, \pi_4)$

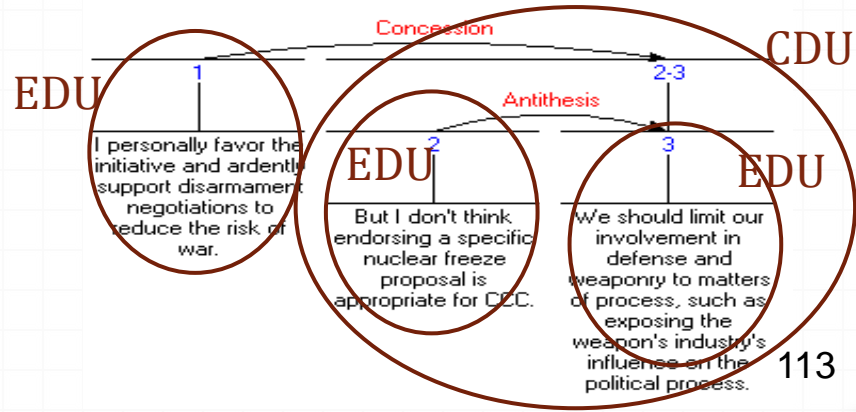
If RST is interpreted indirectly, we should be able to describe how the unpacking of the real arguments in the tree structure works!!!



# RST vs. SDRT

Venant -Asher - Muller - Pascal - Afantenos: **Expressivity and comparison of models of discourse structures**, SigDial 2013.

- 0 Comparisons require a language expressive enough to express semantic scopes for all discourse theories and interpretative mappings from the different structures into this language.
- 0 Venant et al. 2013 propose a language (formal semantic description) in which all theories can be described. Then we can look and compare the commonalities and restrictions, incl. decoding one from the other
- 0 Immediate interpretation of a RS-Tree can be decoded back to a SDRS (it gives lots of CDUs)
- 0 Relaxed Nuclearity Principle → RS Tree can be decoded into a set of SDRSs
- 0 The same with dependency trees
- 0 **DP and RST-tress are both underspecified version of a fully specified structure 😊**



## Comparative attempts

- Taboada, M. (2015) Rhetorical relations are relations of coherence: What discourse coherence means, and how we can find it. Conference of the COST Action TextLink: Structuring Discourse in Multilingual Europe . Louvain-la-Neuve, Belgium. January 2015. (Plenary presentation) ✓
- Rehbein, Ines, Merel Scholman and Vera Demberg. Annotating discourse relations in spoken language: A comparison of the PDTB and CCR frameworks. DiSpoL meeting of TextLink, Saarbrucken, September, 2015. ✓
- Anja Nedoluzhko, Ekaterina Lapshinova, Kerstin Kunz: PDT vs. GECCo (2015: LAW-IX NAACL, TextLink meetings in Louvain-la-Neuve and Saarbrucken, LREC subm.) ✓
- Venant –Asher – Muller – Pascal – Afantenos: Expressivity and comparison of models discourse structures, Sigdial 2013. ✓

## Creating common standard

- ISO standard for the annotation of semantic relations in discourse
- applying Sanders' dimensions (April 2015, Fribourg) ✓

Rashmi Prasad and Harry Bunt. **Semantic Relations in Discourse: The Current State of ISO 24617-8**. Proceedings of the 11th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-11). 2015



compared approaches



created an ISO, very close to current PDTB 3.0 version

- 0 Types of objects connected by discourse relations,
- 0 explicit connective/implicit/AltLex
- 0 representation of discourse structure
- 0 semantic description of discourse relations
- 0 pragmatic variants of discourse relations
- 0 hierarchical classification of discourse relations
- 0 representation of (a)symmetry of relations
- 0 relative importance of arguments for text meaning/structure

RST  
DGB  
PDTB  
SDRT  
CCR  
H&H



ISO 24617-8

# Types of objects connected by discourse relations

RST

A span of text, usually a clause, but in general ranging from minimally a (nominalization) NP to maximally a sentence.

PDTB

PDTB 2.0 - sentences (or where were connectives), PDTB 3.0 - also intrasentential, incl. free adjuncts, VP-coordination without connectives, etc.

DGB

=RST

*Constraints are placed on the semantic nature of arguments rather than on their syntactic form. An argument of a discourse relation must denote a certain type of abstract object. The ISO scheme remains neutral on this issue and does not specify any constraints on the extent or adjacency of argument realizations*

ISO 24617-8

SDRT

embedded, min. clause

CCR

text segments,  
minimally clauses

H&H

sentences



compared approaches



created an ISO, very close to current PDTB 3.0 version

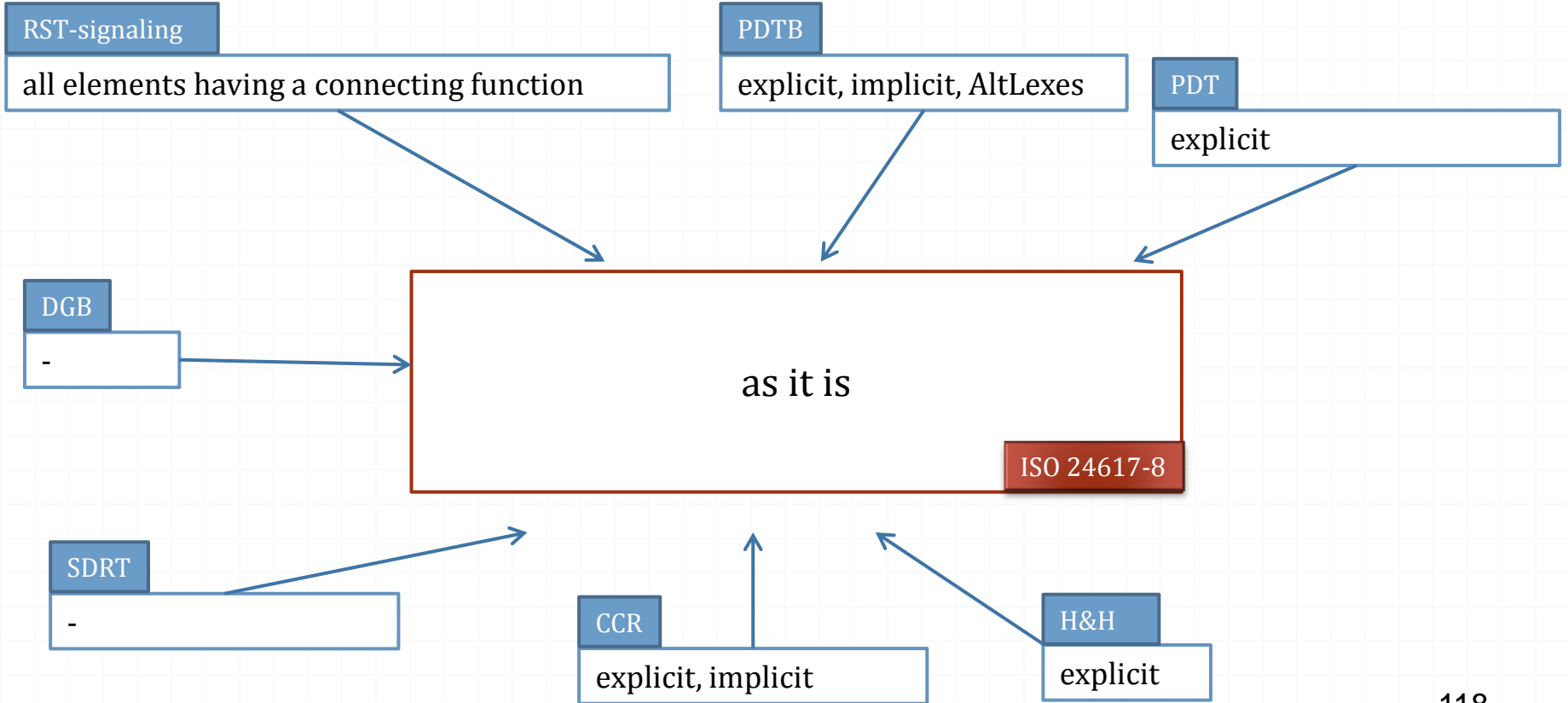
- 0 Types of objects connected by discourse relations ✓
- 0 explicit connective/implicit/AltLex
- 0 representation of discourse structure
- 0 semantic description of discourse relations
- 0 pragmatic variants of discourse relations
- 0 hierarchical classification of discourse relations
- 0 representation of (a)symmetry of relations
- 0 relative importance of arguments for text meaning/structure

RST  
DGB  
PDTB  
SDRT  
CCR  
H&H



ISO 24617-8

# Explicit connectives/Implicit/AltLex





compared approaches



created an ISO, very close to current PDTB 3.0 version

- 0 Types of objects connected by discourse relations, ✓
- 0 explicit connective/implicit/AltLex ✓
- 0 representation of discourse structure
- 0 semantic description of discourse relations
- 0 pragmatic variants of discourse relations
- 0 hierarchical classification of discourse relations
- 0 representation of (a)symmetry of relations
- 0 relative importance of arguments for text meaning/structure

RST  
DGB  
PDTB  
SDRT  
CCR  
H&H



ISO 24617-8

# Representation of Discourse Structure

RST

a tree representation to subsume the complete text of the discourse

PDTB

relations along with their arguments are annotated without being combined to form a structure that encompasses the entire text

DGB

allow general graphs that allow multiple parents and crossing

*ISO takes a pre-theoretical stance involving low-level annotation of discourse; individual relations can then be annotated further to project a higher-level tree or graph structure, depending on one's theoretical preferences*

ISO 24617-8

SDRT

allow directed acyclic graphs that allow for multiple parents, but not for crossing

CCR

relations along with their arguments are annotated without being combined to form a structure that encompasses the entire text

H&H

relations along with their arguments are annotated without being combined to form a structure that encompasses the entire text





compared approaches



created an ISO, very close to current PDTB 3.0 version

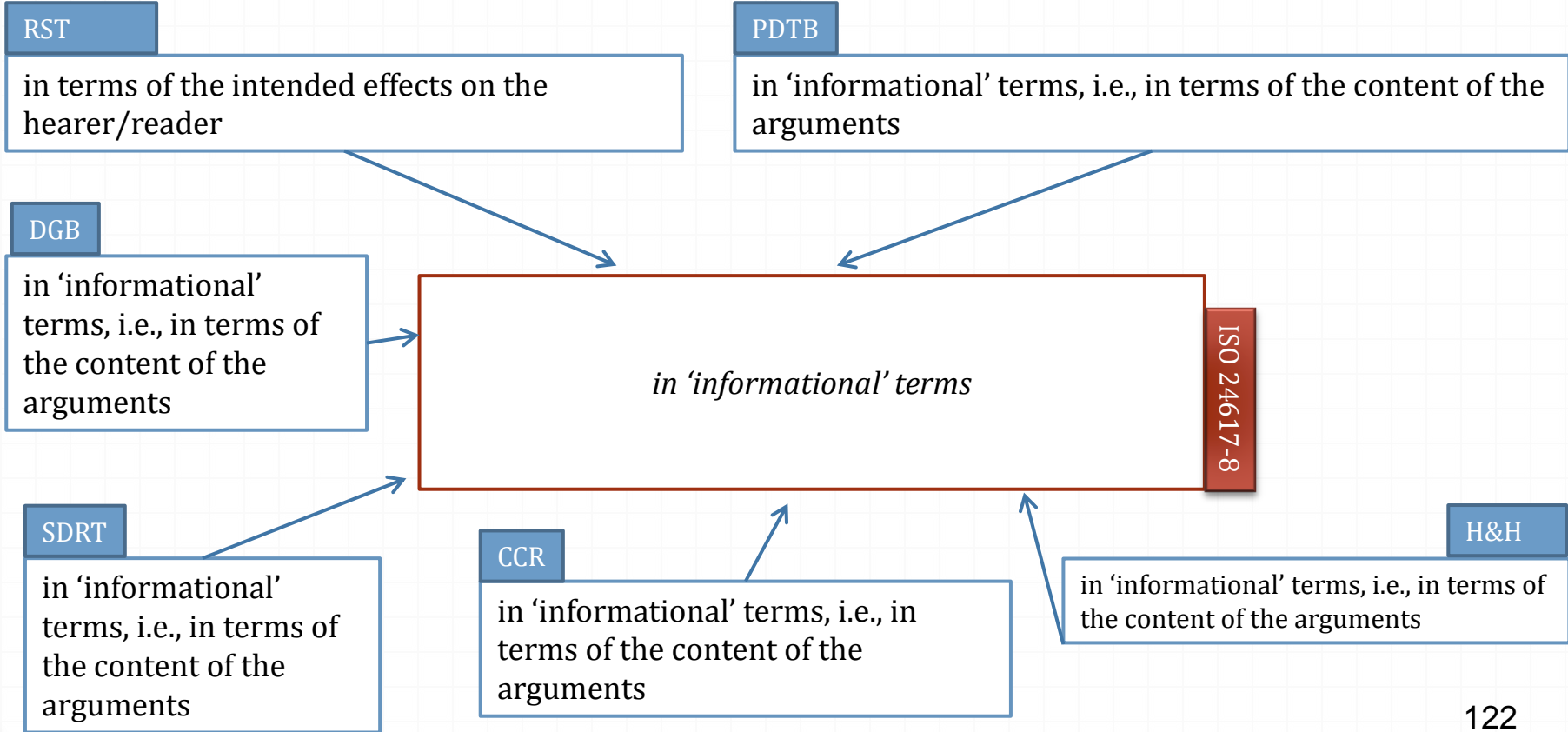
- 0 Types of objects connected by discourse relations, ✓
- 0 explicit connective/implicit/AltLex ✓
- 0 representation of discourse structure ✓
- 0 semantic description of discourse relations
- 0 pragmatic variants of discourse relations
- 0 hierarchical classification of discourse relations
- 0 representation of (a)symmetry of relations
- 0 relative importance of arguments for text meaning/structure

RST  
DGB  
PDTB  
SDRT  
CCR  
H&H



ISO 24617-8

# Semantic Description of Discourse Relations





compared approaches



created an ISO, very close to current PDTB 3.0 version

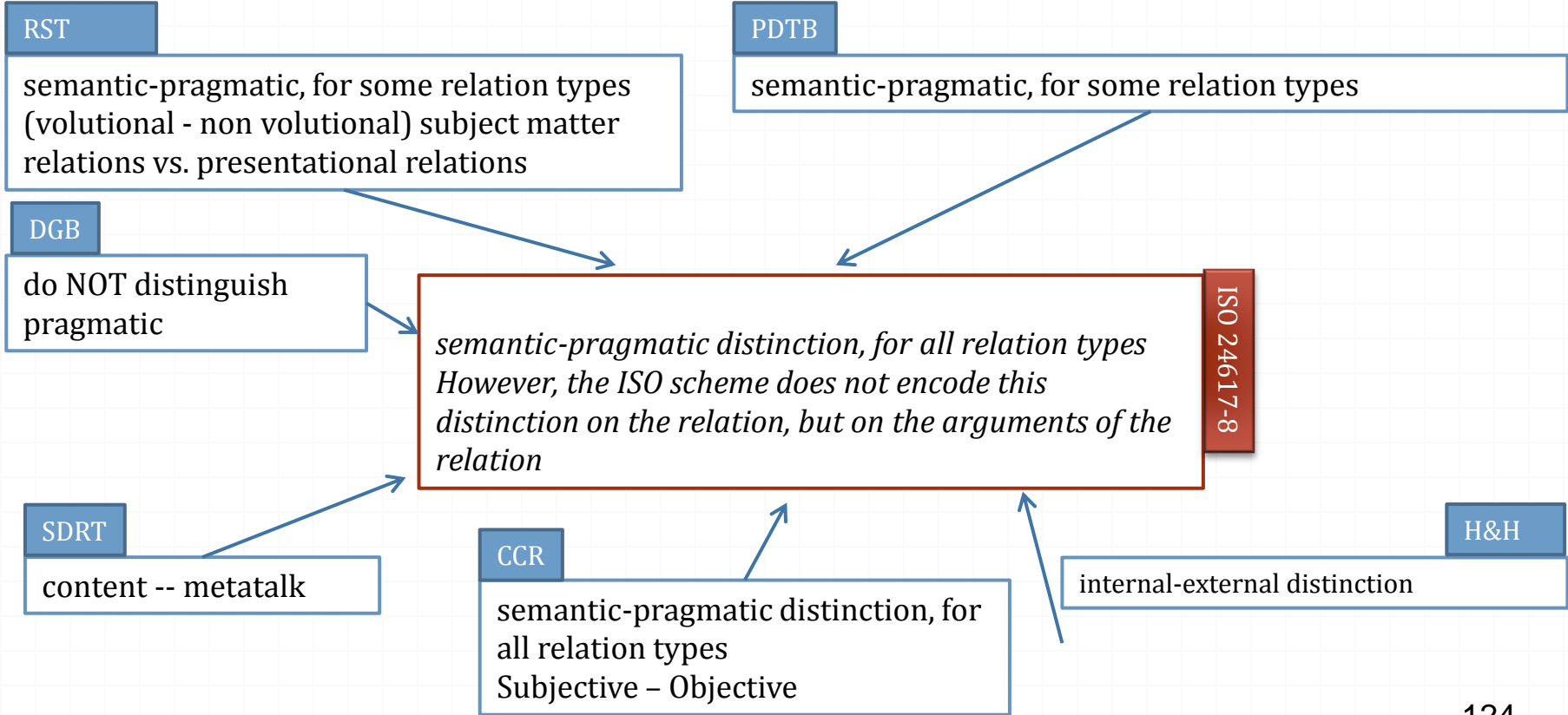
- 0 Types of objects connected by discourse relations, ✓
- 0 explicit connective/implicit/AltLex ✓
- 0 representation of discourse structure ✓
- 0 semantic description of discourse relations ✓
- 0 pragmatic variants of discourse relations
- 0 hierarchical classification of discourse relations
- 0 representation of (a)symmetry of relations
- 0 relative importance of arguments for text meaning/structure

RST  
DGB  
PDTB  
SDRT  
CCR  
H&H



ISO 24617-8

# Pragmatic Variants of Discourse Relations





compared approaches



created an ISO, very close to current PDTB 3.0 version

- 0 Types of objects connected by discourse relations, ✓
- 0 explicit connective/implicit/AltLex ✓
- 0 representation of discourse structure ✓
- 0 semantic description of discourse relations ✓
- 0 pragmatic variants of discourse relations ✓

0 hierarchical classification of discourse relations

0 representation of (a)symmetry of relations

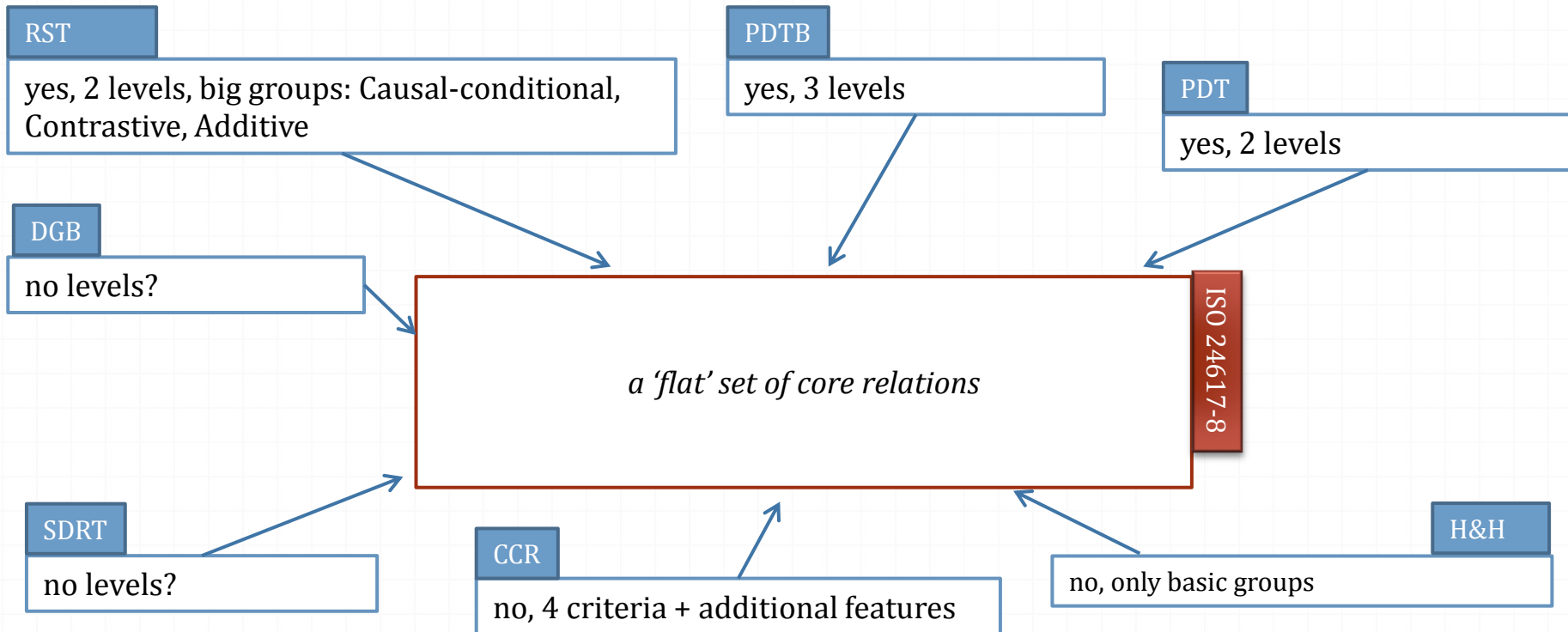
0 relative importance of arguments for text meaning/structure

RST  
DGB  
PDTB  
SDRT  
CCR  
H&H



ISO 24617-8

# Hierarchical Classification of Discourse Relations





compared approaches



created an ISO, very close to current PDTB 3.0 version

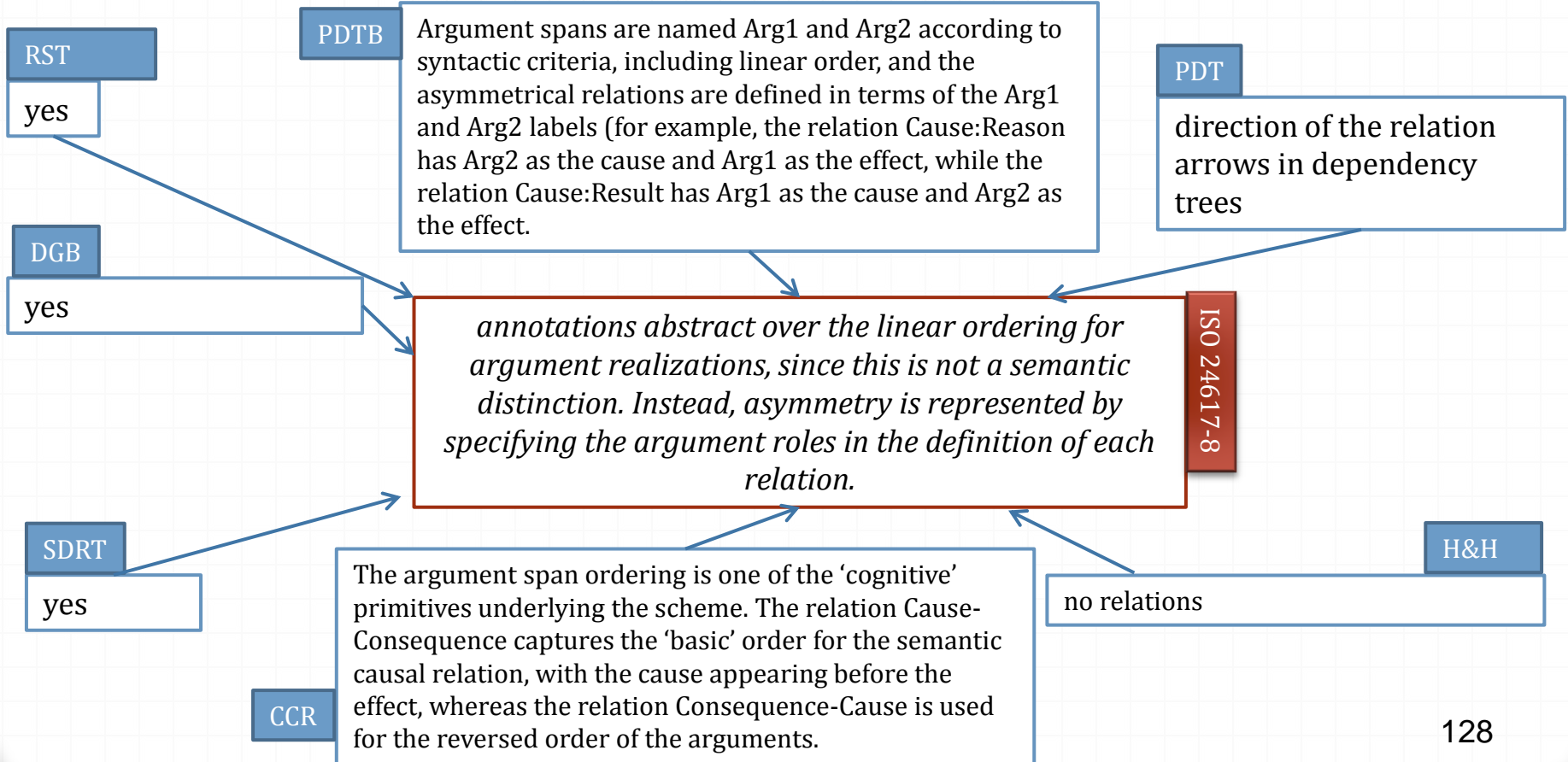
- 0 Types of objects connected by discourse relations, ✓
- 0 explicit connective/implicit/AltLex ✓
- 0 representation of discourse structure ✓
- 0 semantic description of discourse relations ✓
- 0 pragmatic variants of discourse relations ✓
- 0 hierarchical classification of discourse relations ✓
- 0 representation of (a)symmetry of relations
- 0 relative importance of arguments for text meaning/structure

RST  
DGB  
PDTB  
SDRT  
CCR  
H&H



ISO 24617-8

# Representation of (a)symmetry of Relations







compared approaches



created an ISO, very close to current PDTB 3.0 version

- 0 Types of objects connected by discourse relations, ✓
- 0 explicit connective/implicit/AltLex ✓
- 0 representation of discourse structure ✓
- 0 semantic description of discourse relations ✓
- 0 pragmatic variants of discourse relations ✓
- 0 hierarchical classification of discourse relations ✓
- 0 representation of (a)symmetry of relations ✓
- 0 relative importance of arguments for text meaning/structure

RST  
DGB  
PDTB  
SDRT  
CCR  
H&H



ISO 24617-8

# Relative Importance of Arguments for Text meaning/structure

RST

yes (nuclear-satellite)

DGB

yes, 'dominance', deriving a single assertion from a discourse relation that connects two segments, and distinguishing relations in terms of how this single assertion should be derived. In subordinating relations, in particular, the assertion associated with the relation is obtained from the 'dominant' segment, as specified in the relation definitions.

PDTB

Argument spans are named Arg1 and Arg2 according to syntactic criteria, including linear order, and the asymmetrical relations are defined in terms of the Arg1 and Arg2 labels (for example, the relation Cause:Reason has Arg2 as the cause and Arg1 as the effect, while the relation Cause:Result has Arg1 as the cause and Arg2 as the effect).

PDT

direction of the relation arrows in dependency trees

*the relative role of arguments for the text (meaning or structure) as a whole is not represented directly*

ISO 24617-8

SDRT

yes, classifies a relation as 'subordinating' or 'coordinating', depending on what structural configuration the arguments create in the discourse graph (Asher and Vieu, 2005)

H&H

no arguments

no

CCR

130

# Not addressed in **ISO 24617-8**

- 0 How many discourse functions can a single unit have?  
*We bought the apartment, but then we rented it out. RST – one, PDTB-like, SDRT, DGB: both, other theories?*
- 0 Can substructures of discourse serve as arguments of coherence relations? (e.g. yes for SDRT)
- 0 Attachment principles - where can we attach a new information:
  - 0 attachment to the last in a queue, to a preceding sentence (dynamic semantics)
  - 0 long distance dependencies (RST-treebank, SDRT - prove that there are long-distance dependencies: 35-40% of the attachments are outside the sentence and beyond the previous sentence) - there are some principles how you can attach, graph, positions of the graph, primary attachment portions (right frontier constraint)
  - 0 dialogues give interesting constraints (to backwards attachment - *Although I'm tired, I'm enjoying it here*)

**Is that all? Of course not, but ...**



**Thank you**

**for attention!**

# Acknowledgements

- 0 I would like to thank my colleagues, whose materials published in articles and slides I used: Nicholas Asher, Vera Demberg, Markus Egg, Annemarie Friedrich, Jacques Jayez, Alex Lascarides, Lynn Carlson, Daniel Marcu, Gisela Redeker, Ines Rehbein, Ted Sanders, Merel Scholman, Manfrede Stede, Maite Taboada, etc.
- 0 Many thanks to my discourse colleagues (Eva Hajičová, Lucie Poláková, Pavlína Jínová, Šárka Zikánová, Jiří Mírovský, etc.)
- 0 The research was supported from the Grant Agency of Czech Republic (grant GA 16-05394S, Structure of coreferential chains in parallel language data )