

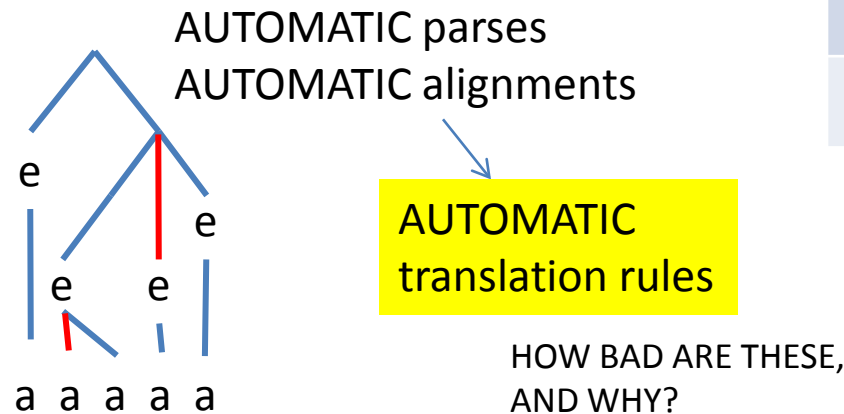
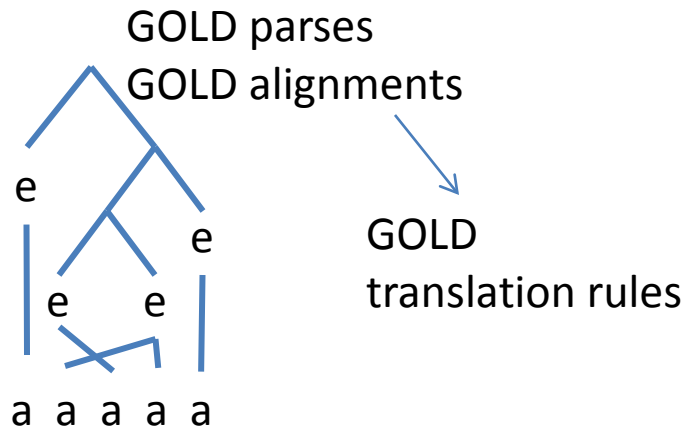
Manually Parsed, Manually Aligned Bilingual Data for MT

Kevin Knight, USC/ISI
Member of BAC

Parsed, Aligned, Bilingual Data

- Understanding weak points of automatic parsers and aligners
- Building better parsers and aligners
- Designing radically new MT models

Understanding Weak Points



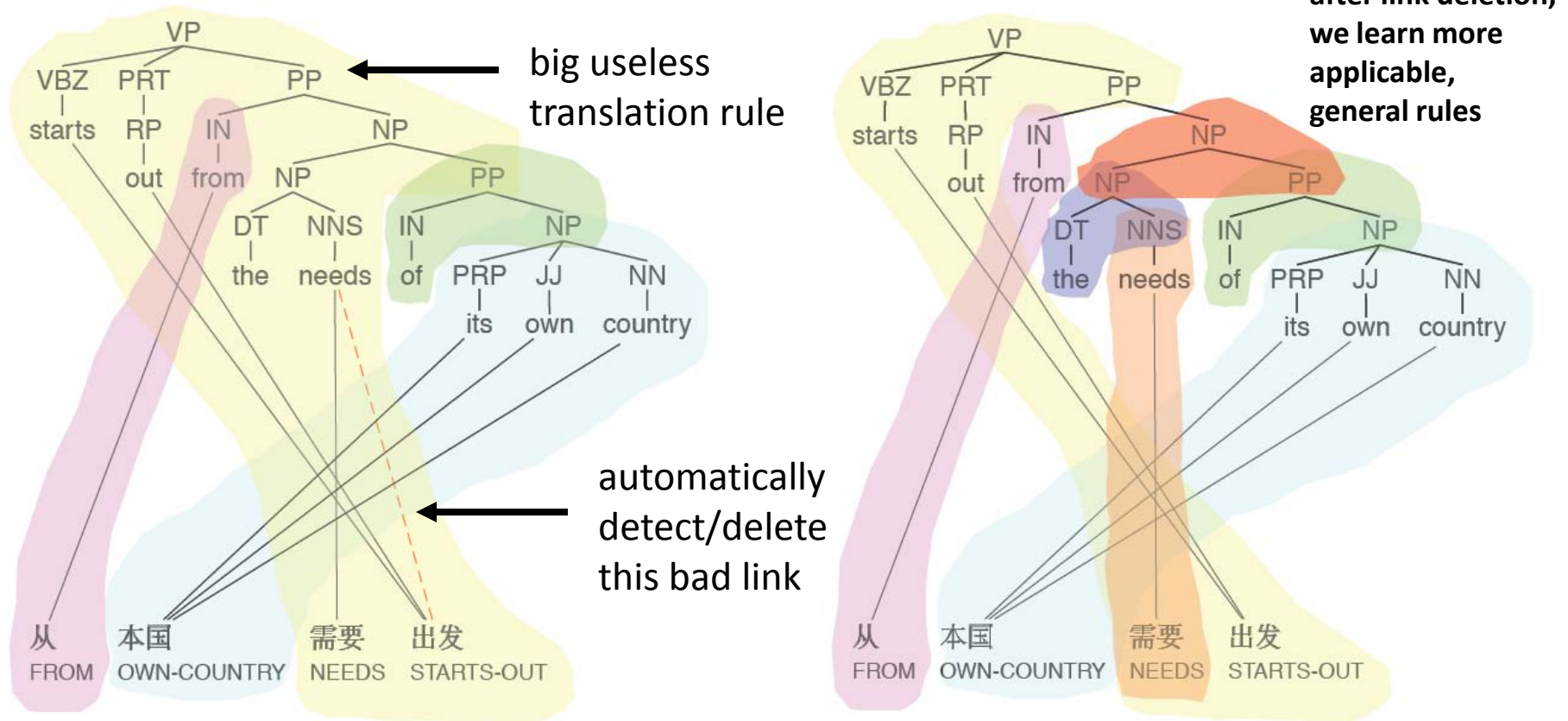
Chinese/English

Alignments	Parses	Rule Quality (f-score vs. GOLD rules)
GOLD	GOLD	100
GOLD	Automatic (Collins)	75
Automatic (GIZA union)	GOLD	57
Automatic	Automatic	47

Building Better Aligners and Parsers

Problem: GIZA only aligns string pairs.

Idea: Use parse trees to improve alignments.



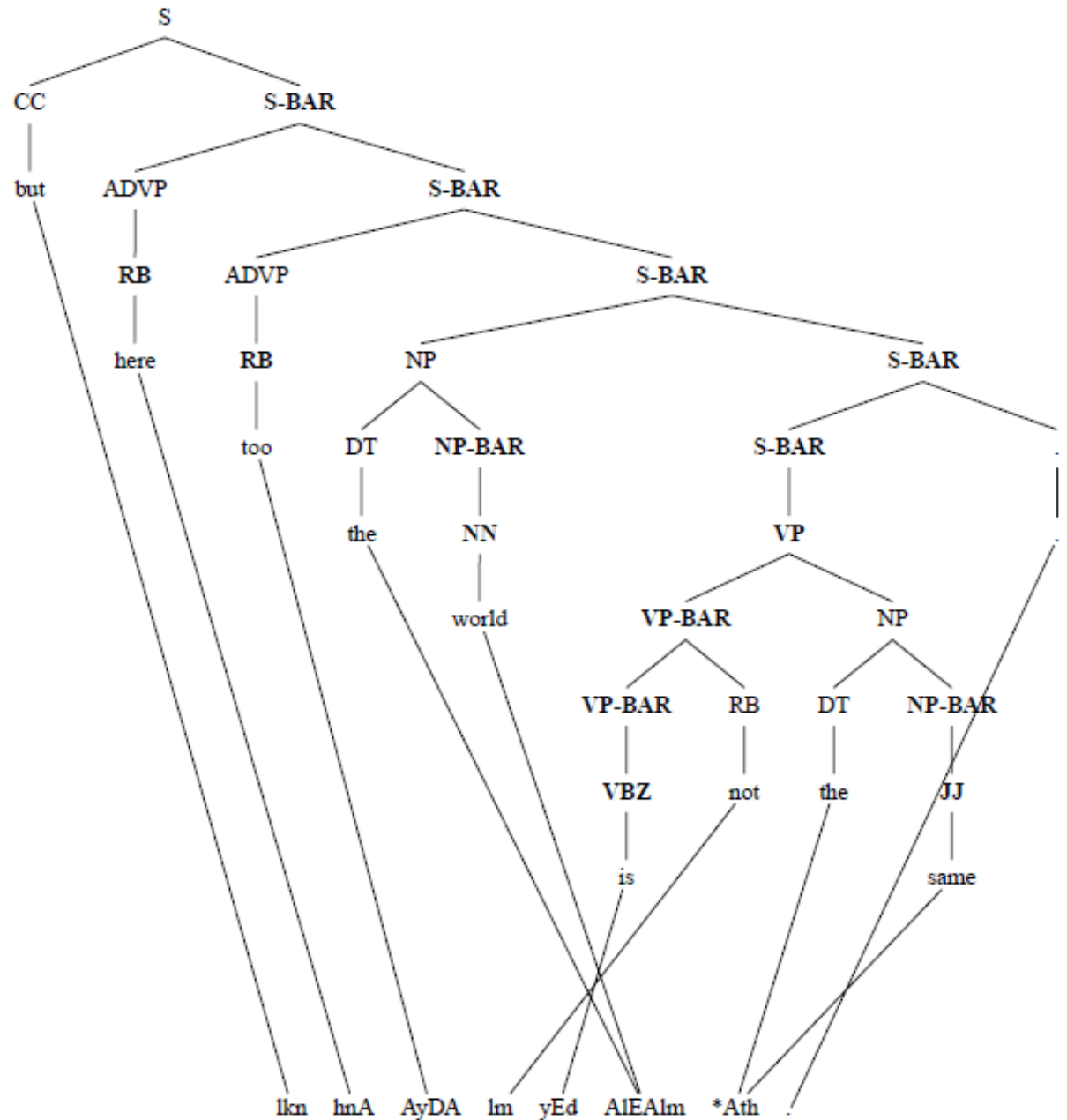
Impact: +0.7 Bleu on Chinese/English

Designing radically new MT models

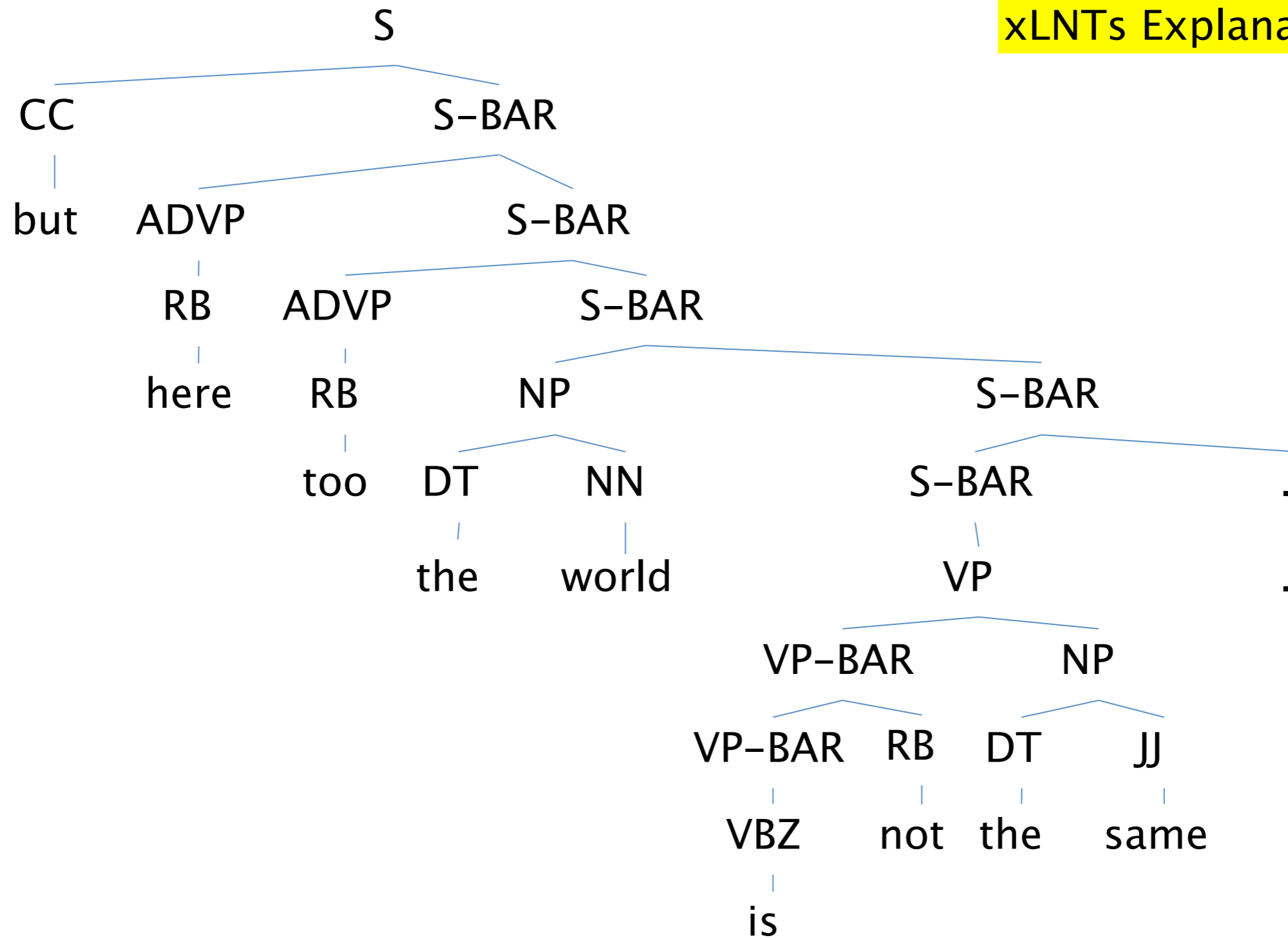
- There are **lots** of ways to envision syntactic/semantic translation patterns
 - Mathematical considerations
 - Linguistic considerations
- Which models best fit the data?
- Best way:
 - See how new models physically work on clean data
 - Clean = manually parsed, manually aligned

Look at many examples.

Ask:
“What is going on here?”

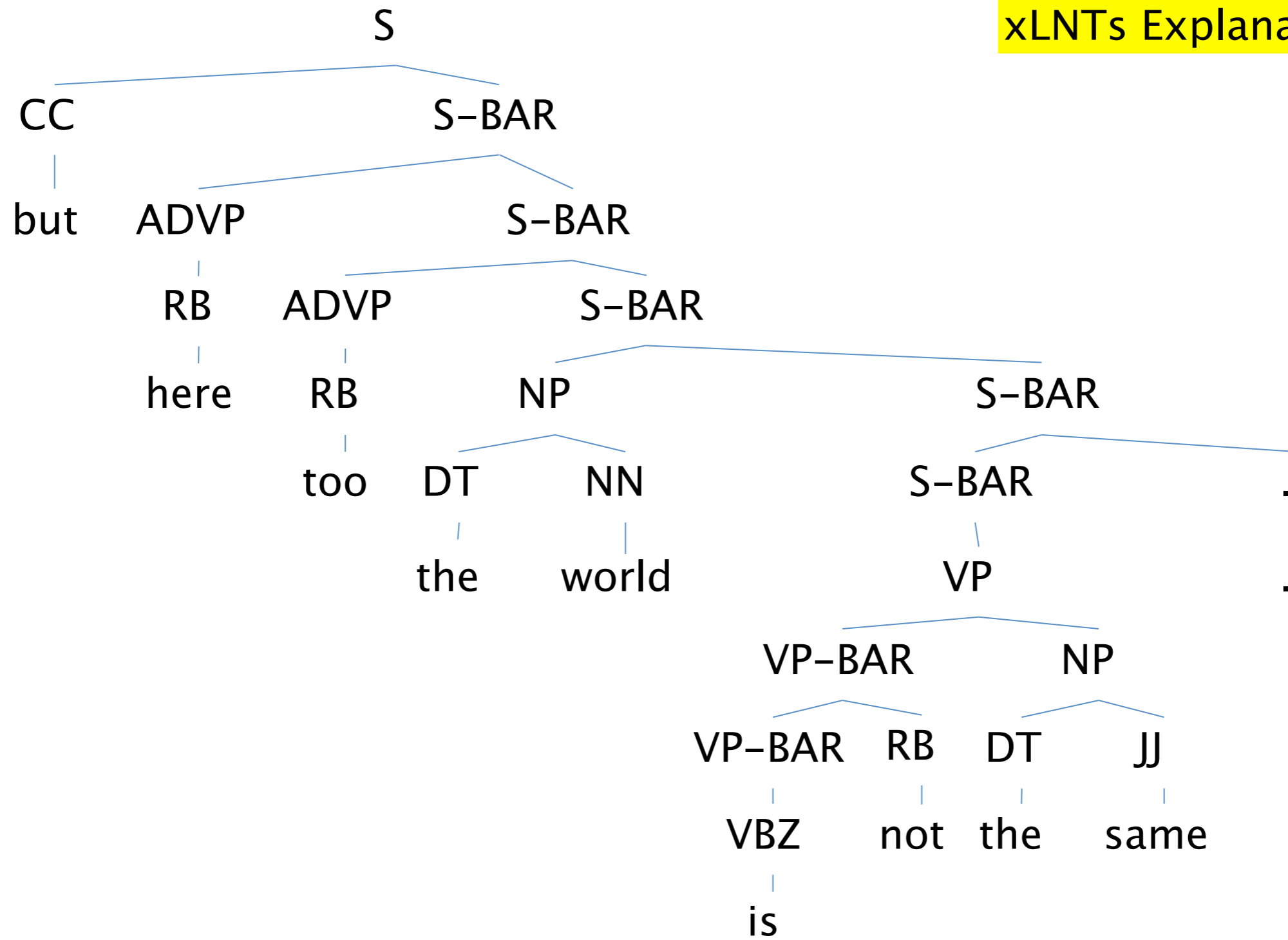


xLNTs Explanation



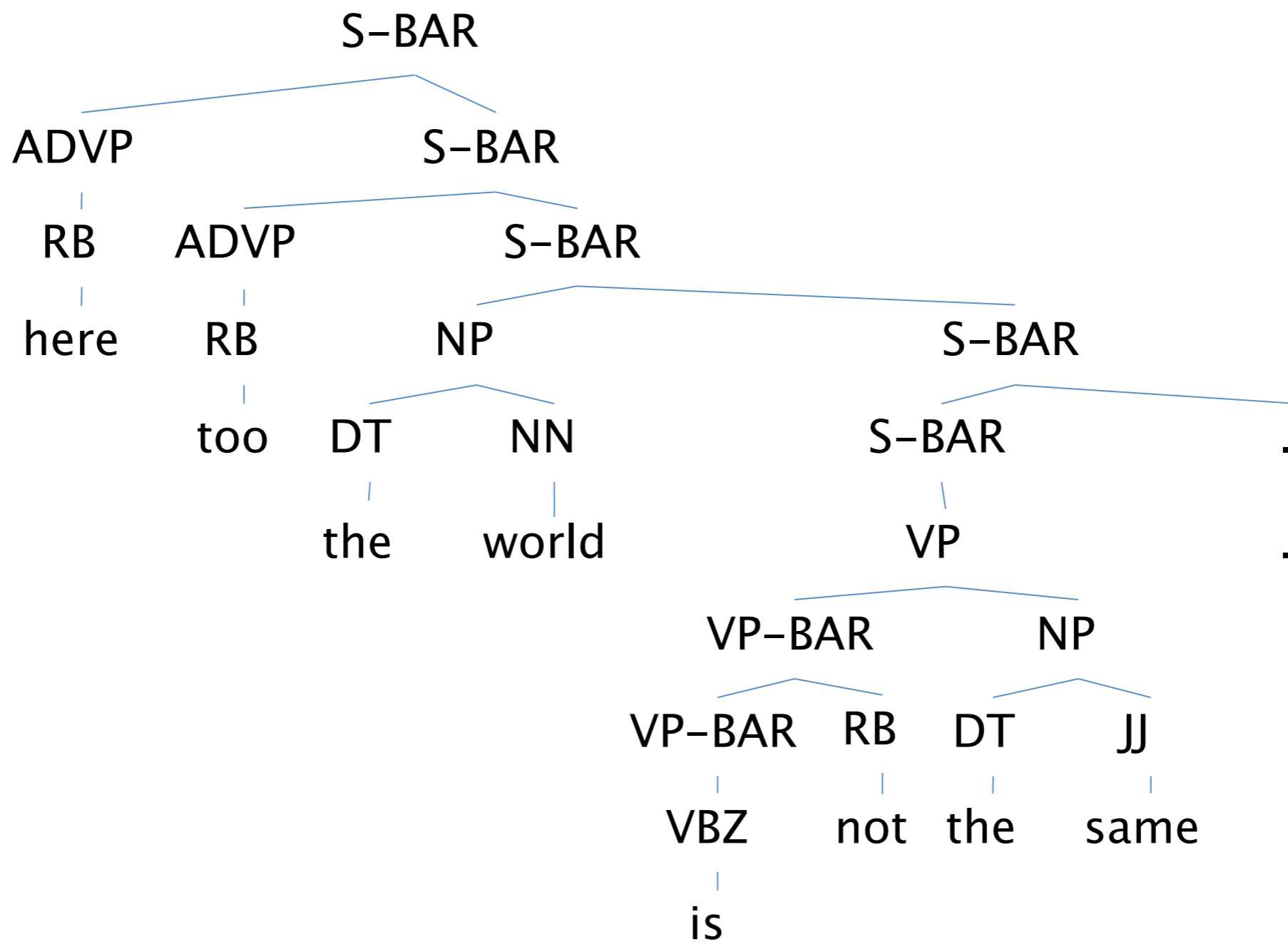
lkn hnA AyDA Im yEd AIEAlm *Ath .

xLNTs Explanation



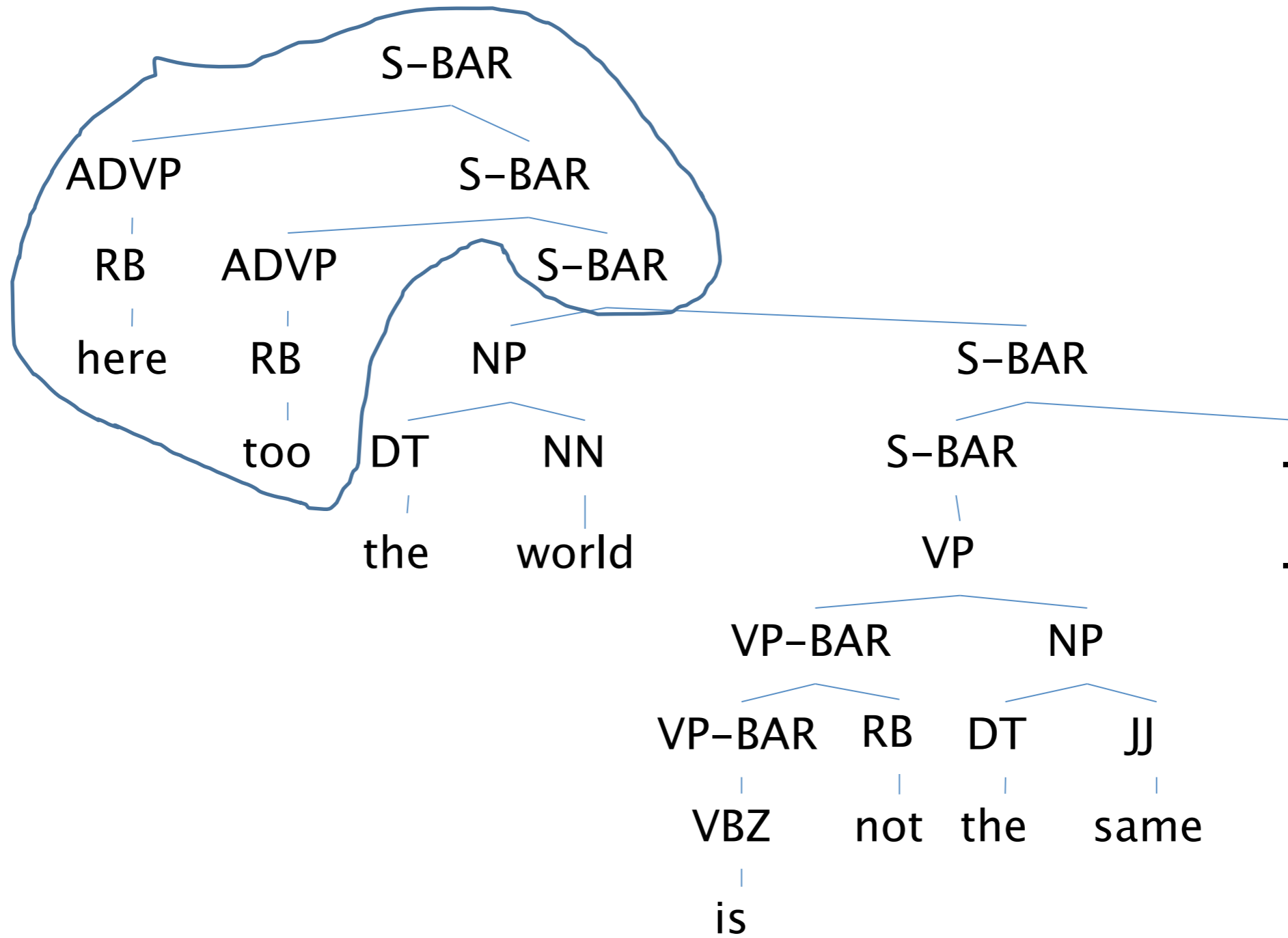
xLNTs Explanation

Ikn ,



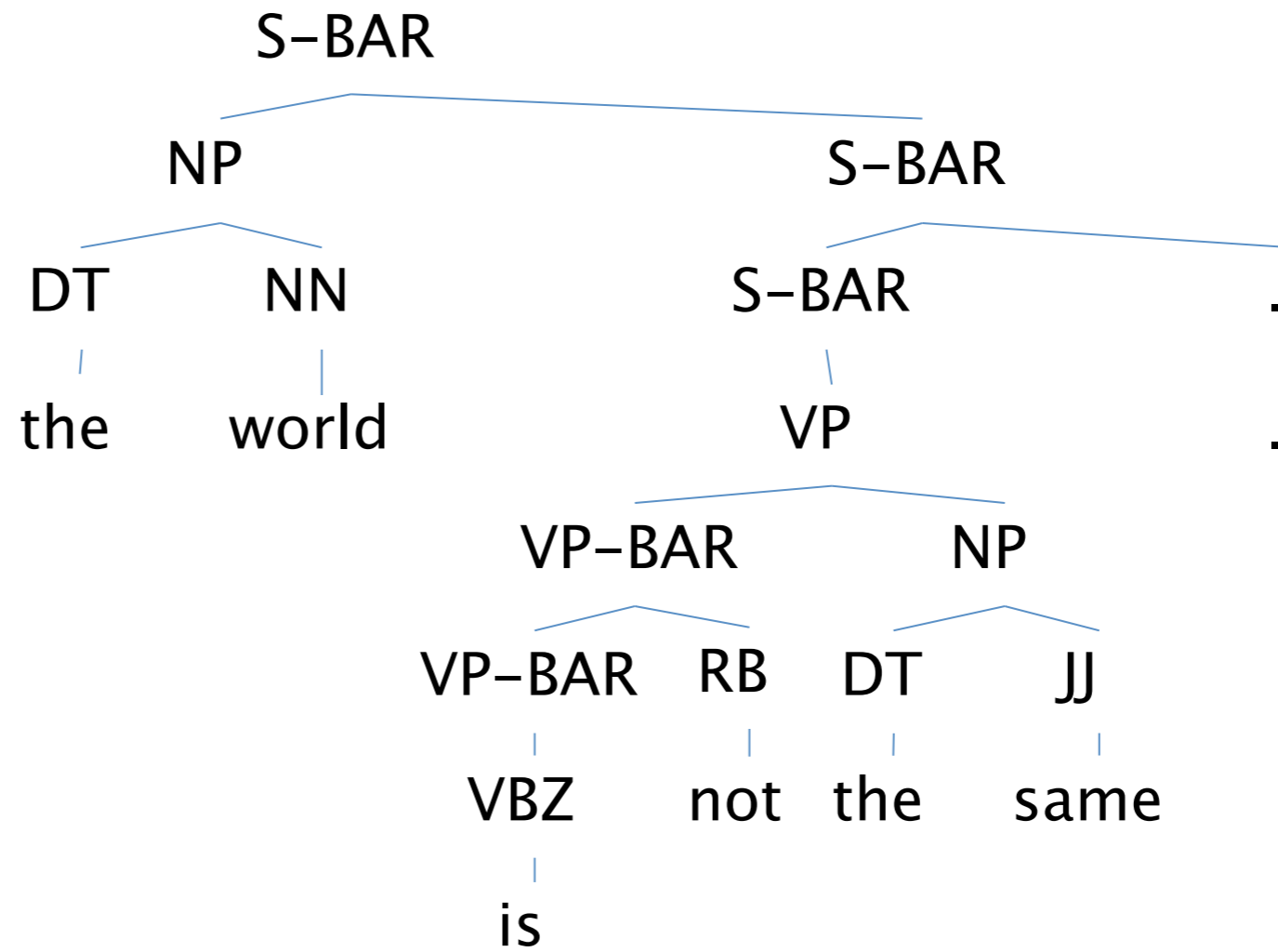
xLNTs Explanation

lkn ,

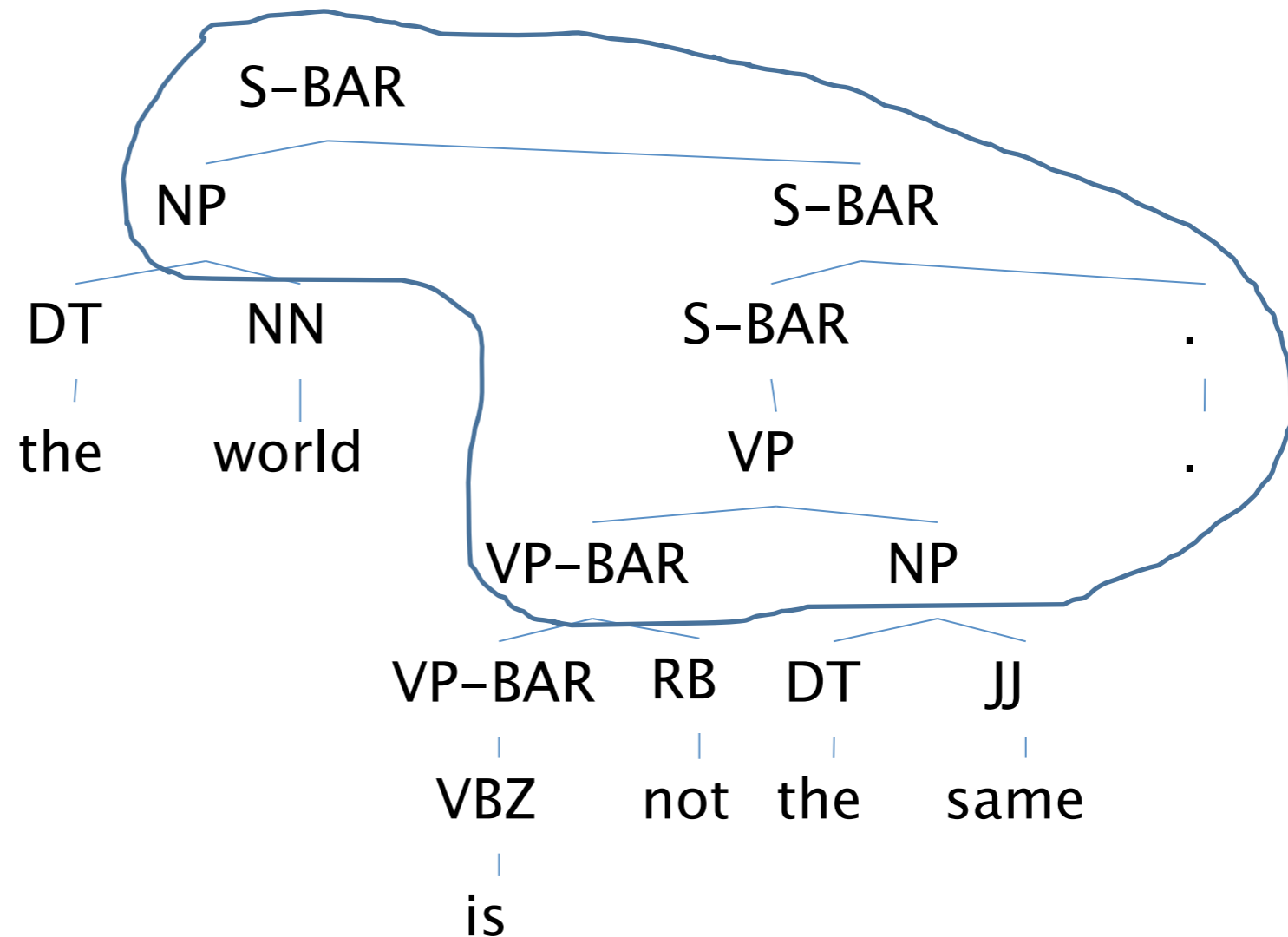


xLNTs Explanation

Ikn , hnA , AyDA ,



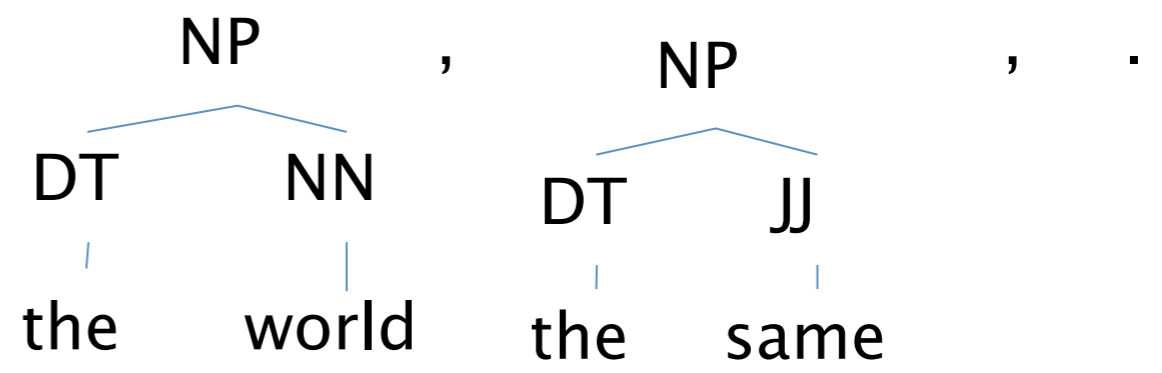
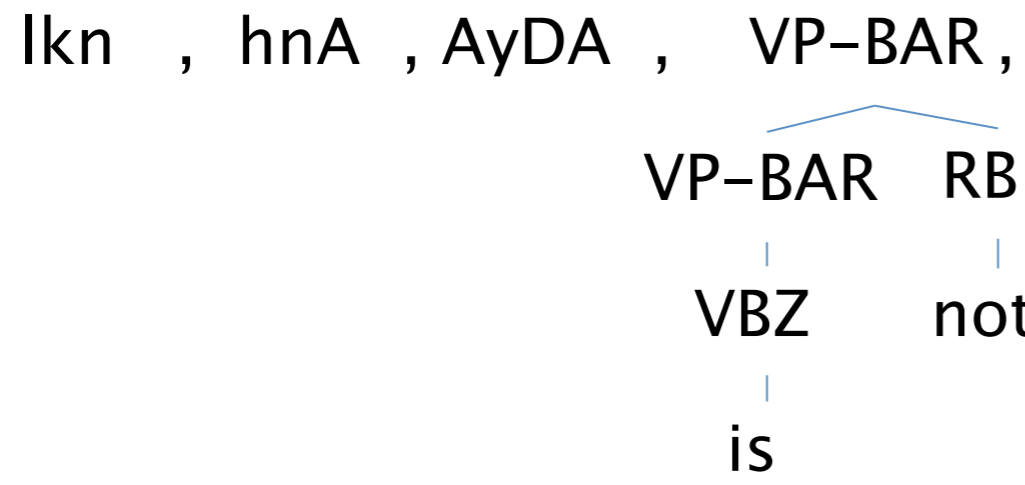
lkn , hnA , AyDA ,

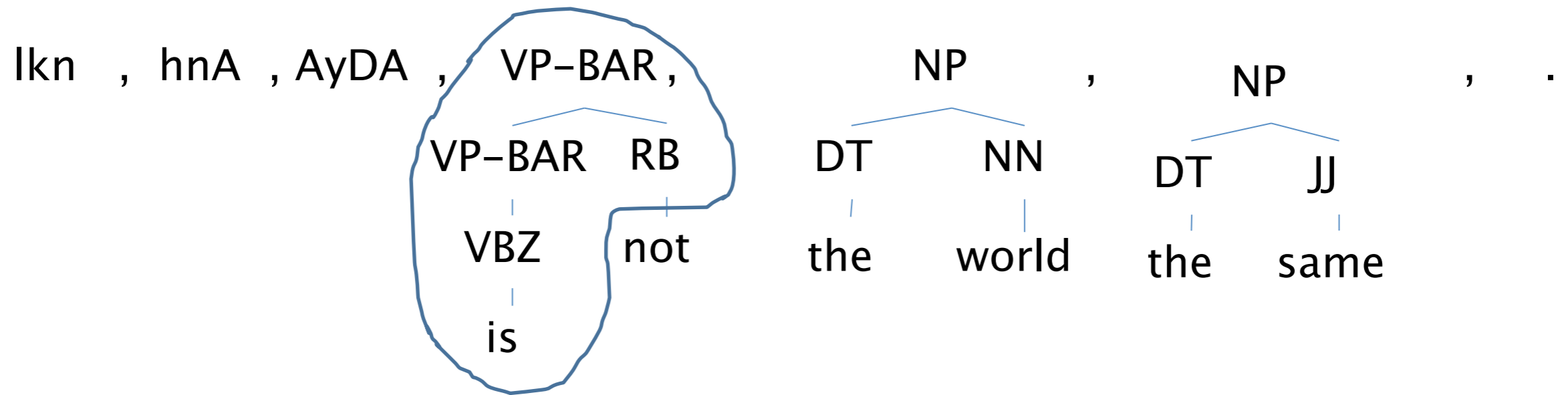


This rule takes care of SVO→VSO re-ordering.
Problems

- Period is part of rule: lots of SVO→VSO rules are needed
- Even if period were higher, there might be PPs
- Verb still unknown, may control re-ordering?

xLNTs Explanation



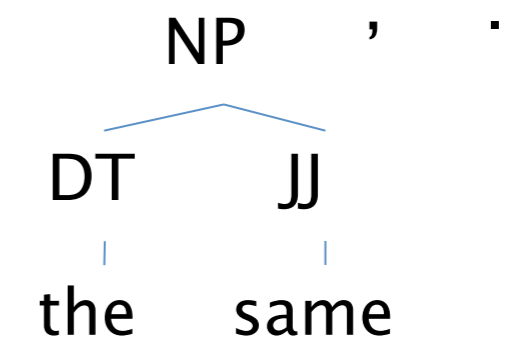
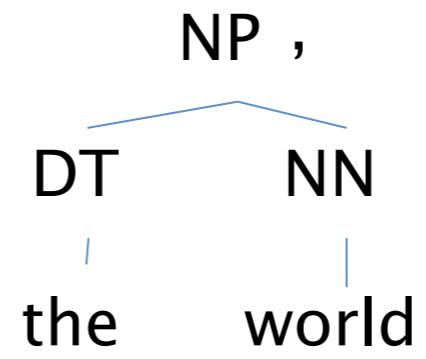


Problem

- Translation of "is" by this rule requires an RB to the right
- So, there are zillions of rules that translate "is" to the same Arabic word

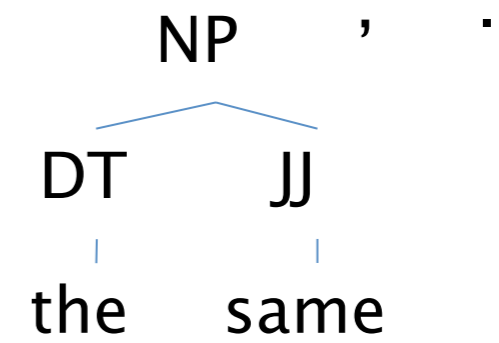
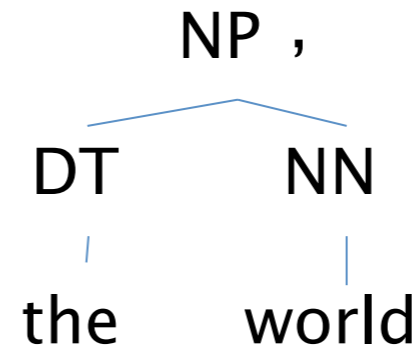
xLNTs Explanation

lkn , hnA , AyDA , RB , yEd ,
|
not



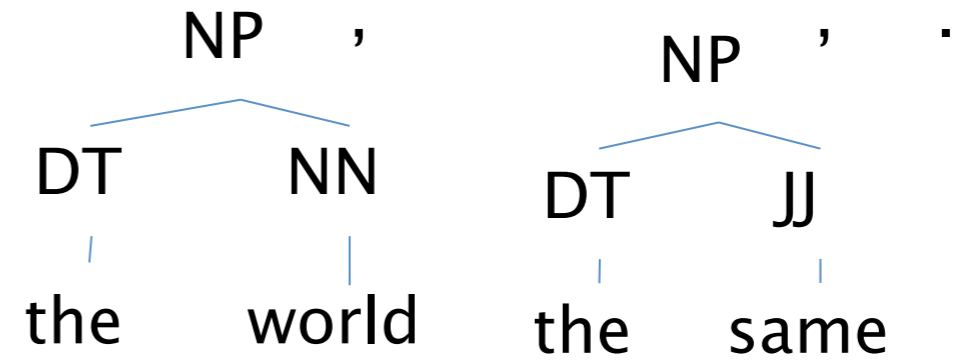
xLNTs Explanation

lkn , hnA , AyDA , RB
|
not , yEd ,



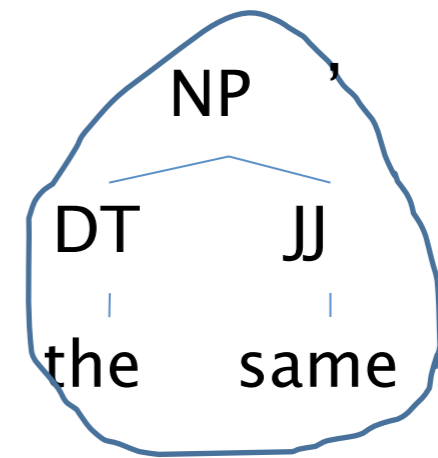
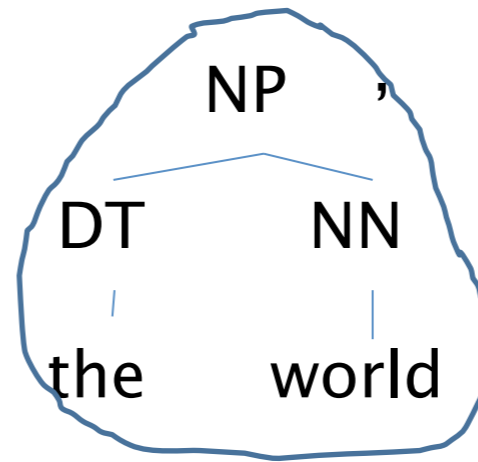
xLNTs Explanation

lkn , hnA , AyDA , Im , yEd ,



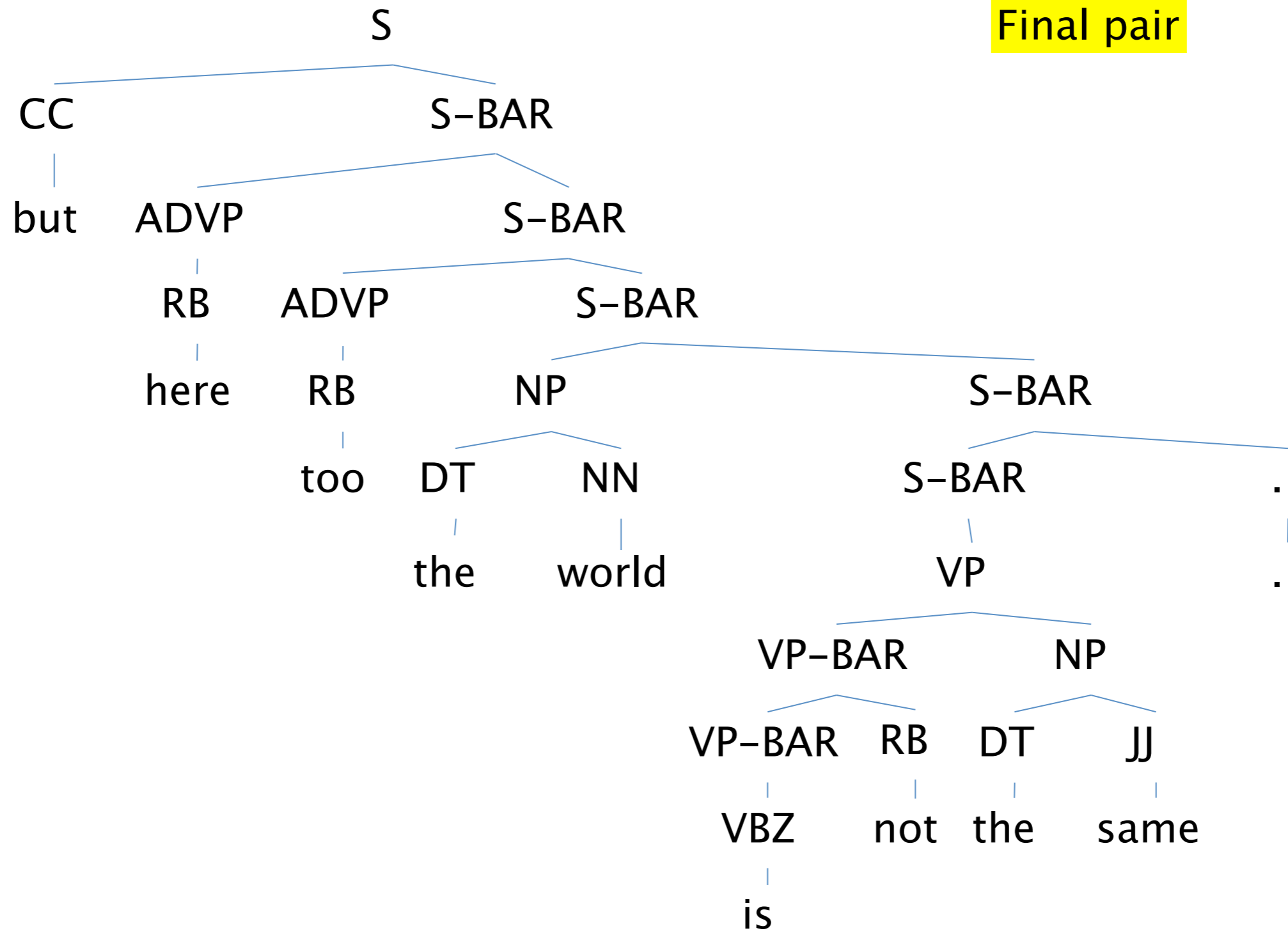
xLNTs Explanation

lkn , hnA , AyDA , Im , yEd ,



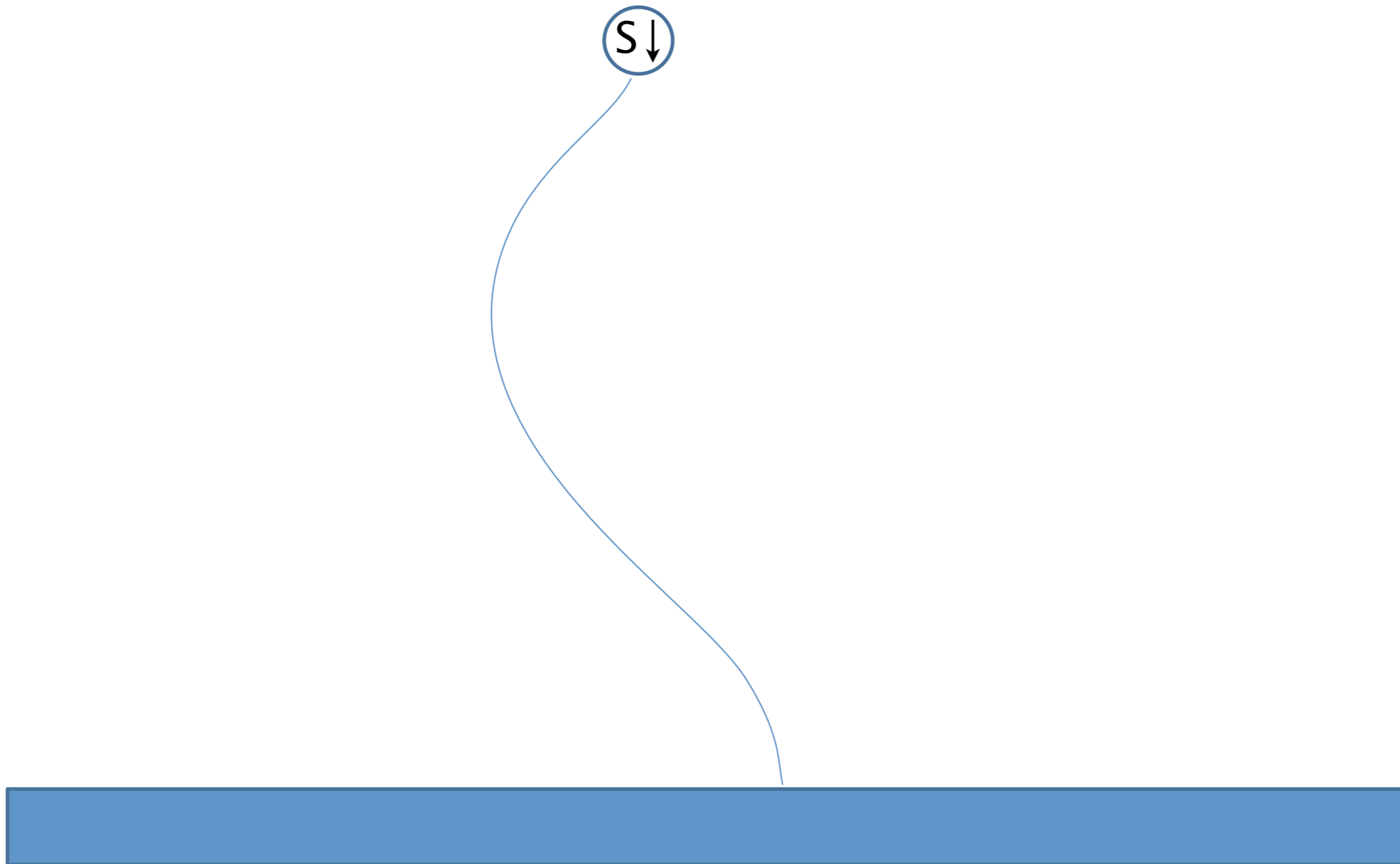
lkn , hnA , AyDA , Im , yEd , AIEAlm , *Ath , .

Final pair

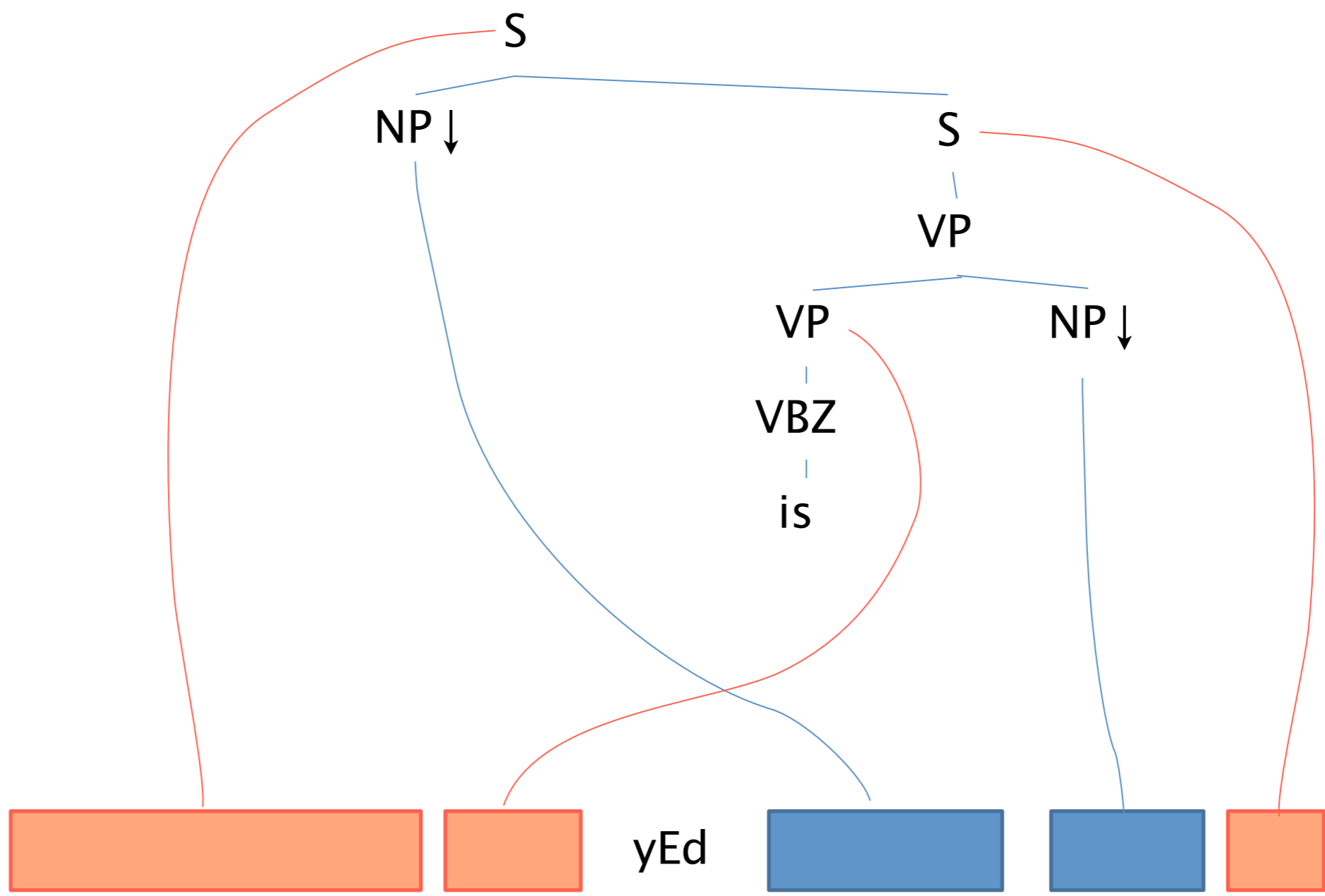


lkn hnA AyDA Im yEd AIEAlm *Ath .

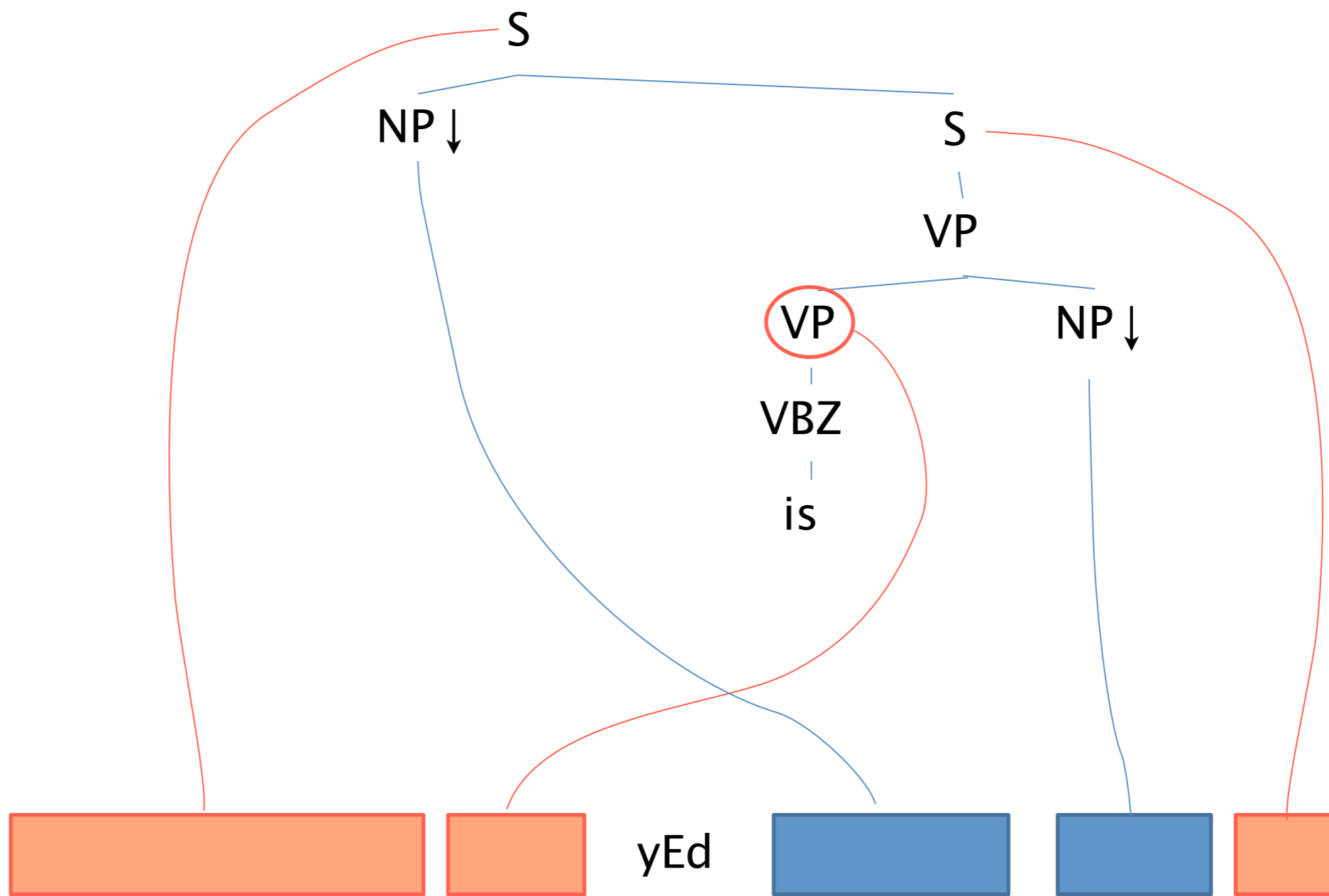
STAG Explanation



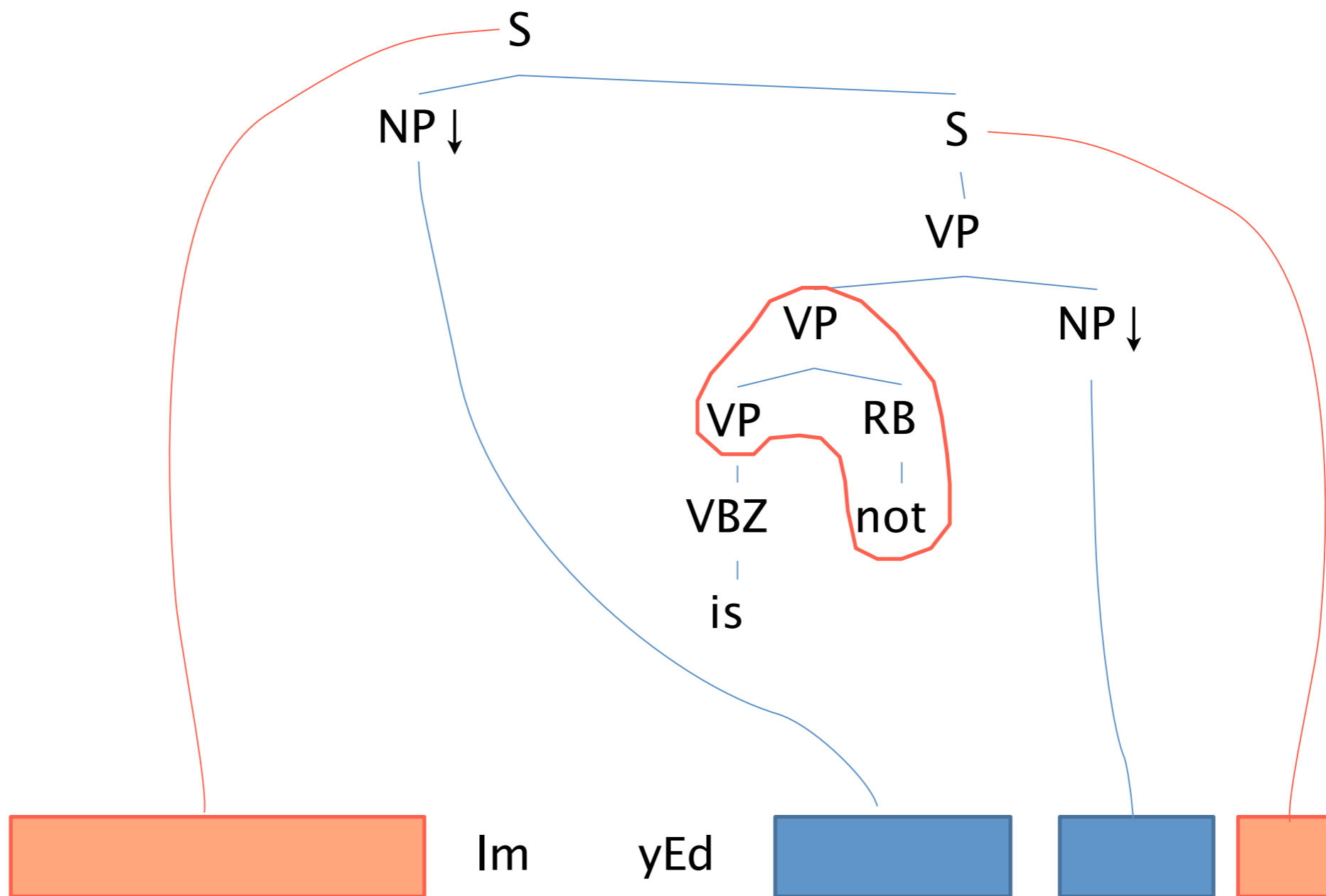
STAG Explanation



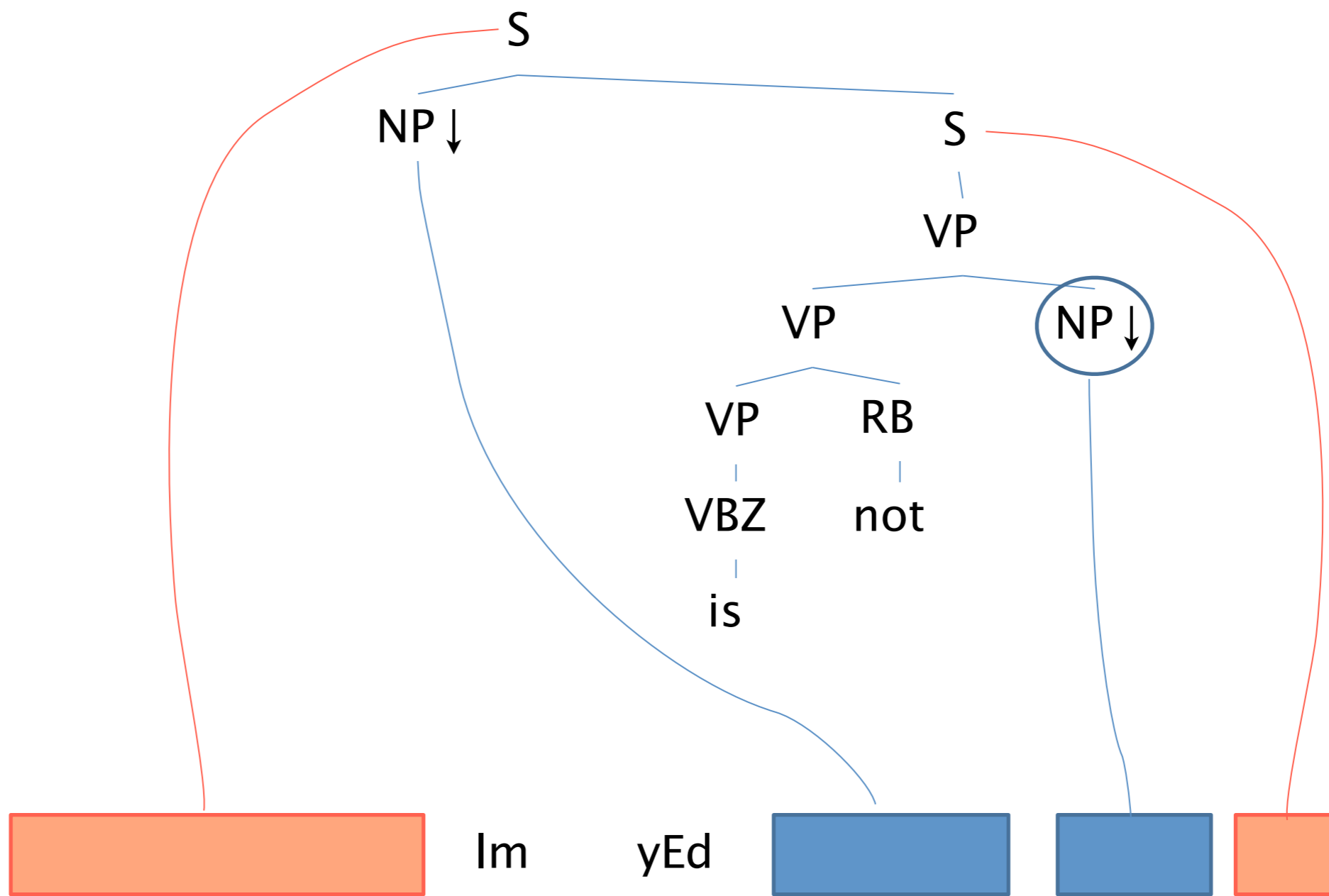
STAG Explanation



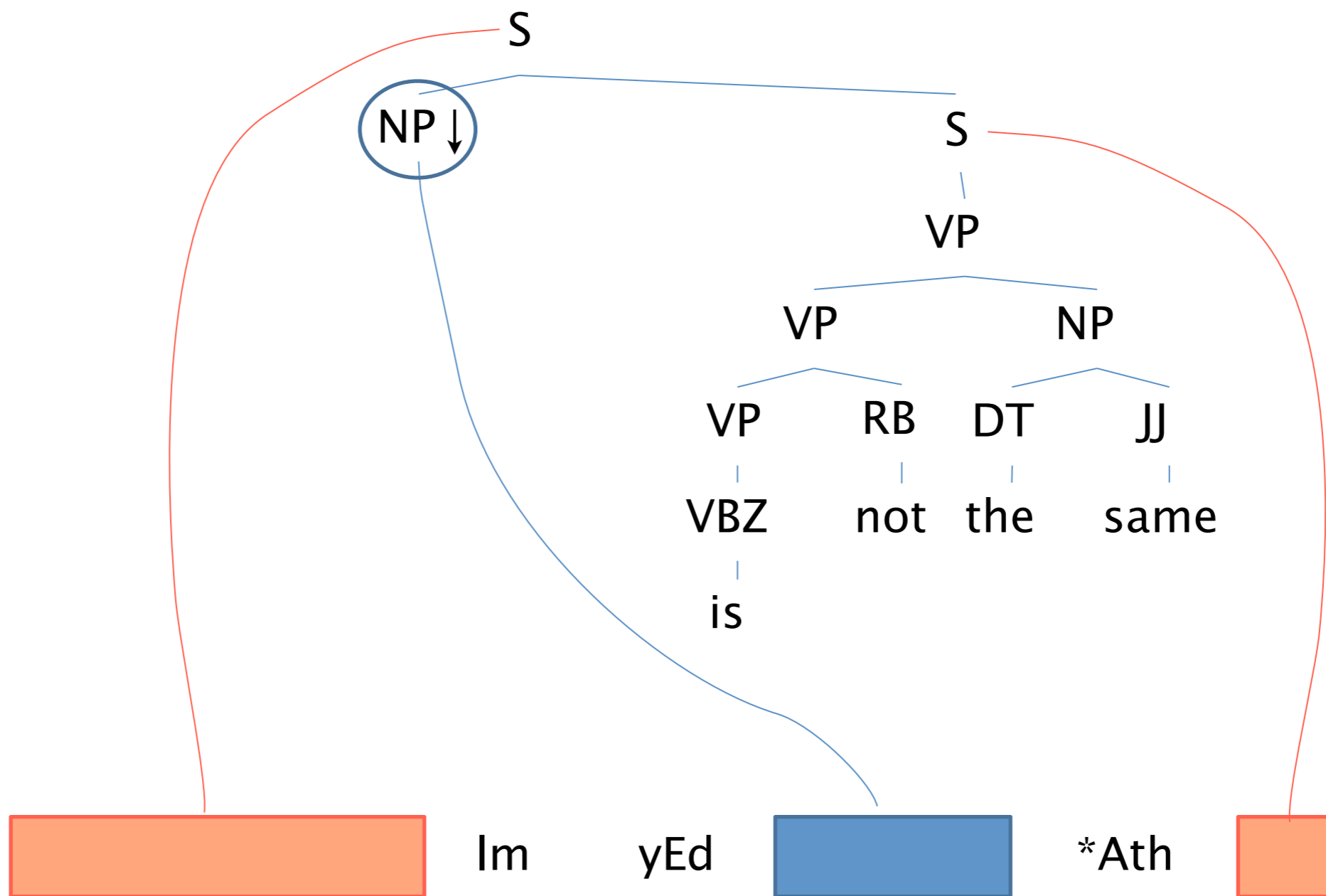
STAG Explanation



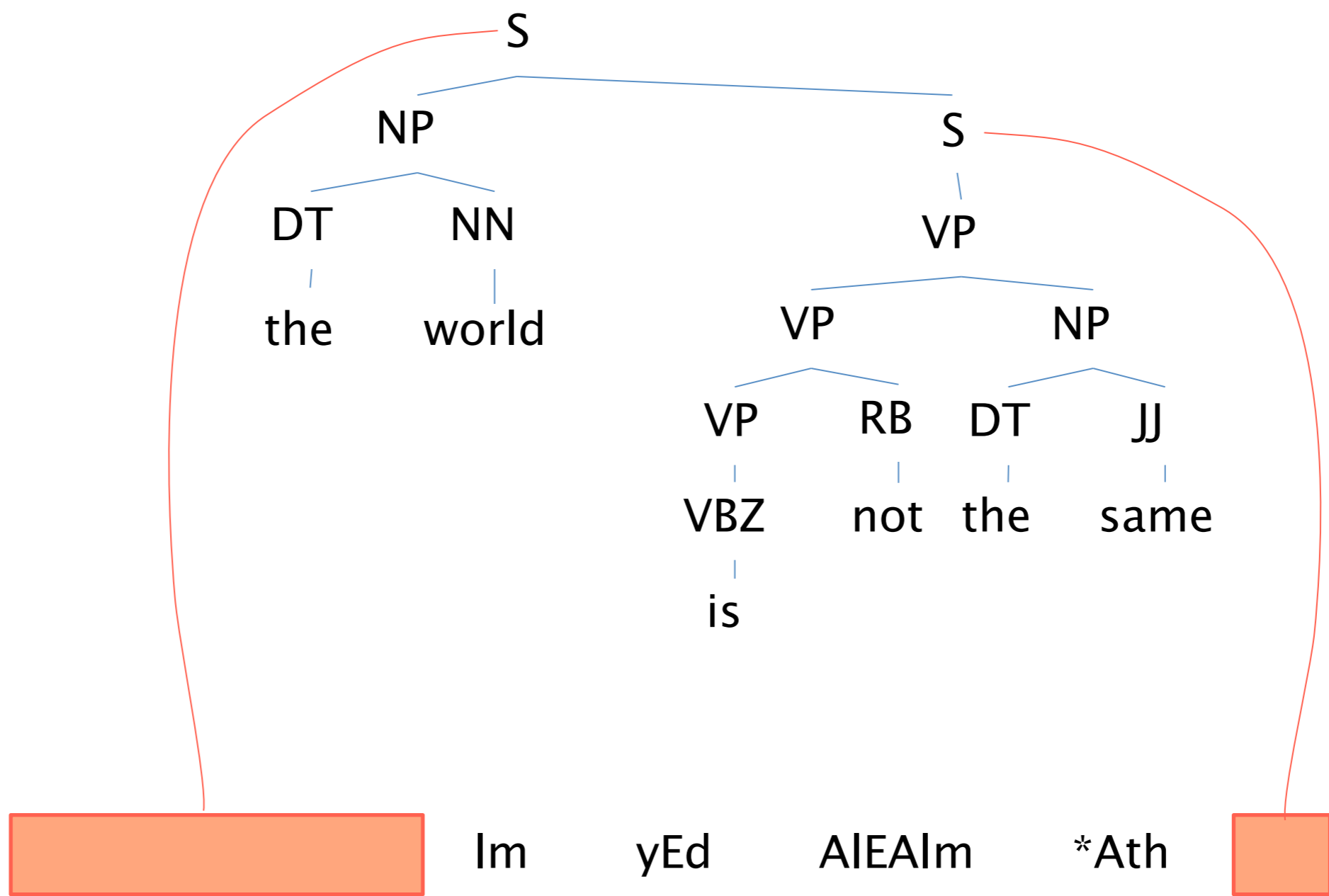
STAG Explanation



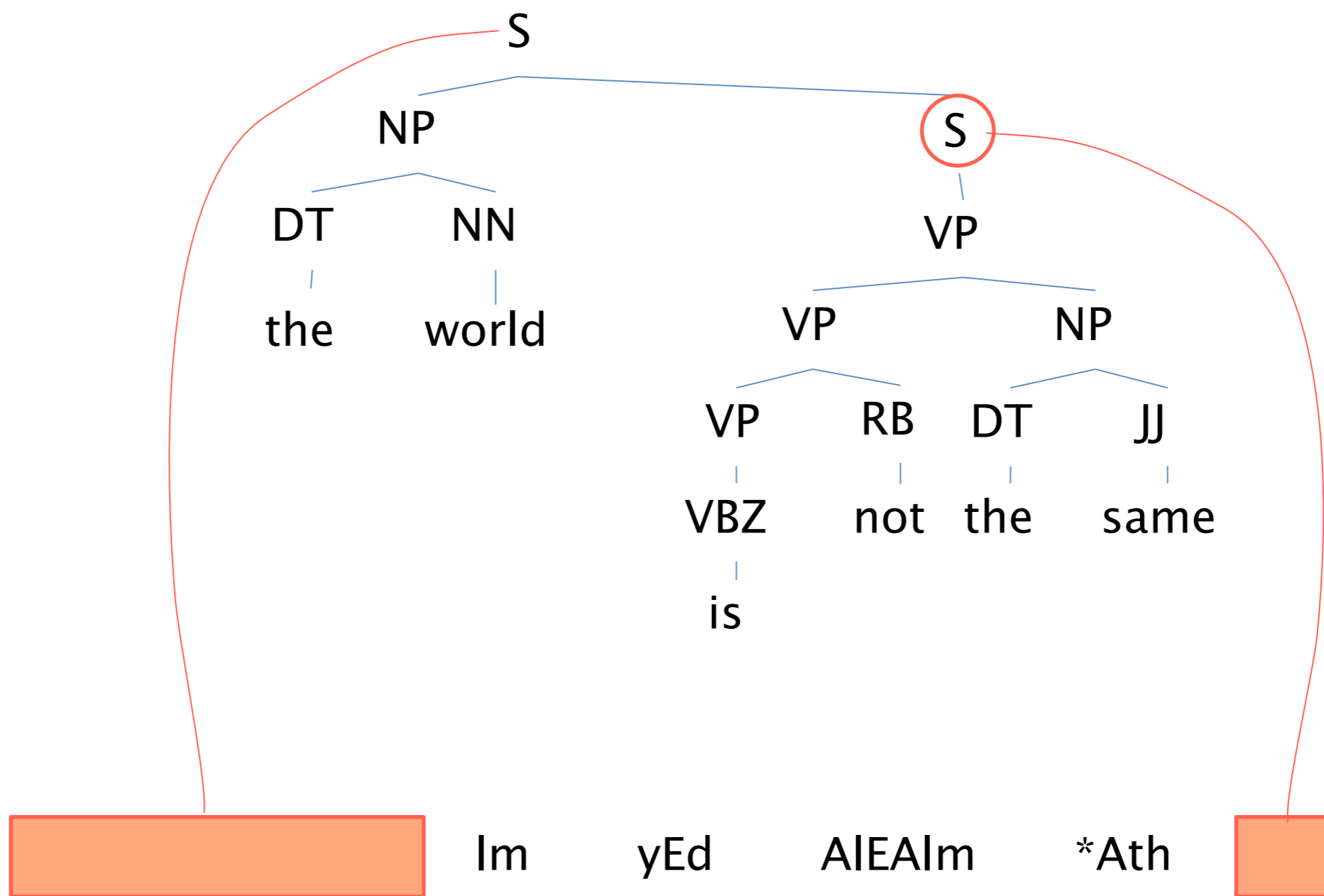
STAG Explanation



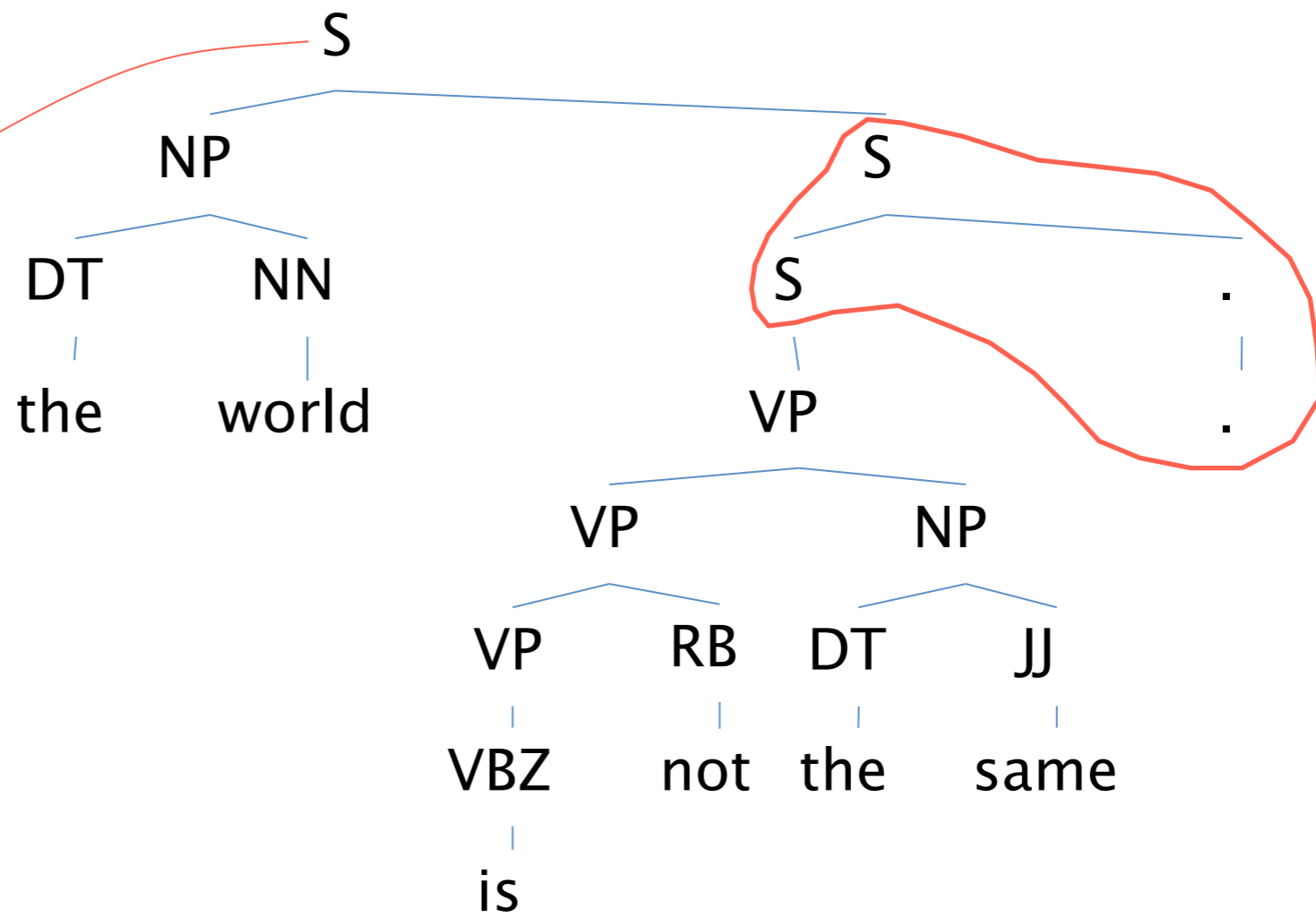
STAG Explanation



STAG Explanation

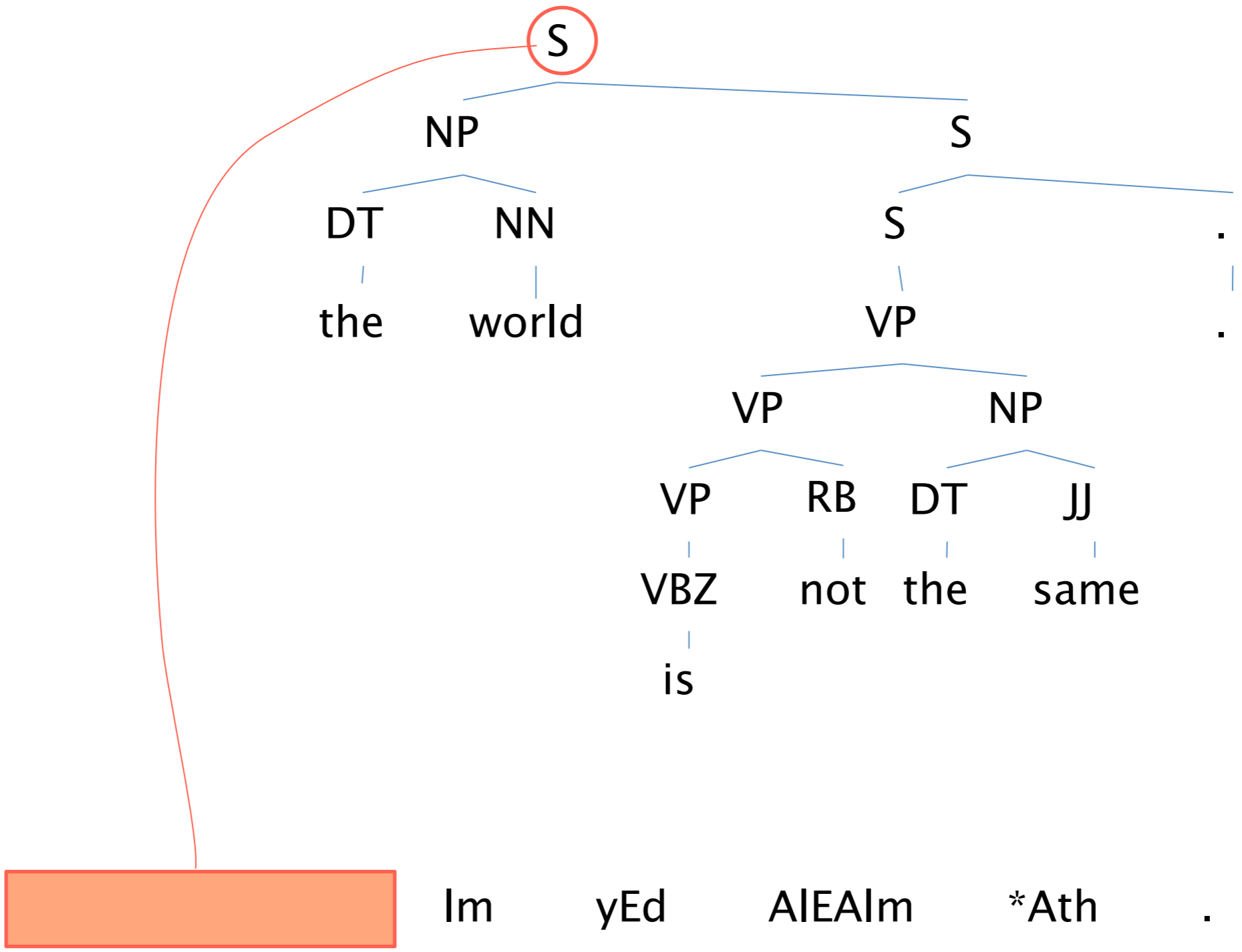


STAG Explanation

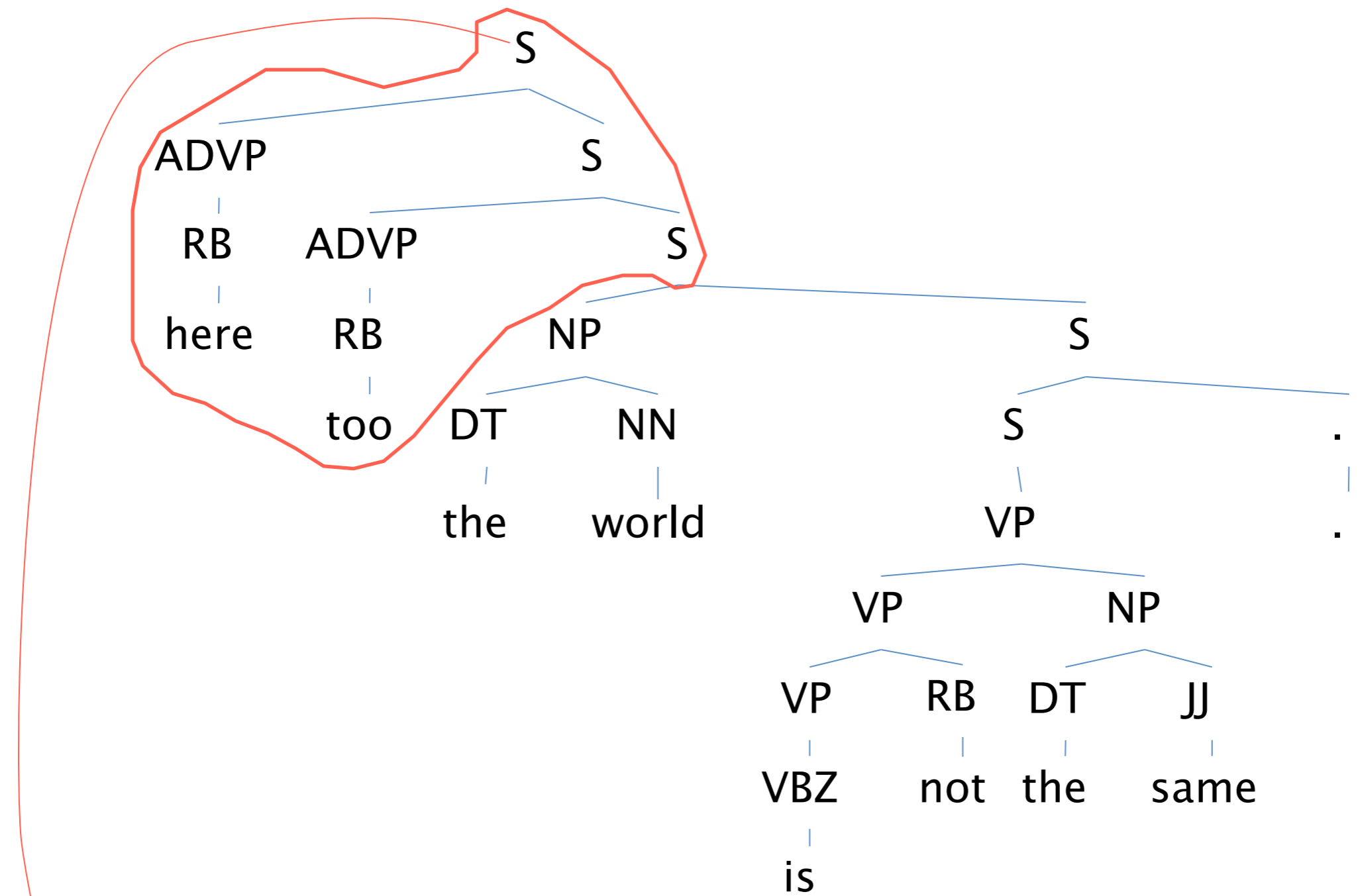


Im yEd AIEAlm *Ath .

STAG Explanation

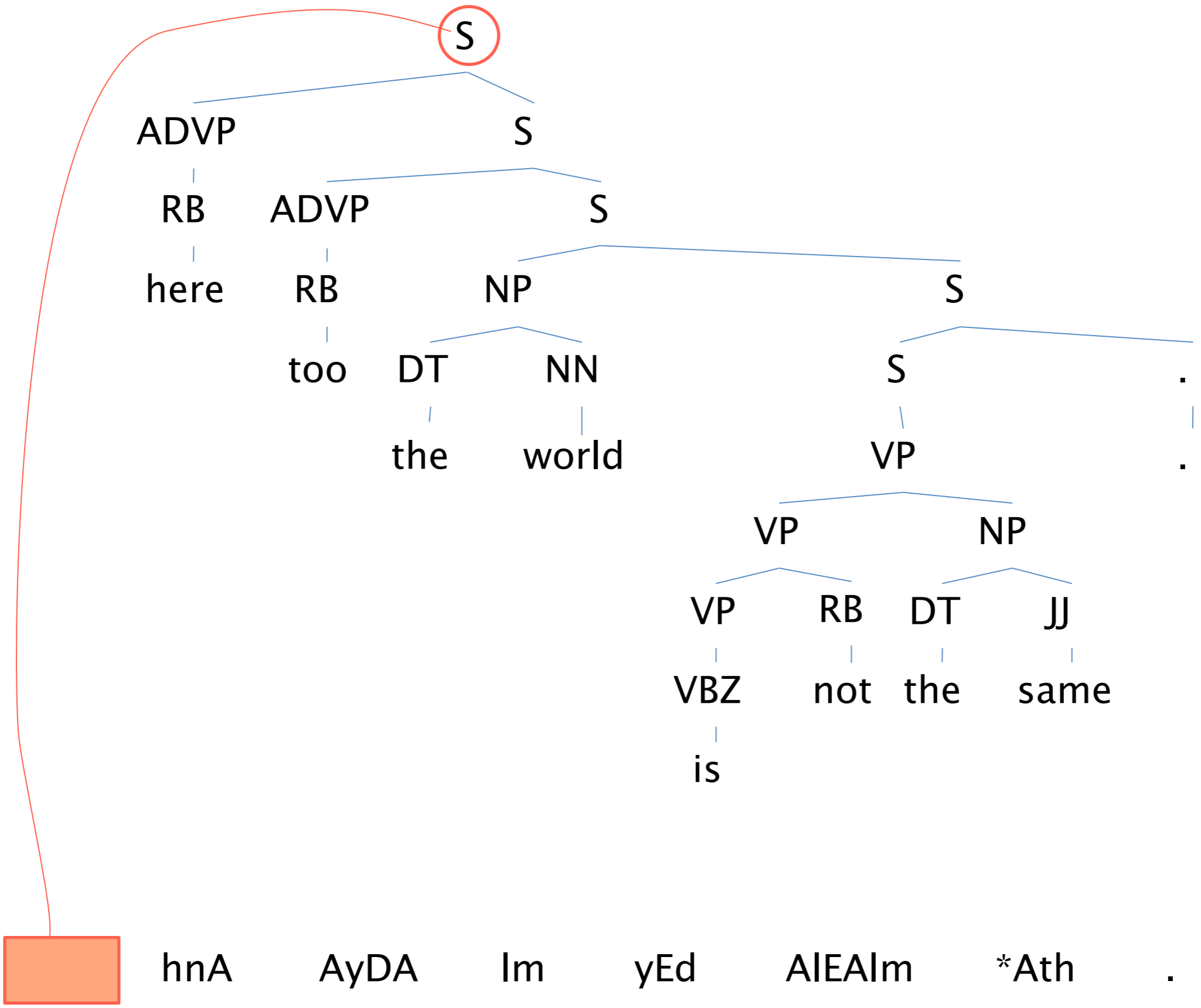


STAG Explanation

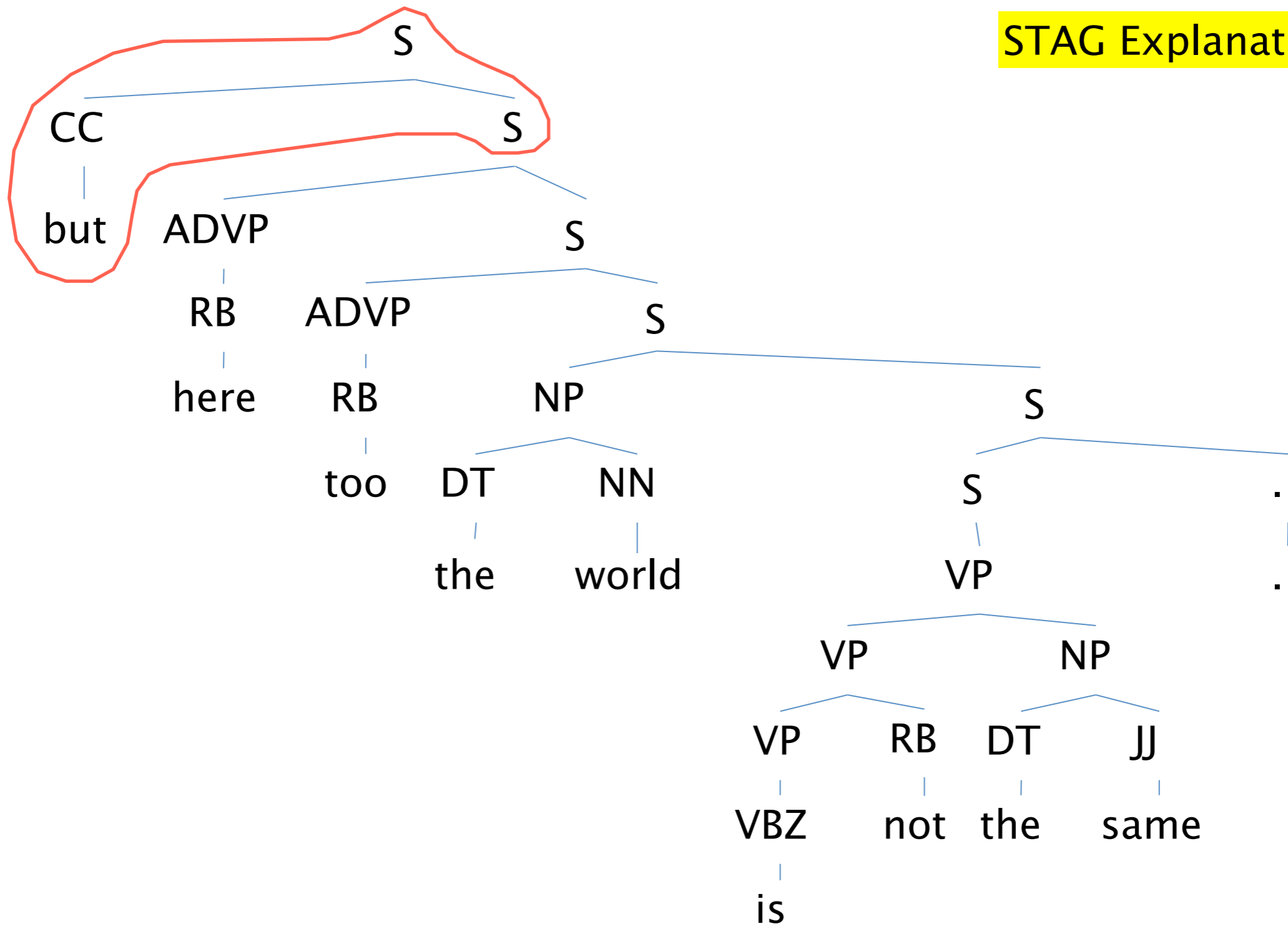


hnA AyDA Im yEd AIEAlm *Ath .

STAG Explanation

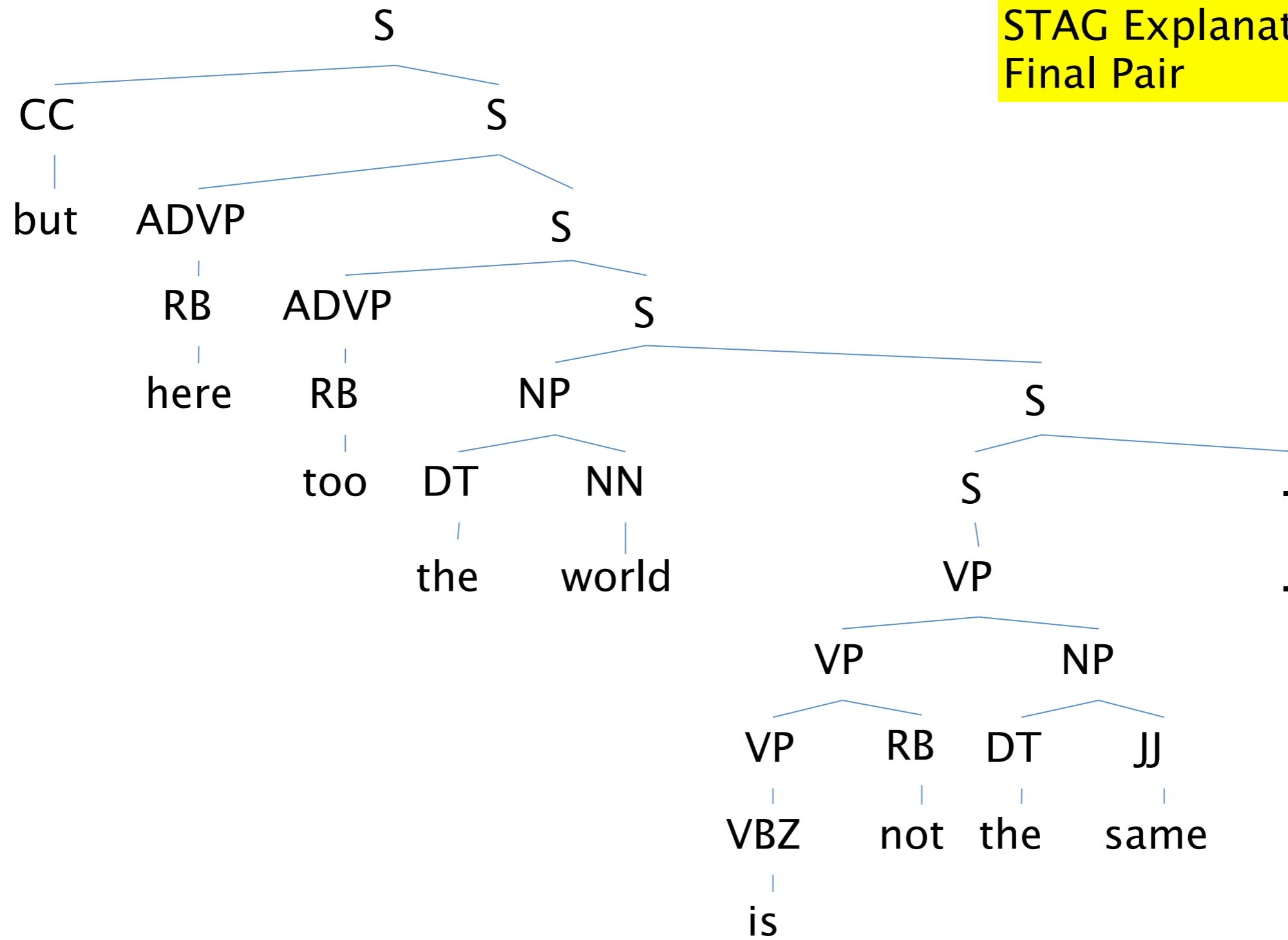


STAG Explanation



lkn hnA AyDA Im yEd AIEAlm *Ath .

STAG Explanation
Final Pair



lkn hnA AyDA Im yEd AIEAlm *Ath .

Parsed, Aligned, Bilingual Data

- Understanding weak points of automatic parsers and aligners
- Building better parsers and aligners
- Designing radically new MT models