



The DialogBank

Harry Bunt¹, Volha Petukhova², Andrei Malchanau², Alex Chengyu Fang³ and Kars Wijnhoven¹
¹Tilburg University (NL), ²Universität des Saarlandes (D), ³City University of Hong Kong

The DialogBank is a collection of dialogues annotated according to the ISO 24617-2 standard for dialogue act annotation.

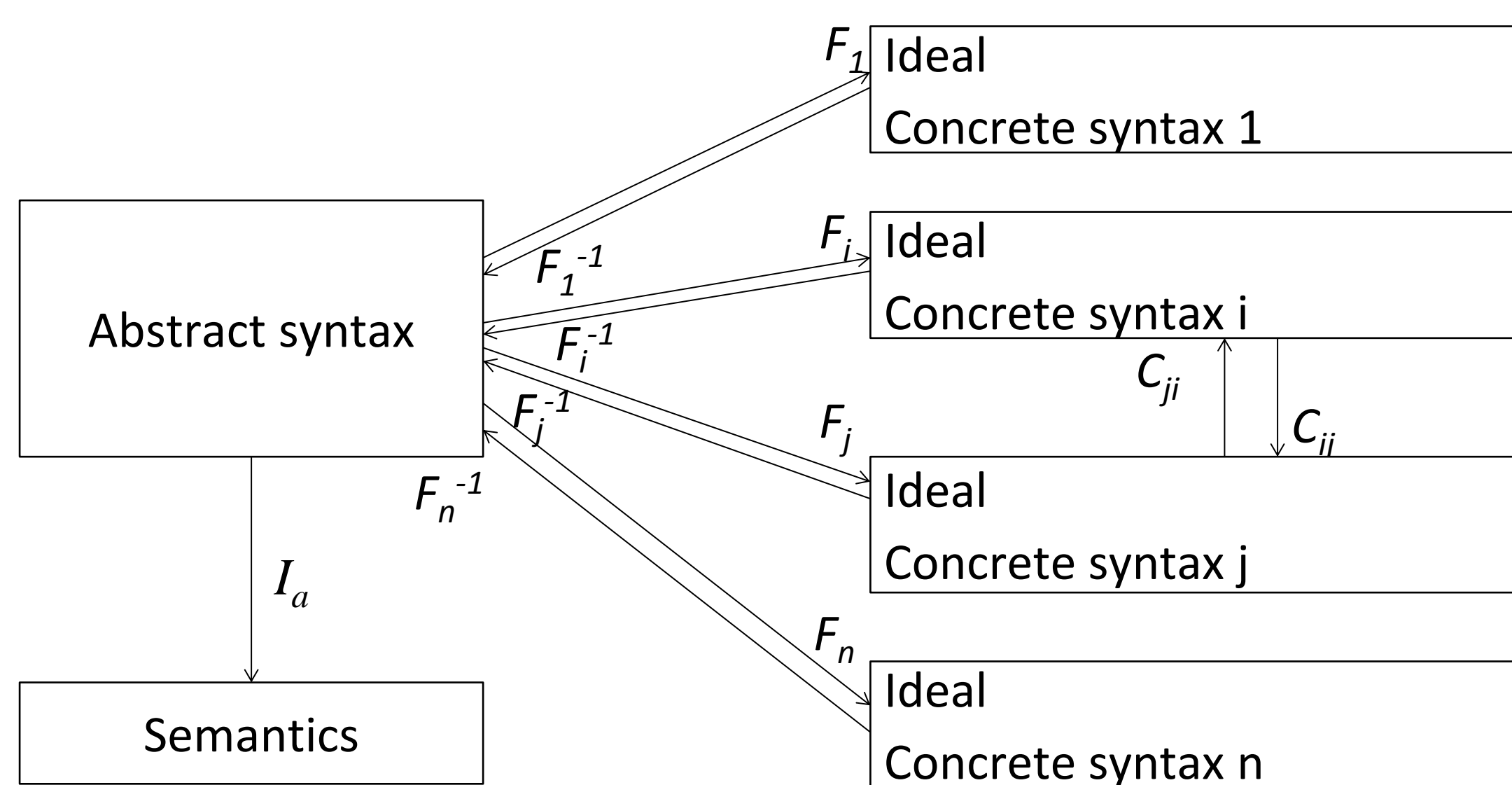
It contains newly annotated and re-annotated dialogues from various corpora, including Switchboard-DA, HCRC Map Task, TRAINS, DBOX, DIAMOND (in Dutch), OVIS (in Dutch), and Schiphol (in Dutch). <https://dialogbank.uvt.nl>

ISO 24617-2 standard for dialogue act annotation:

Multidimensional annotation: multiple communicative functions may be assigned to 'functional segments' including. In addition, annotations in DiAML include:

- Dimensions**, or categories of semantic content. Nine dimensions are distinguished on empirical and theoretical grounds (inherited from DIT++).
- Qualifiers**, for expressing that a dialogue act is performed conditionally, with uncertainty, or with a particular sentiment.
- Dependence relations** for expressing semantic relations between dialogue acts, e.g. for indicating which question is answered by a certain answer act.
- Rhetorical relations** between dialogue acts.

DiAML abstract syntax (triples, n -tuples of concepts) **supports alternative concrete representation formats:** ('ideal' = complete and unambiguous)



Switchboard-DAMSL annotation:

Slash unit ID	Function	Transcript
sw01-0105-0001-A001-01	qw	A.1 utt1: Jimmy, {D so } how do you get most of your news? /
sw01-0105-0002-B002-01	sd	B.1 utt1: {D Well, [I kind of, + fF uh, } I] watch the {F uh, } national news every day, for one /
sw01-0105-0003-B002-02	sd	B.2 utt1 I also read one or two papers a day /
sw01-0105-0004-B002-03	sd	B.3 utt1: {C and } [I'm a, + I'm pretty much a] news junkie /
sw01-0105-0005-B002-04	sd	B.4 utt1: {C and } I tune in to CNN a lot. /
sw01-0105-0006-A003-01	ba	A.3 utt1: {F Oh, } wow. /

Example: core part of ISO 24617-2 annotation in DiAML-XML format:

- G: go south and you'll pass some cliffs on your right
- F: uhm...
- G: and some adobe huts on your left
- F: oh okay

```
<diaml xmlns="http://www.iso.org/diaml">
<dialogueAct xml:id="da1" target="#fs1" sender="#g" addressee="#f" dimension="task"
communicativeFunction="instruct"/>
<dialogueAct xml:id="da2" target="#fs2" sender="#f" addressee="#f"
dimension="turnManagement" communicativeFunction="turnTake"/>
<dialogueAct xml:id="da3" target="#fs2" sender="#f" addressee="#g"
dimension="timeManagement" communicativeFunction="stalling"/>
<dialogueAct xml:id="da4" target="#fs3" sender="#g" addressee="#f" dimension="task"
communicativeFunction="inform"/>
<rhetoricalLink dact="#da4" rhetoAntecedent="#da1" rhetoRel="elaborate"/>
<dialogueAct xml:id="da5" target="#fs4" sender="#f" addressee="#g"
dimension="autoFeedback" communicativeFunction="autoPositive"
feedbackDependence="#da1" "#da4"/>
</diaml>
```

Full XML representation is very hard to read, inspect or correct.

Equivalent, better human-readable tabular representation formats: DiAML-TabSW and DiAML-MultiTab.

DiAML-TabSW format, derived from Switchboard-DA format:

Func.segment ID	DA-ID	Dialogue acts	Sp	Functional segment text	Turn transcript
sw01-0105-fs.1	da1	Ta:setQuestion	A	Jimmy, so how do you get most of your news?	Jimmy. {D so } how do you get most of your news?/
			B		{D Well, } [I kind of, + {F uh, } I] watch the, national news every day, for one. / I also read one or two papers a day / {C and } [I'm a, + I'm pretty much a] news junkie / {C and } I tune in to CNN a lot. /
sw01-0105-fs.2	da2 da3	TiM:stalling TuM:turnTake	B	Well,	
			B	I kind of, I	
sw01-0105-fs.3	da4	OCM:selfCorrection	B	uh	
sw01-0105-fs.4	da5	TiM:stalling	B	uh	
sw01-0105-fs.5	da6	Ta:answer (da2)	B	I watch the national news every day, for one	
sw01-0105-fs.6	da7	TiM:stalling	B	uh	
sw01-0105-fs.7	da8	Ta:answer (da2) (Expansion da7)	B	I also read on or two papers a day	
sw01-0105-fs.8	da9	TuM:turnKeep	B	and	
sw01-0105-fs.9	da10	Ta:inform	B	I'm pretty much a news junkie	
sw01-0105-fs.10	da11	OCM:selfCorrection	B	I'm a, I'm pretty much a	
sw01-0105-fs.11	da12	TuM:turnKeep	B	and	
sw01-0105-fs.12	da13	Ta:answer (da2) (Expansion da7, d9)	B	I tune in to CNN a lot	
sw01-0105-fs.13	da14	AuF:autoPositive	A	Oh, wow.	Oh, wow.

DiAML-MultiTab representation format:

Func. segment ID	Sp	Functional segment text	Turn transcript	Task	AutoFeedback	Turn Management	Time Management	Discourse Structuring	Social Obligation Management
			hello can I help you						
TR1-fs.1	s	hello							da1:Initial Greeting
TR1-fs.2	s	can I help you						da2:Offer	
			uhm, yes hello, maybe, I'd like to take a tanker with orange juice from...						
TR1-fs.3	u	uhm				da3:Turn Take	da4:Stalling		
TR1-fs.4	u	yes hello			da5: Positive(da1)				da6: Return Greeting (da1)
TR1-fs.5	u	yes maybe						da7: Accept Offer(da2) [uncertain]	
TR1-fs.6	u	I'd like to take a tanker with		da8: Inform					

DialogBank contents:

Origin	Language	Original representation	Original annotation	DiAML format of ISO 24617-2 annotation
HCRC MapTask	English	NITE XML	HCRC MapTask communicative functions	DiAML-XML
Switchboard	English	3-column tabular	SWBD-DAMSL communicative functions	DiAML-TabSW
TRAINS	English	13-column tabular	DAMSL communicative functions	DiAML-MultiTab
DBOX	English	DiAML-XML	ISO 24617-2 annotations	DiAML-XML
Dutch MapTask	Dutch	plain text transcript	no dialogue act annotation	DiAML-MultiTab
DIAMOND	Dutch	13-column tabular	DIT++ communicative functions and dimensions	DiAML-MultiTab
OVIS	Dutch	plain text transcript	no dialogue act annotation	DiAML-MultiTab
Schiphol Airport	Dutch	plain text transcript	no dialogue act annotation	DiAML-XML

Ongoing and future work:

- Addition of more dialogues with gold standard ISO 24617-2 annotations in various DiAML formats
- Experimental investigation of usability of alternative DiAML representation formats
- Integration of ISO 24617-2 annotation with discourse relation annotation according to ISO 24617-8
- Implementation of convertors between alternative DiAML formats, allowing users to view annotations in optimally convenient ways