Let's not lose any information: mapping discourse relations

Vera Demberg
Universität des Saarlandes, Germany

WG2/WG3 meeting Fribourg

What are our goals?

Goals and use cases:

- **language learners and translators:** easily identifiable advice on how a discourse connector translates

- **NLP:** more resources, being able to adapt tools to another language more easily

- **language science:** crosslingual studies
    - check how some discourse relation is marked in another language
    - on a larger scale, compare how discourse relations are marked in one language vs. another
    - check your hypotheses about discourse relation usage and marking in different languages etc.

- **the PORTAL:** one can put in one relation in one language / framework and query for the same relation in other resources (plus information about known mismatches!)
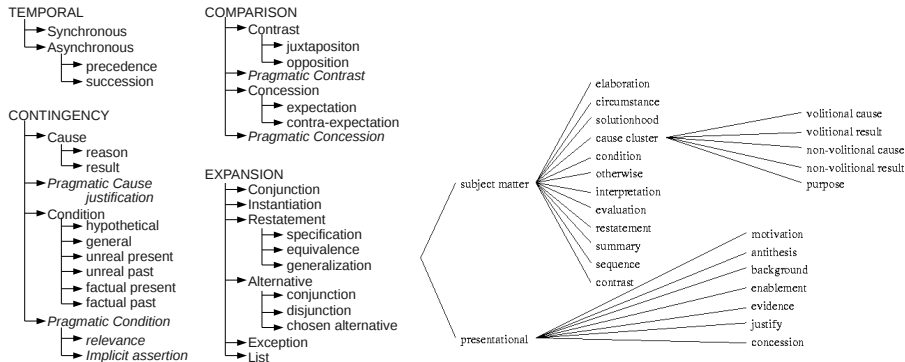
Current state of annotation schemes

TEMPORAL
- Synchronous
- Asynchronous
  - precedence
  - succession

CONTINGENCY
- Cause
  - reason
  - result
- *Pragmatic Cause*
  - *justification*
- Condition
  - hypothetical
  - general
  - unreal present
  - unreal past
  - factual present
  - factual past
- *Pragmatic Condition*
  - *relevance*
  - *Implicit assertion*

COMPARISON
- Contrast
  - juxtapositon
  - opposition
- *Pragmatic Contrast*
- Concession
  - expectation
  - contra-expectation
- *Pragmatic Concession*

EXPANSION
- Conjunction
- Instantiation
- Restatement
  - specification
  - equivalence
  - generalization
- Alternative
  - conjunction
  - disjunction
  - chosen alternative
- Exception
- List

subject matter
- elaboration
- circumstance
- solutionhood
- cause cluster
  - volitional cause
  - volitional result
  - non-volitional cause
  - non-volitional result
  - purpose
- condition
- otherwise
- interpretation
- evaluation
- restatement
- summary
- sequence
- contrast

presentational
- motivation
- antithesis
- background
- enablement
- evidence
- justify
- concession

Table 1
Contentful conjunctions used to illustrate coherence relations.

Cause–effect	because; and so
Violated expectation	although; but; while
Condition	if … (then); as long as; while
Similarity	and; (and) similarly
Contrast	by contrast; but
Temporal sequence	(and) then; first, second, …; before; after; while
Attribution	according to …; … said; claim that …; maintain that …; stated that …
Example	for example; for instance
Elaboration	also; furthermore; in addition; note (furthermore) that; (for, in, on, against, with, …) which; who; (for, in, on, against, with, …) whom
Generalization	in general

Across languages

annotation efforts in other languages might

- add relations / distinctions
- modify the annotation scheme
- what do we want to mark? (between-clausal? nominalizations?)

Example: Porting PDTB to Turkish

Modifier Class	Example	Gloss	Count	Percent
Focus	dA	focus particle (FP)	265	49.07
Temporal	*üç gün* sonra	*three days* later	170	31.48
Intensifier	*tam* aksine	*just* to the contrary	26	4.81
Counterfactuality	*sanki* … gibi	as though	25	4.63
Epistemic	*belki de* bunun için	*perhaps FP* because of this	17	3.15
Interrogative	bu yüzden *mi*	*is this* the reason	14	2.59
Quantifier	*bütün* bunlara rağmen	despite *all* these	9	1.67
Condition	*ancak* bundan sonra	*only* after this	5	0.93
Negation	için *değil*	*not* because of this	5	0.93
Qualifier	*çarpıcı* örnek olarak	as a *striking* example	3	0.56
Pragmatic	*peki* o zaman	well, *ok* then.	1	0.19
Total			540	100.00

Table 4. The frequency of the modifier tags in the TDB

Zeyrek, Deniz, et al. "Turkish Discourse Bank: Porting a discourse annotation style to a morphologically rich language." Dialogue & Discourse 4.2 (2013): 174–184.

Portal use cases

the portal will be most useful, if we can give as much info as possible about what is returned from each resource

- ▶ is a "superset" returned from the point of view of the question?
- ▶ what qualifies that superset?

Task: query for chosen_alternative in German

want to find other language examples of PDTB chosen alternative

It's time *business leaders and the general public learn* **that mankind does not rule over this natural environment but is** rather **the integral, symbiotic player within nature's workings**.

Conn/AltLex   Conn/AltLex Attr   Arg1   Arg1 Attr   Arg2   Arg2 Attr   Sup1   Sup2

in Potsdam Commentary Corpus: annotated as contrast
Immer mehr verantwortungslose Zeitgenossen versuchen, ihren Müll illegal loszuwerden statt ihn ordnungsgemäß zu entsorgen.

in RST (Marcu 1999): annotated as preference
Rather than go there by air, I'd take the slowest train.

Portal use cases

the portal will be most useful, if we can give as much info as possible about what is returned from each resource

- ▶ is a "superset" returned from the point of view of the question?
- ▶ what qualifies that superset?

Task: query for chosen_alternative in German

- ▶ are several subsets returned? What distinction does that other resource make?

Task: want to find causals!

find volitional and non-volitional causals.

She went home early because she promised her husband she would.
"Ze kwam vroeg thuis omdat ze haar man beloofd had dat ze dat zou doen."

She arrived home early because her plane landed early.
"Ze kwam vroeg thuis doordat haar vliegtuig eerder dan gepland was geland."

Portal use cases

the portal will be most useful, if we can give as much info as possible about what is returned from each resource

- ▶ is a "superset" returned from the point of view of the question?
- ▶ what qualifies that superset?

Task: query for chosen_alternative in German

- ▶ are several subsets returned? What distinction does that other resource make?

Task: want to find causals!

- ▶ both explicit and implicit ones returned?
- ▶ examples of relations between full sentences / clauses / NPs / ..?

Example

An example for nominalized clauses based on the factive –DIK[8]:

(a) **Üzül-DÜĞ-Ü kadar** *şaşırmıştı da.*

   She/He was surprised **as much as** *she/he was saddened* [sad-PASS-DIK].

How can we achieve a mapping?

How can we achieve a mapping?

- ▶ definitions must be compatible.
- ▶ instructions must be clear so that annotation is consistent.
- ▶ we need to know about cases where two schemes would differ.

Definitions

Example: Concession

PDTB The type Concession applies when the connective indicates that one of the arguments describes a situation A which causes C, while the other asserts (or implies) ¬C. (Then goes on to distinguish expt vs. contra-expt.)

RST The situation indicated in the nucleus is contrary to expectation in the light of the information presented in the satellite. In other words, a concessive relation is always characterized by a violated expectation. In some cases, which text span is the satellite and which is the nucleus do not depend on the semantics of the spans, but rather on the intention of the writer.

Hobbs / Wolf and Gibson 2005: In the violated expectation relation (also violated expectation in Hobbs [1985]), a causal relation between two discourse segments that normally would be present is absent.

Example

The new software worked great, but nobody was happy.
The new software worked great, although it was programmed by a novice.

Separate problems

Two orthogonal problems:

1) consistent notions and good annotation practices
  ▶ defining discourse relations well enough to cover all cases where we think they should apply
  ▶ getting people to define and annotate consistently, given that we have the same intention.
→ Ted's talk

2) how to represent the mapping.

Separate problems

Two orthogonal problems:

1) consistent notions and good annotation practices

- defining discourse relations well enough to cover all cases where we think they should apply
- getting people to define and annotate consistently, given that we have the same intention.

$\rightarrow$ Ted's talk

2) **how to represent the mapping.**

Different ways to go about the mapping

- all to all mapping

- identify a small set of most general concepts that we can all agree on and use those for mapping

- use a representation that reflects all the distinctions that have been made in the schemes / languages

all to all mapping

for all pairs of resources, someone needs to create a mapping.

- ▶ too much work now, and even more work in the future.
- ▶ unrealistic that we can keep this up to date.

Small set of most general concepts

1. come up with a small set of things everybody can agree on
2. all try to map all relations that were annotated onto this set

unfortunately, we lose information

- if two languages have been distinguishing something which is not considered as part of the core relations, this information is lost, even though both resources have gone through a lot of pain to annotate it
  e.g., volitional cause
- we might find that some resource uses different connectors for something that only has one connector in English. Then if we only keep main distinctions, we can't represent that difference.
- lots of work has to be re-done every time, to figure out what things were annotated in a resource, and which ones weren't.

Maximally detailed relations

Two step approach:

1. collect (from each resource, what distinctions are made?
   - Does the distinction "translate" into one that's already present? (e.g., concession vs. contra-expectation)
   - if there is a distinction that doesn't map onto existing dimensions, add it.
2. organize (find common dimensions, decide about status)
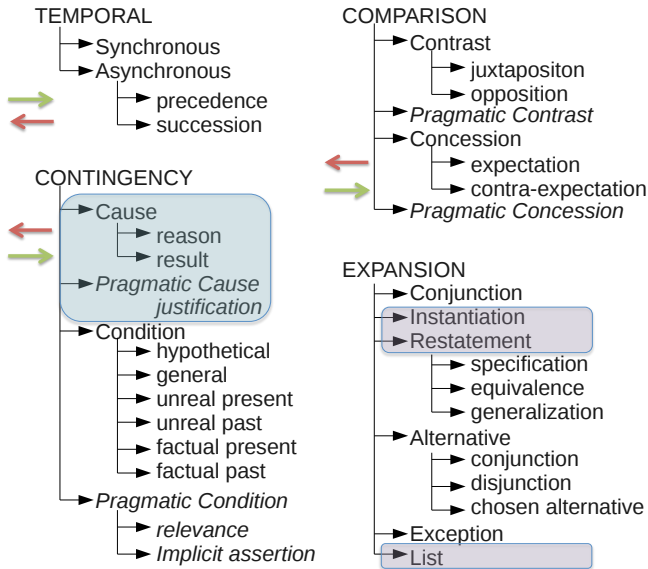
Maximally detailed relations

Two step approach:

1. collect (from each resource, what distinctions are made?
   - Does the distinction "translate" into one that's already present? (e.g., concession vs. contra-expectation)
   - if there is a distinction that doesn't map onto existing dimensions, add it.
2. organize (find common dimensions, decide about status)

How to represent the distinctions?

- set of relation names without structure
- hierarchy
- "dimensions"

Hierarchy

TEMPORAL
→ Synchronous
→ Asynchronous
  → precedence
  → succession

CONTINGENCY
→ Cause
  → reason
  → result
→ *Pragmatic Cause justification*
→ Condition
  → hypothetical
  → general
  → unreal present
  → unreal past
  → factual present
  → factual past
→ *Pragmatic Condition*
  → *relevance*
  → *Implicit assertion*

COMPARISON
→ Contrast
  → juxtapositon
  → opposition
→ *Pragmatic Contrast*
→ Concession
  → expectation
  → contra-expectation
→ *Pragmatic Concession*

EXPANSION
→ Conjunction
→ Instantiation
→ Restatement
  → specification
  → equivalence
  → generalization
→ Alternative
  → conjunction
  → disjunction
  → chosen alternative
→ Exception
→ List

In favour of dimensions

- ▶ better conceptualization? → don't repeat same distinction at different leaves
- ▶ more internally-consistent discourse hierarchies

```
Software was great    because      it was written by an expert    cause.reason
Software was great    therefore,    everybody was happy            cause.result
```

In favour of dimensions

- better conceptualization? → don't repeat same distinction at different leaves
- more internally-consistent discourse hierarchies

Software was great	because	it was written by an expert	`cause.reason`
Software was great	therefore,	everybody was happy	`cause.result`
Software was great	but	everybody was annoyed	`conc.contra-expt`
Software was great	although	it was written by a novice	`conc.expt`

In favour of dimensions

- better conceptualization? → don't repeat same distinction at different leaves
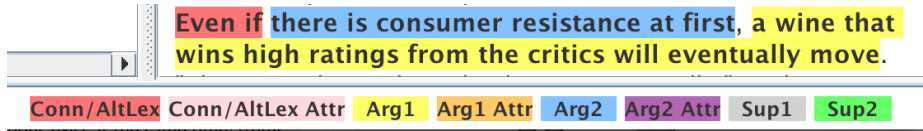- more internally-consistent discourse hierarchies

Software was great	because	it was written by an expert	`cause.reason`
Software was great	therefore,	everybody was happy	`cause.result`
Software was great	but	everybody was annoyed	`conc.contra-expt`
Software was great	although	it was written by a novice	`conc.expt`

RST distinguishes

- many types of causals (justify, non-volitional cause, non-volitional result, volitional cause, volitional result)
- but only one type of concession
- considering dimensions might have drawn attention to this.

Even if

PDTB annotation: **Comparison.Concession.Expectation**

| Conn/AltLex | Conn/AltLex Attr | Arg1 | Arg1 Attr | Arg2 | Arg2 Attr | Sup1 | Sup2 |

Even if there is consumer resistance at first, a wine that wins high ratings from the critics will eventually move.

shouldn't these be distinguished from concessives in the same way as contingencies (if) are distinguished from causals?

suggested dimension: modal status – actual vs. hypothetical or conditional

Unclean categories

Expansion.Conjunction is quite a messy category in PDTB.

4.6.5 Type: "Conjunction"

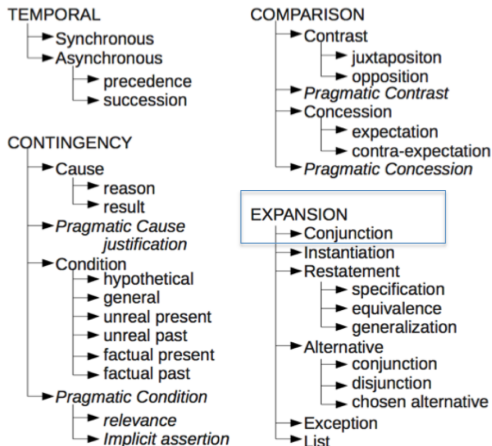The Type "Conjunction" is used when the connective indicates that the situation described in `Arg2` provides additional, discourse new, information that is related to the situation described in `Arg1`, but is not related to `Arg1` in any of the ways described for other types of "EXPANSION". (That is, the rough semantics of "Conjunction" is simply ||Arg1|| ∧ ||Arg2||.) An example of "Conjunction" is shown in (134). Typical connectives for "Conjunction" are *also*, *in addition*, *additionally*, *further*, etc.

(134) *Food prices are expected to be unchanged, but energy costs jumped as much as 4%, said Gary Ciminero, economist at Fleet/Norstar Financial Group.* **He** <u>also</u> **says he thinks "core inflation,"** which excludes the volatile food and energy prices, **was strong last month**. (EXPANSION:Conjunction) (2400)

Would it be cleaner if existing dimensions were applied to split up this category into subtypes?

Conjunction in PDTB

– Level-2 sense label
– ~25% of all relations
  - 5212 explicit
  - 3440 implicit

TEMPORAL
► Synchronous
► Asynchronous
  ► precedence
  ► succession

CONTINGENCY
► Cause
  ► reason
  ► result
► *Pragmatic Cause*
    justification
► Condition
  ► hypothetical
  ► general
  ► unreal present
  ► unreal past
  ► factual present
  ► factual past
► *Pragmatic Condition*
  ► *relevance*
  ► *Implicit assertion*

COMPARISON
► Contrast
  ► juxtapositon
  ► opposition
► *Pragmatic Contrast*
► Concession
  ► expectation
  ► contra-expectation
► *Pragmatic Concession*

EXPANSION
► Conjunction
► Instantiation
► Restatement
  ► specification
  ► equivalence
  ► generalization
► Alternative
  ► conjunction
  ► disjunction
  ► chosen alternative
► Exception
► List

Are these really conjunctions?

Possible to be grouped into finer grained classes:

additionally : Conjunction (7) **emphasize**
in addition: Conjunction (165)
moreover: Conjunction (100), List (1)

similarly: COMPARISON/Conjunction (2), Conjunction (16)
likewise: Conjunction (8) **Comparison**
as well: Conjunction (6)

Other more diverse connectives:

▶ Frequent but also appearing in other specific relations:
but (63), finally (11), in fact (33) , indeed (53), meanwhile (25), separately (69), then (9), while (39)

▶ Infrequent (possibly errors):
however (2), in the end (1), overall (3), neither..nor (1), yet (2), nonetheless (1), nor (25), on the other hand (1), or (5), later (1), in turn (4),...

Possible dimensions

- ▶ semantic / pragmatic (objective / subjective)
- ▶ causal / additive / temporal
- ▶ negative / positive
- ▶ surface order
- ▶ order of events
- ▶ pragmatic order (e.g., reason before result)
- ▶ modal status (actual vs. hypothetical/conditional)
- ▶ anchor or focus or nucleus vs. satelite
- ▶ instantiation / specification / generalization
- ▶ disjunctive (or vs. xor)

Example

pragmatic contrast:

That explains why the number of these wines is expanding so rapidly.
But consumers who buy at this level are also more knowledgeable than they were a few years ago.

Conn/AltLex Conn/AltLex Attr Arg1 Arg1 Attr Arg2 Arg2 Attr Sup1 Sup2

semantic contrast:

The House has voted to raise the ceiling to $3.1 trillion, but
the Senate isn't expected to act until next week at the earliest.

Possible dimensions

- ▶ semantic / pragmatic (objective / subjective)
- ▶ causal / additive / temporal
- ▶ negative / positive
- ▶ surface order
- ▶ order of events
- ▶ pragmatic order (e.g., reason before result)
- ▶ modal status (actual vs. hypothetical/conditional)
- ▶ anchor or focus or nucleus vs. satelite
- ▶ instantiation / specification / generalization
- ▶ disjunctive (or vs. xor)

Example

Possible dimensions

- ▶ semantic / pragmatic (objective / subjective)
- ▶ causal / additive / temporal
- ▶ negative / positive
- ▶ surface order
- ▶ order of events
- ▶ pragmatic order (e.g., reason before result)
- ▶ modal status (actual vs. hypothetical/conditional)
- ▶ anchor or focus or nucleus vs. satelite
- ▶ instantiation / specification / generalization
- ▶ disjunctive (or vs. xor)

Example

surface order:

Although Peter was tired, he didn't sleep.
Peter didn't sleep, although he was tired.

Possible dimensions

- ▶ semantic / pragmatic (objective / subjective)
- ▶ causal / additive / temporal
- ▶ negative / positive
- ▶ surface order
- ▶ order of events
- ▶ pragmatic order (e.g., reason before result)
- ▶ modal status (actual vs. hypothetical/conditional)
- ▶ anchor or focus or nucleus vs. satelite
- ▶ instantiation / specification / generalization
- ▶ disjunctive (or vs. xor)

Example

the direction of causality is not necessarily equivalent to the temporal relation:
"Mary didn't go to the party because she will have an exam tomorrow".

- ▶ semantic temporal: party avoidance $\rightarrow$ exam
- ▶ pragmatic causal: exam $\rightarrow$ party avoidance

Relation in PDTB	Basic operation	Order	Source of coherence	Polarity	rule type (causal/inductive)	pattern of instantiation (bilateral/unilateral)	modal status (actual/hypotheti
TEMPORAL.Asynchronous.precedence	additive 1	basic	semantic	positive	?	?	actual
TEMPORAL.Asynchronous.succession	additive 1	reversed	semantic	positive	?	?	actual
TEMPORAL.Synchronous	additive 1	none 2	semantic	positive	?	?	actual
CONTINGENCY.Cause.reason	causal	revereesd	semantic	positive	causal	bilateral	actual
CONTINGENCY.Cause.result	causal	basic	semantic	positive	causal	bilateral	actual
CONTINGENCY.Pragmatic cause.justification	causal	reversed	pragmatic	positive	causal	bilateral	actual
CONTINGENCY.Condition.hypothetical	causal	basic	semantic	positive	causal	none	hypothetical
CONTINGENCY.Condition.general	causal	basic	semantic	positive	causal	none	hypothetical
CONTINGENCY.Condition.unreal past	causal	basic	semantic	positive	causal	none	hypothetical
CONTINGENCY.Condition.unreal present	causal	basic	semantic	positive	causal	none	hypothetical
CONTINGENCY.Condition.factual past	causal	basic	semantic	positive	causal	none	actual ?
CONTINGENCY.Condition.factual present	causal	basic	semantic	positive	causal	none	actual ?
CONTINGENCY.Pragmatic condition.relevance	causal	basic	pragmatic	positive	causal	none	actual ?
CONTINGENCY.Pragmatic condition.implicit assert	causal	basic	pragmatic	positive	causal	none	actual

Advantages of dimensions

- structuring into hierarchy on demand is possible.
- no fixed hierarchy
- for a task that needs to do e.g. sentiment analysis, can structure with negation at first level
- generate a coarser hierarchy with fewer distinctions if desired

Thank you for your attention!

and thanks also to Fatemeh Torabi Asr