Automatic Labeling of Semantic Roles

Presentation by
Philip John Gorinski

Seminar
„Recent Developments in Computational Semantics“

Saarland University, Summer Term 2011

Outline

- Introduction
  - The concept of Semantic Roles
  - FrameNet
- Automatic Role Assignment
  - The task
  - Features
  - Probability Estimation
  - Results
- Generalizing to abstract Sematic Roles
- Conclusion

Outline

- Introduction
  - The concept of Semantic Roles
  - FrameNet
- Automatic Role Assignment
  - The task
  - Features
  - Probability Estimation
  - Results
- Generalizing to abstract Sematic Roles
- Conclusion

The concept of Semantic Roles

- Concept of Semantic Roles dates back thousands of years
- High variety of theories
- From very specific
  - Domain dependent Information Extraction
  - Domain dependent Dialogue Understanding
- To very general
  - e.g., as part of linking theory

The concept of Semantic Roles

- Proto-Roles (Van Valin, 1993; Dowty, 1991)
  - **`Proto-Agent`**
  - **`Proto-Patient`**

"$[\text{Jane}]_{AG}$ hit $[\text{Tommy}]_{PAT}$ $[\text{with a bat}]_{PAT}$."

- Typically proposed by linguists, e.g., concerned with integration into linking theory

The concept of Semantic Roles

- Domain specific slots (e.g. travel booking)
  - "I would like to fly [from Chicago]$_{\texttt{orig\_city}}$ [to New York]$_{\texttt{dest\_city}}$ [on the 19$^{\texttt{th}}$]$_{\texttt{depart\_date}}$."
- Verb specific roles.
  - "[I]$_{\text{eater}}$ ate [an apple]$_{\text{eaten}}$."
  - "[I]$_{\text{devourer}}$ devoured [an apple]$_{\text{devoured}}$."
- Typically used by computer scientists for implementation

Outline

- Introduction
  - The concept of Semantic Roles
  - FrameNet
- Automatic Role Assignment
  - The task
  - Features
  - Probability Estimation
  - Results
- Generalizing to abstract Sematic Roles
- Conclusion

FrameNet

- Developed at ICSI Berkeley
- Aims to be in the middle of the spectrum
  - Neither too general
  - Nor too specific
- FrameNet **roles** defined as **Frame Elements**
  - part of the Frame they are defined for
- FrameNet **Frames** defined on **situations** involving various roles
  - invoked by numerous predicates (verbs, nouns, adjectives)

FrameNet

Example Frame: Discussion

Description: Two (or more) people talk to one another. No person is construed as only a speaker or only an addressee. Rather, it is understood that both (or all) participants do some speaking and some listening – the process is understood to be symmetrical or reciprocal.

Frame Elements:

`Interlocutor_1`
`Interlocutor_2`
`Interlocutors`
`Topic`

FrameNet

Example Frame: Discussion

Frame Elements:
Interlocutor_1
Interlocutor_2
Interlocutors
Topic

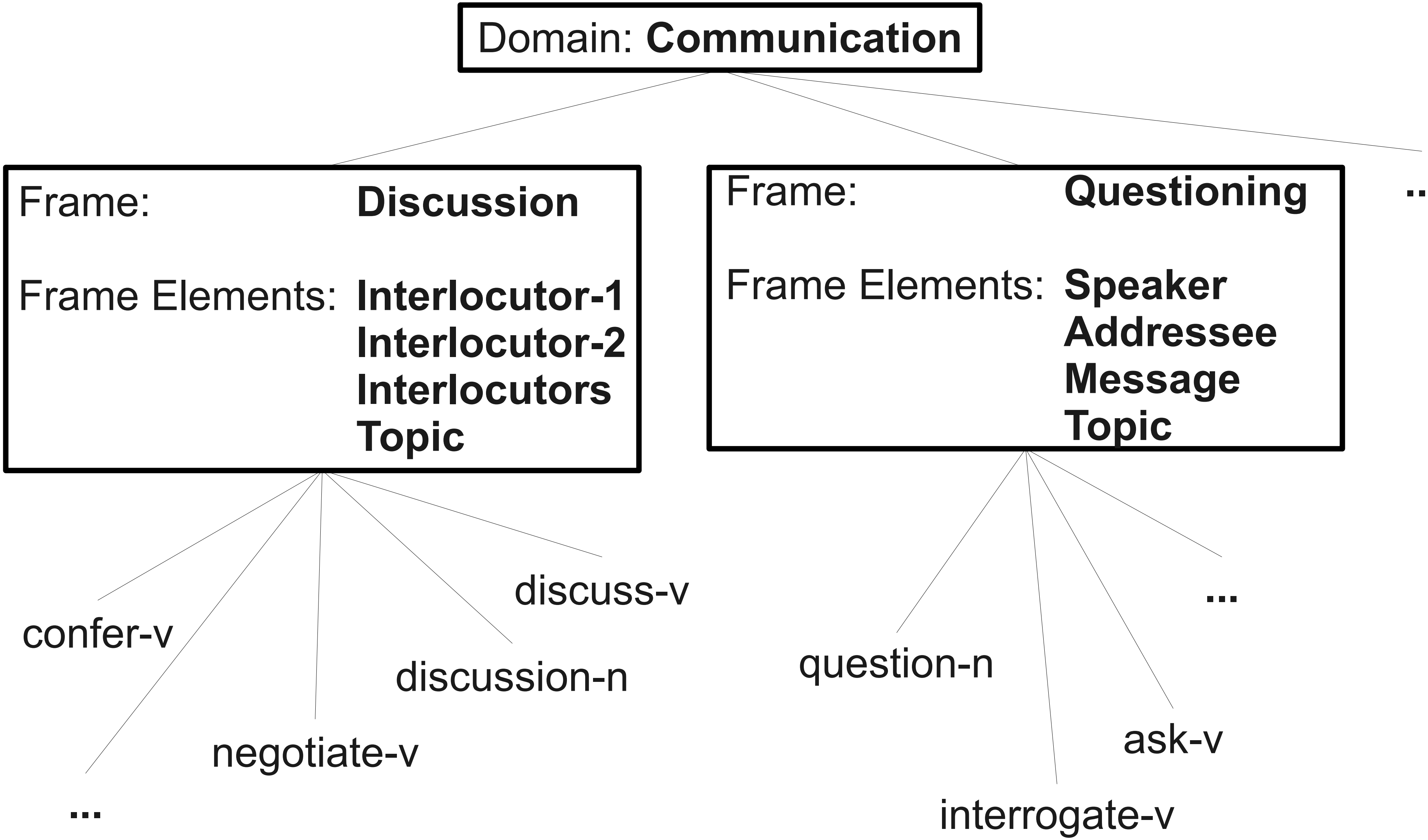
”[Peter]$_{\text{Interlocutor\_1}}$ and [Mary]$_{\text{Interlocutor\_2}}$ *discussed* [the women's soccer wold cup]$_{\text{Topic}}$.”

“[The companies]$_{\text{Interlocutors}}$ *negotiated* [the contract]$_{\text{Topic}}$.”

FrameNet

- Frames form a hierarchy
  - from more general to more specific
  - allows for inheritance of Frame Elements
- Abstract thematic roles such as **`Proto-Agent`** can be seen as defined by abstract frames
  - inheriting Frames add semantics to the more general Frames
- Currently ~800 Frames, 10.000 lexical units, 120.000 sample sentences taken from the BNC
  - hand annotated

FrameNet


Domain: **Communication**
Frame: **Discussion**
Frame Elements: **Interlocutor-1** **Interlocutor-2** **Interlocutors** **Topic**
Frame: **Questioning**
Frame Elements: **Speaker** **Addressee** **Message** **Topic**
...
confer-v
discuss-v
discussion-n
negotiate-v
...
question-n
...
ask-v
interrogate-v

FrameNet

- Different predicates sharing the same Frame can provide very useful information
- e.g., in question answering, using the Sending-Frame for "send-v" and "receive-v"

"[Which party]$_{\text{Sender}}$ *sent* [absentee ballots]$_{\text{Theme}}$ to [voters]$_{\text{Recipient}}$?"

"[Democratic and Republican voters]$_{\text{Recipient}}$ *received* [absentee ballots]$_{\text{Theme}}$ [from their parties]$_{\text{Sender}}$."

Outline

- Introduction
  - The concept of Semantic Roles
  - FrameNet
- Automatic Role Assignment
  - The task
  - Features
  - Probability Estimation
  - Results
- Generalizing to abstract Sematic Roles
- Conclusion

The Task

- Systems using semantic Frames typically are
- → hand-crafted
  - human engineers “make up” and annotate frames
- → domain dependent
  - e.g., flight booking systems use slots like **`Orig_City`**, **`Dest_City`**, **`Depart_Date`** [...]
  - Merger and acquisition systems use slots **`Products`**, **`Joint_Venture_Company`**, **`Amount`** [...]
- Can we automatically label data with more domain-independent semantic information?

The Task

- Preliminary version of FrameNet Corpus
  - 67 frame types
  - 12 semantic domains
  - 1.462 target words
  - 49.013 annotated sentences
  - 99.232 annotated frame elements
- For each target word, corpus is split into
  - 10% test data
  - 10% tuning set
  - 80% training data

The Task

- Given a sentence with
  - target word, FrameNet Frame, Frame Element boundaries
- Rank the semantic roles which the syntactic constituents are likely to fill
  - By means of a probabilistic classifier
- Automatically assign FrameNet roles to sentence constituents
  - most likely role, according to a certain set of features

The Task

Frame: Discussion

Frame Elements: `Interlocutor_1`
`Interlocutor_2`
`Interlocutors`
`Topic`

“[The companies]$_{\texttt{element1?}}$ *negotiated* [the contract]$_{\texttt{element2?}}$.”

Estimate and rank:

*P(**Interlocutor_1**| [the companies])*
*P(**Interlocutor_2**| [the companies])*
*P(**Interlocutors**| [the companies])*
*P(**Topic**| [the companies])*
*P(**Interlocutor_1**| [the contract])*

...

The Task

- Automatic labeling is challenging
- Not always direct correspondence between syntactic category and semantic role

  "[We]$_{\text{Judge}}$ *praised* [her apple pie]$_{\text{Evaluee}}$."

  "[The actor]$_{\text{Evaluee}}$ received [critical]$_{\text{Judge}}$ *praise*."

- Same role for different syntactic functions

  "[We]$_{\text{Judge}}$ *praised* [her apple pie]$_{\text{Evaluee}}$."

  "[Her apple pie]$_{\text{Evaluee}}$ was *praised*."

Outline

- Introduction
  - The concept of Semantic Roles
  - FrameNet
- Automatic Role Assignment
  - The task
  - Features
  - Probability Estimation
  - Results
- Generalizing to abstract Sematic Roles
- Conclusion

Features

- Linking theory suggests correspondence between a sentence's syntax and semantics
- Original FrameNet Corpus is not syntactically annotated
- Automated syntactic parser used to analyze the training set
  - parse sentences
  - match syntactic constituents to frame element boundaries

Features

- 6 features extracted for probability estimation
  - Phrase Type
  - Governing Category
  - Parse Tree Path
  - Position
  - Voice
  - Head Word

Features

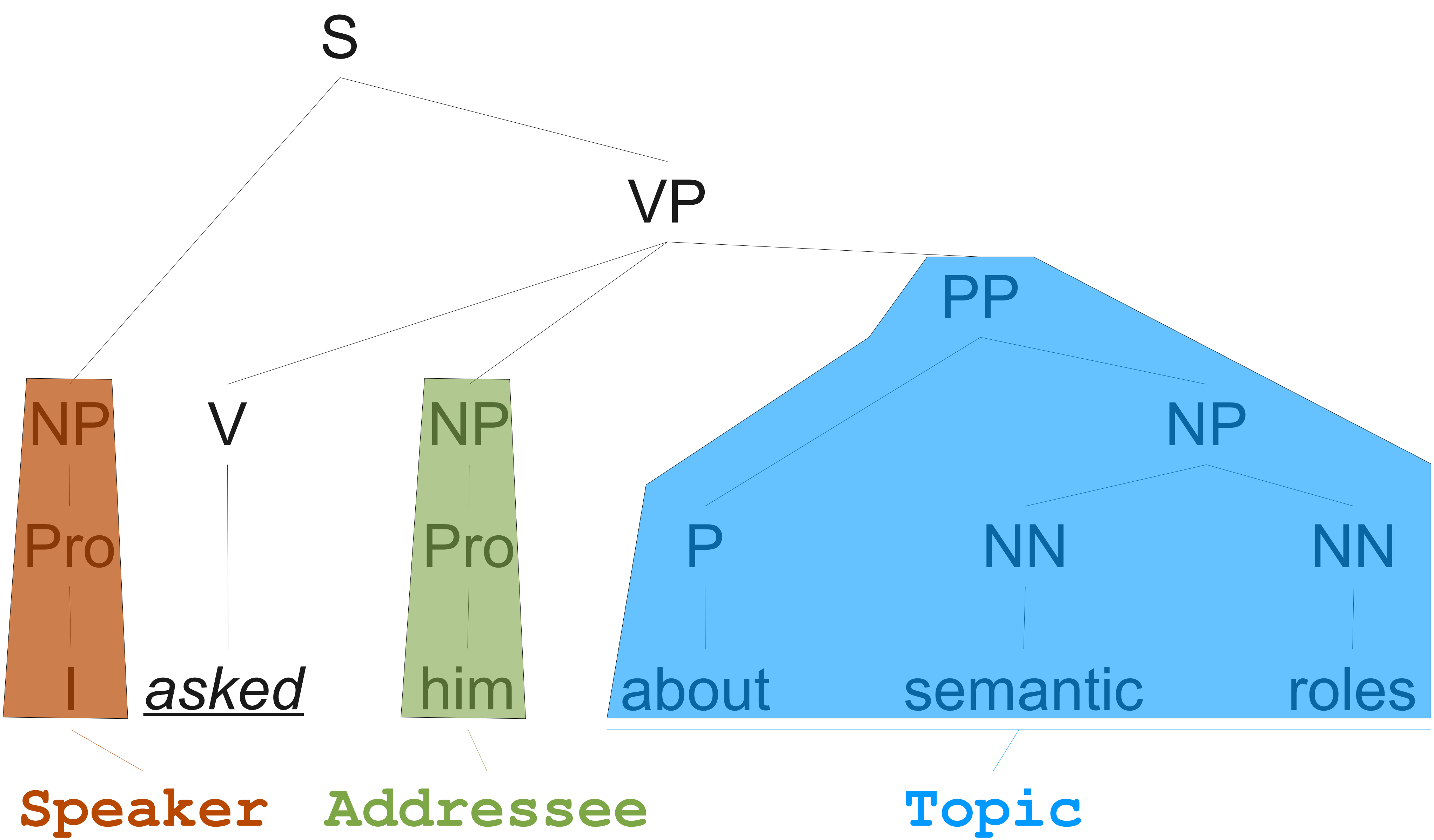
- **Phrase Type**

- Different roles ~ Different syntactic categories
  - For Frame Questioning

    `Speaker` ~ Noun Phrase

    `Topic` ~ Prepositional Phrase
- Extracted from parse trees
- Most common categories for Frame Elements:
  - Noun Phrases (47%)
  - Prepositional Phrases (22%)

Features

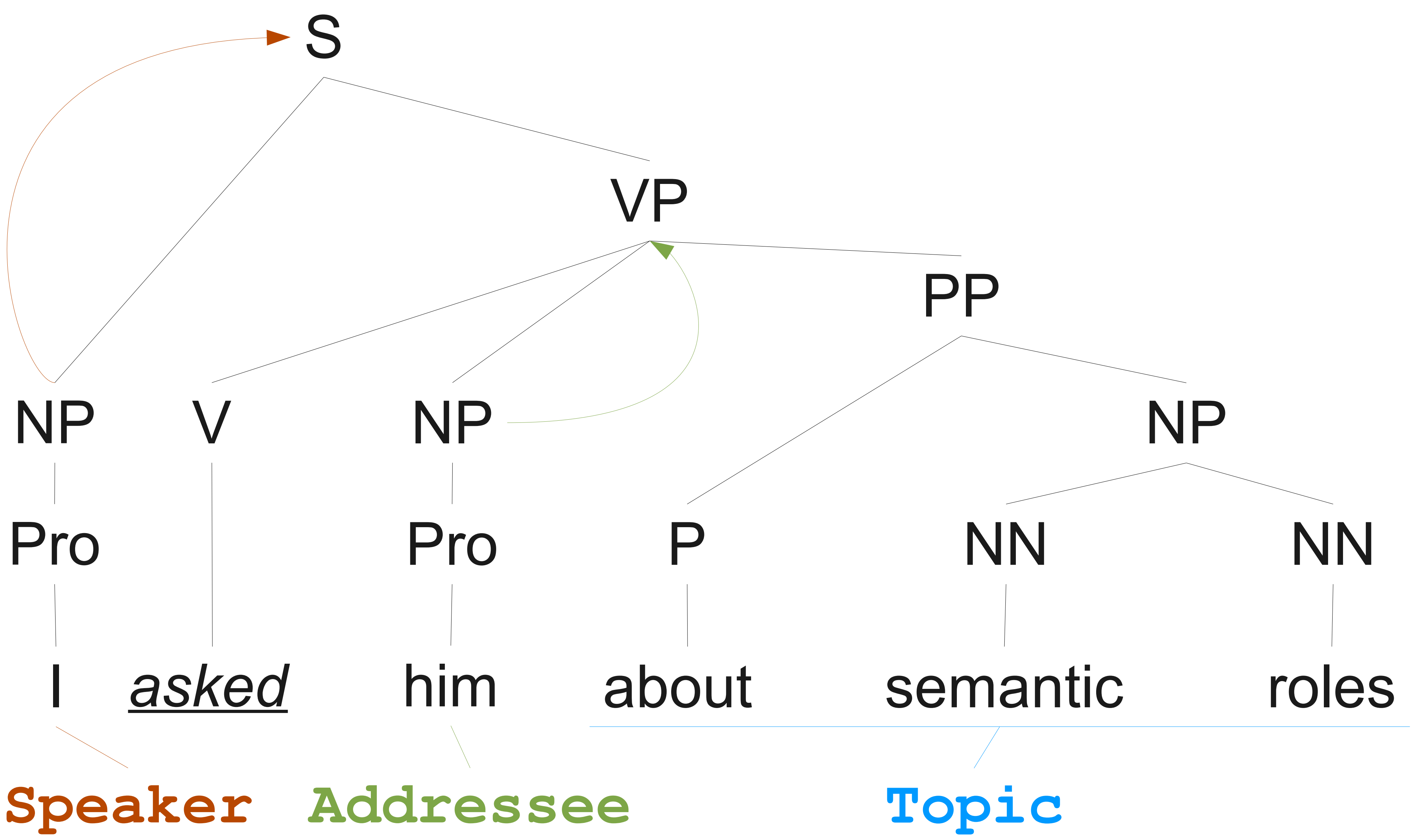

S
VP
PP
NP
V
NP
NP
Pro
Pro
P
NN
NN
I
asked
him
about
semantic
roles
Speaker
Addressee
Topic

Features

- **Governing Category**

- Semantic role ~ Syntactic subject / direct object
- “He drove the car over the cliff.”
  - Subject NP more likely to be `Agent` than the two other NPs
- Two values: S (subject) & VP (object)
- Only applied to Noun Phrases
- First S or VP reachable from target NP

Features

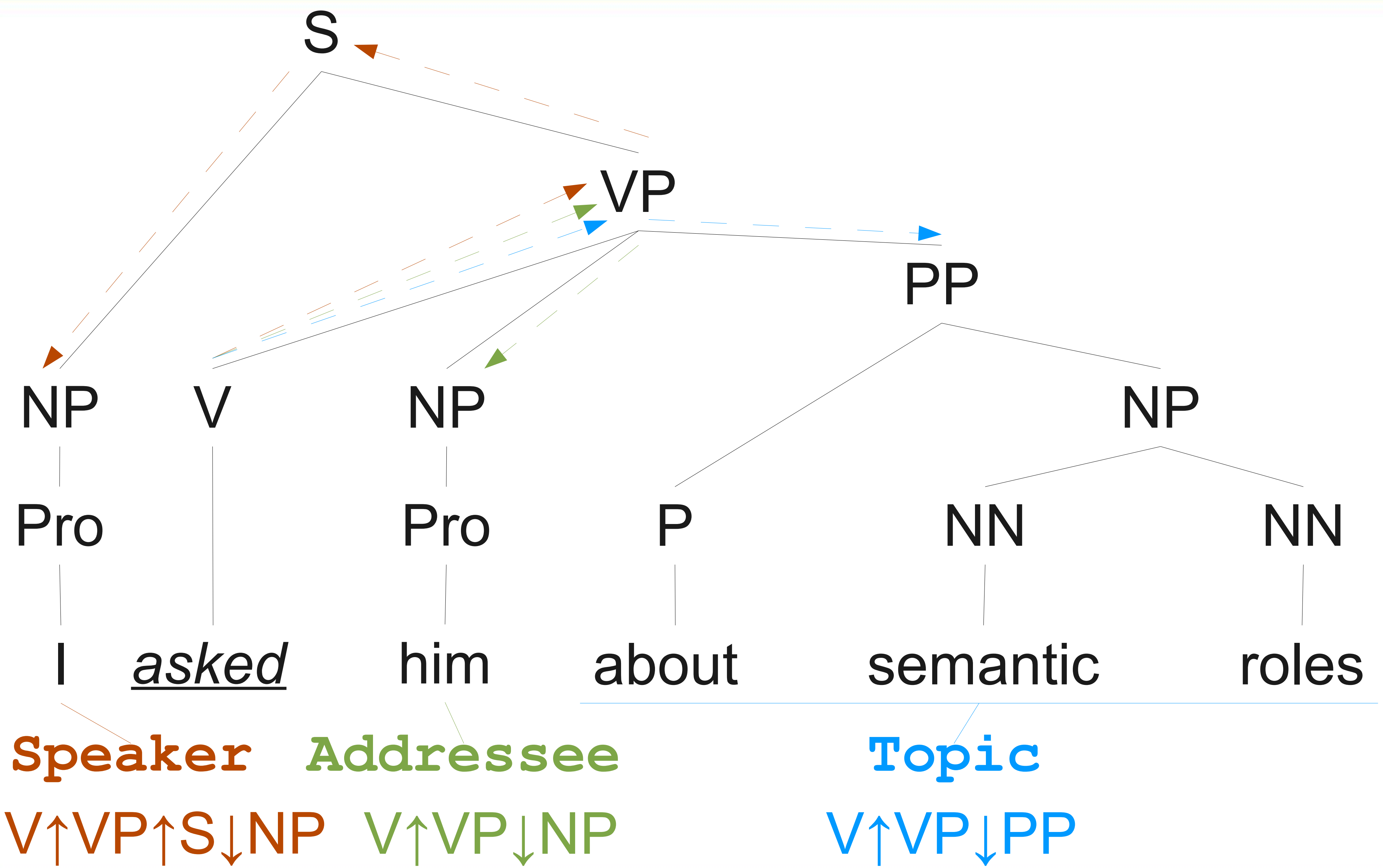


S
VP
PP
NP
V
NP
NP
Pro
Pro
P
NN
NN
I
asked
him
about
semantic
roles
Speaker
Addressee
Topic

Features

- **Parse Tree Path**
- Similar to Governing Category, but
- Path from target word to parse constituent
  - thus potentially unlimited values
- Represented as string using categories and up/down arrows
  - e.g., Prototypical Subject: V↑VP↑S↓NP
- Not limited to NPs

Features



S
VP
PP
NP
V
NP
NP
Pro
Pro
P
NN
NN
I
asked
him
about
semantic
roles
Speaker
Addressee
Topic
V↑VP↑S↓NP
V↑VP↓NP
V↑VP↓PP

Features

- **Position**

- Independent of parse tree
- Indicates whether constituent is to the left or to the right of the target word
  - in general, left ~ subject; right ~ object
- Used primarily to overcome parse errors

Features

- **Voice**

- Distinction between active and passive verbs
- Generally, direct object role of active verbs corresponds to subject role of passive verbs

  [I]$_{\text{Agent}}$ <u>*broke*</u> [the window]$_{\text{Whole\_patient}}$.

  [The window]$_{\text{Whole\_patient}}$ was <u>*broken*</u>.

Features

- **Head Word**

- Head words of Noun Phrases can indicate selectional restrictions for roles
  - e.g., in the Discussion Frame

    I, they, Bill ~ `Speaker`

    him, them ~ `Addressee`

    proposal, contract ~ `Topic`

Outline

- Introduction
  - The concept of Semantic Roles
  - FrameNet
- Automatic Role Assignment
  - The task
  - Features
  - Probability Estimation
  - Results
- Generalizing to abstract Sematic Roles
- Conclusion

Probability Estimation

- For a given constituent, estimate the probability of it filling a certain role *r*

$$P(r \mid h, pt, gov, pos, voice, t)$$

h: head word; pt: phrase type, t: target word

- Intuitively, this is

$$P(r \mid \ldots) = \frac{\#(r, h, pt, gov, pos, voice, t)}{\#(h, pt, gov, pos, voice, t)}$$

Times role *r* occurs with given features in training data, divided by times the given features occur in total

Probability Estimation

- Problems with initial model
- High amount of sparse data
  - small number of sentences per target word
  - large number of values for certain features, especially Head Word
  - most combinations of features will never occur, or only occur very few times in the training data
- Solution: Estimate probabilities for distributions over several feature subsets

Probability Estimation

Distribution	Coverage	Accuracy	Performance
$P(r \mid t)$	100%	40.9%	40.9%
$P(r \mid pt, t)$	92.5	60.1	55.6
$P(r \mid pt, gov, t)$	92.0	66.6	61.3
$P(r \mid pt, pos, voice)$	98.8	57.1	56.4
$P(r \mid pt, pos, voice, t)$	90.8	70.1	63.7
$P(r \mid h)$	80.3	73.6	59.1
$P(r \mid h, t)$	56.0	86.6	48.5
$P(r \mid h, pt, t)$	50.1	87.4	43.8

Probability Estimation

- Trade-off between coverage and accuracy
  - more general distributions cover more data, show lower accuracy
  - more specific distributions cover less data, show higher accuracy
- Combine distributions to achieve broad coverage as well as high accuracy
  - Linear Interpolation
  - Geometric Mean
  - Backoff Lattice

Linear Interpolation

$$
\begin{aligned}
P(r \mid constituent) = {} & \lambda_1 P(r \mid t) + \lambda_2 P(r \mid pt, pos, voice) \\
& + \lambda_3 P(r \mid pt, t) + \lambda_4 P(r \mid pt, gov, t) \\
& + \lambda_5 P(r \mid pt, pos, voice, t) + \lambda_6 P(r \mid h) \\
& + \lambda_7 P(r \mid h, t) + \lambda_8 P(r \mid h, pt, t)
\end{aligned}
$$

with $\sum_i \lambda_i = 1$

- λ_i provide weights for the different distributions
- can all be the same (Equal Linear Interpolation) or estimated using EM training (EM Linear Interpolation)

Geometric Mean

$$P(r \mid constituent) = \frac{1}{Z}\exp(\lambda_1 logP(r \mid t) + \lambda_2 logP(r \mid pt, pos, voice) + \lambda_3 logP(r \mid pt, t) + \lambda_4 logP(r \mid pt, gov, t) + \lambda_5 logP(r \mid pt, pos, voice, t) + \lambda_6 logP(r \mid h) + \lambda_7 logP(r \mid h, t) + \lambda_8 logP(r \mid h, pt, t))$$

with Z being a normalization constant ensuring that

$$\sum_r P(r \mid constituent) = 1$$

Backoff Lattice

- Construct a lattice over different distributions
  - starting at more specific events
  - ending at more general events
- Use general distribution only if no data is available for more specific distribution
- Select only distributions for which instances have been seen in training
- Combine selected distributions with Linear Interpolation and Geometric Mean

Backoff Lattice

$$
\begin{tikzcd}
P(r \mid h, pt, t) \arrow[d] \arrow[dr] & P(r \mid pt, gf, t) \arrow[d] & P(r \mid pt, pos, voice, t) \arrow[dl] \arrow[d] \\
P(r \mid h, t) \arrow[d] \arrow[dr] & P(r \mid pt, t) \arrow[d] & P(r \mid pt, pos, voice) \\
P(r \mid h) & P(r \mid t) &
\end{tikzcd}
$$

Outline

- Introduction
  - The concept of Semantic Roles
  - FrameNet
- Automatic Role Assignment
  - The task
  - Features
  - Probability Estimation
  - Results
- Generalizing to abstract Sematic Roles
- Conclusion

Results

- All classifiers using the different estimation methods where trained on 80% of the FrameNet Corpus
- Tested on 10% tuning and 10% test sets
- Baseline: Assigning most common role of a Frame to all candidate constituents

Results

Combining Method	**Correct**
development set	
Equal Linear Interpolation	79.5%
EM Linear Interpolation	79.3
Geometric Mean	79.6
Backoff, Linear Interpolation	**80.4**
Backoff, Geometric Mean	79.6
Baseline: Most common role	40.9
test set	
EM Linear Interpolation	78.5
Backoff, Linear Interpolation	76.9
Baseline: Most common role	40.6

Feature Interaction

- 3 features for capturing syntactic relation between target word and candidate constituent
  - position, gov, path
- Do these features have significant effect on the performance combined with other features?
- Construct lattices using either of the 3 features and
  - containing no voice information
  - with independent voice
  - with conjunction of voice and grammatical function

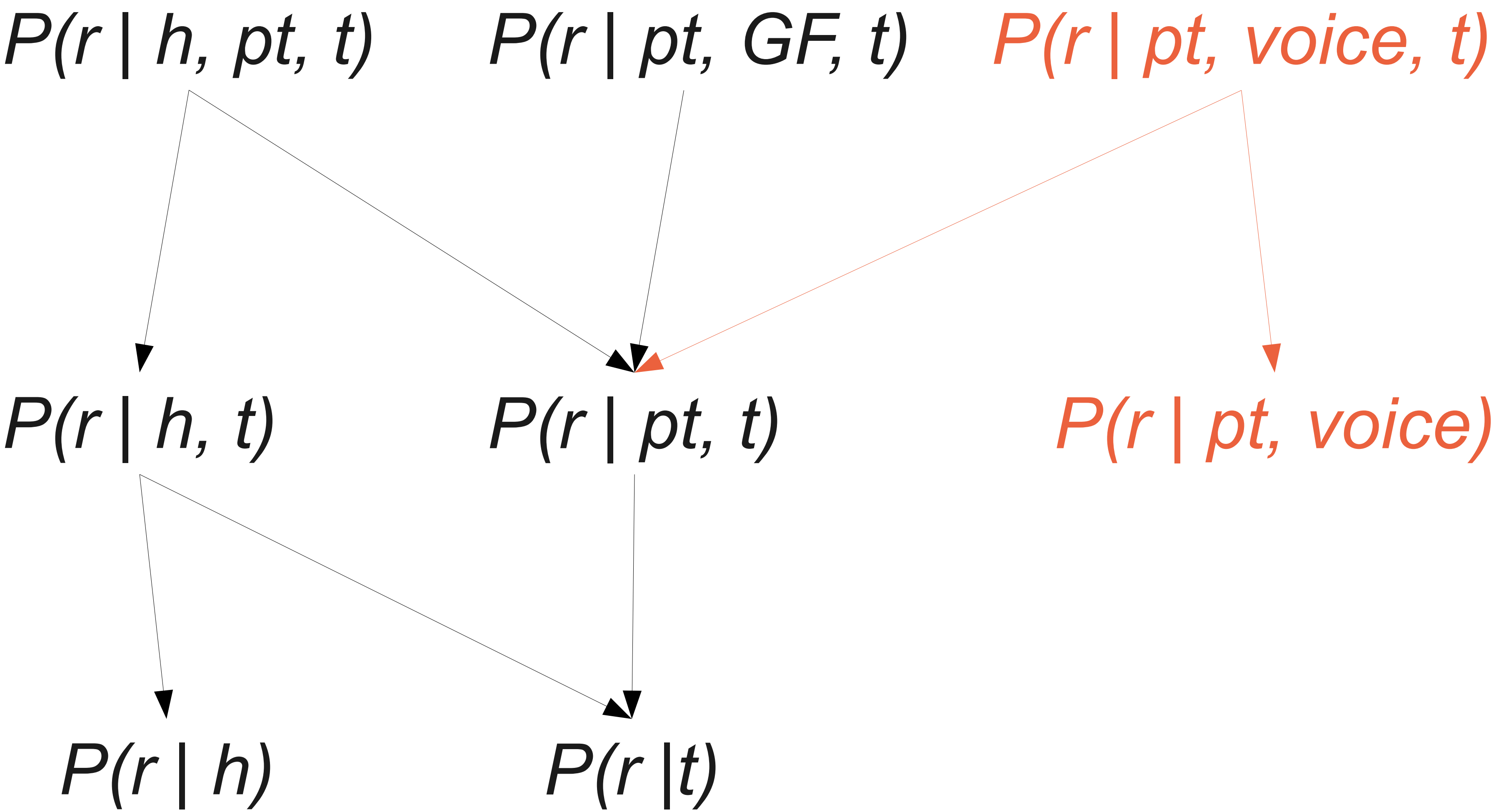
Feature Interaction

No voice information (GF=Grammatical Function)

$$
\begin{tikzcd}
P(r \mid h, pt, t) \arrow[d] \arrow[dr] & P(r \mid pt, GF, t) \arrow[d] \\
P(r \mid h, t) \arrow[d] \arrow[dr] & P(r \mid pt, t) \arrow[d] \\
P(r \mid h) & P(r \mid t)
\end{tikzcd}
$$

Feature Interaction

Independent voice information



Feature Interaction

Conjunction of voice and grammatical function

$P(r \mid h, pt, t)$ $P(r \mid pt, GF, voice, t)$

$P(r \mid h, t)$ $P(r \mid pt, t)$ $P(r \mid pt, GF, voice)$

$P(r \mid h)$ $P(r \mid t)$

Feature Interaction

Feature	No voice	Independent voice	Conjunction
path	79.4%	79.2%	80.4%
gov	79.1	79.2	80.7
position	79.9	79.7	80.5
--	76.3	76.0	76.0

- Even simple position information (candidate left/right of target word) performs similar to information extracted from parse trees
- Grammatical Function interacts with voice
- Using no Grammatical Function at all still yields good results

Outline

- Introduction
  - The concept of Semantic Roles
  - FrameNet
- Automatic Role Assignment
  - The task
  - Features
  - Probability Estimation
  - Results
- Generalizing to abstract Sematic Roles
- Conclusion

Generalizing to abstract roles

- How dependent on the given set of semantic roles is the classifier?
- For unseen frames, how can the data be generalized s.th. automatic role labeling is still possible?
- Use **thematic roles** such as **`Agent`**, **`Patient`**, **`Goal`**
- Allows for generalization over semantic domains
  - “If a sentence has an **`Agent`**, the **`Agent`** will occupy the subject position.”

Generalizing to abstract roles

- To achieve generalization
  - find correspondence from frame-specific roles to abstract thematic roles
  - assign an abstract role to each Frame Element of each FrameNet Frame

- To test the generalized version
  - replace all Frame Element occurrences in the corpus by their abstract roles
  - train and test classifier as before

Generalizing to abstract roles

- General classifier's performance equal to assigning frame-specific roles
  - 82.1% vs. 80.4%
- Shows that the underlying set of roles has little effect on classification
  - roughly 1-to-1 mapping between specific and abstract roles
- Could be useful for annotating unknown frames, independent of semantic domain

Outline

- Introduction
  - The concept of Semantic Roles
  - FrameNet
- Automatic Role Assignment
  - The task
  - Features
  - Probability Estimation
  - Results
- Generalizing to abstract Sematic Roles
- Conclusion

Conclusion

- Automatic labeling of semantic roles is a feasible task
  - 80.4% accuracy in case Frame Element boundaries are known
- Provided classification methods are relatively independent of granularity of semantic roles
- Room for improvement
  - unseen predicates
  - unknown frames
  - finding roles without being given Frame Element boundaries

Thank you very much
for your attention!