Combining Referring Expression Generation and Surface Realization: A Corpus-Based Investigation of Architectures

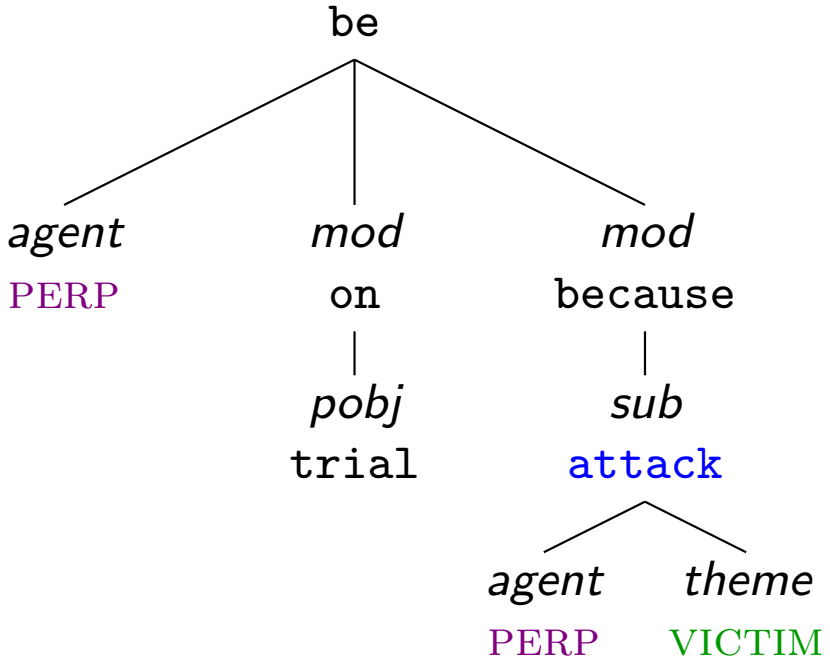
Sina Zarrieß Jonas Kuhn

Institute for Natural Language Processing (IMS), University of Stuttgart (Germany)

Generating Coherent Text

Example: realizing a transitive relation and its referential arguments

Abstract Input:



Linguistic Output:

natural:
Two brothers are on trial because of an attack on a young man.

incoherent:
The masked robbers are on trial because two brothers attacked their victim.

Corpus-based Generation: Learning Choices From Data

Observed corpus data

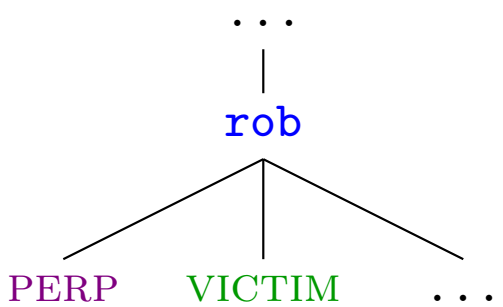
... If you have been robbed , there are a lot of things ...
... Initial reports indicated that the robbery had happened inside ...
... decision to send a teenager who violently robbed and attacked an elderly woman ...

Corpus-based Generation: Learning Choices From Data

Observed corpus data

... If you have been robbed implicit_perp, there are a lot of things ...
... Initial reports indicated that the robbery implicit_perp implicit_victim had happened inside ...
... decision to send a teenager who violently robbed and attacked an elderly woman ...

Annotation

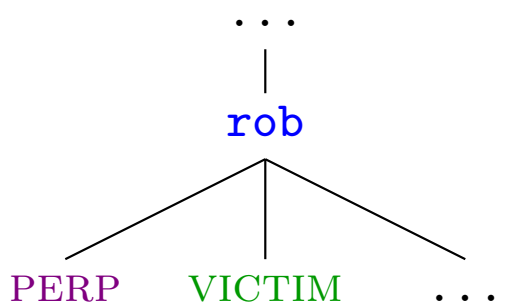


Corpus-based Generation: Learning Choices From Data

Observed corpus data

... If you have been robbed implicit_perp, there are a lot of things ...
... Initial reports indicated that the robbery implicit_perp implicit_victim had happened inside ...
... decision to send a teenager who violently robbed and attacked an elderly woman ...

Annotation

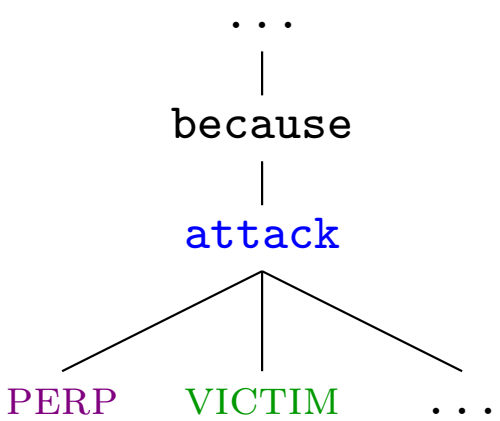


Modeling

v=pronoun+passive+p=implicit
v=implicit+nominalized+p=implicit
v=indefinite+active+p=rel-pron

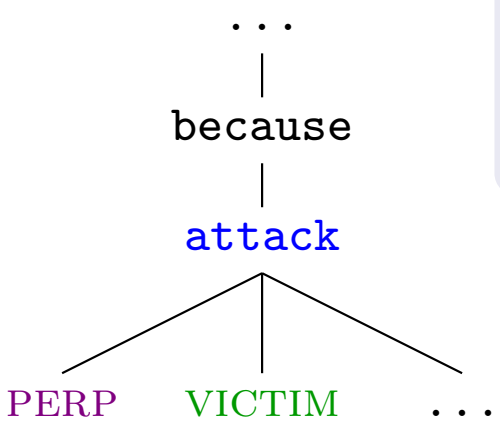
Predicting Choices

Unseen Input

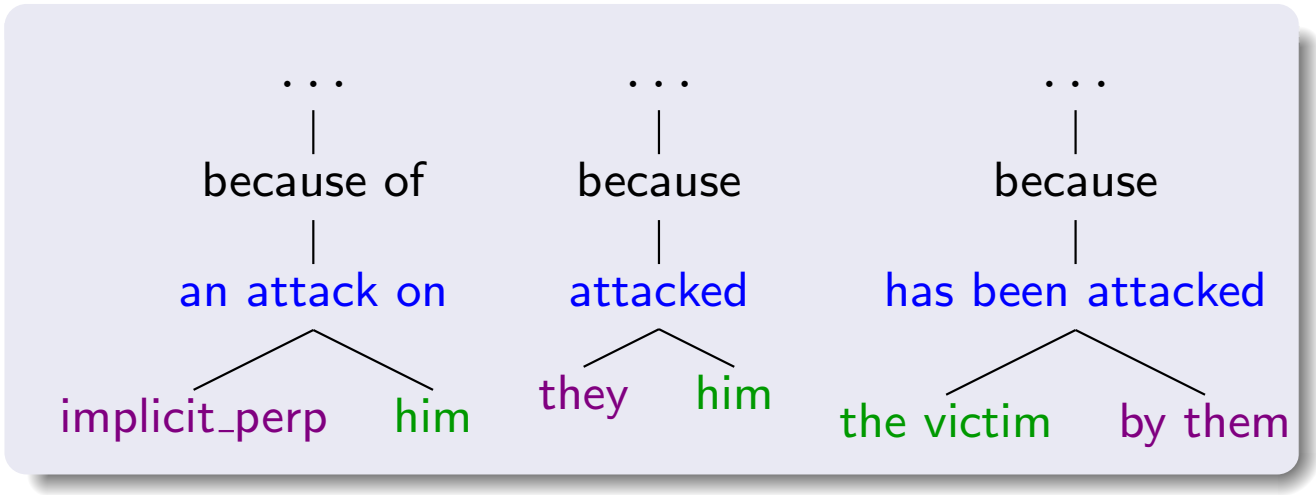


Predicting Choices

Unseen Input

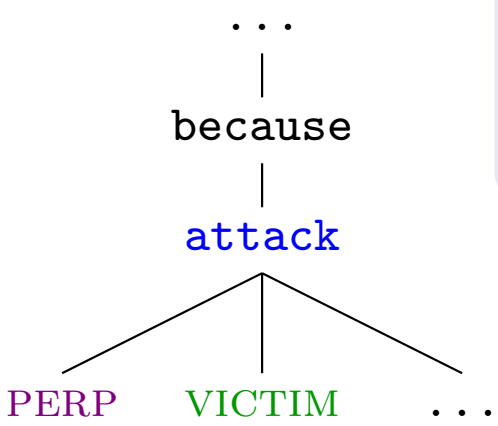


Candidates

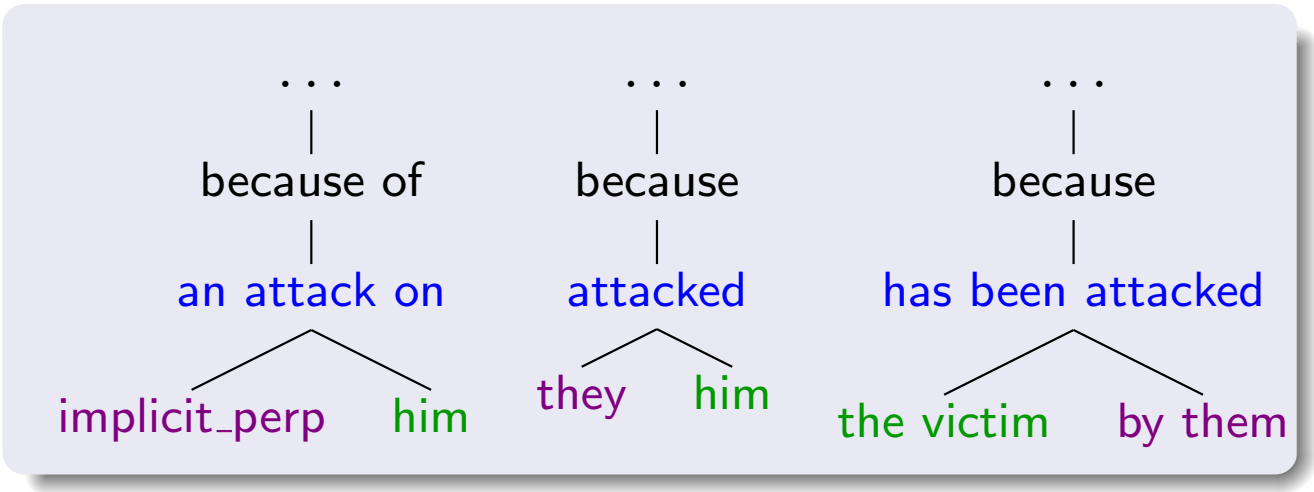


Predicting Choices

Unseen Input



Candidates

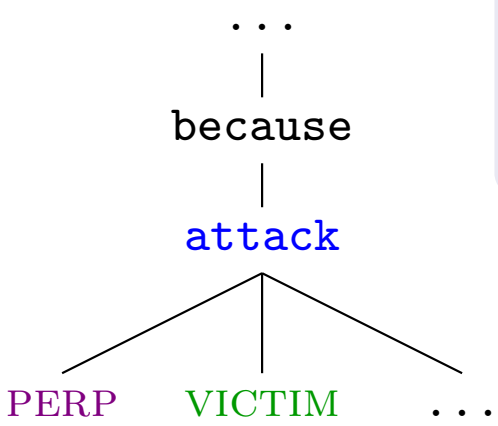


Ranking

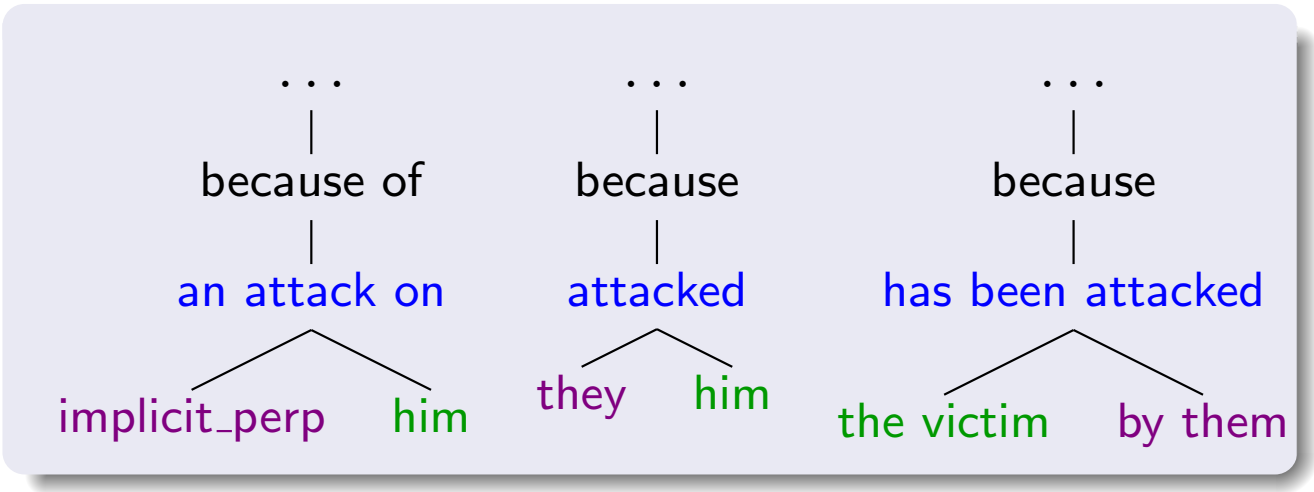
v=pronoun+nominalization+p=implicit
v=pronoun+active+p=pronoun
v=definite+passive+p=pronoun

Predicting Choices

Unseen Input



Candidates



Ranking

v=pronoun+nominalization+p=implicit

v=pronoun+active+p=pronoun

v=definite+passive+p=pronoun

Natural Output

... because of an attack on the victim ...

General Background

- Corpus-based generation for empirical studies of linguistic variation (Langkilde and Knight, 1998; Ratnaparkhi, 2000; Ringger et al., 2004; Filippova and Strube, 2007; Cahill and Riester, 2009)
- Two well-studied paradigms:
  - **Surface realization (SR):** predict word order/syntactic realization on existing treebanks, shared task (Belz et al., 2011)
  - **Referring expression generation (REG):** predict pronominalization a.o. on texts with entity annotations, GREC shared tasks (Belz and Kow, 2010)

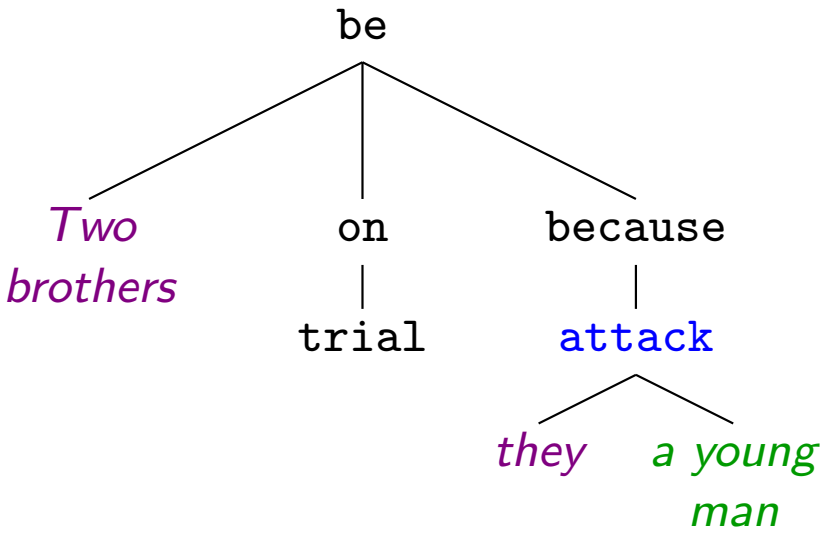
General Background

- Corpus-based generation for empirical studies of linguistic variation (Langkilde and Knight, 1998; Ratnaparkhi, 2000; Ringger et al., 2004; Filippova and Strube, 2007; Cahill and Riester, 2009)
- Two well-studied paradigms:
  - **Surface realization (SR):** predict word order/syntactic realization on existing treebanks, shared task (Belz et al., 2011)
  - **Referring expression generation (REG):** predict pronominalization a.o. on texts with entity annotations, GREC shared tasks (Belz and Kow, 2010)

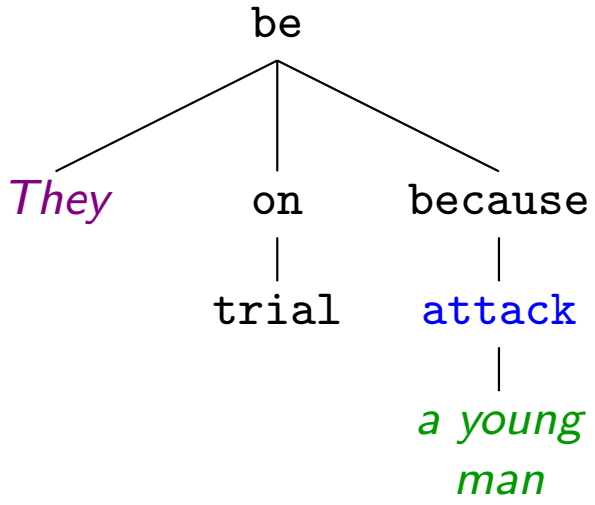
But: SR and REG are often treated as isolated tasks

Choices Depend on Each Other

Standard set-up: surface realization is determined by referring expressions



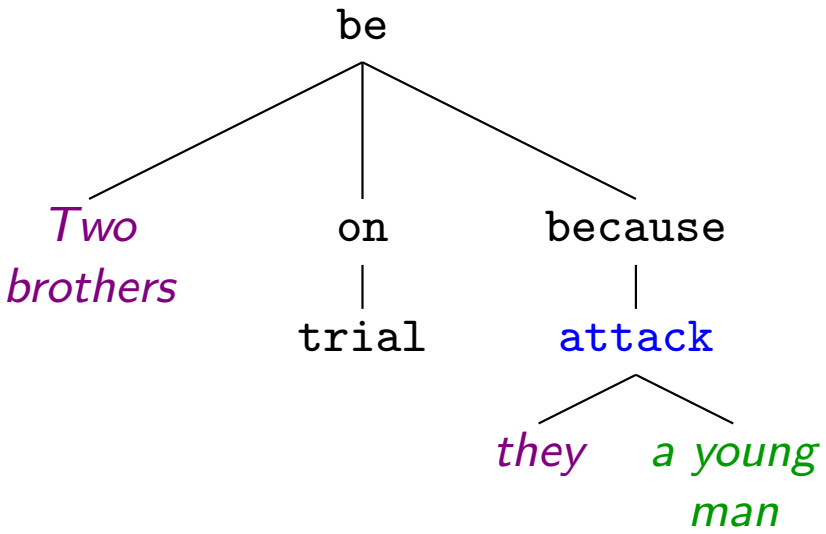
- Likely: *attacked*



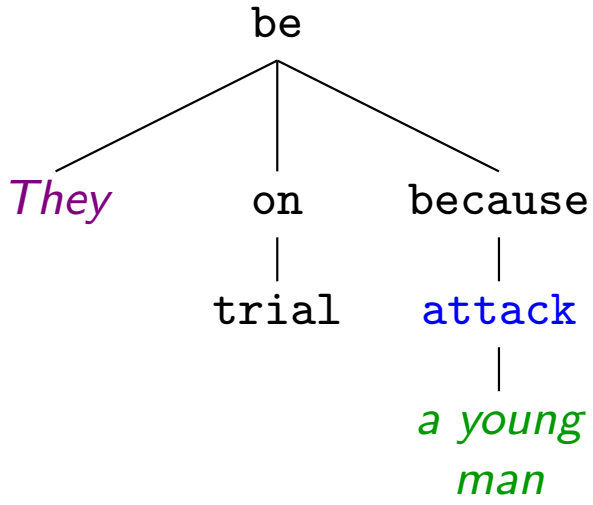
- Likely: *an attack on, was attacked*

Choices Depend on Each Other

Standard set-up: surface realization is determined by referring expressions



- Likely: *attacked*

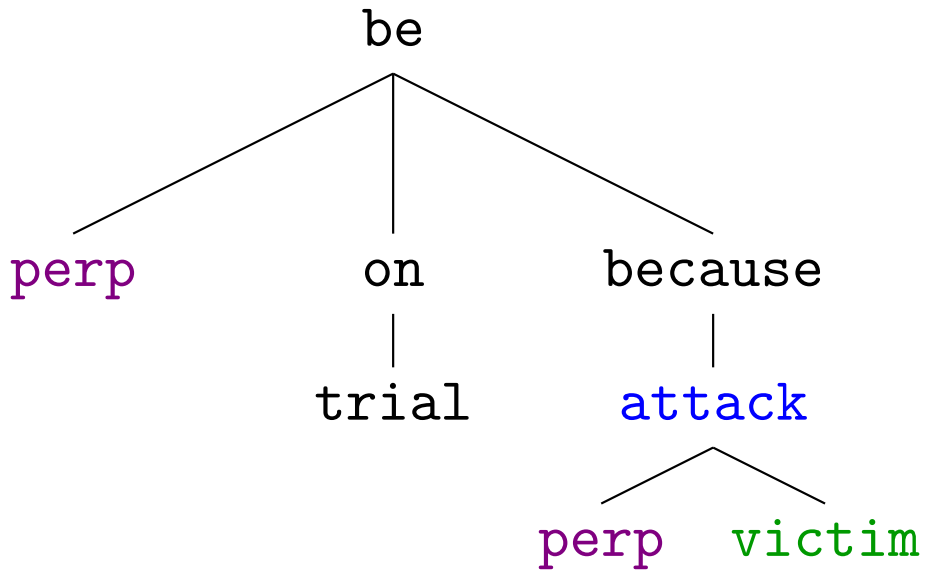


- Likely: *an attack on, was attacked*

If original REs and arguments are given, they reflect syntactic choices and their contexts to a large extent. (Zarrieß, Cahill and Kuhn, 2011;2012)

Combining Choices

- We propose a task that combines surface realization and referring expression generation



- Challenge: Modeling interaction between choices

Research Questions

- Standard approach for complex, multi-level generation problems
  - **Pipeline:** (Reiter and Dale, 1997)
    document planning → sentence-planning → surface realization
- Well-known problems
  - Error propagation
  - **Generation Gap:** Discourse-level decisions can depend on sentence-level decisions (Meteer, 1991)
- Place of REG in the pipeline has been a notorious problem
  (Cahill et al., 1999; Mellish et al., 2000)

Goal

- Corpus-based framework for investigating architectures and interactions between choices
- Integrate existing methods for annotation and data set creation (GREC and SR shared tasks)
  add more types of implicit referents

Generation Approach

- Combined SR and REG annotations
- Generator is based on three trainable modules
  - SYN: Deep $\rightarrow$ shallow dependencies
  - REG: insert RE candidates
  - LIN: linearize dependency tree
- Experimental parameters
  - **order of modules**
  - treatment of implicit referents
  - models of discourse context

A Data Set for Combined REG and SR

- Language: German
- Size: 200 texts, 2030 sentences
- Text type: newspaper articles about *robberies*
- Referents: frequent mentions of *victim*, *perpetrator*
- Syntax: unrestricted newspaper-style constructions, free word order

Available on my homepage: `www.ims.uni-stuttgart.de/~zarriesa`

Combining Annotations for REG and SR

Junge Familie auf dem Heimweg ausgeraubt
Young family robbed on the way home

Die Polizei sucht nach zwei Männern im Alter von 25 Jahren.
The Police looks for two 25-year-old men.

Sie sollen am Montag gegen 20 Uhr eine junge Familie überfallen haben
They are said to have attacked a young family on Monday around 20 o'clock.

Combining Annotations for REG and SR

Manual RE annotation

Junge Familie auf dem Heimweg ausgeraubt

Young family robbed on the way home

implicit perp head

implicit victim head

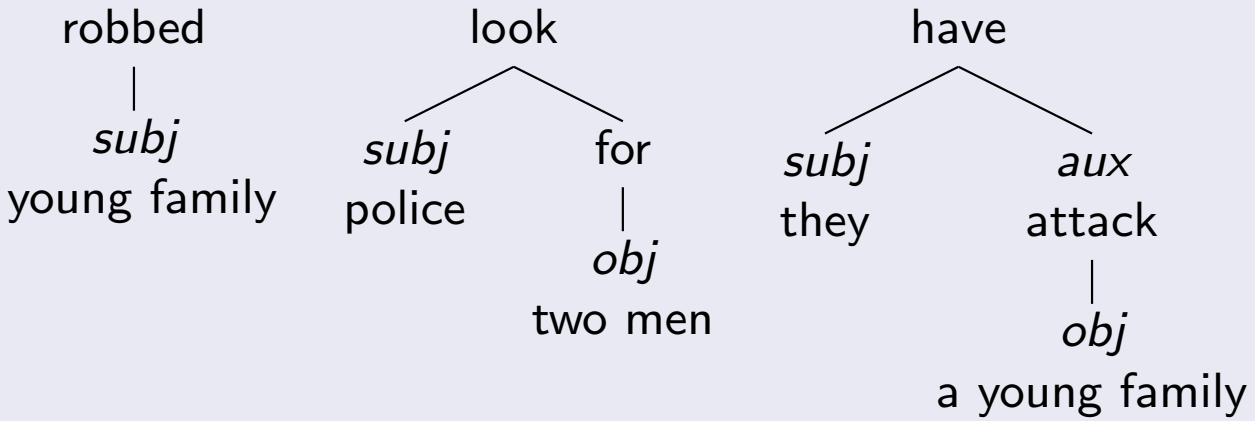
Die Polizei sucht nach zwei Männern im Alter von 25 Jahren. PERP

The Police looks for two 25-year-old men. VICTIM

Sie sollen am Montag gegen 20 Uhr eine junge Familie überfallen haben

They are said to have attacked a young family on Monday around 20 o'clock.

Shallow dependency annotation



- Bohnet (2010) dependency parser
- remove surface order of nodes

Combining Annotations for REG and SR

Manual RE annotation

Junge Familie auf dem Heimweg ausgeraubt

Young family robbed on the way home

implicit perp head

implicit victim head

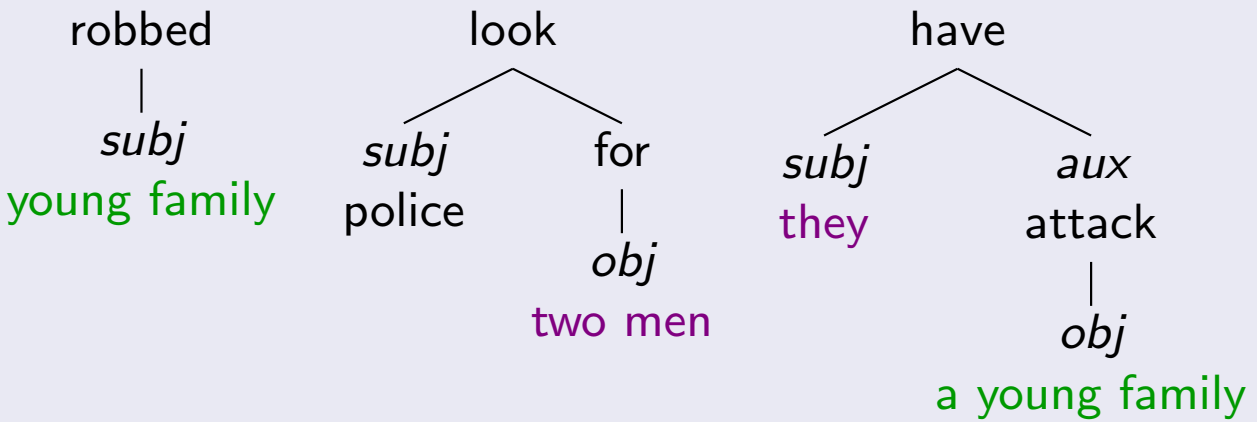
Die Polizei sucht nach zwei Männern im Alter von 25 Jahren. PERP

The Police looks for two 25-year-old men. VICTIM

Sie sollen am Montag gegen 20 Uhr eine junge Familie überfallen haben

They are said to have attacked a young family on Monday around 20 o'clock.

Shallow dependency annotation



- Bohnet (2010) dependency parser
- remove surface order of nodes

Combining Annotations for REG and SR

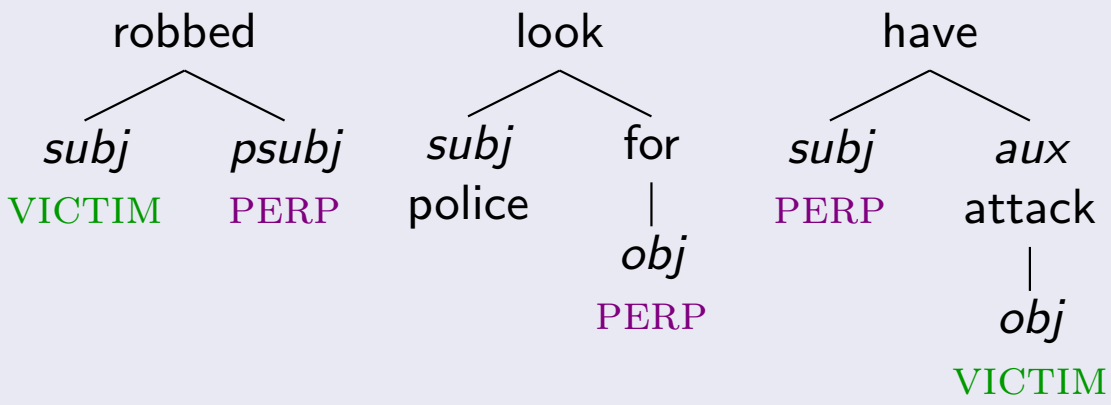
Manual RE annotation

Junge Familie auf dem Heimweg ausgeraubt
Young family robbed on the way home

Die Polizei sucht nach zwei Männern im Alter von 25 Jahren.
The Police looks for two 25-year-old men.

Sie sollen am Montag gegen 20 Uhr eine junge Familie überfallen haben
They are said to have attacked a young family on Monday around 20 o'clock.

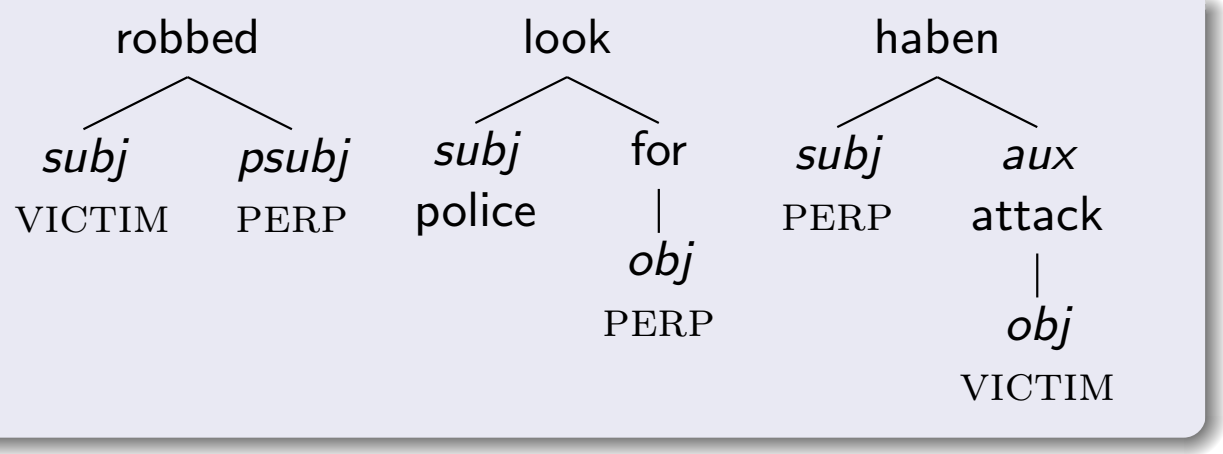
Shallow dependencies with RE slots



RE candidates

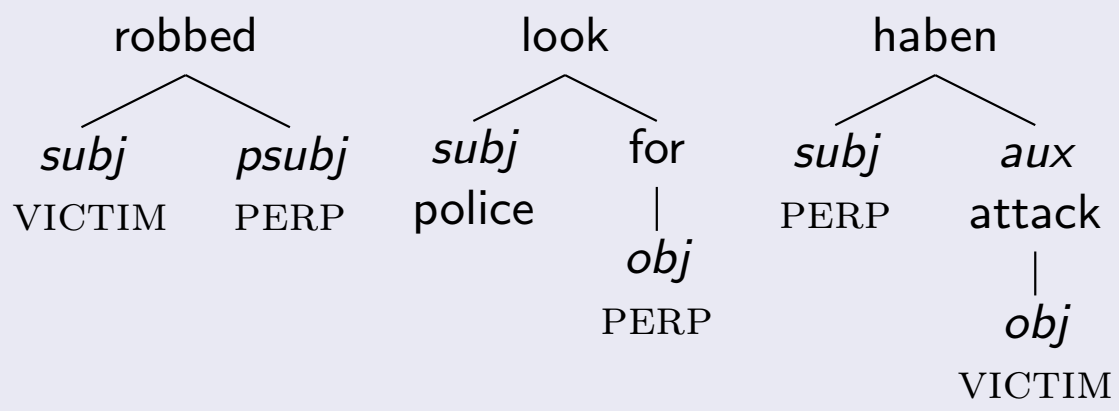
PERP: "two men" | "they" | empty
VICTIM: "young family" | "they" | "a young family" | empty

Shallow dependencies with RE slots



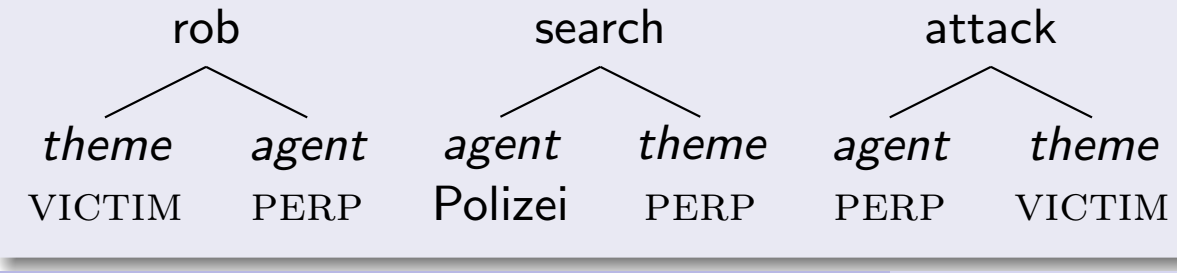
- rule-based transformations for passives, nominalizations
- remove auxiliaries, map syntactic functions

Shallow dependencies with RE slots

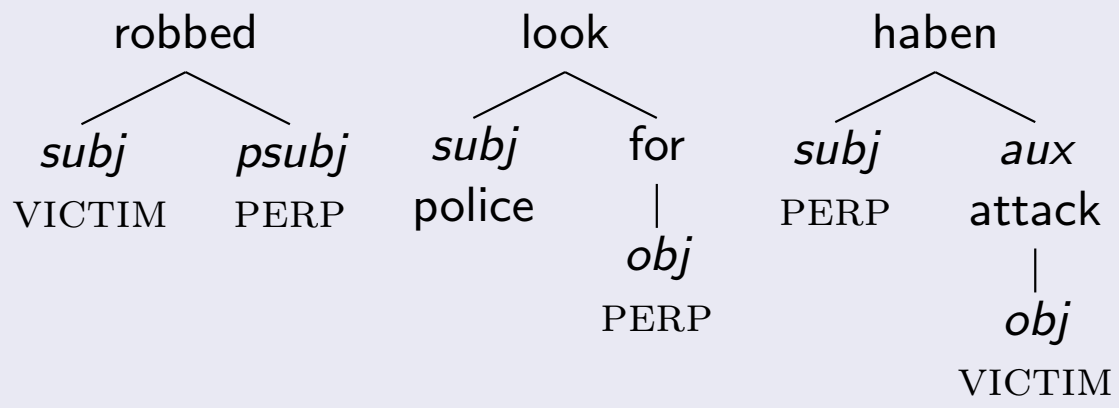


- rule-based transformations for passives, nominalizations
- remove auxiliaries, map syntactic functions

Deep dependencies with RE slots

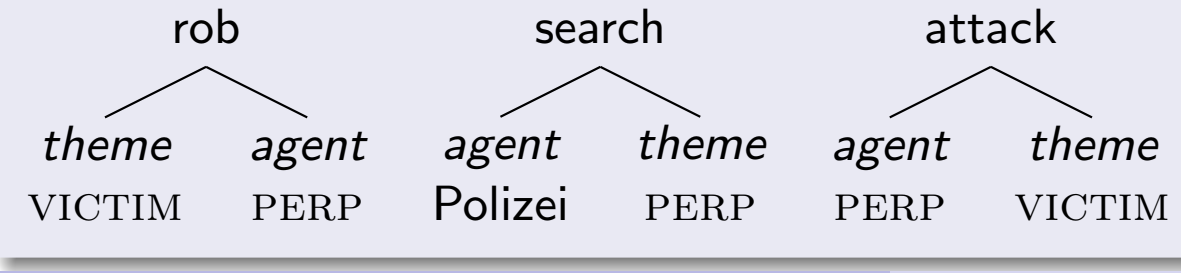


Shallow dependencies with RE slots



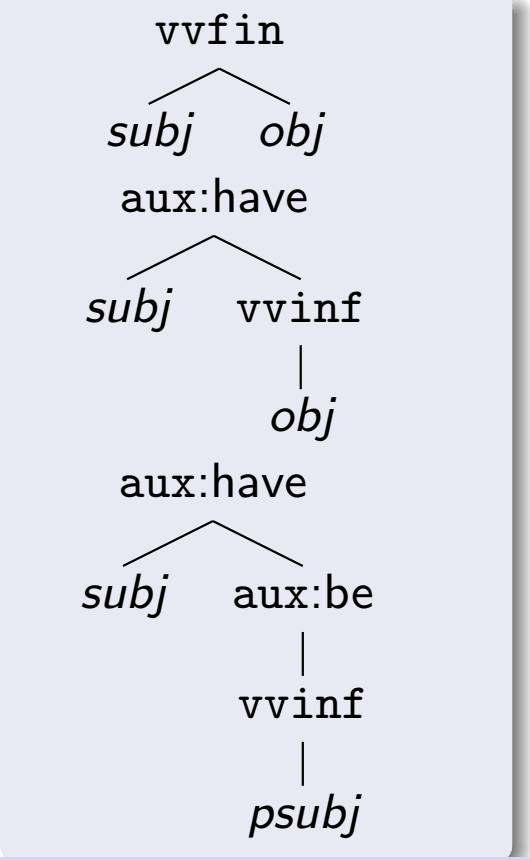
- rule-based transformations for passives, nominalizations
- remove auxiliaries, map syntactic functions

Deep dependencies with RE slots



Shallow verb candidates

deep → shallow alignments



Generation Modules

REG
(based on SVMrank, Joachims 2006)

rank RE candidates for a given slot in tree
($\rightarrow$ set of all original REs for a referent/text)

label=subj+RE=PRON head=attack+RE=PRON
label=subj+RE=DEF head=attack+RE=DEF
label=subj+RE=IMPL head=attack+RE=IMPL

SYN
(based on SVMrank, Joachims 2006)

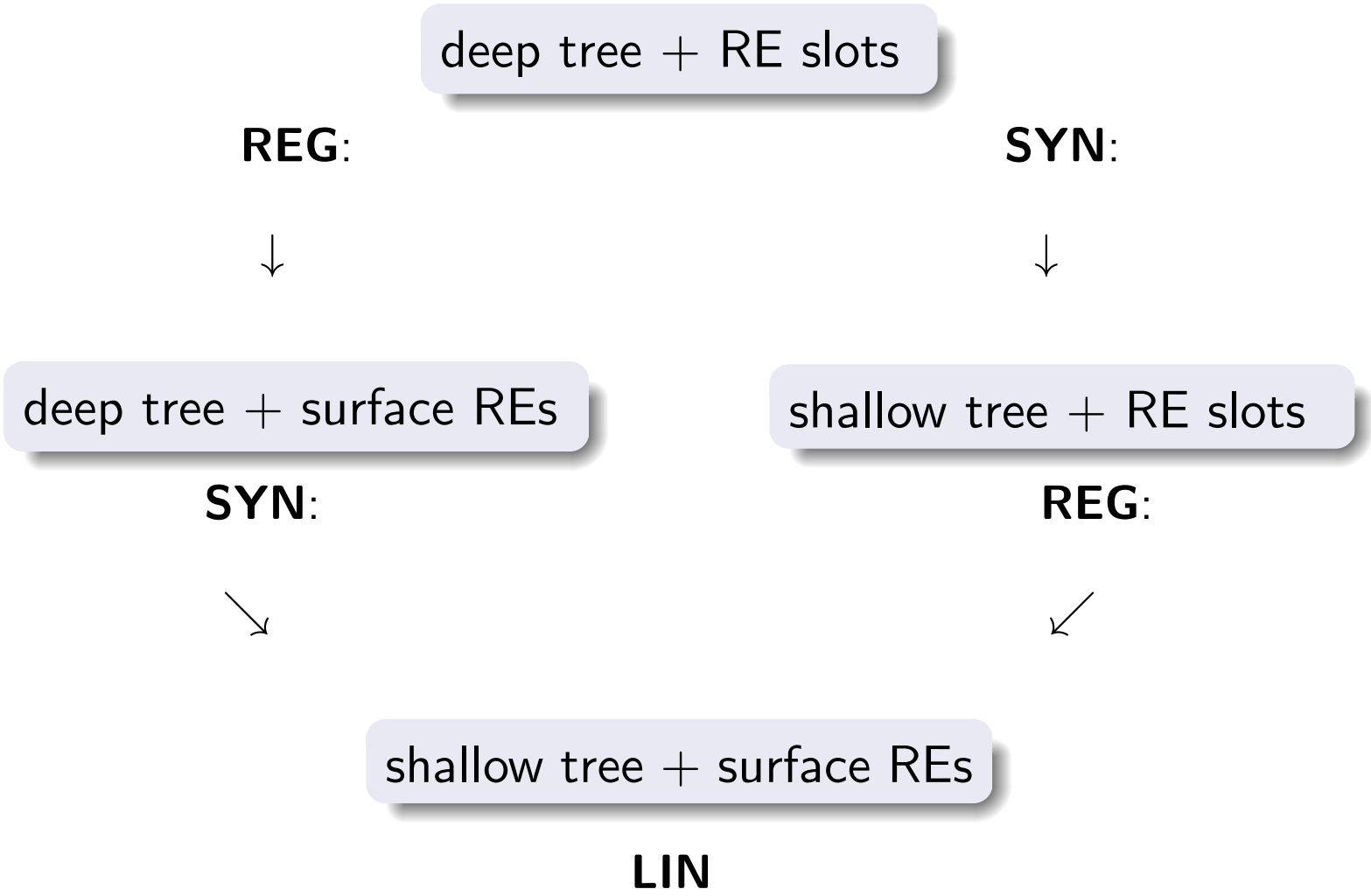
rank shallow candidates for a verb node in deep tree
($\rightarrow$ set of verb transformations/training set)

agent=victim+V=ACT theme=perp+V=ACT
agent=victim+V=NOM theme=perp+V=NOM
agent=victim+V=PASS theme=perp+V=PASS

LIN

Bohnet et al. (2012), trained on German TIGER corpus

Ordering Modules: Pipelines



Ordering Modules: Pipelines

deep tree + RE slots

REG:

label=agent+RE=PRONOUN

label=agent+RE=DEFINITE

SYN:

↓

deep tree + surface REs

SYN:

↘

shallow tree + RE slots

REG:

label=subj+RE=PRONOUN

label=subj+RE=DEFINITE

shallow tree + surface REs

LIN

Ordering Modules: Pipelines

deep tree + RE slots

REG:

↓

SYN:

AGENT=VICTIM+shallow=finite

AGENT=VICTIM+shallow=nominal

deep tree + surface REs

SYN:

AGENT=PRONOUN+shallow=finite

AGENT=PRONOUN+shallow=nominal

shallow tree + RE slots

REG:

↙

shallow tree + surface REs

LIN

First Results for Standard Pipelines

- 10-fold cross-validation on 200 texts
- sentences are lemmatized
- SYN/RE accuracy:
  proportion of exactly matching shallow/RE subtrees
- Baseline: rule-based heuristics for REG and SYN

	BLEU	**SYN Accuracy**	**RE Accuracy**
Baseline			
REG ... SYN			
SYN ... REG			
LIN on original shallow			

First Results for Standard Pipelines

- Pipelines perform very similarly

	BLEU	SYN Accuracy	RE Accuracy
Baseline			
REG ... SYN	54.65		
SYN ... REG	54.28		
LIN on original shallow			

First Results for Standard Pipelines

- SYN suffers from REG errors

	BLEU	SYN Accuracy	RE Accuracy
Baseline			
REG ... SYN	54.65	57.09	
SYN ... REG	54.28	59.14	
LIN on original shallow			

First Results for Standard Pipelines

- REG suffers from SYN errors

	BLEU	SYN Accuracy	RE Accuracy
Baseline			
REG ... SYN	54.65	57.09	54.61
SYN ... REG	54.28	59.14	52.21
LIN on original shallow			

First Results for Standard Pipelines

- REG and SYN introduce considerable uncertainty

	BLEU	SYN Accuracy	RE Accuracy
Baseline	42.38	35.66	33.3
REG ... SYN	54.65	57.09	54.61
SYN ... REG	54.28	59.14	52.21
LIN on original shallow	79.17	100	100

More Results

- How can we address the effect of error propagation?

	BLEU	SYN Accuracy	RE Accuracy
REG ... SYN	54.65	57.09	54.61
SYN ... REG	54.28	59.14	52.21

More Results

- Parallel system: Combine independent predictions of SYN and REG

	BLEU	SYN Accuracy	RE Accuracy
REG ... SYN	54.65	57.09	54.61
SYN ... REG	54.28	59.14	52.21
Parallel		59.14	54.61

More Results

- Parallel system improves only marginally over pipelines

	BLEU	SYN Accuracy	RE Accuracy
REG ... SYN	54.65	57.09	54.61
SYN ... REG	54.28	59.14	52.21
Parallel	54.78	59.14	54.61

Towards Capturing Interactions: Revision-based Architecture

Intermediate linearization

many known interactions between REG and linearization
Centering Theory (Grosz et al., 1995)

approximates integrated architecture
e.g. (Robin, 1993)

deep tree + RE slots

SYN

shallow tree + RE slots

LIN:
preliminary surface order of nodes

shallow tree + RE slots + order

REG:
additional positional features

shallow tree + REs + order

LIN

Final Results

	BLEU	SYN Accuracy	RE Accuracy
REG ... SYN	54.65	57.09	54.61
SYN ... REG	54.28	59.14	52.21
Parallel	54.78	59.14	54.61
Revision			

Final Results

- Revision-based system leads to clear over-all improvements
- No error propagation from SYN to REG if shallow trees are linearized

	BLEU	SYN Accuracy	RE Accuracy
REG ... SYN	54.65	57.09	54.61
SYN ... REG	54.28	59.14	52.21
Parallel	54.78	59.14	54.61
Revision	56.31	59.14	58.81

Final Results

- Is the effect only due to higher lexical overlap because of better RE accuracy?
- $BLEU_r$:
  replace REs by placeholder in gold and predicted sentence, BLEU score factoring out lexical RE overlap

	BLEU	**SYN Accuracy**	**RE Accuracy**	**$BLEU_r$**
REG ... SYN	54.65	57.09	54.61	
SYN ... REG	54.28	59.14	52.21	
Parallel	54.78	59.14	54.61	
Revision	56.31	59.14	58.81	

Final Results

- Improvements in $BLEU_r$ indicate generally better linearization/sentence quality

	BLEU	**SYN Accuracy**	**RE Accuracy**	**$BLEU_r$**
REG ... SYN	54.65	57.09	54.61	
SYN ... REG	54.28	59.14	52.21	
Parallel	54.78	59.14	54.61	60.05
Revision	56.31	59.14	58.81	61.30

Conclusions

- Existing annotation standards for REG and SR can be easily integrated
- Data-driven study of generation architectures
- First results point to shortcomings of standard pipeline set-up
- Systems need to capture interactions (e.g. via revision)