




A Hybrid Approach to Finding Negated and Uncertain Expressions in Biomedical Documents

K. Fujikawa, K. Seki, and K. Uehara
Kobe University



Background

- More and more biomedical documents are digitally written and stored
 - It is important to accurately find documents and/or information pertinent to user's need
 - One of the obstacles in finding information in natural language text is **negations**, which deny or reverse the meaning of a sentence or clause.

Background

- More and more biomedical documents are digitally written and stored

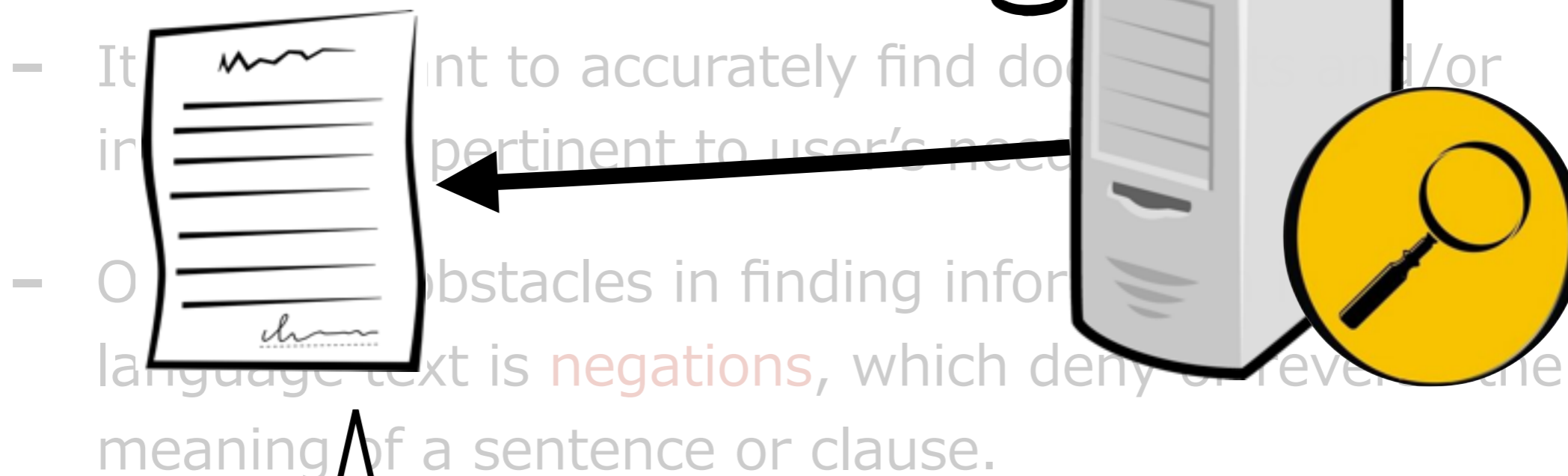
- It is difficult to accurately find documents and/or information pertinent to user's need
- One of the major obstacles in finding information in natural language text is **negations**, which deny or reverse the meaning of a sentence or clause.

ex) There is **no evidence of cervical lymph node enlargement.**

Background

- More and more documents are digitally written and stored.

This document associated with cervical lymph node enlargement



ex) There is **no evidence of cervical lymph node enlargement.**

Background

- More articles associated with cervical lymph node enlargement are digitally written and analyzed.

13.5% of the sentences in biological paper abstracts have negated expressions. (Szarvas et al.)

Obstacles in finding information in natural language text is **negations**, which deny or reverse the meaning of a sentence or clause.

ex) There is **no evidence of cervical lymph node enlargement.**

Related Work

- Manually crafted rule-based approaches
 - A hybrid approach, combining regular expression with grammatical parsing (Huang et al., 2007)
- Supervised classification-based approaches
 - A metalearning approach, combining several classifiers (Morante et al., 2009)

Related Work (Huang, 2007)

Design: lexico-syntactic patterns peculiar to negated expressions so as to spot them

1. Identification of sentences with negation using regular expressions
2. Identification of the scope of negation using heuristics

Related Work (Huang, 2007)

ex) There is no evidence of cervical lymph node enlargement.

Regular expressions (10 patterns)

- N [JJ] NN|NNS {of|for|to suggest} → 【Adjective-like Negation】
- {is|am|are} {not|no longer} {present|existent} → 【Adverb Negation】
- {none|absence|lack|resolution} of → 【Noun Negation】

:

Developed and validated using
500 radiology reports.

Related Work (Huang, 2007)

ex) There is **no evidence of** cervical lymph node enlargement.

Regular expressions (10 patterns)

- **N [JJ] NN|NNS {of|for|to suggest} → [Adjective-like Negation]**
- {is|am|are} {not|no longer} {present|existent} → [Adverb Negation]
- {none|absence|lack|resolution} of → [Noun Negation]

:

Related Work (Huang, 2007)

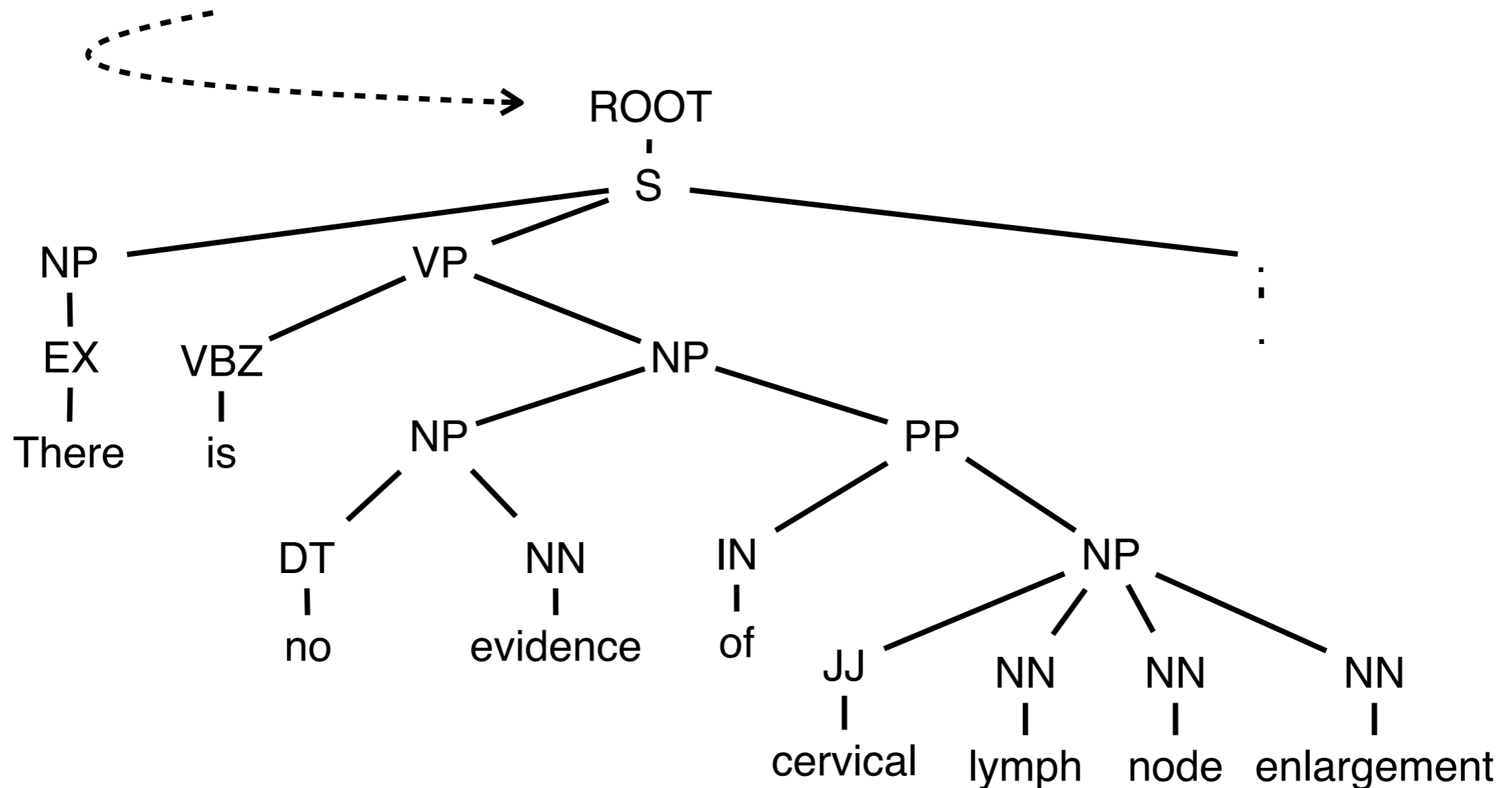
ex) There is **no evidence of cervical lymph node enlargement.**

Grammar rules (manually developed)

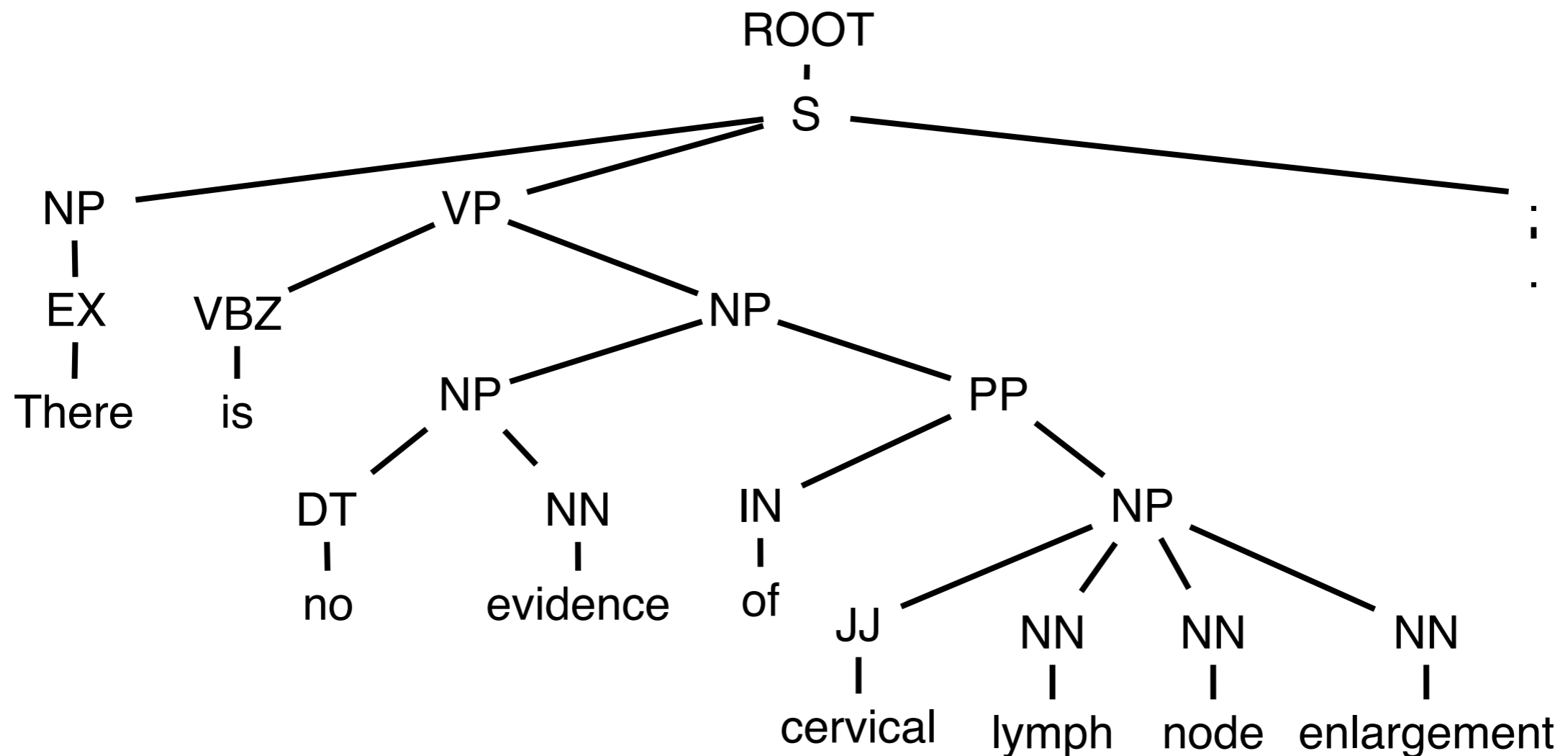
```
Adjective-like Negation : Phrasal : NounPhrase
Expression Pattern : N [JJ] NN|NNS {of|for|to suggest}
NegdPhr
N → {no|without|absent}
JJ → {mammographic|significant}
NN → {evidence|feature|area|pattern|history|sign}
NNS → {features|areas|patterns|signs}
Grammar:
PP → INO NP
INO → {without}
NP → NP PP
NP → DT|JJ0 [JJ1] NN|NNS
NP → [JJ1] NN|NNS
DT → {no}
JJ0 → {absent}
JJ1 → {mammography|significant}
NN → {evidence|feature|area|pattern|history|sign}
NNS → {features|areas|patterns|signs}
PP → IN NP
IN → {of|for}
NP → NegdPhr
```

Related Work (Huang, 2007)

ex) There is no evidence of cervical lymph node enlargement.

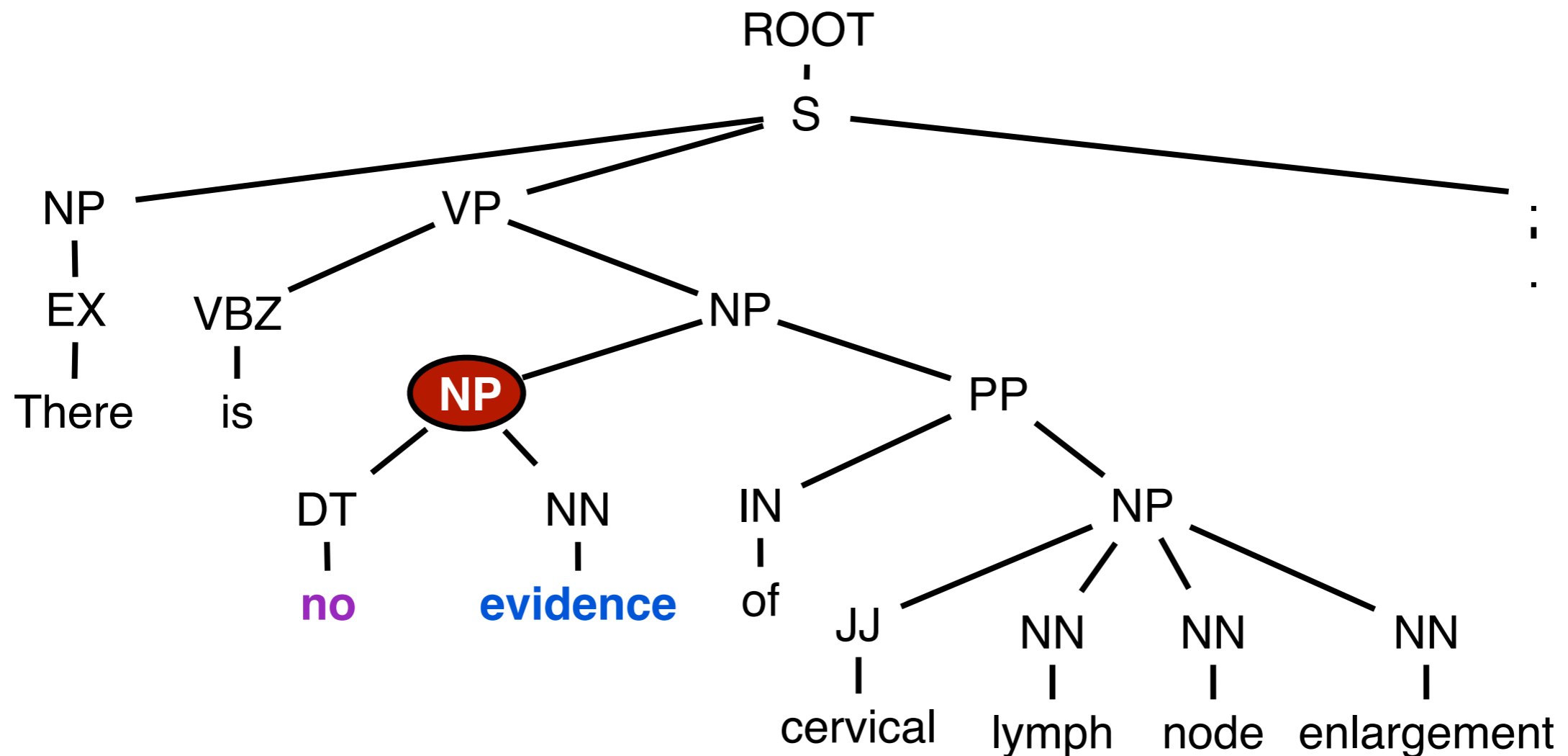


Related Work (Huang, 2007)



Adjective-like negation pattern

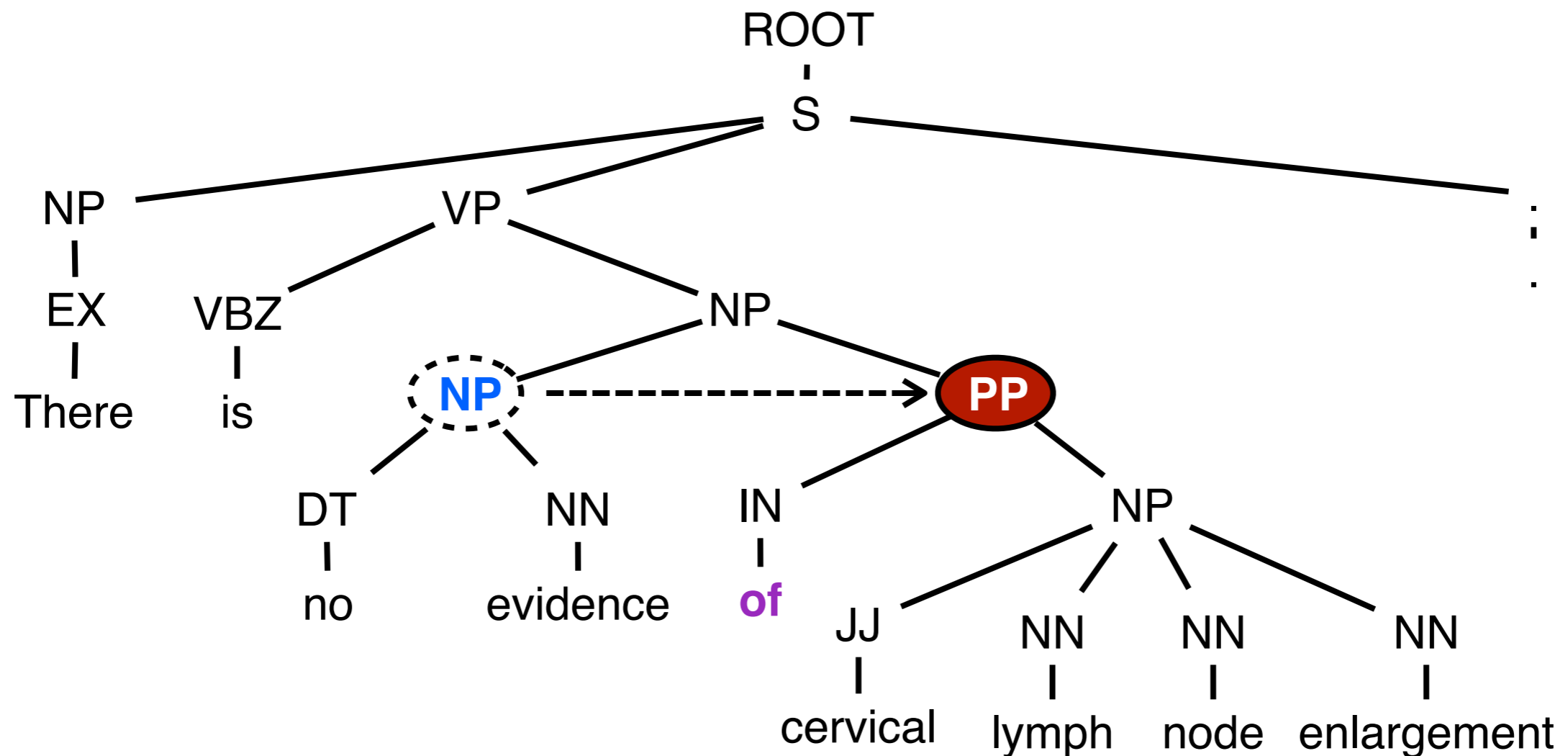
Related Work (Huang, 2007)



Adjective-like negation pattern

1. Locate **NP** with a head from a small set of nouns (e.g. "evidence") and modified by words (e.g. "no", "without", "absent")

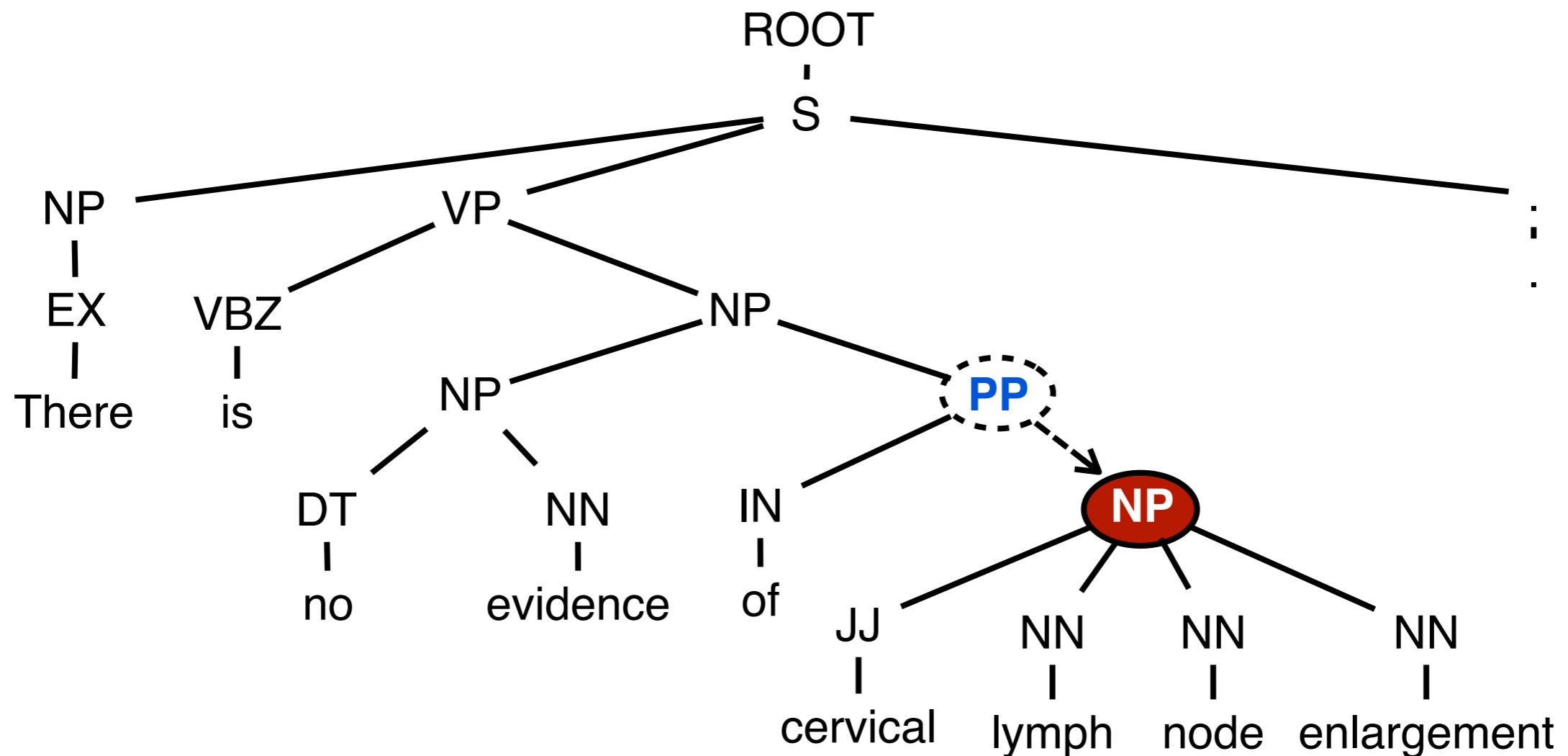
Related Work (Huang, 2007)



Adjective-like negation pattern

2. Locate **PP** by "of" or "for" following the above **NP**

Related Work (Huang, 2007)



Adjective-like negation pattern

3. Extract **NP** under the above **PP**, which contains the negated phrase

Related Work (Morante, 2009)

Design: cascading, 2-step classification procedure

ex) There is no evidence of cervical lymph node enlargement.

Related Work (Morante, 2009)

Design: cascading, 2-step classification procedure

ex) There is no evidence of cervical lymph node enlargement.



Identified by a supervised classifier

There is **no evidence** of cervical lymph node enlargement.

Negation signal

Related Work (Morante, 2009)

Design: cascading, 2-step classification procedure

ex) There is no evidence of cervical lymph node enlargement.

↓ Identified by a supervised classifier

There is **no evidence** of cervical lymph node enlargement.

Negation signal

↓ **Identified by ensemble of classifiers**

There is **no evidence of cervical lymph node enlargement.**

Negation scope

Related Work (Morante, 2009)

- Supervised classification
 - Machine learning task
 - Use labeled training data
 - Classify unlabeled data using feature sets
- Decision tree
 - One of the supervised classification methods (also used by Morante et al.)

Problems

- Huang et al.'s approach has limitation in adaptability
 - Their rules contain domain-specific words (e.g. "evidence")
 - They are extensive but may not be effective in other domains
- Morante et al.'s approach sometimes leads to the grammatically odd solutions

Problems

- Huang et al.'s approach has limitation in adaptability
 - Their rules contain domain-specific words (e.g. "evidence")
 - They are extensive but may not be effective in other domains
- Morante et al.'s approach sometimes leads to the grammatically odd solutions

Problems

- Huang et al.'s approach has limitation in adaptability
 - Their rules contain domain-specific words (e.g. “evidence”)
 - They are extensive but may not be effective in other domains
- Morante et al.'s approach sometimes leads to the grammatically odd solutions

There is **no evidence of cervical lymph node** enlargement;

Ungrammatical negation scope

Proposed approach

Objective: combining supervised classification and parsing

1. Identification of negation signals
2. Identification of negation scopes
3. Adjustment of negation scopes

Proposed approach

Objective: combining supervised classification and parsing

1. Identification of negation signals

2. Identification of negation scopes

3. Adjusting negation

Based on supervised classifiers, IGTre
(similarly to Morante et al.)

Proposed approach

Objective: combining supervised classification and parsing

1. Identification of negation signals

2. Identification of negation scopes

3. Adjusting negation scopes

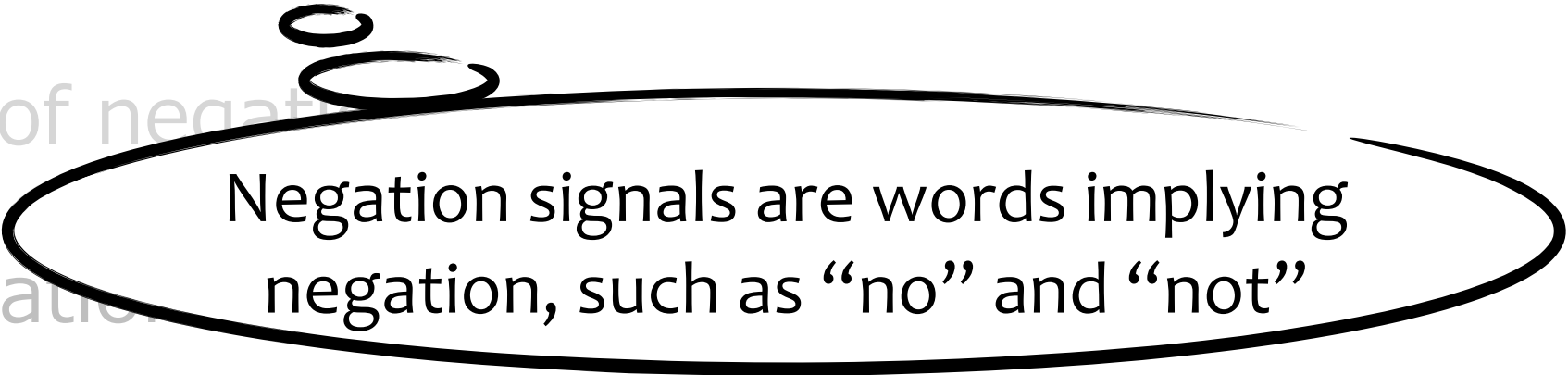
Proposed approach

Objective: combining supervised classification and parsing

1. Identification of **negation signals**

2. Identification of negated expressions


3. Adjusting negated expressions



Negation signals are words implying negation, such as “no” and “not”


Identification of negation signals

- Each token in an input is classified as following classes
 - **Beginning of a negation signal (B)**
 - **Inside of a negation signal (I)**
 - **Outside of a negation signal (O)**

ex) There is  no evidence of cervical lymph node enlargement.
Negation signal

Identification of negation signals

- Each token in an input is classified as following classes
 - **Beginning of a negation signal (B)**
 - Inside of a negation signal (I)
 - Outside of a negation signal (O)

ex) There is  evidence of cervical lymph node enlargement.

Identification of negation signals

- Each token in an input is classified as following classes
 - Beginning of a negation signal (B)
 - **Inside of a negation signal (I)**
 - Outside of a negation signal (O)

ex) There is no  of cervical lymph node enlargement.

Identification of negation signals

- Each token in an input is classified as following classes
 - Beginning of a negation signal (B)
 - Inside of a negation signal (I)
 - Outside of a negation signal (O)

ex) There is no evidence of cervical lymph node enlargement.

Identification of negation signals

- Primary features
 - Raw word, root form, part-of-speech (POS), chunk IOB tags
ex) There is no evidence of cervical lymph node enlargement.

Word	Root form	POS tag	Chunk	Chunk tag	Class
There	there	EX	There	B-NP	O
is	is	VBZ	is	B-VP	O
no	no	DT	no-evidence	B-NP	B
evidence	evidence	NN	no-evidence	I-NP	I
of	of	IN	of	B-PP	O
cervical	cervical	JJ	cervical-lymph-node-enlargement	B-NP	O
lymph	lymph	NN	cervical-lymph-node-enlargement	I-NP	O
node	node	NN	cervical-lymph-node-enlargement	I-NP	O
enlargement	enlargement	NN	cervical-lymph-node-enlargement	I-NP	O
.	.	.	.	O	O

Identification of negation signals

- Primary features
 - Raw word, root form, part-of-speech (POS), chunk IOB tags
ex) There is no evidence of cervical lymph node enlargement.

Word	Root form	POS tag	Chunk	Chunk tag	Class
There	there	EX	There	B-NP	O
is	is	VBZ	is	B-VP	O
no	no	DT	no-evidence	B-NP	B
evidence	evidence	NN	no-evidence	I-NP	I
of	of	IN	of	B-PP	O
cervical	cervical	JJ	cervical-lymph-node-enlargement	B-NP	O
lymph	lymph	NN	cervical-lymph-node-enlargement	I-NP	O
node	node	NN	cervical-lymph-node-enlargement	I-NP	O
enlargement	enlargement	NN	cervical-lymph-node-enlargement	I-NP	O
.	.	.	.	O	O

Identification of negation signals

- Primary features
 - Raw word, root form, part-of-speech (POS), chunk IOB tags
ex) There is no evidence of cervical lymph node enlargement.

Word	Root form	POS tag	Chunk	Chunk tag	Class
There	there	EX	There	B-NP	O
is	is	VBZ	is	B-VP	O
no	no	DT	no-evidence	B-NP	B
evidence	evidence	NN	no-evidence	I-NP	I
of	of	IN	of	B-PP	O
cervical	cervical	JJ	cervical-lymph-node-enlargement	B-NP	O
lymph	lymph	NN	cervical-lymph-node-enlargement	I-NP	O
node	node	NN	cervical-lymph-node-enlargement	I-NP	O
enlargement	enlargement	NN	cervical-lymph-node-enlargement	I-NP	O
.	.	.	.	O	O

Identification of negation signals

- Primary features
 - Raw word, root form, part-of-speech (POS), chunk IOB tags
ex) There is no evidence of cervical lymph node enlargement.

Word	Root form	POS tag	Chunk	Chunk tag	Class
There	there	EX	There	B-NP	O
is	is	VBZ	is	B-VP	O
no	no	DT	no-evidence	B-NP	B
evidence	evidence	NN	no-evidence	I-NP	I
of	of	IN	of	B-PP	O
cervical	cervical	JJ	cervical-lymph-node-enlargement	B-NP	O
lymph	lymph	NN	cervical-lymph-node-enlargement	I-NP	O
node	node	NN	cervical-lymph-node-enlargement	I-NP	O
enlargement	enlargement	NN	cervical-lymph-node-enlargement	I-NP	O
.	.	.	.	O	O


Identification of negation scopes

Objective: combining supervised classification and parsing

1. Identification of negation signals
- 2. Identification of negation scopes**
3. Adjusting negation scopes

Identification of negation scopes

- Each token in an input is classified as following classes
 - **First token of a negation scope (F)**
 - **Last token of a negation scope (L)**
 - **Neither (N)**

ex) There is  no evidence of cervical lymph node enlargement.

Negation scope

Identification of negation scopes

- Each token in an input is classified as following classes
 - **First token of a negation scope (F)**
 - Last token of a negation scope (L)
 - Neither (N)

ex) There is **no** evidence of cervical lymph node enlargement.



Identification of negation scopes

- Each token in an input is classified as following classes
 - First token of a negation scope (F)
 - Last token of a negation scope (L)
 - Neither (N)

ex) There is no evidence of cervical lymph node enlargement

L

Identification of negation scopes

- Each token in an input is classified as following classes
 - First token of a negation scope (F)
 - Last token of a negation scope (L)
 - Neither (N)

ex) There is no evidence of cervical lymph node enlargement

N N N

Identification of negation scopes

- Primary features (follow Morante et al.'s work)
 - Features regarding a detected negation signal
 - Features regarding the token to be classified
 - Features regarding a chunk containing the token to be classified

Identification of negation scopes

- Primary features (follow Morante et al.'s work)
 - **Features regarding a detected negation signal**
 - ▶ Raw word, the relative position of the token in question with respect to the negation signal; Distance to the token in question counted as the number of words; Whether or not the token is a negation signal
 - Features regarding the token to be classified
 - Features regarding a chunk containing the token to be classified

Identification of negation scopes

● P

Word	Nearest neg signal	Position	Distance	Is neg signal?
There	no	PRE	2	FALSE
is	no	PRE	1	FALSE
no	no	SAME	0	TRUE
evidence	no	POST	1	FALSE
of	no	POST	2	FALSE
cervical	no	POST	3	FALSE
lymph	no	POST	4	FALSE
node	no	POST	5	FALSE
enlargement	no	POST	6	FALSE
.	no	POST	7	FALSE

respect to
as the
al
classified

Table: Features regarding a detected negation signal

Identification of negation scopes

● P

Word	Nearest neg signal	Position	Distance	Is neg signal?
There	no	PRE	2	FALSE
is	no	PRE	1	FALSE
no	no	SAME	0	TRUE
evidence	no	POST	1	FALSE
of	no	POST	2	FALSE
cervical	no	POST	3	FALSE
lymph	no	POST	4	FALSE
node	no	POST	5	FALSE
enlargement	no	POST	6	FALSE
.	no	POST	7	FALSE

respect to
as the
al
classified

Table: Features regarding a detected negation signal

Identification of negation scopes

● P

Word	Nearest neg signal	Position	Distance	Is neg signal?
There	no	PRE	2	FALSE
is	no	PRE	1	FALSE
no	no	SAME	0	TRUE
evidence	no	POST	1	FALSE
of	no	POST	2	FALSE
cervical	no	POST	3	FALSE
lymph	no	POST	4	FALSE
node	no	POST	5	FALSE
enlargement	no	POST	6	FALSE
.	no	POST	7	FALSE

respect to
as the
al
classified

Table: Features regarding a detected negation signal

Identification of negation scopes

● P

Word	Nearest neg signal	Position	Distance	Is neg signal?
There	no	PRE	2	FALSE
is	no	PRE	1	FALSE
no	no	SAME	0	TRUE
evidence	no	POST	1	FALSE
of	no	POST	2	FALSE
cervical	no	POST	3	FALSE
lymph	no	POST	4	FALSE
node	no	POST	5	FALSE
enlargement	no	POST	6	FALSE
.	no	POST	7	FALSE

respect to
as the
al
classified

Table: Features regarding a detected negation signal

Identification of negation scopes

- Primary features (follow Morante et al.'s work)
 - Features regarding a detected negation signal
 - **Features regarding the token to be classified**
 - ▶ Raw word and root form, POS, and chunk IOB tag; Root form, POS, and chunk IOB tag of one token to the left and to the right; Root form of the second token to the left and to the right
 - Features regarding a chunk containing the token to be classified

Identification of negation scopes

Word	POS	POS of 1 token to the left	POS of 2 tokens to left	POS of 1 token to the right	POS of 2 tokens to right
There	EX	-	-	VBZ	DT
is	VBZ	EX	-	DT	NN
no	DT	VBZ	EX	NN	IN
evidence	NN	DT	VBZ	IN	JJ
of	IN	NN	DT	JJ	NN
cervical	JJ	IN	NN	NN	NN
lymph	NN	JJ	IN	NN	NN
node	NN	NN	JJ	NN	.
enlargement	NN	NN	NN	.	-
.	.	NN	NN	-	-

Table: Features regarding the token to be classified

Identification of negation scopes

Word	POS	POS of 1 token to the left	POS of 2 tokens to left	POS of 1 token to the right	POS of 2 tokens to right
There	EX	-	-	VBZ	DT
is	VBZ	EX	-	DT	NN
no	DT	VBZ	EX	NN	IN
evidence	NN	DT	VBZ	IN	JJ
of	IN	NN	DT	JJ	NN
cervical	JJ	IN	NN	NN	NN
lymph	NN	JJ	IN	NN	NN
node	NN	NN	JJ	NN	.
enlargement	NN	NN	NN	.	-
.	.	NN	NN	-	-

Table: Features regarding the token to be classified

Identification of negation scopes

Word	POS	POS of 1 token to the left	POS of 2 tokens to left	POS of 1 token to the right	POS of 2 tokens to right
There	EX	-	-	VBZ	DT
is	VBZ	EX	-	DT	NN
no	DT	VBZ	EX	NN	IN
evidence	NN	DT	VBZ	IN	JJ
of	IN	NN	DT	JJ	NN
cervical	JJ	IN	NN	NN	NN
lymph	NN	JJ	IN	NN	NN
node	NN	NN	JJ	NN	.
enlargement	NN	NN	NN	.	-
.	.	NN	NN	-	-

Table: Features regarding the token to be classified

Identification of negation scopes

Word	POS	POS of 1 token to the left	POS of 2 tokens to left	POS of 1 token to the right	POS of 2 tokens to right
There	EX	-	-	VBZ	DT
is	VBZ	EX	-	DT	NN
no	DT	VBZ	EX	NN	IN
evidence	NN	DT	VBZ	IN	JJ
of	IN	NN	DT	JJ	NN
cervical	JJ	IN	NN	NN	NN
lymph	NN	JJ	IN	NN	NN
node	NN	NN	JJ	NN	.
enlargement	NN	NN	NN	.	-
.	.	NN	NN	-	-

Table: Features regarding the token to be classified

Identification of negation scopes

Word	POS	POS of 1 token to the left	POS of 2 tokens to left	POS of 1 token to the right	POS of 2 tokens to right
There	EX	-	-	VBZ	DT
is	VBZ	EX	-	DT	NN
no	DT	VBZ	EX	NN	IN
evidence	NN	DT	VBZ	IN	JJ
of	IN	NN	DT	JJ	NN
cervical	JJ	IN	NN	NN	NN
lymph	NN	JJ	IN	NN	NN
node	NN	NN	JJ	NN	.
enlargement	NN	NN	NN	.	-
.	.	NN	NN	-	-

Table: Features regarding the token to be classified

Identification of negation scopes

- Primary features (follow Morante et al.'s work)
 - Features regarding a detected negation signal
 - Features regarding the token to be classified
 - **Features regarding a chunk containing the token to be classified**
 - ▶ The first and last token in the chunk; Sequence of the tokens in the chunk; Sequence of the POS tags in the chunk; The first and last token, hyphenated all tokens, and hyphenated all POS tags of two chunks to the left and two chunks to the right

Word	Chunk	First token in the chunk	Last token in the chunk	Chunk of 1 chunk to the left
There	There	There	There	-
is	is	is	is	There
no	no-evidence	no	evidence	is
evidence	no-evidence	no	evidence	is
of	of	of	of	no-evidence
cervical	cervical-lymph-node-enlargement	cervical	enlargement	of
lymph	cervical-lymph-node-enlargement	cervical	enlargement	of
node	cervical-lymph-node-enlargement	cervical	enlargement	of
enlargement	cervical-lymph-node-enlargement	cervical	enlargement	of
.	.	.	.	cervical-lymph-node-enlargement

Table: Features regarding a chunk containing the token to be classified

Word	Chunk	First token in the chunk	Last token in the chunk	Chunk of 1 chunk to the left
There	There	There	There	-
is	is	is	is	There
no	no-evidence	no	evidence	is
evidence	no-evidence	no	evidence	is
of	of	of	of	no-evidence
cervical	cervical-lymph-node-enlargement	cervical	enlargement	of
lymph	cervical-lymph-node-enlargement	cervical	enlargement	of
node	cervical-lymph-node-enlargement	cervical	enlargement	of
enlargement	cervical-lymph-node-enlargement	cervical	enlargement	of
.	.	.	.	cervical-lymph-node-enlargement

Table: Features regarding a chunk containing the token to be classified

Word	Chunk	First token in the chunk	Last token in the chunk	Chunk of 1 chunk to the left
There	There	There	There	-
is	is	is	is	There
no	no-evidence	no	evidence	is
evidence	no-evidence	no	evidence	is
of	of	of	of	no-evidence
cervical	cervical-lymph-node-enlargement	cervical	enlargement	of
lymph	cervical-lymph-node-enlargement	cervical	enlargement	of
node	cervical-lymph-node-enlargement	cervical	enlargement	of
enlargement	cervical-lymph-node-enlargement	cervical	enlargement	of
.	.	.	.	cervical-lymph-node-enlargement

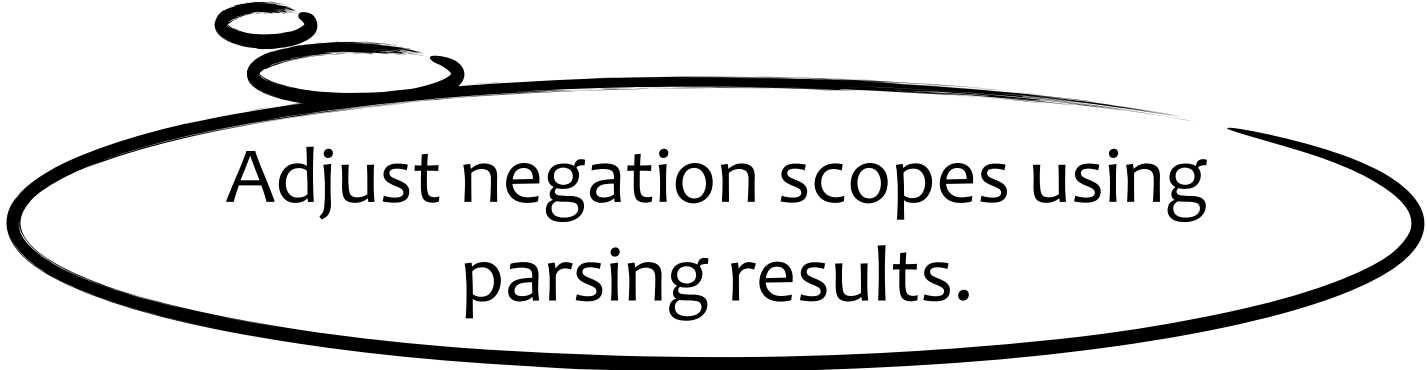
Table: Features regarding a chunk containing the token to be classified

Proposed approach

Objective: combining supervised classification and parsing

1. Identification of negation signals
2. Identification of negation scopes

3. Adjustment of negation scopes



Adjust negation scopes using parsing results.

Proposed approach

Objective: combining supervised classification and parsing

1. Identification of negation signals
2. Identification of negation scopes

3. Adjustment of negation scopes

<Preliminary experiment>

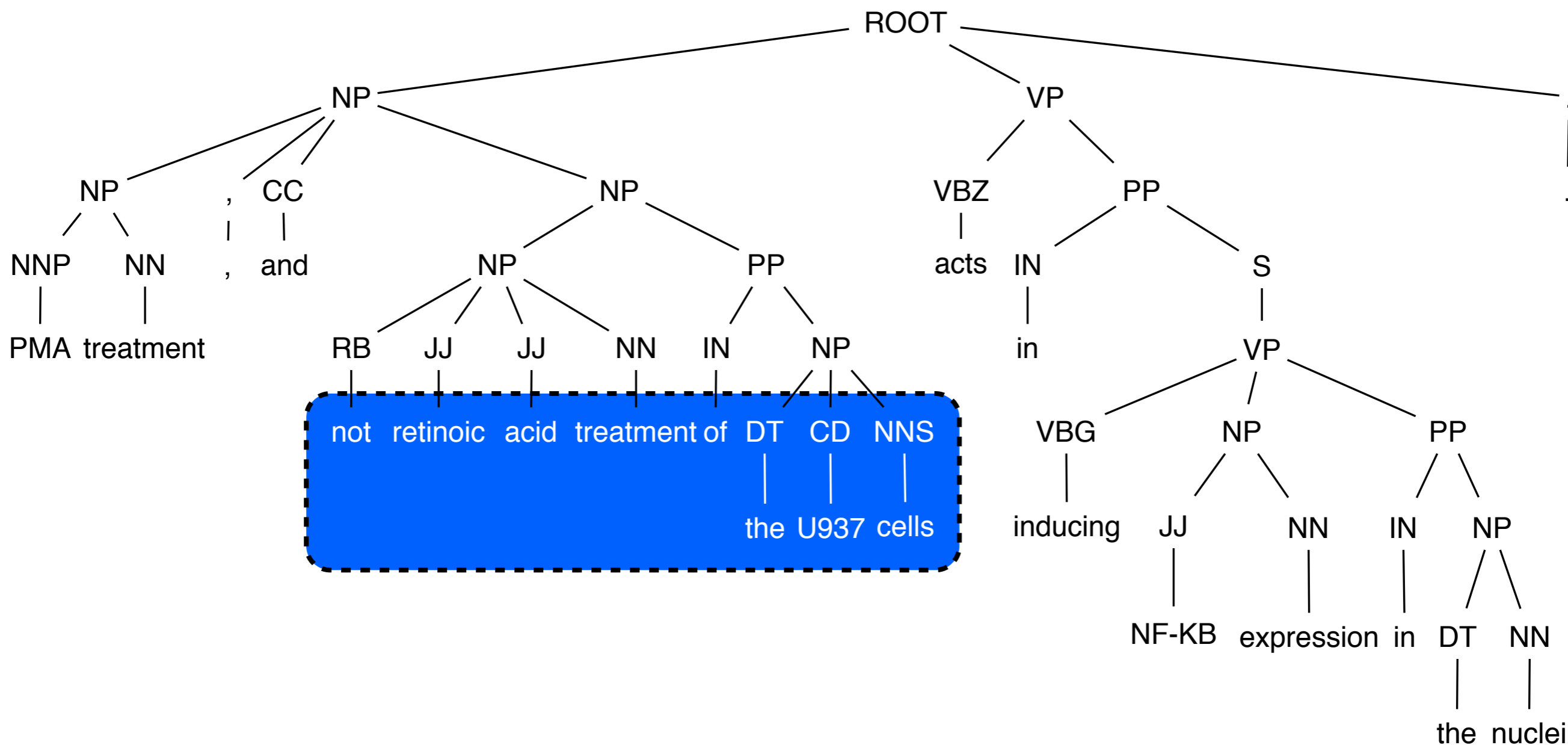
ex) There is **no evidence of cervical lymph node** enlargement.

→ **Worse at the end than at the beginning**

Adjusting negation scopes

correct scope

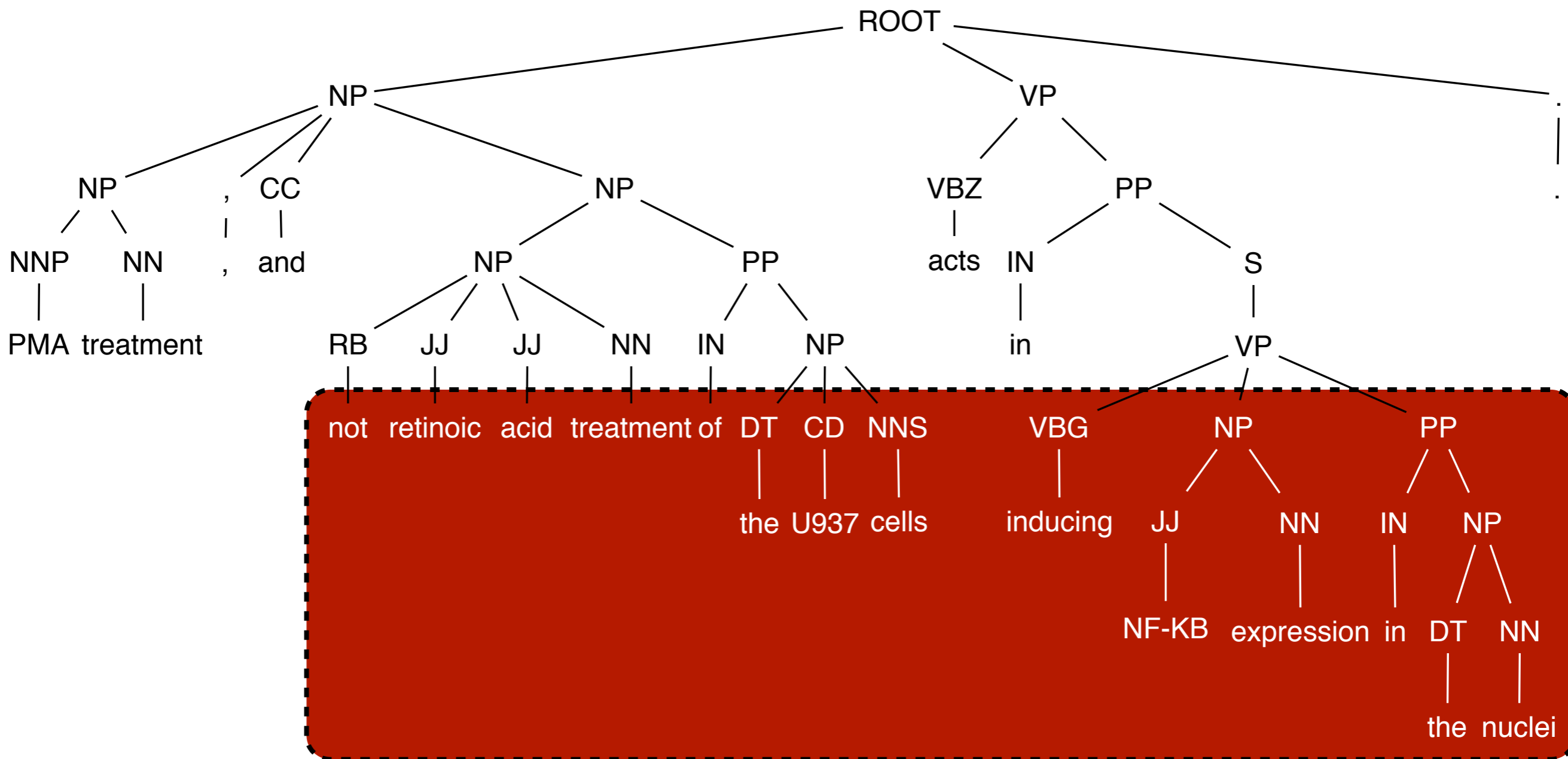
PMA treatment, and **not retinoic acid treatment of the U937 cells** acts in inducing NF-KB expression in the nuclei.



Adjusting negation scopes

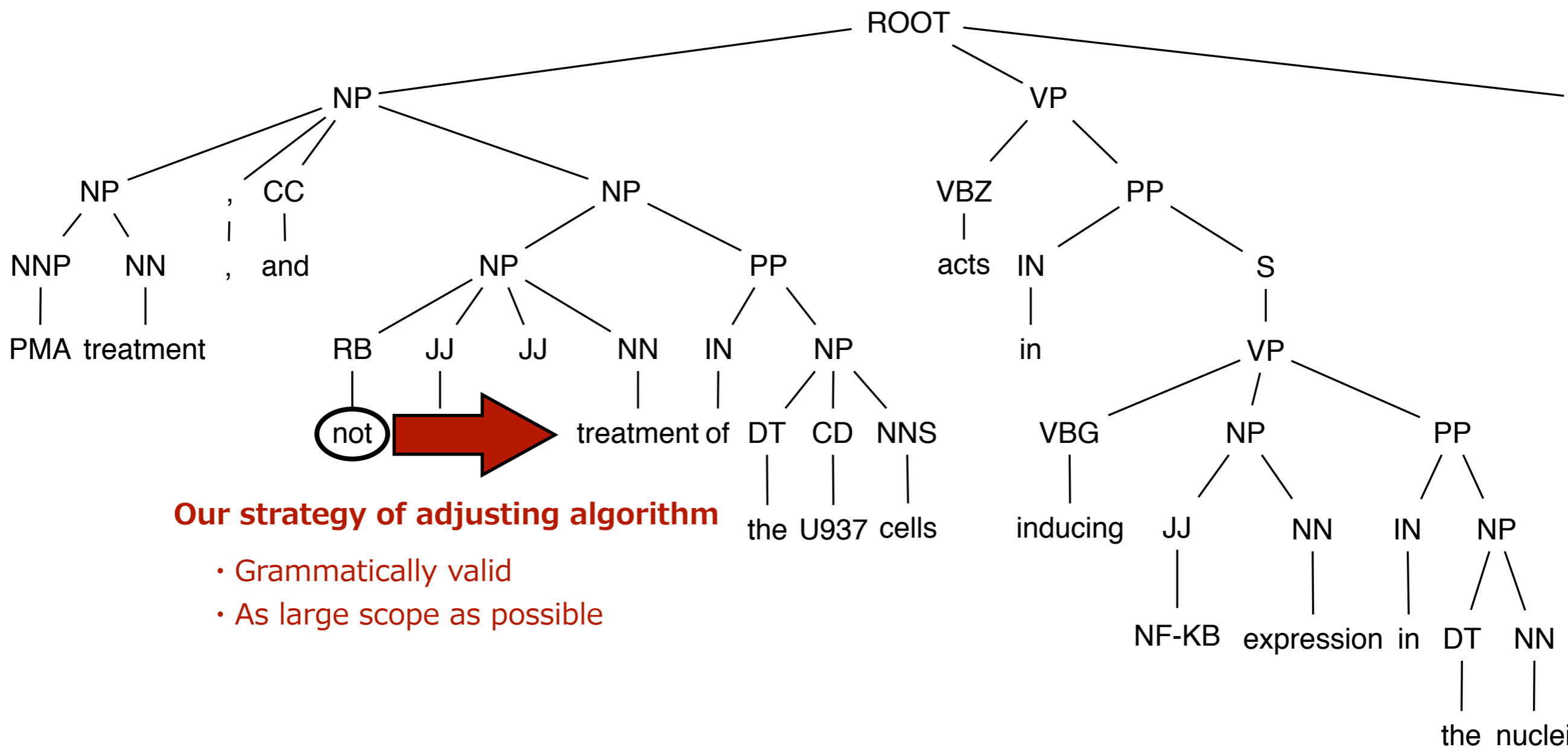
wrong scope

PMA treatment, and **not retinoic acid treatment of the U937 cells acts in inducing NF-KB expression in the nuclei.**



Adjusting negation scopes

PMA treatment, and not retinoic acid treatment of the U937 cells acts in inducing NF-KB expression in the nuclei.

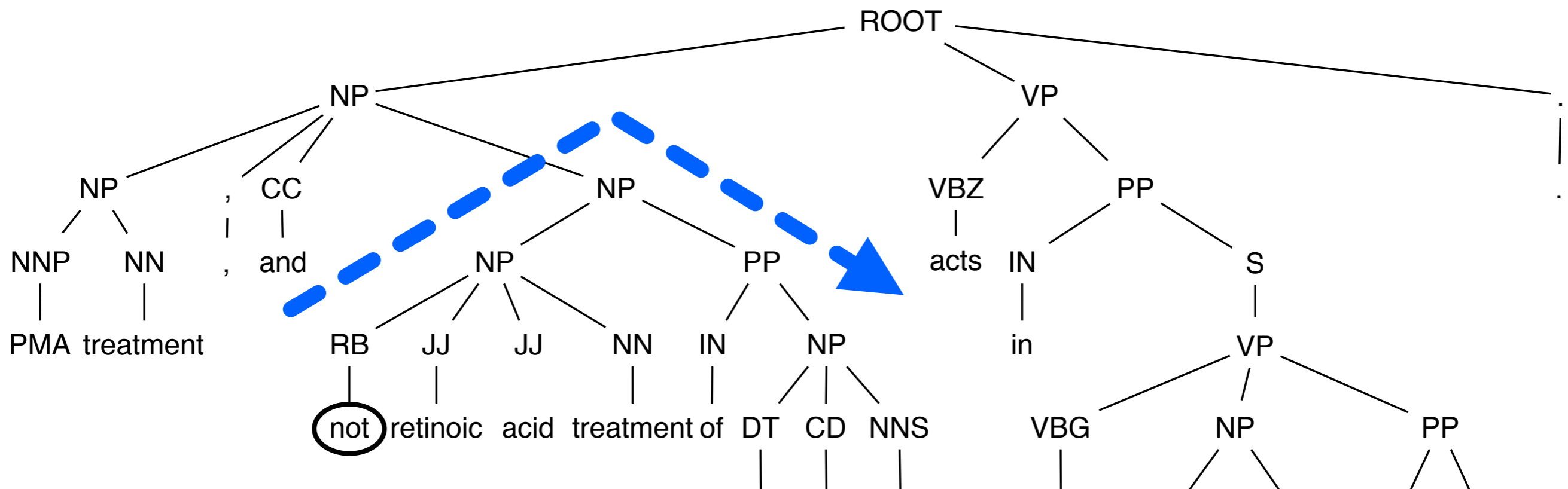


Our strategy of adjusting algorithm

- Grammatically valid
- As large scope as possible

Adjusting negation scopes

PMA treatment, and not retinoic acid treatment of the U937 cells acts in inducing NF-KB expression in the nuclei.

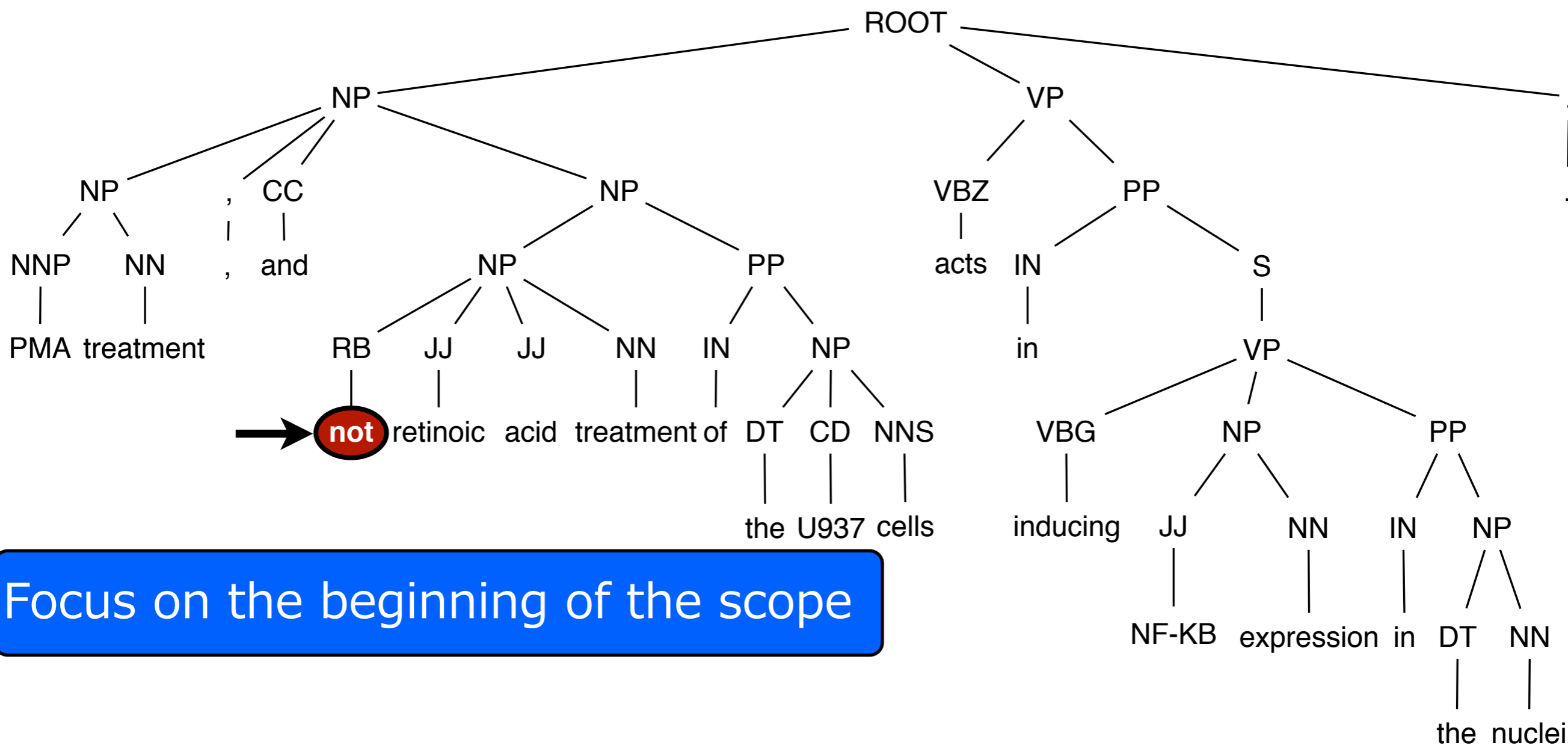


The right-most boundary is the last descendant node which contains the beginning as the first descendant node

the nuclei

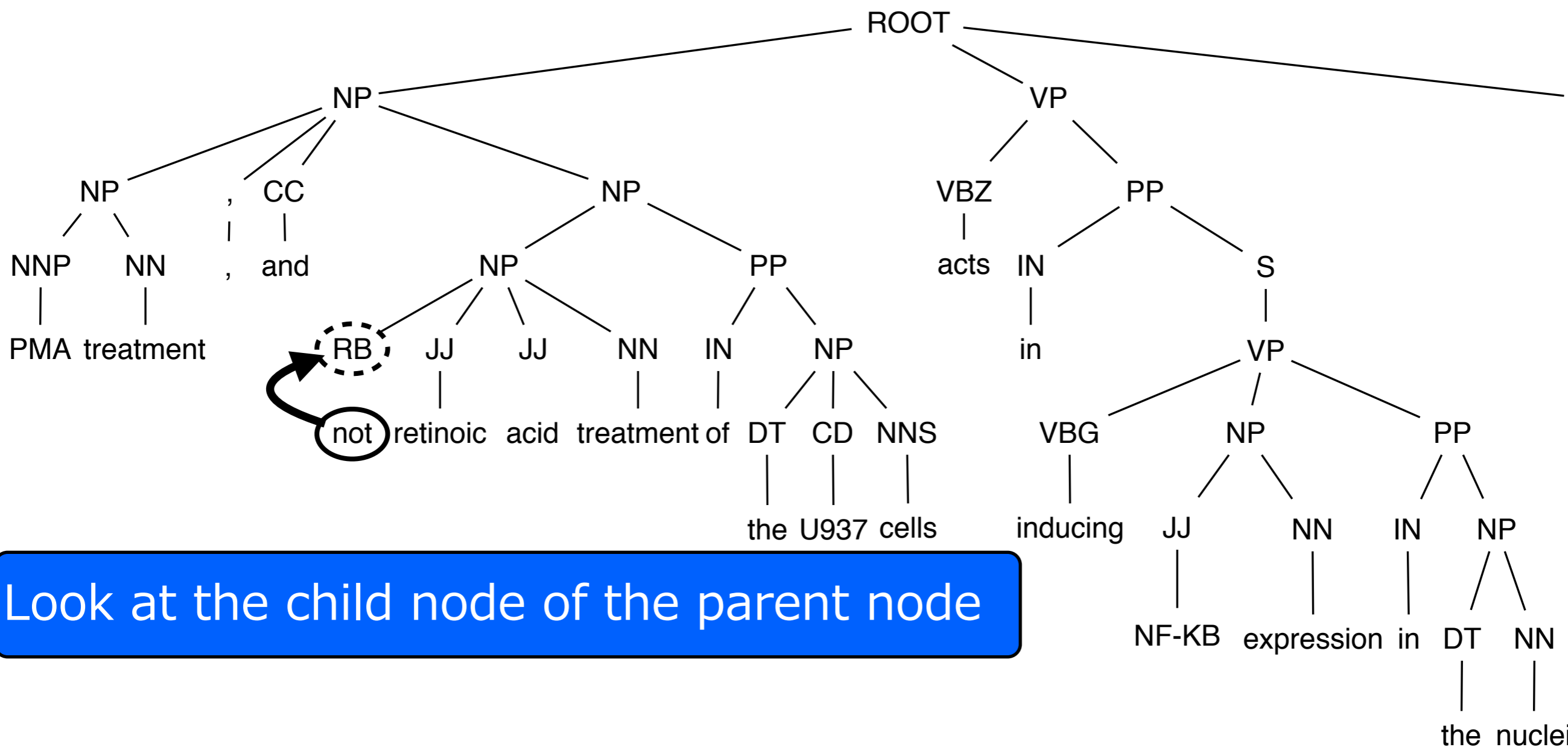
Adjusting negation scopes

PMA treatment, and not retinoic acid treatment of the U937 cells acts in inducing NF-KB expression in the nuclei.



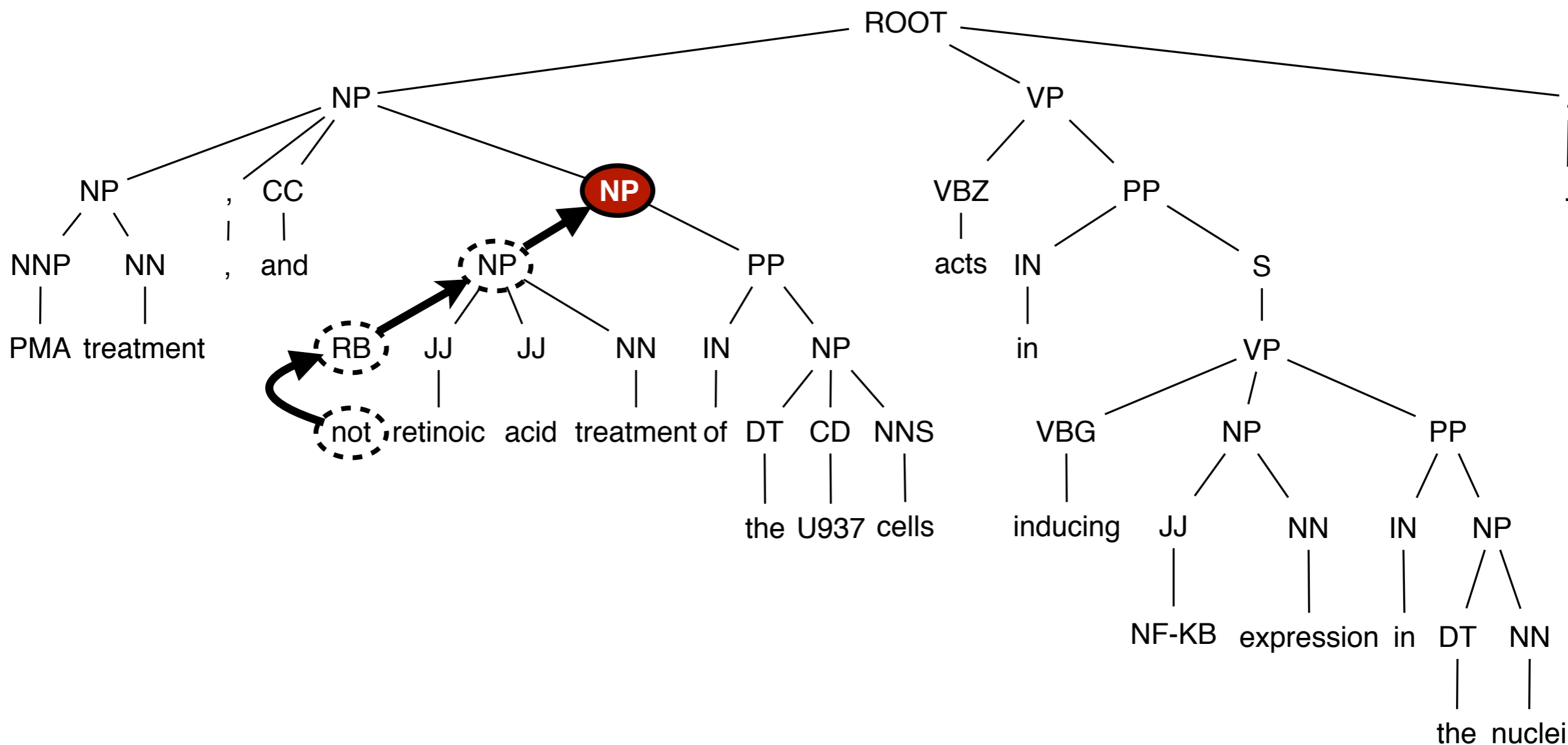
Adjusting negation scopes

PMA treatment, and not retinoic acid treatment of the U937 cells acts in inducing NF-KB expression in the nuclei.



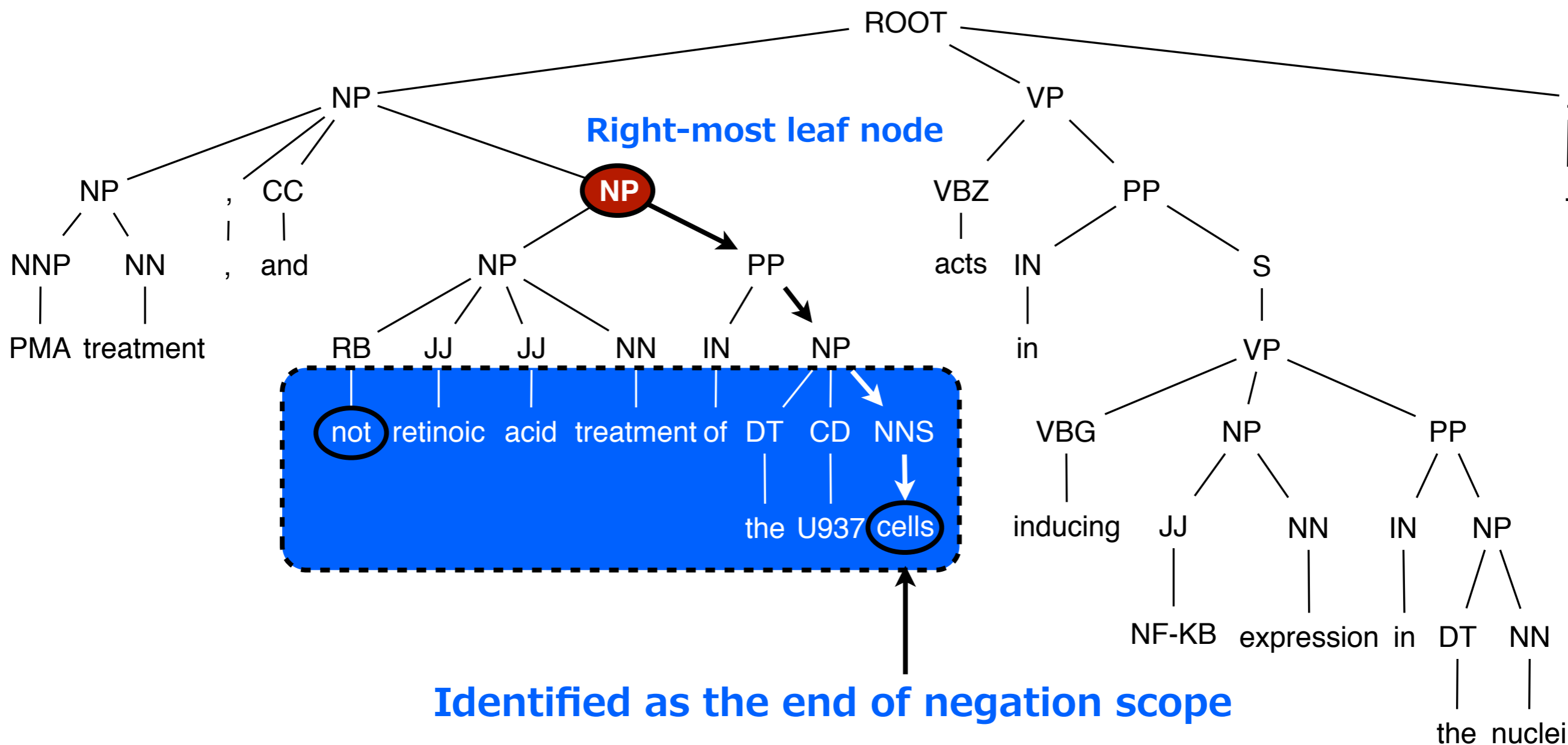
Adjusting negation scopes

PMA treatment, and not retinoic acid treatment of the U937 cells acts in inducing NF-KB expression in the nuclei.



Adjusting negation scopes

PMA treatment, and **not retinoic acid treatment of the U937 cells** acts in inducing NF-KB expression in the nuclei.



Experimental settings

- Data
 - BioScope Corpus
 - ▶ Each sentence is annotated with information about **negation and uncertain expressions**.
 - ▶ Consist of three data sets: **biological paper abstracts, biological full-text, and clinical free text**
 - ▶ Annotation was done by two linguists following the guidelines written by a expert.

```
<sentence id="S43.1">
  Stable exam
  <xcope id="X43.1.1">
    <cue type="negation" ref="X43.1.1">without
    </cue>
    radiographic evidence of acute
    cardiopulmonary disease
  </xcope>.
</sentence>
```

Experimental settings

- Descriptive statistics of the three data sets

	Abstracts	Full text	Clinical
# of documents	1273	9	1954
# of sentences	11871	2670	6383
# of words	282243	60935	41985
Avg. sentence length	26.4	26.2	7.7
# of negation signals	1848	389	877
IAA	90.7	79.4	91.5

Experimental settings

- Descriptive statistics of the three data sets

	Abstracts	Full text	Clinical
# of documents	1273	9	1954
# of sentences	11871	2670	6383
# of words	282243	60935	41985
Avg. sentence length	26.4	26.2	7.7
# of negation signals	1848	389	877
IAA	90.7	79.4	91.5

Experimental settings

- Descriptive statistics of the three data sets

	Abstracts	Full text	Clinical
# of documents	1273	9	1954
# of sentences	11871	2670	6383
# of words	282243	60935	41985
A	It is difficult even for experts to correctly identify negation signal and scopes		7.7
# of negation signals		89	877
IAA	90.7	79.4	91.5

Experimental settings

- Evaluation

$$\text{Precision} = \frac{\# \text{ correctly classified tokens}}{\# \text{ tokens classified as } \textit{positive}}$$

$$\text{Recall} = \frac{\# \text{ correctly classified tokens}}{\# \text{ tokens annotated as } \textit{positive}}$$

- Baseline

- Supervised learning approach
- Roughly correspond to the previous work using supervised learning approach

Experimental settings

- Evaluation

$$\text{Precision} = \frac{\# \text{ correctly classified tokens}}{\# \text{ tokens classified as } \textit{positive}}$$

$$\text{Recall} = \frac{\# \text{ correctly classified tokens}}{\# \text{ tokens annotated as } \textit{positive}}$$

- Baseline

- Supervised learning approach
- Roughly correspond to Morante et al.'s approach

Results

- Identification of negation scopes

Approach	Corpus	Precision	Recall	F-score
Baseline	Abstracts	79.5	71.5	75.3
	Full text	69.9	47.1	56.2
	Clinical	90.4	85.0	87.6
	Overall	79.9	67.9	73.0
Proposed	Abstracts	83.0	71.4	76.8
	Full text	73.0	54.5	62.4
	Clinical	89.9	82.3	85.9
	Overall	82.0	69.4	75.0

Results

- Identification of negation scopes

Approach	Cornus	Precision	Recall	F-score
Baseline				75.3
				56.2
	Clinical	90.4	85.0	87.6
	Overall	79.9	67.9	73.0
Proposed	Abstracts	83.0	71.4	76.8
	Full text	73.0	54.5	62.4
	Clinical	89.9	82.3	85.9
	Overall	82.0	69.4	75.0

The improvement was statistically significant at the $p = 0.05$ level

Results

- Identification of negation scope improved for both abstracts and full text

Approach	Corpus	Precision	Recall	F-score
Baseline	Abstracts	79.5	71.5	75.3
	Full text	69.9	47.1	56.2
	Clinical	90.4	85.0	87.6
	Overall	79.9	67.9	73.0
Proposed	Abstracts	83.0	71.4	76.8
	Full text	73.0	54.5	62.4
	Clinical	89.9	82.3	85.9
	Overall	82.0	69.4	75.0

Results

- Identification of negation scopes

The decrease of the performance in the clinical records requires more in-depth investigation

				F-score
Baseline	Full text	69.9	47.1	56.2
	Clinical	90.4	85.0	87.6
	Overall	79.9	67.9	73.0
	Abstracts	83.0	71.4	76.8
Proposed	Full text	73.0	54.5	62.4
	Clinical	89.9	82.3	85.9
	Overall	82.0	69.4	75.0

Error Analysis

- Clinical records are not written to follow strict grammatical rules ?

	All sentences	Sentences with verbs	Rate
Abstracts	11988	11369	94.8%
Full texts	2719	2456	90.3%
Clinical	6387	2261	35.4%

Table: Proportion of sentences with verbs

Error Analysis

- Clinical records are not written to follow strict grammatical rules ?

	All sentences	Sentences with verbs	Rate
Abstracts	11988	11369	94.8%
Full texts	2719	2456	90.3%
Clinical	6387	2261	35.4%

This result may have caused the decrease of the performance in the clinical domain

Error Analysis

- Scope finding accuracy for most frequent negation signals

Signal	Abstracts	Full text	Clinical
no	83.4	71.2	98.6
without	80.2	71.4	89.8
or	55.6	52.2	71.4
not	70.6	49.0	61.7
likely	54.3	28.8	76.5

Error Analysis

- Scope finding accuracy for most frequent negation signals

Signal	Abstracts	Full text	Clinical
no	83.4	71.2	98.6
without	80.2	71.4	89.8
or	55.6	52.2	71.4
not	70.6	49.0	61.7
likely	54.3	28.8	76.5

Error Analysis

- Scope finding accuracy for most frequent negation signals

Signal	Abstracts	Full text	Clinical
no	83.4	71.2	98.6
without	80.2	71.4	89.8
or	55.6	52.2	71.4
not	70.6	49.0	61.7
likely	54.3	28.8	76.5

Comparison with existing approach

- Heuristic approach (Huang, 2007)
 - Implement identification of negated sentences using regular expressions

Corpus	Accuracy
Abstracts	29.7%
Full texts	40.0%
Clinical	95.4%

Comparison with existing approach

- Heuristic approach (Huang, 2007)
 - Implement identification of negated sentences using regular expressions

Corpus	Accuracy
Abstracts	29.7%
Full texts	40.0%
Clinical	95.4%

Comparison with existing approach

- Heuristic approach (Huang, 2007)
 - Implement identification of negated sentences using regular expressions

Corpus	Accuracy
Abstracts	29.7%
Full texts	40.0%
Clinical	95.4%

Comparison with existing approach

- Heuristic approach (Huang, 2007)

This result shows this system is not effective in documents except for clinical records

Corpus	Accuracy
Abstracts	29.7%
Full texts	40.0%
Clinical	95.4%

Conclusions

- Developed a hybrid approach to identifying the scope of negated and uncertain expressions
 - Combining statistical and heuristic approaches
 - Took advantage of the syntactic structure of an input sentence and **adjusted the right-most boundary of negation scope**
- Findings
 - The grammatical rule was shown effective to improve the overall performance
 - Our approach was not effective to improve the performance of the documents which were not written to follow strict grammatical rules

Conclusions

- Developed a hybrid approach to identifying the scope of negated and uncertain expressions
 - Combining statistical and heuristic approaches
 - Took advantage of syntactic structure of an input sentence and adjusted the right-most boundary of negation scope
- Findings
 - The grammatical rule was shown effective to improve the overall performance
 - Our approach was not effective to improve the performance for Clinical records

Future work

- To further improve the performance of negation identification
 - Refine our heuristics for negation signals which led to low accuracy in scope identification
 - Deal with ungrammatical sentences
- To examine the impact of our approach on real-world application (e.g. information retrieval)

Any Questions?