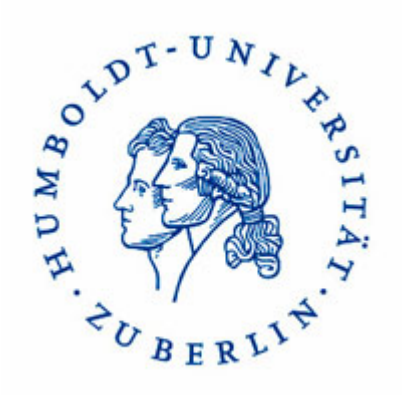
Automated Generation of Business Process Models from Natural Language Input

Fabian Friedrich

School of Business and Economics
Humboldt University of Berlin

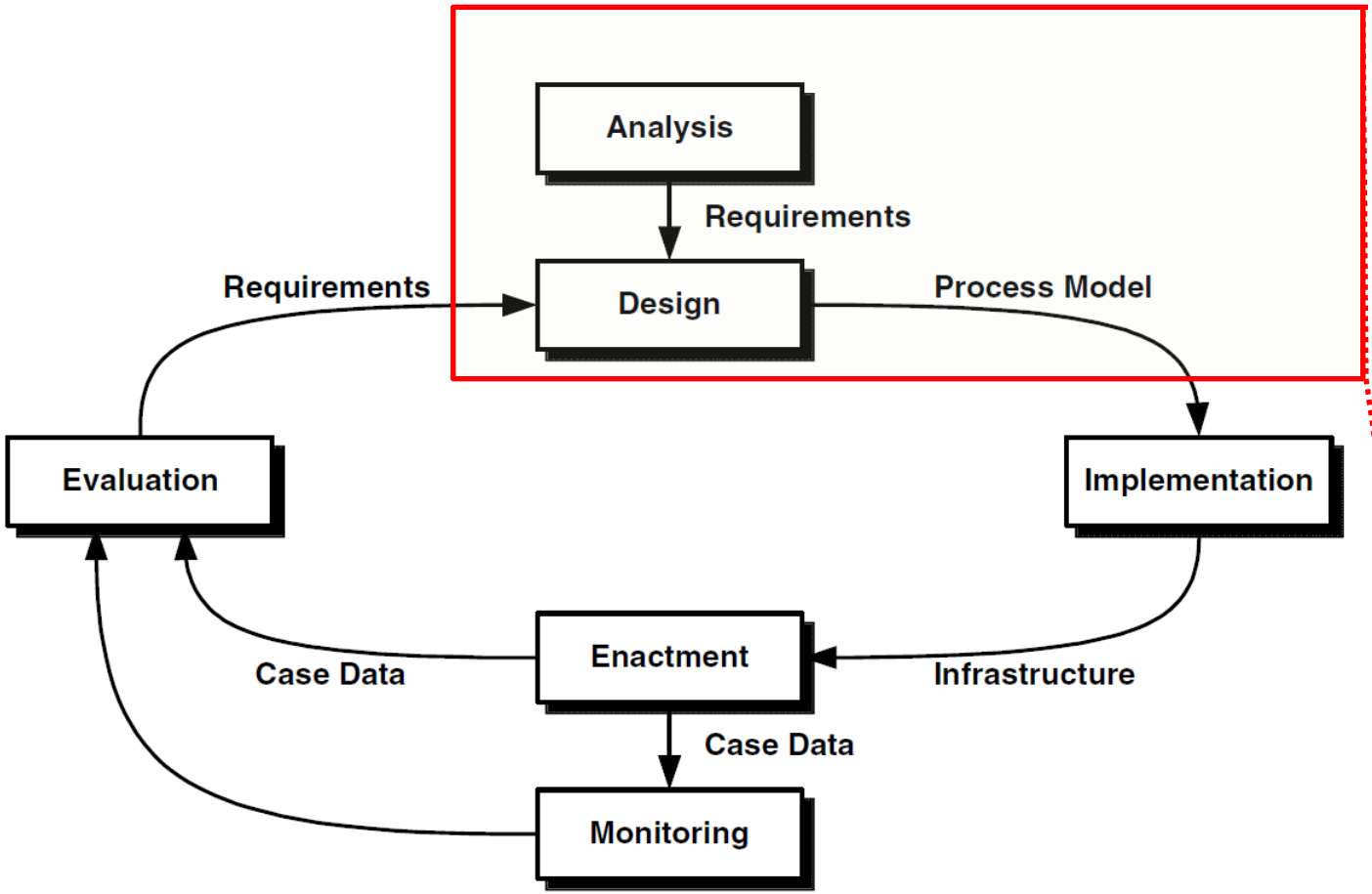
2010-10-27



Today's Agenda

- Introduction – The Idea
- Background
- Transformation Approach
- Evaluation & Demo

Initial modeling is expensive



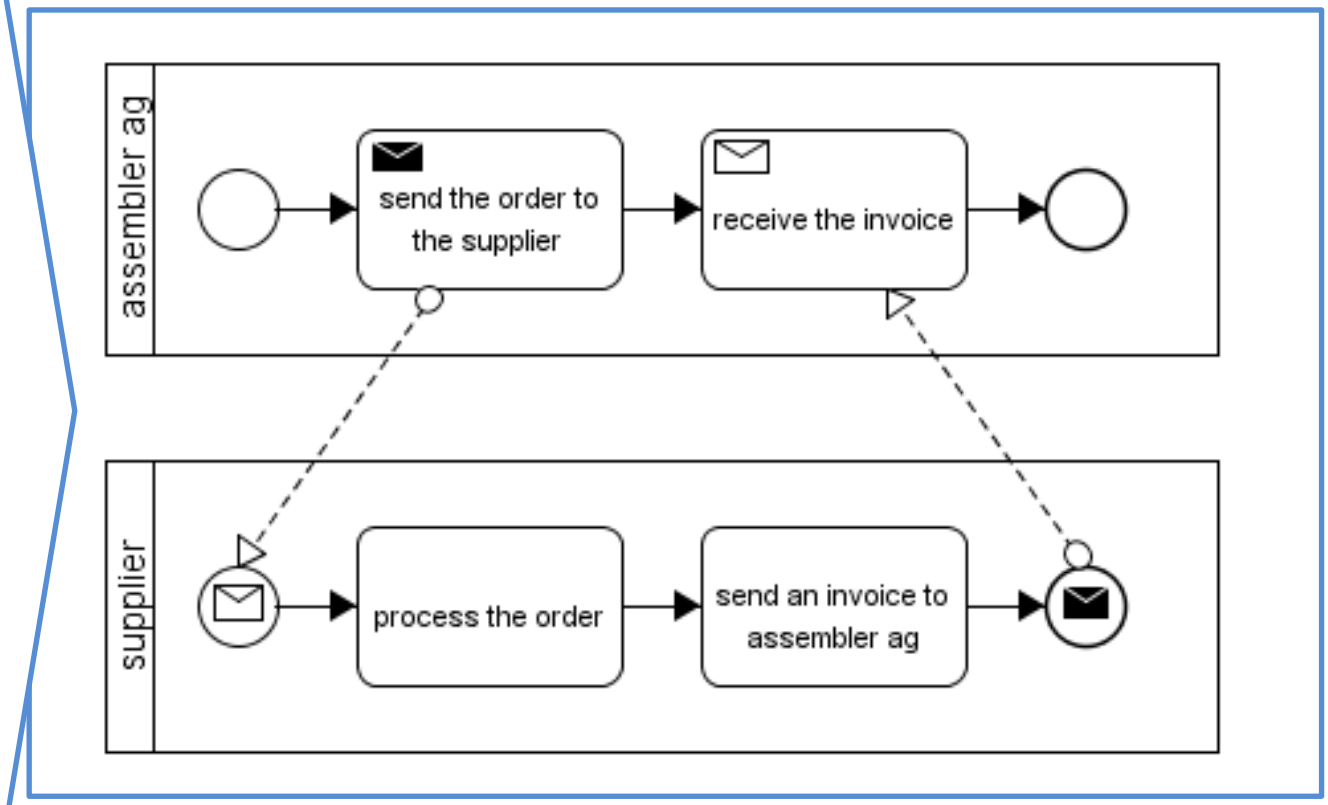
- Documents are present, but unstructured
- Reading the text and creating process models requires substantial amounts of work
- Creating the AS-IS model can take 60% of the time[1]

1 Herbst, J. & Karagiannis, D.: *An inductive approach to the acquisition and adaptation of workflow models* Proceedings of the IJCAI, **1999**.

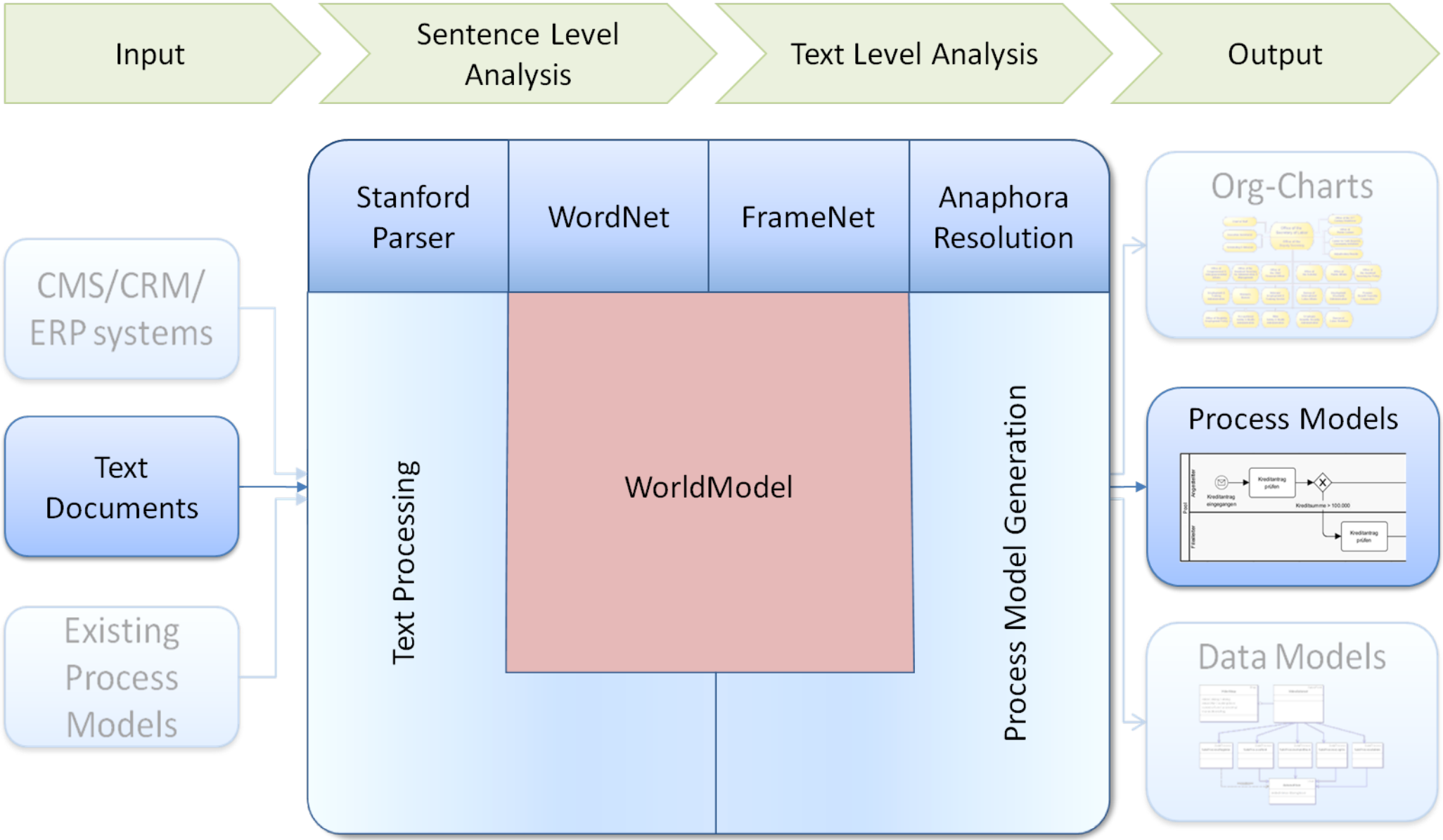
The Idea: Automated Transformation

The process starts periodically on the first of each month, when Assembler AG places an order with the supplier in order to request more product parts.

a) Assembler AG sends the order to the supplier.
b) The supplier processes the order.
c) The supplier sends an invoice to Assembler AG.
d) Assembler AG receives the invoice.

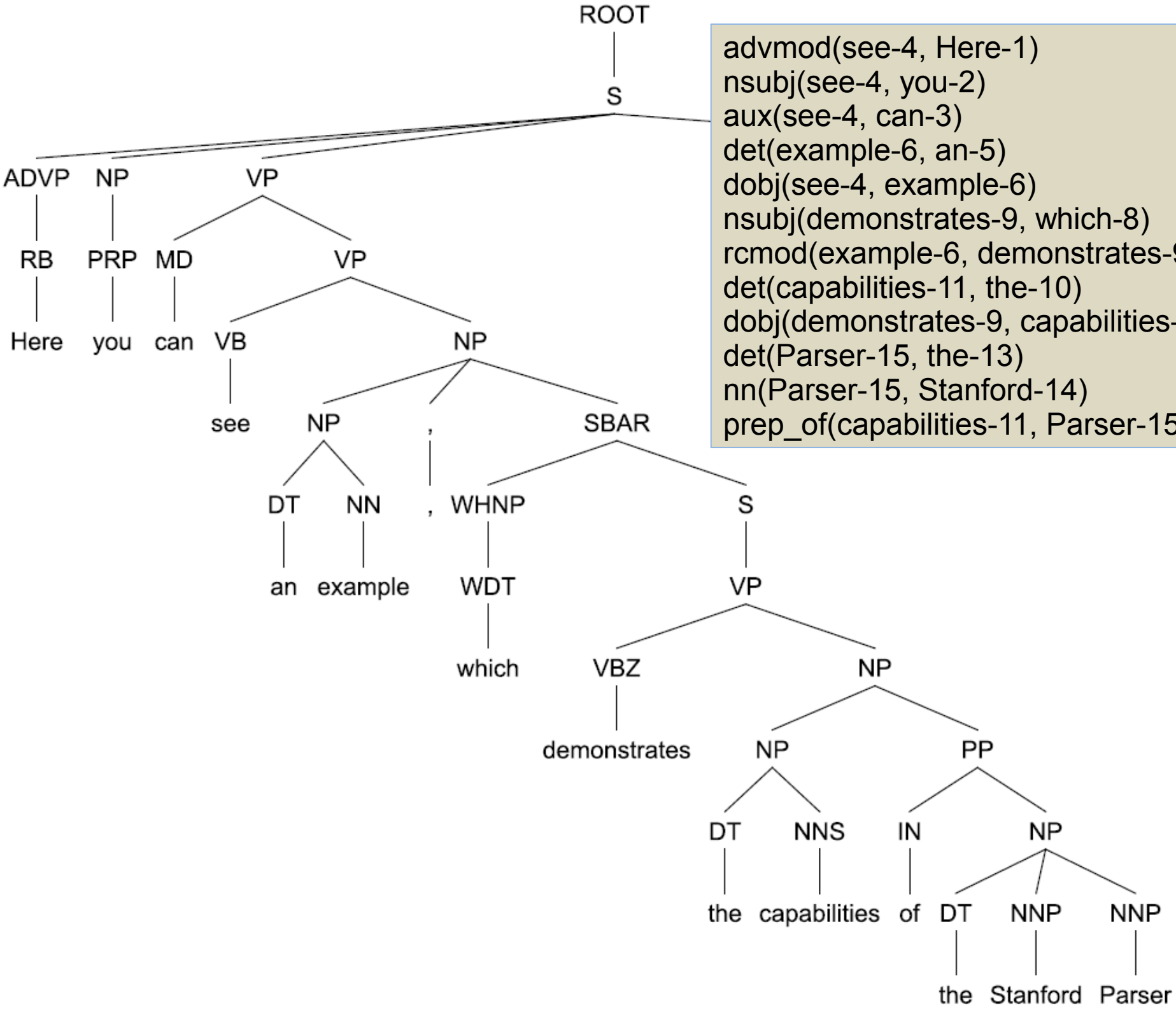


Approach: Automated Transformation



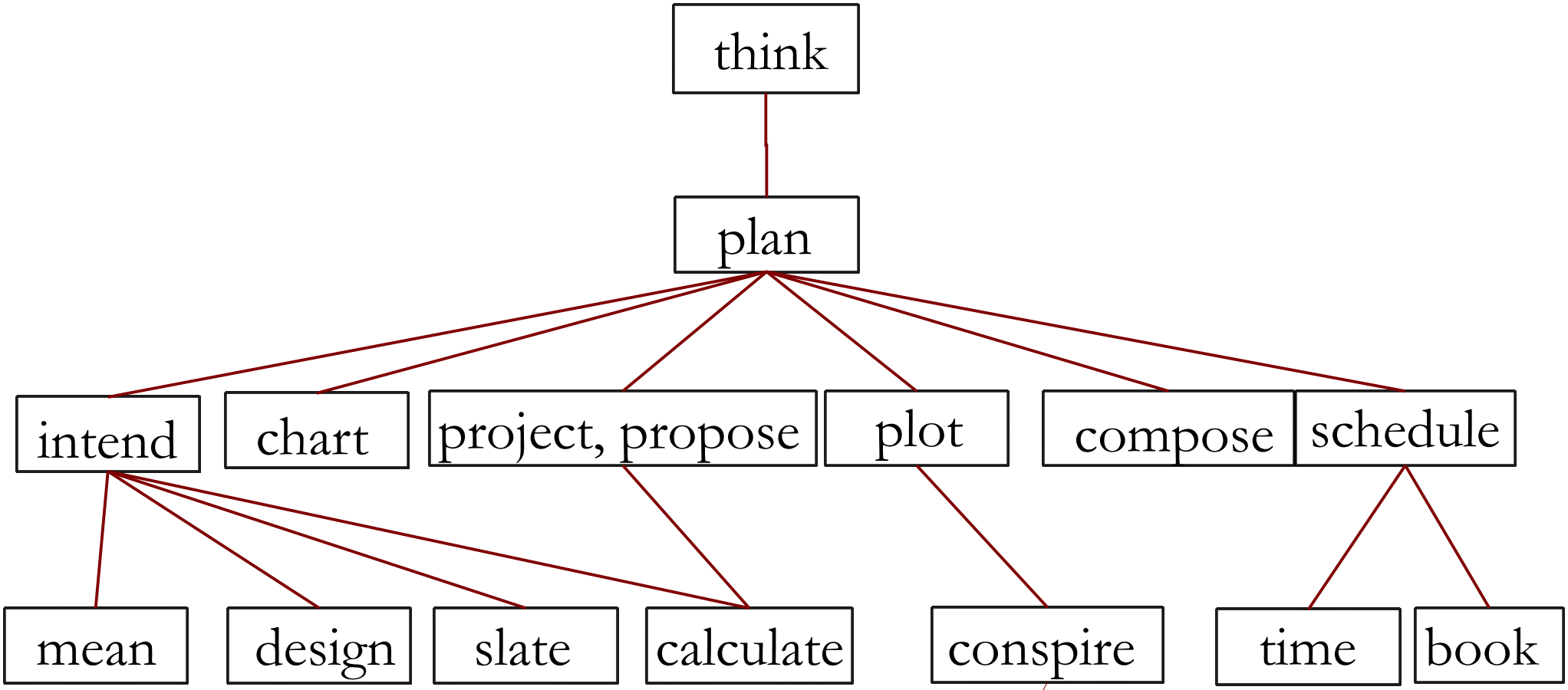
Today's Agenda

- Introduction – The Idea
- Background
- Transformation Approach
- Evaluation & Demo



advmod(see-4, Here-1)
nsubj(see-4, you-2)
aux(see-4, can-3)
det(example-6, an-5)
dobj(see-4, example-6)
nsubj(demonstrates-9, which-8)
rcmod(example-6, demonstrates-9)
det(capabilities-11, the-10)
dobj(demonstrates-9, capabilities-11)
det(Parser-15, the-13)
nn(Parser-15, Stanford-14)
prep_of(capabilities-11, Parser-15)

Wordnet provides a Hypernym Structure



General Purpose Ontology

FrameNet – Assignment of Semantic Frames

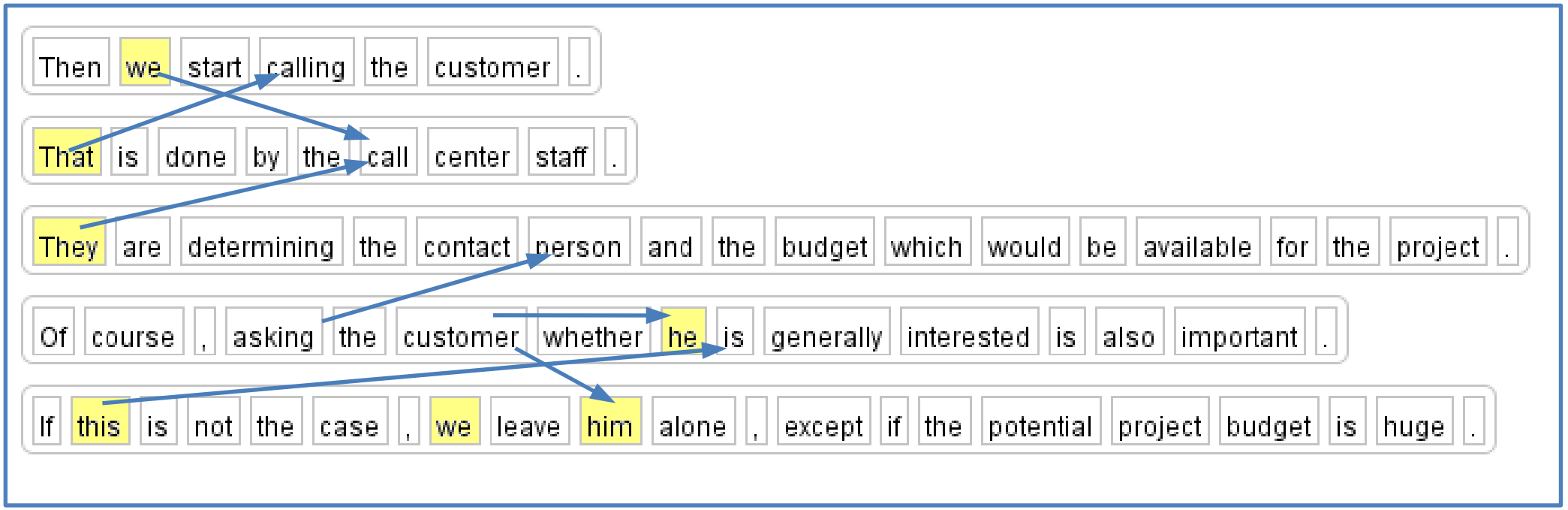
Frame: Telling

FrameElements:
Core: Addressee, Medium, Message, Speaker, Topic
Non-Core: Descriptor, Epistemic_Stance, Iteration, Manner, Means, Place, Time

Example Annotation for Telling.Inform

Time	In 2002,
Speaker	the U.S State Department
Target	INFORMED
Addressee	North Korea
Message	that the U.S. was aware of this program, and regards it as a violation of Pyongyang's nonproliferation commitments

Anaphorisms have to be Resolved



Prepositions and Determiners have to be resolved to their respective Actions/Resources/Actors

Standard Properties:

- Distance
- Gender/number/person agreement
- Syntactic role
- Number of previous occurrences

Today's Agenda

- Introduction – The Idea
- Background
- Transformation Approach
- Evaluation & Demo

Identified Problem-Patterns

Semantics ≠ Syntax

Active vs. Passive Voice

Word Order/Rephrasing

Implicit rethoric structure

Text Relevance

Relative clauses

"Meta"-Sentences

Example Sentences

Atomicity

Distributed actions

Complex sentences

Relative clauses

References

Anaphorisms

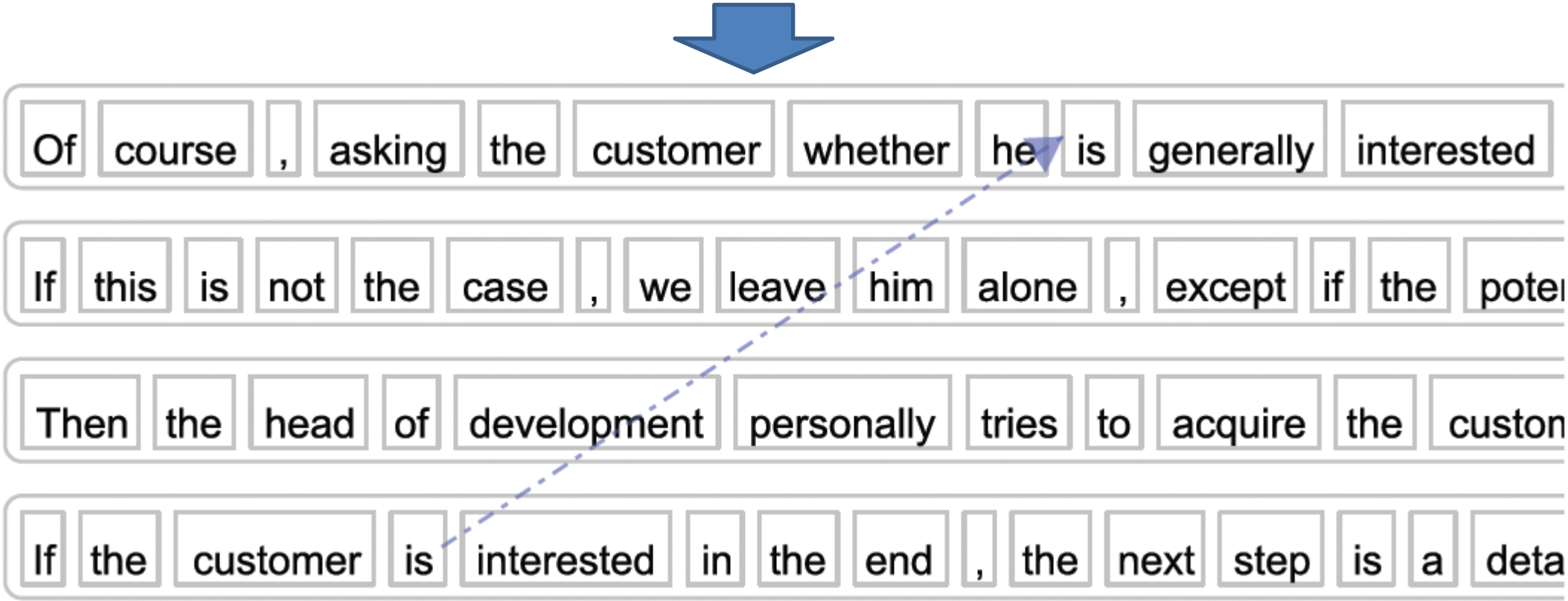
Backward/forward references and jumps

End-of-block recognition

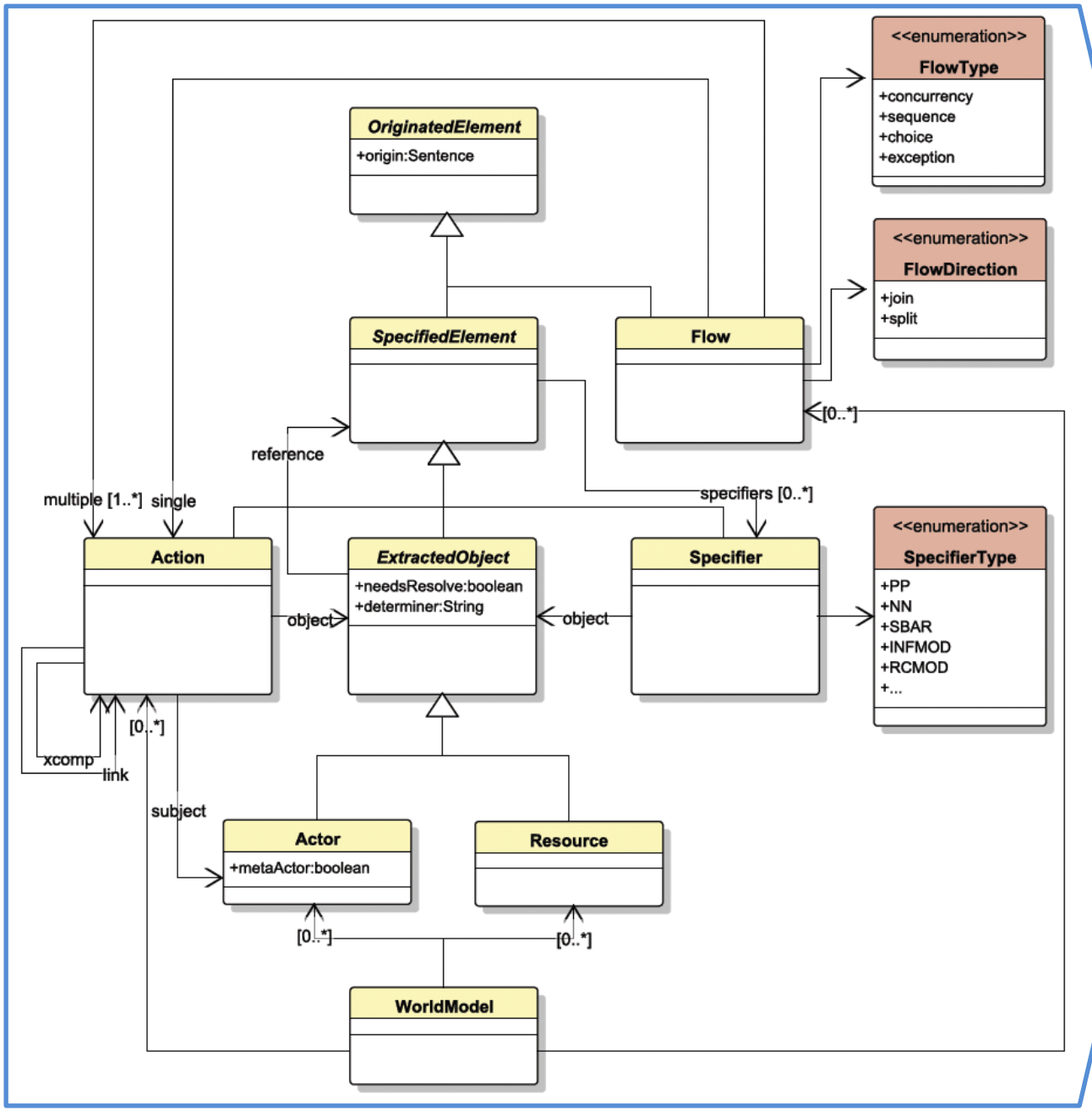
Textual „Links“

To define more than sequential structure „Links“ are needed
Three types were identified:

- Forward Links
- Backward Links (Loops)
- Jumps



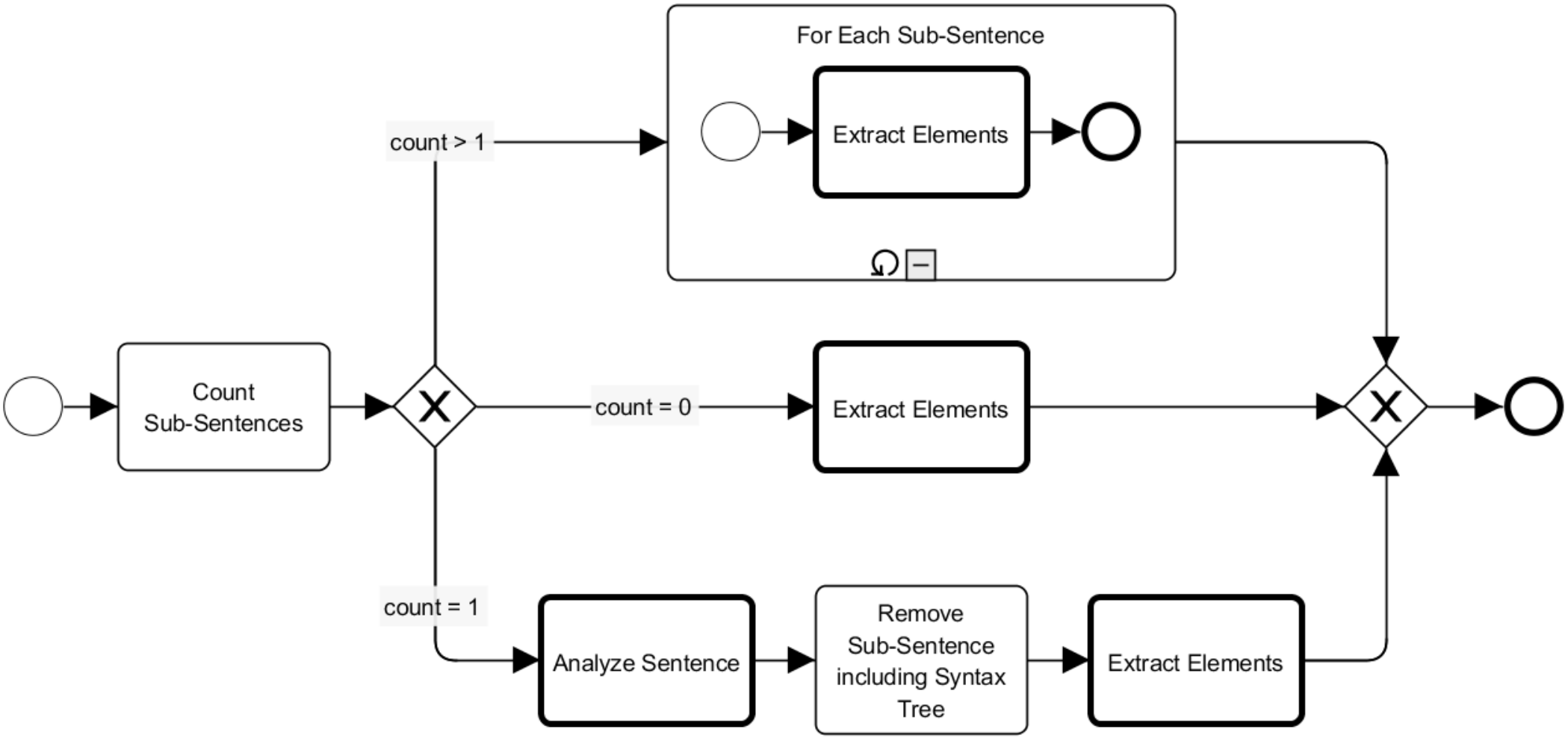
We defined an Intermediate Data Structure



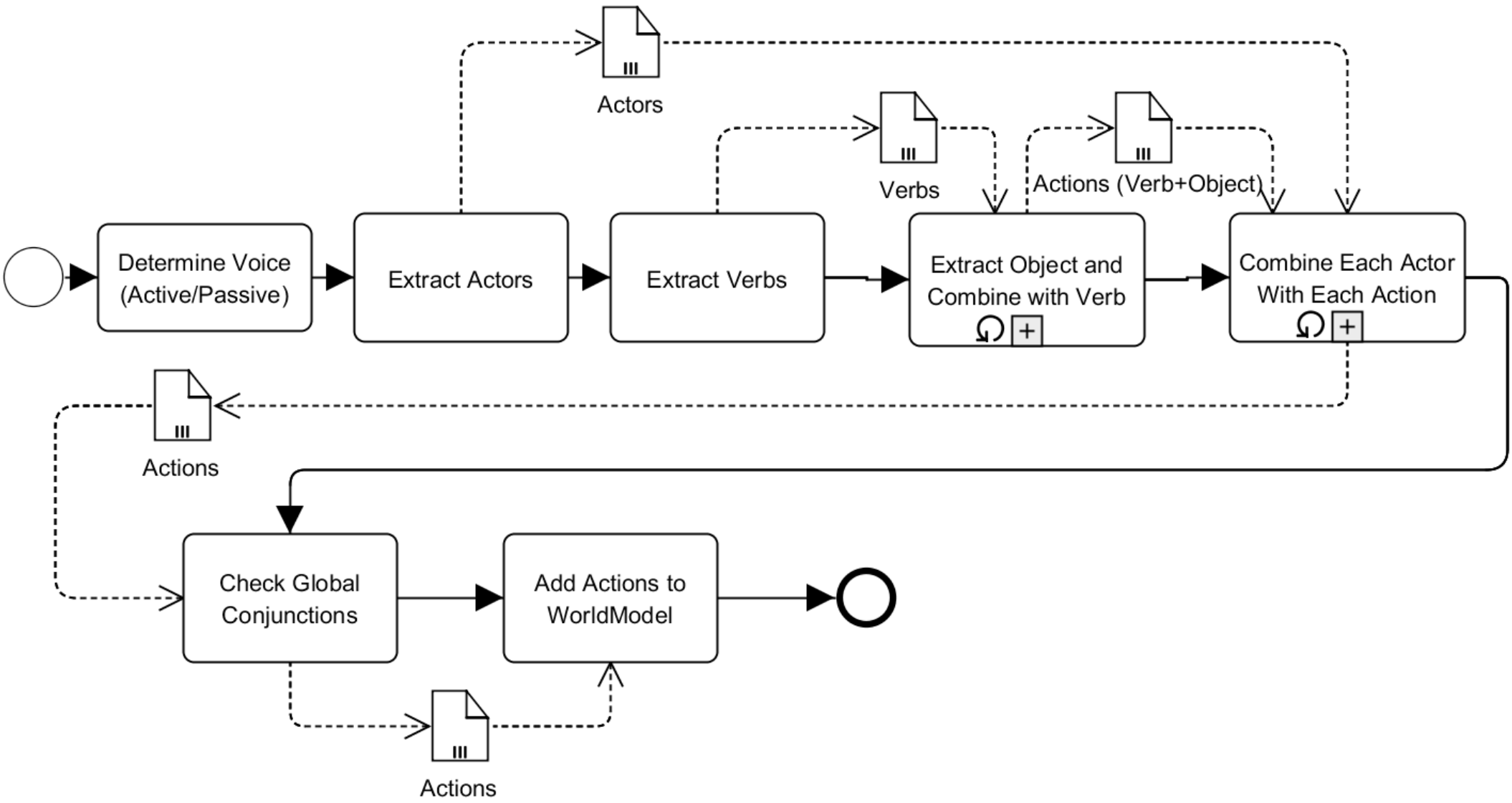
Features

- Mainly syntax oriented
- Action centric
- Used during all stages of the transformation
- Main Parts:
  - Action
  - Resource
  - Actor
  - Flow

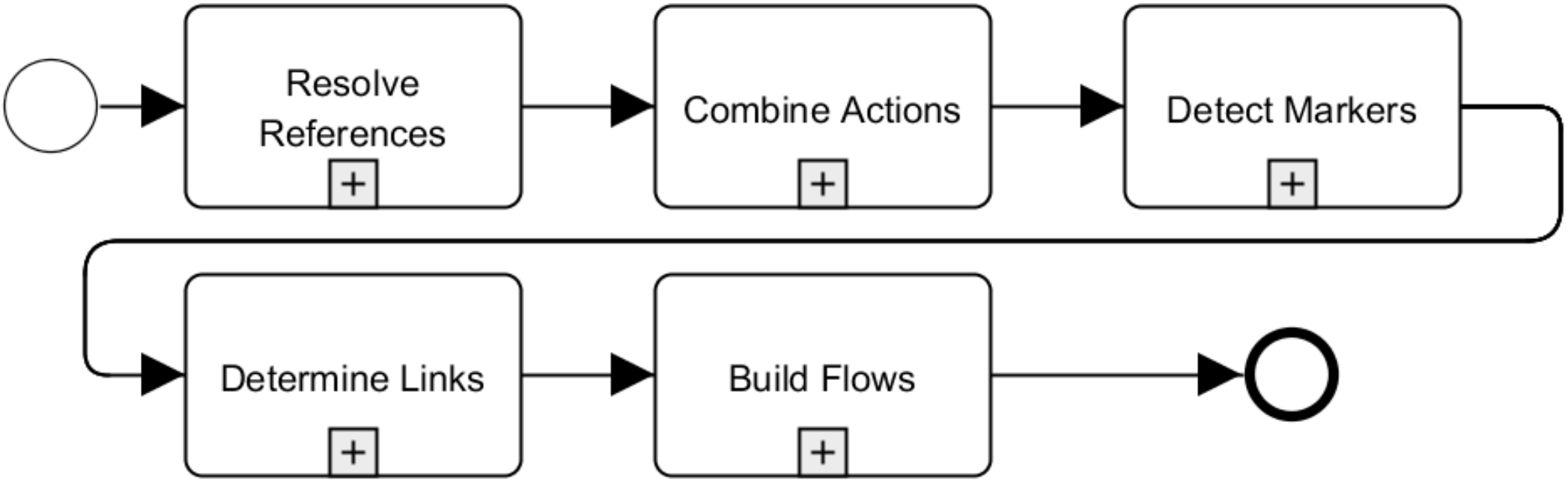
First, each Sentence is Analyzed in Isolation



Syntactic Elements are Extracted and Combined

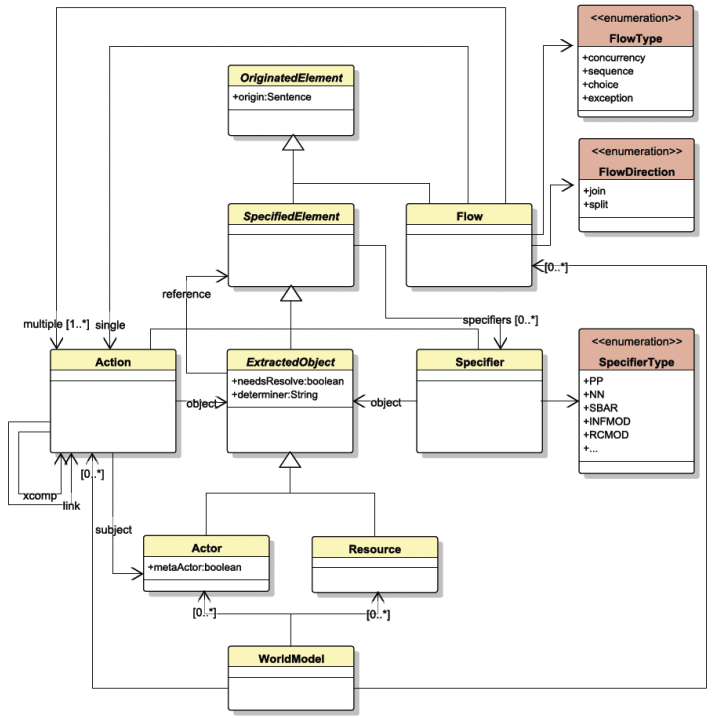


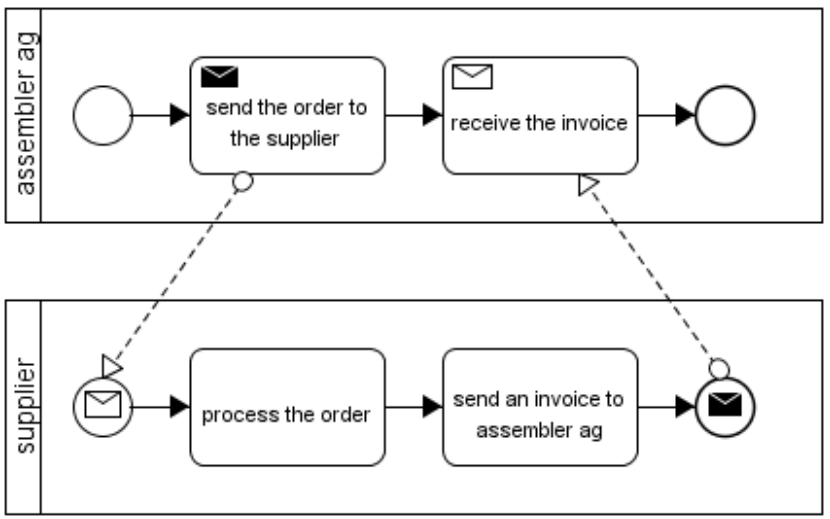
Then, all Actions are analyzed wholeistically



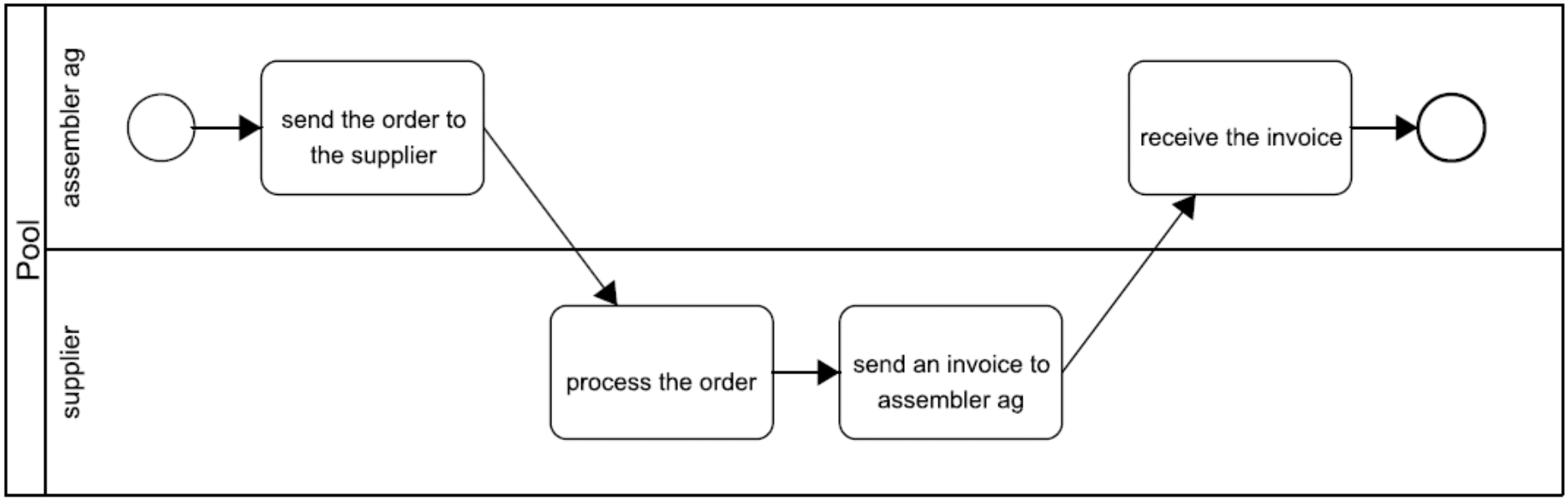
The Last Step: Process Model Generation

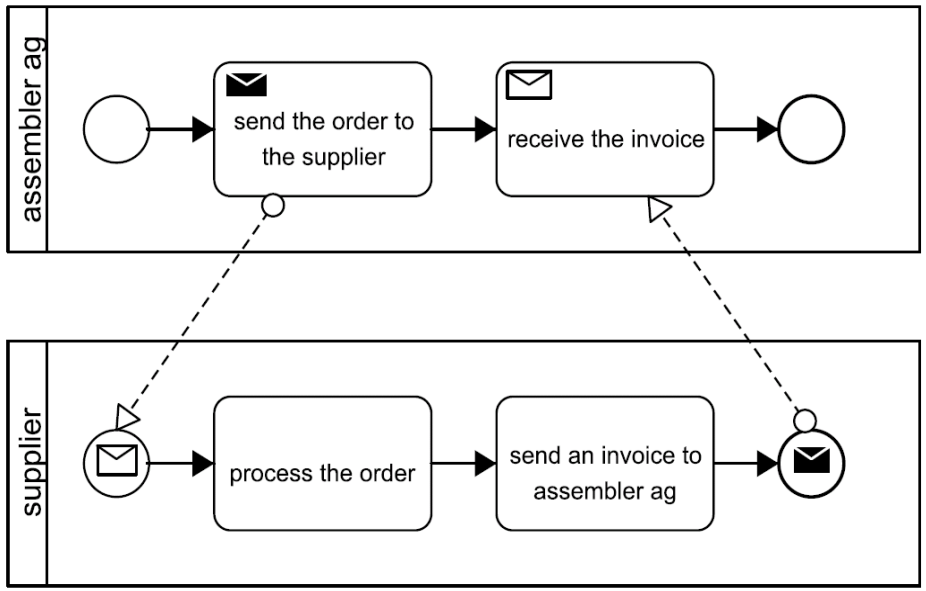
1) Create Actions/Events
2) Build Sequence Flows
3) Add Start/End Events
4) Events → Labels
5) Handle „Meta“-Activities
6) Create BlackBox Pools
7) Create Data Objects
8) Layouting





Post-Processing: Splitting of Lanes





Today's Agenda

- Introduction – The Idea
- Background
- Transformation Approach
- Evaluation & Demo

So far 43 Test-Data Sets were Collected

Source	# of Models	Type
TU Berlin	2	academic
QUT	8	academic
BPM academics	1	academic
Vendor Tutorials	4	industry
inubit	4	industry
BPM Practicioners	1	industry
Book - BPMN practical handbook	3	textbooks
Book - BPMN modeling and reference guide	6	textbooks
Federal Network Agency - Messwesen	14	public sector
SUM	**43**	

Number of Models by Type	# of Models	% of Total
academic	11	25,58%
industry	9	20,93%
textbooks	9	20,93%
public sector	14	32,56%

The analysis of 8 characteristics is planned

Spalte1	FNA - B6	inubit University	TU - supplier switch
# of sentences	3	14	30
avg. sentence length	10,00	14,28	22,08
# of meta sent.	0	3	9
# of rel. ref.	0	13	7
# of links	0	1	1
Similarity (Graph Edit Distance)	99,50%	94,20%	74,10%

The End

Thank you very much for your attention!