



Background

Annotating Discourse Relations (DRs)

- ▶ Many different frameworks: PDTB, RST, DRT, SDRT, ...
- ▶ No unified scheme, no interoperability

Idea:

- ▶ Use CCR (Cognitive approach to Coherence Relations, Sanders, Spooren & Noordman 1992) as an intermediate language between different frameworks

PDTB vs. CCR

- ▶ **PDTB**
 - ▶ Hierarchical scheme with three layers/43 sense labels
- ▶ **CCR**
 - ▶ each DR is described according to 4 cognitive primitives: *polarity, basic operation, source of coherence, order*

Mapping PDTB – CCR

Question

To what extent can PDTB relations be analysed consistently using CCR dimensions?

Method:

- ▶ 2 annotators analysed 1197 relations independently using PDTB 3.0 and CCR, respectively
- ▶ Annotations mapped onto each other to investigate consistency of relation meanings across theories

Results:

	<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	

Fig. 1: Distribution (%) of explicit and implicit relations, only labels where $n > 20$ (uspec: underspecified, NA: not applicable, nra: no relation annotated)

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

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.*

Spoken and written genres

- ▶ Penn Discourse Treebank (Prasad et al., 2008)
- ▶ SPICE-Ireland (Kallen & Kirk, 2012)

	<i>written</i>				<i>spoken</i>	
genre	essays	summaries	letters	news	broadcast	telephone
no. sent	6.517	1.667	911	38.963	1.507	2.717
no. words	139.445	31.316	18.207	821.104	24.609	19.707

Fig. 2: Data: 4 WSJ genres (Webber 2009) and 2 SPICE genres

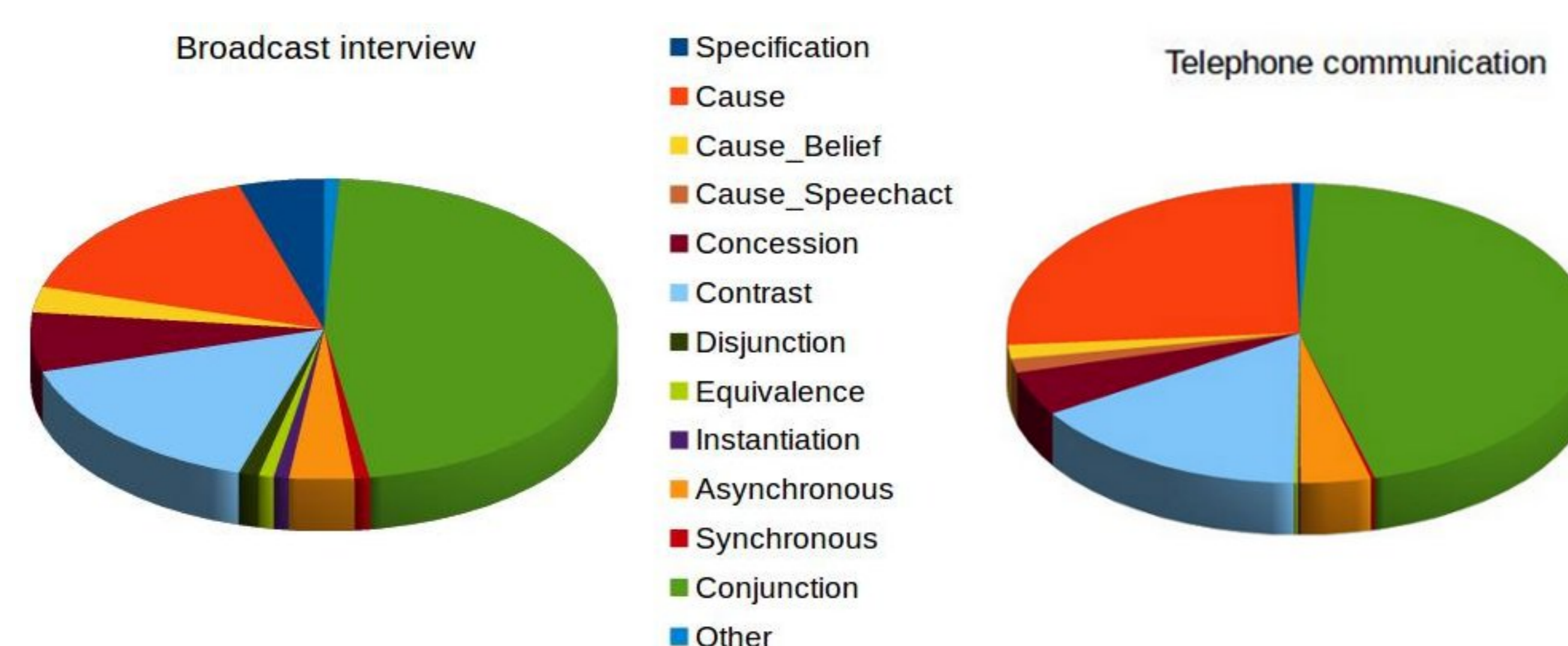
Examples

- ▶ Spoken: more **Cause.Result** relations
 - (3) *I ordered it so I'd better not forget about it* (Telephone)
- ▶ Spoken: more **Expansion.Conjunction** relations
 - (4) *Indeed I would be and I believe that there's tremendous goodwill out there* (Broadcast)
- ▶ Written: more **Comparison.Contrast** relations
 - (5) *A few blue-chip stocks posted strong gains while the majority of shares ended little changed* (WSJ)

Problems for annotating spoken language

- ▶ No clear-cut distinction between connectives and discourse markers (esp. for *and, but, so*)
 - (6) *But you know. So listen, enough of me.*
- ▶ PDTB: no sense for reinforcement relations
 - (7) *That's finished. That story is finished.*
- ▶ PDTB: no difference between conclusions and causal effects
 - (8) *She's a bay mare, fifteen two hands high so she's not very big*

Differences between Broadcast and Telephone



- ▶ **Specification** more frequent in broadcast
- ▶ **Cause** more frequent in telephone conversation

Cognitive dimensions and speech acts

- Representatives:** speaker is committed to the truth of the proposition
 - Directives:** speaker attempts to get the hearer to do something
 - Expressives:** speaker's attitudes/emotions towards the proposition
- ▶ Over 90% of DRs in Representatives in SPICE
 - ▶ Similar distribution across the 3 most frequent speech acts

References

- Kallen, J.L. & Kirk, J.M. (2012). SPICE-Ireland: A User's Guide. Belfast: CI Ollscoil na Banróna.
- Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., & Webber, B. (2008). The Penn Discourse Treebank 2.0. *Proceedings of LREC '08*, Marrakech.
- Sanders, T.J.M., Spooren, W.P.M.S., & Noordman, L.G.M. (1992). Toward a taxonomy of coherence relations. *Discourse Processes*, 15: 1-35.
- Searle, J. (1976). A classification of illocutionary acts. *Language in Society* 5:1-23.
- Webber, B. (2009). Genre distinctions for discourse in the Penn TreeBank. *Proceedings of ACL '09*, Singapore.