Cohesive Constraints in a Beam Search Phrase-based Decoder

Nguyen Bach, Stephan Vogel
Carnegie Mellon University

Colin Cherry
Microsoft Research

interACT
International Center for Advanced Communication Technologies

Carnegie Mellon

lti

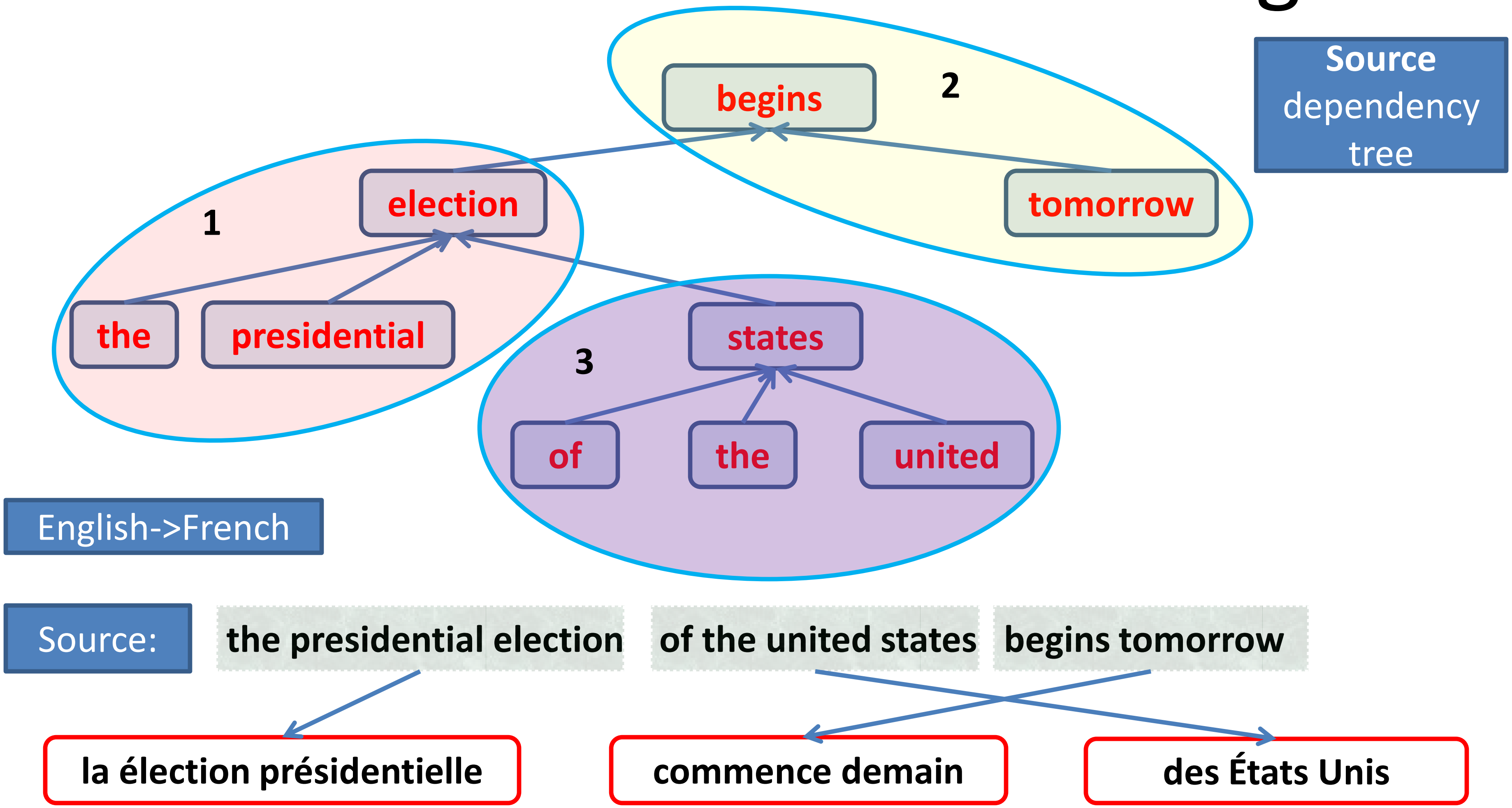
Overview

- Apply **cohesive constraints** during decoding process to consider the **source dependency structures**
- Introduce extensions of the cohesive constraints.
- Analyze the impact of cohesive constraints across language pairs with different reordering models
- Applied to English-Spanish , English-Iraqi and Chinese-English translation tasks
  - Significant improvements on English-Spanish
  - Stable improvements on other pairs

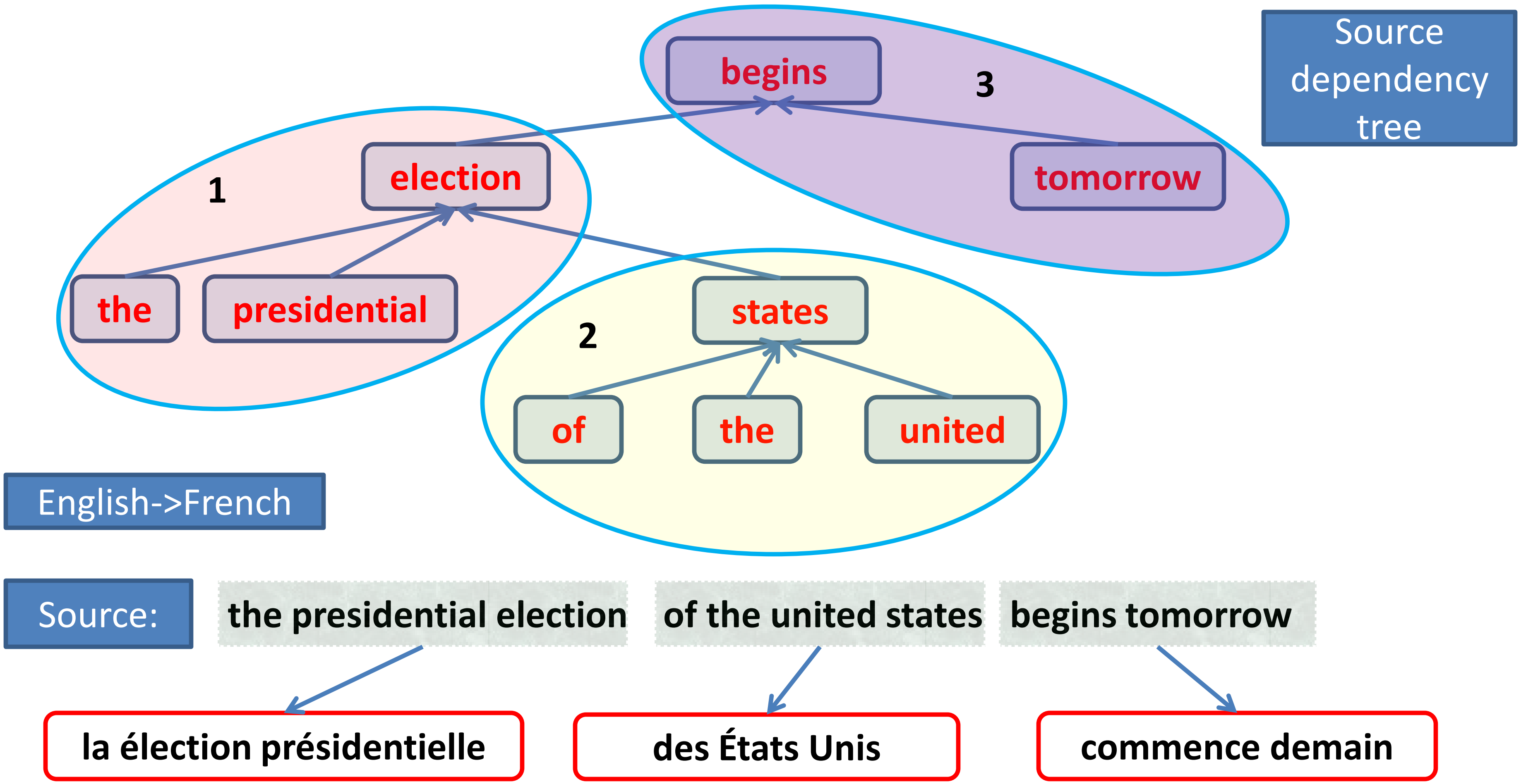
Outline

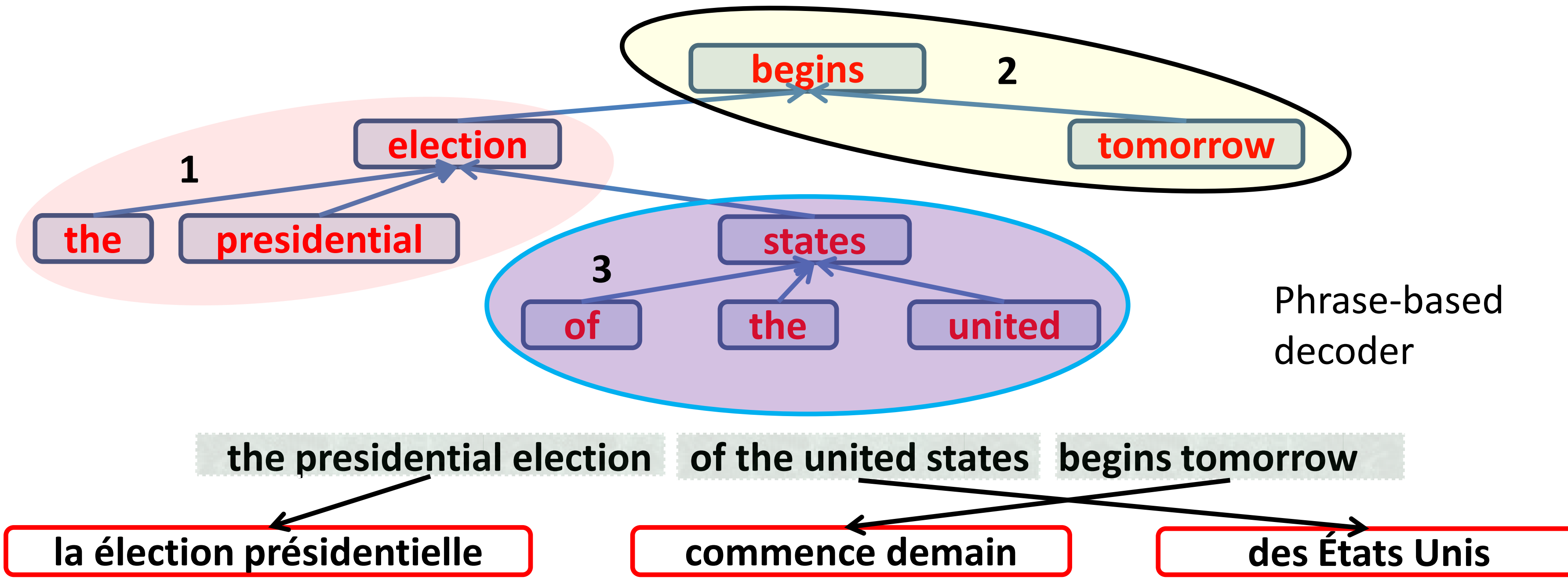
- **Cohesive Decoding Approach**
- Experiments
- Conclusions & Future Work

What is a cohesive decoding?

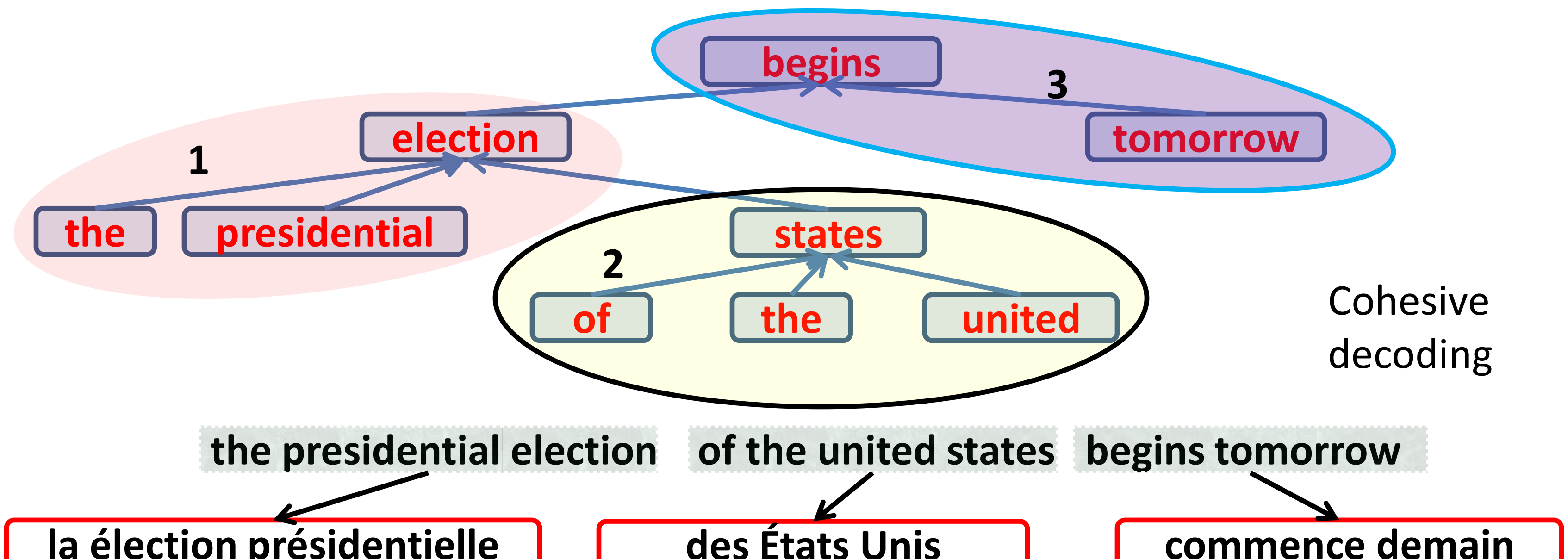


What is a cohesive decoding?



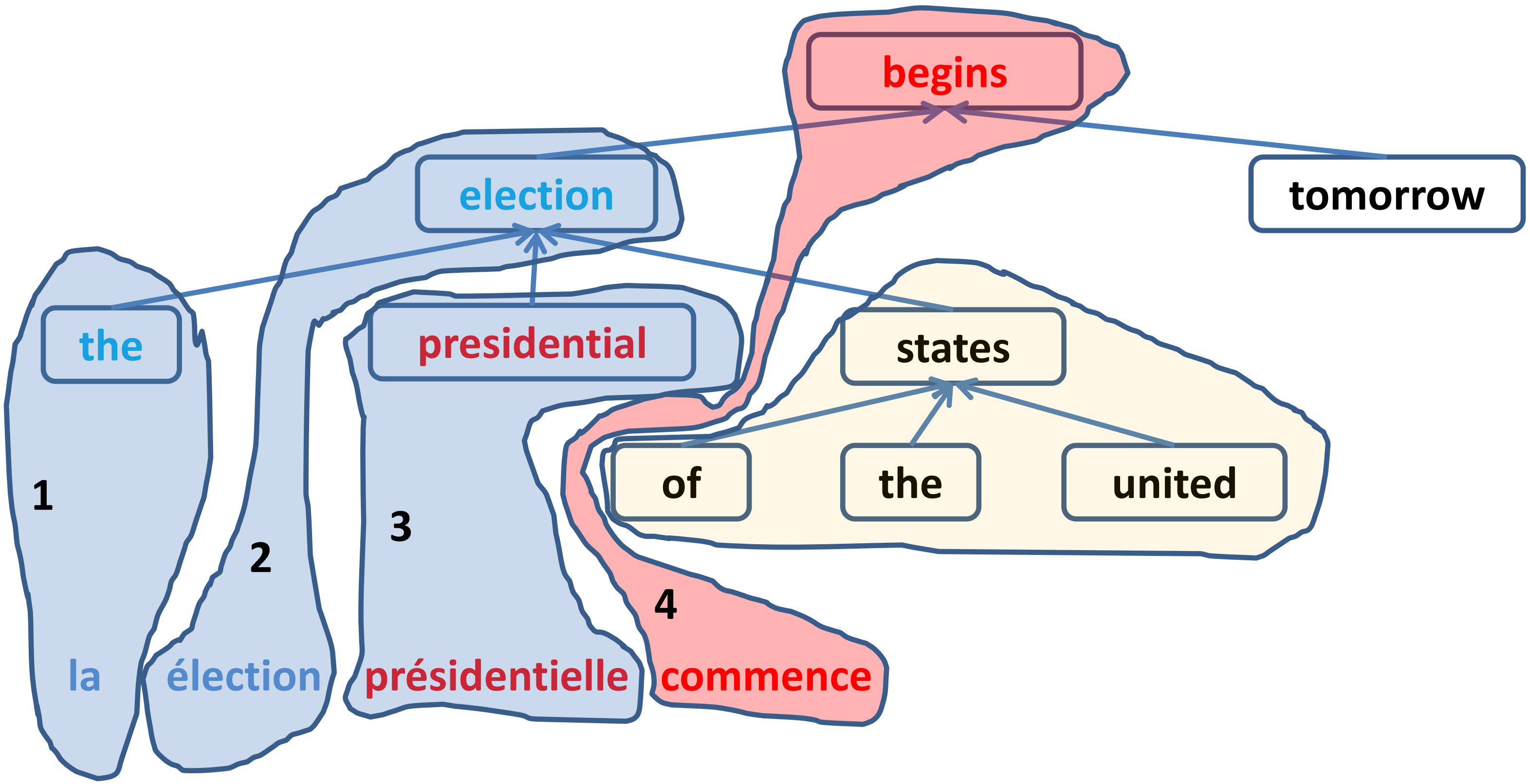


Phrase-based decoder



Cohesive decoding

Interruption Checks (Cherry, 2008)



Two Questions

- How to determine the largest subtree that needs to be completed before the translation process can move elsewhere in the tree?
  - Interruption Check: use **left and right most tokens of *the previous translated source phrase*** and climb up the tree
- If a violation happens, how to constrain the decoder to penalize cohesion violated translation hypothesis?
  - Interruption Check **: Binary event**

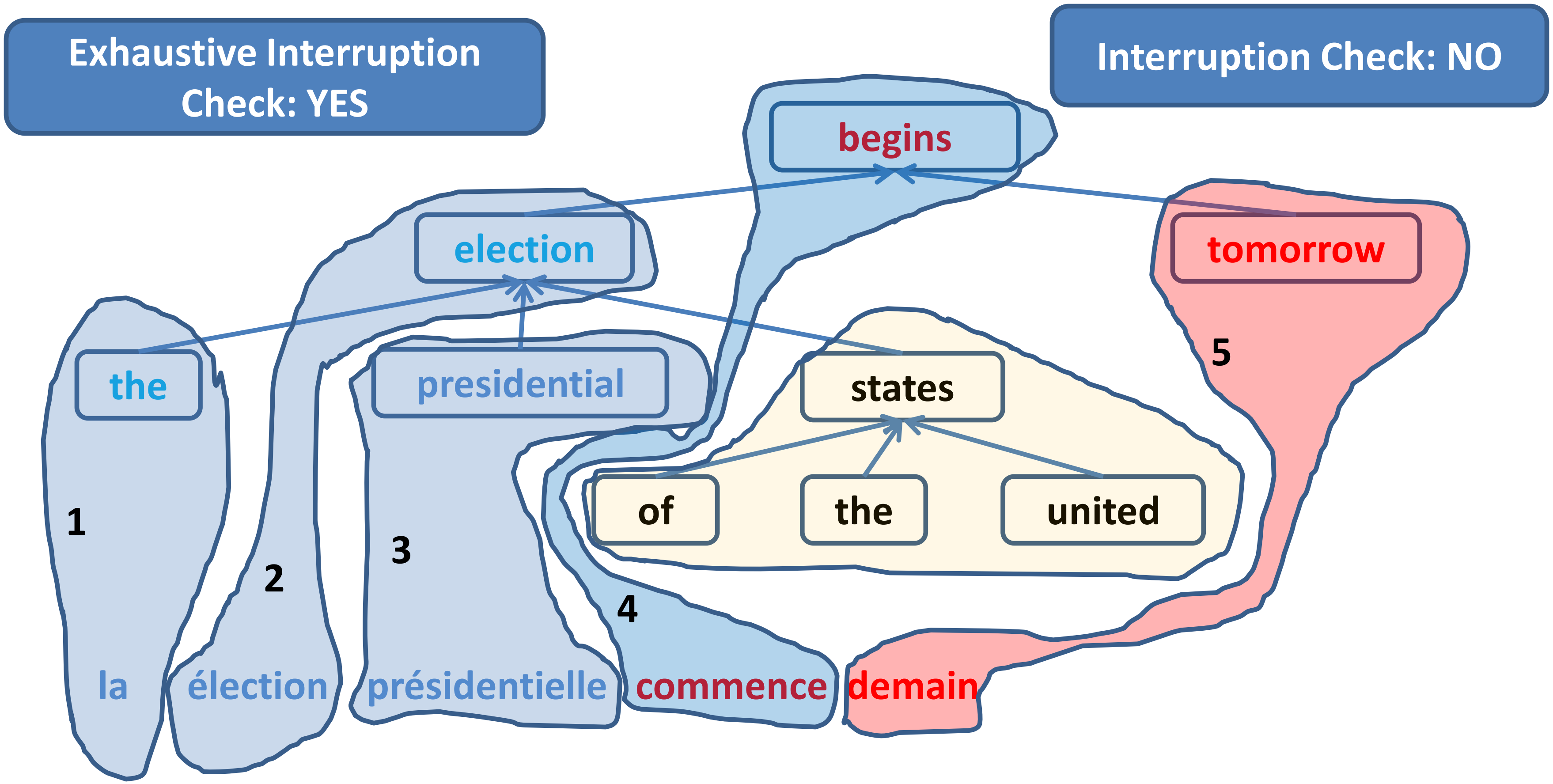
Exhaustive Interruption Check

- Interruption Check only penalizes the cohesion violation 1 time

- **Should penalties persist as long as violations remain unresolved?**

- Exhaustive Interruption Check keeps punishing a cohesion violation until it is fixed.

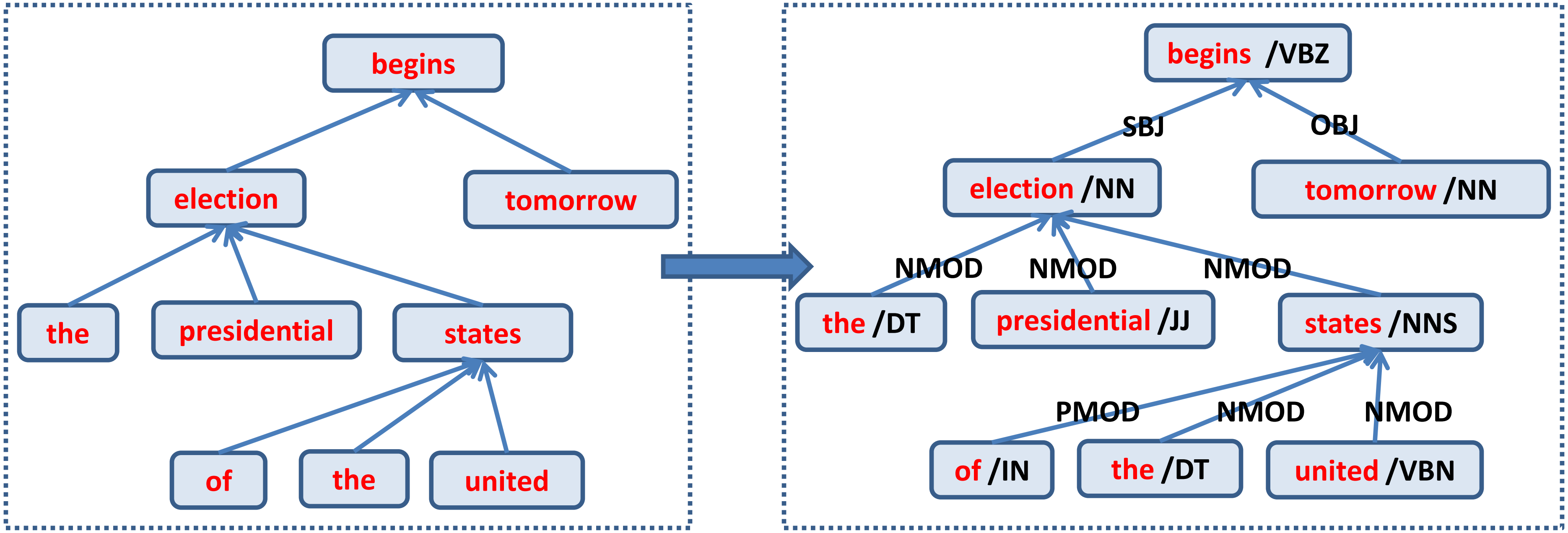
Exhaustive Interruption Check



Cohesion Violation Penalties

- Interruption Check and Exhaustive Interruption Check: binary event
- **Are some violations worse than others?**
- Penalize a cohesion violation by ***the number of untranslated words*** under the largest subtree
  - Interruption Check -> **Interruption Count**
  - Exhaustive Interruption Check -> **Exhaustive Interruption Count**

Rich Interruption Constraints



- Penalize a cohesion violation by **4 constraints**
  - Binary event: violation/not violate
  - Interruption Count: untranslated word count
  - Verb Count: untranslated verb count
  - Noun Count: untranslated noun count

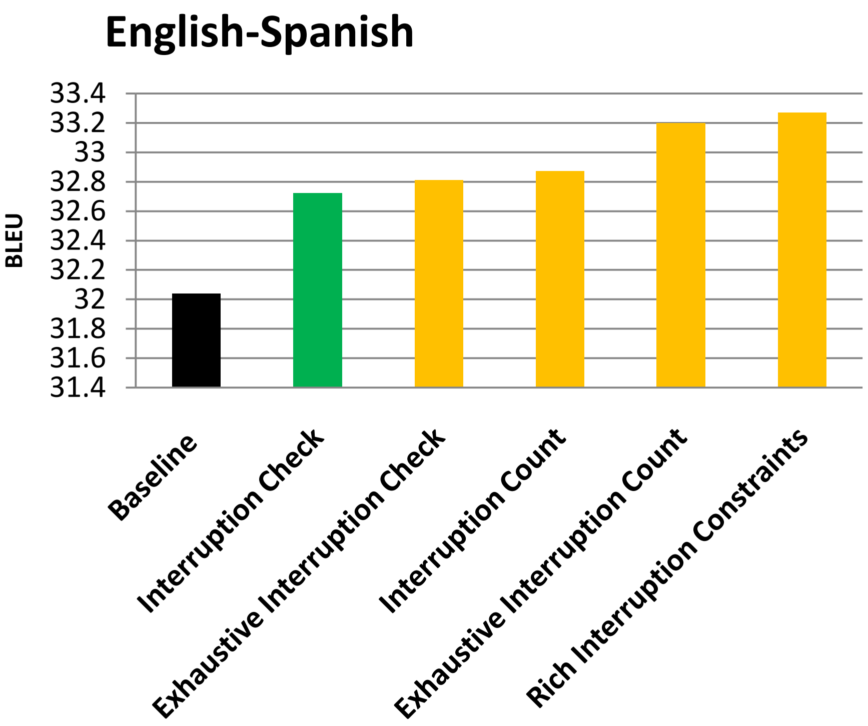
Comparison

		How to penalize a cohesion violation?		
		Binary	Number of untranslated words	Linguistics features
How to detect the largest subtree *T(n)*?	The previous phrase	**Interruption Check**	**Interruption Count**	**Rich Interruption Constraints**
	All previous phrases	**Exhaustive Interruption Check**	**Exhaustive Interruption Count**	N/A

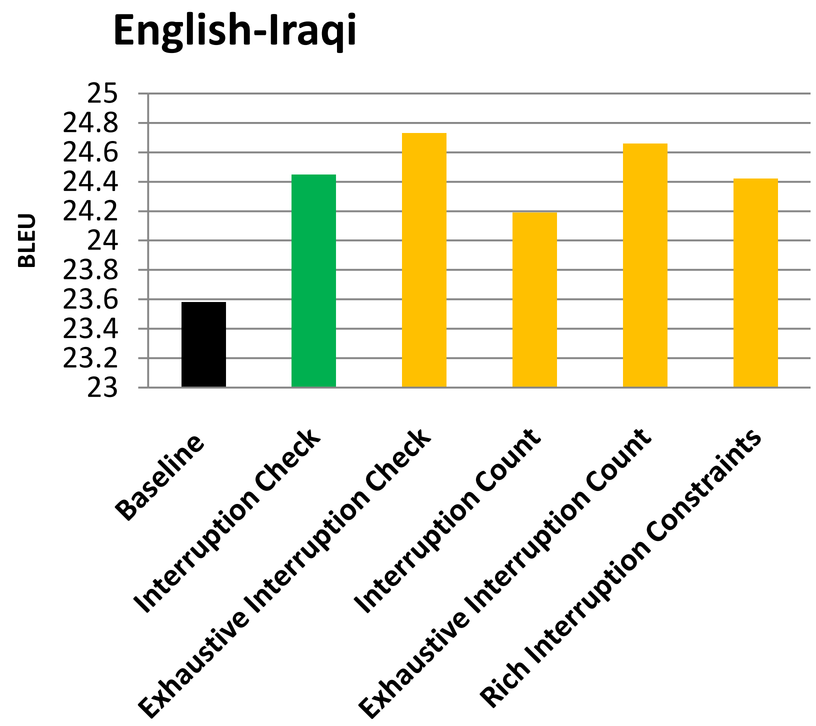
Outline

- Cohesion Decoding Approach
- **Experiments**
- Conclusions & Future Work

English-Spanish; English-Iraqi



Europarl nctest2007

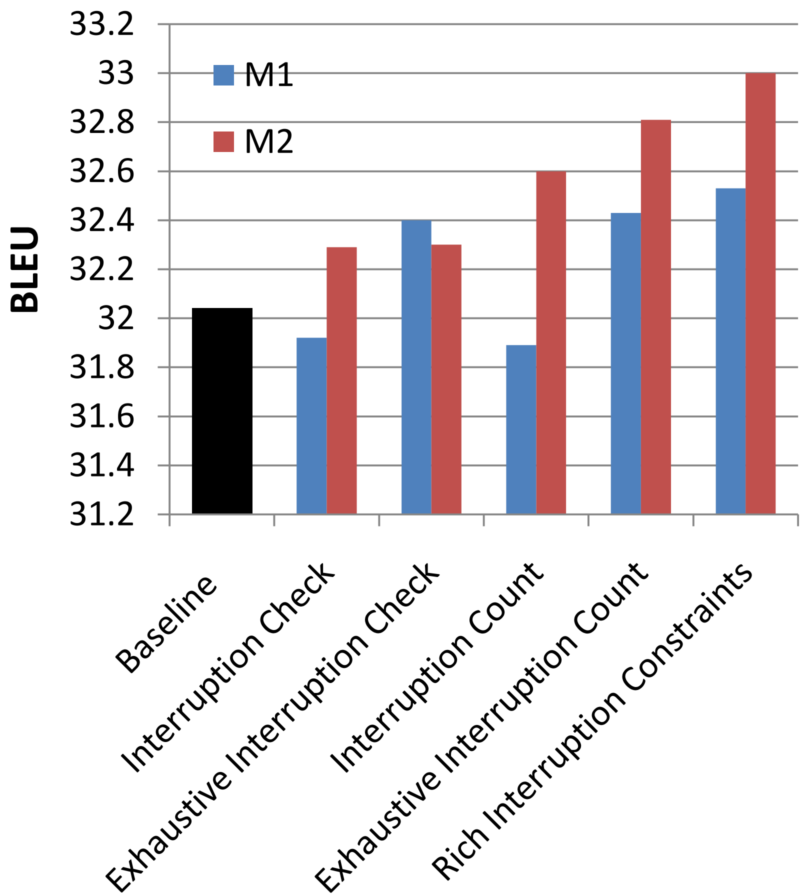


TransTac June08

Cohesive constraints obtained **improvements** over the standard phrase-based decoder.

How does the performance of the dependency parser affect cohesive constraints?

The Role of Dependency Parser on English-Spanish

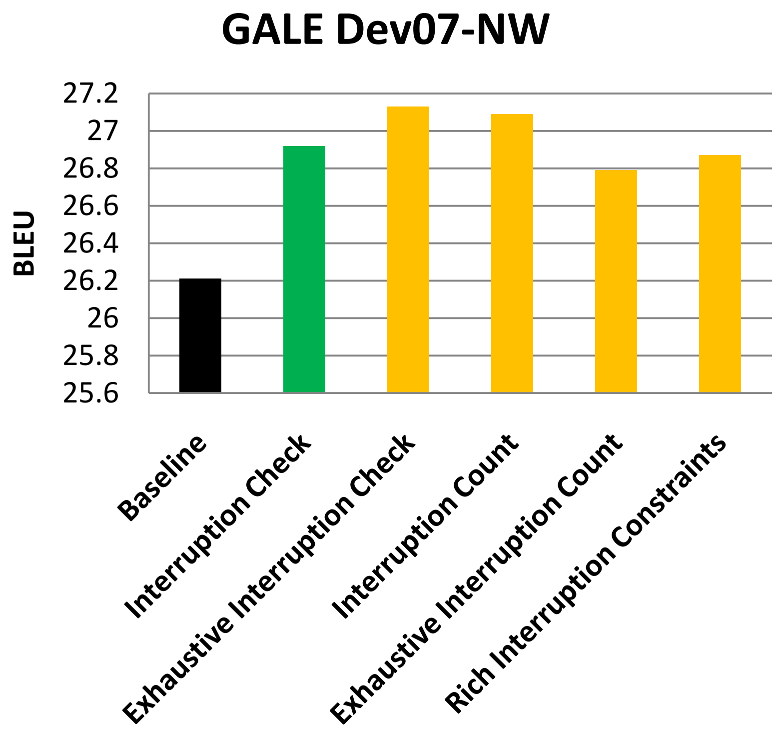


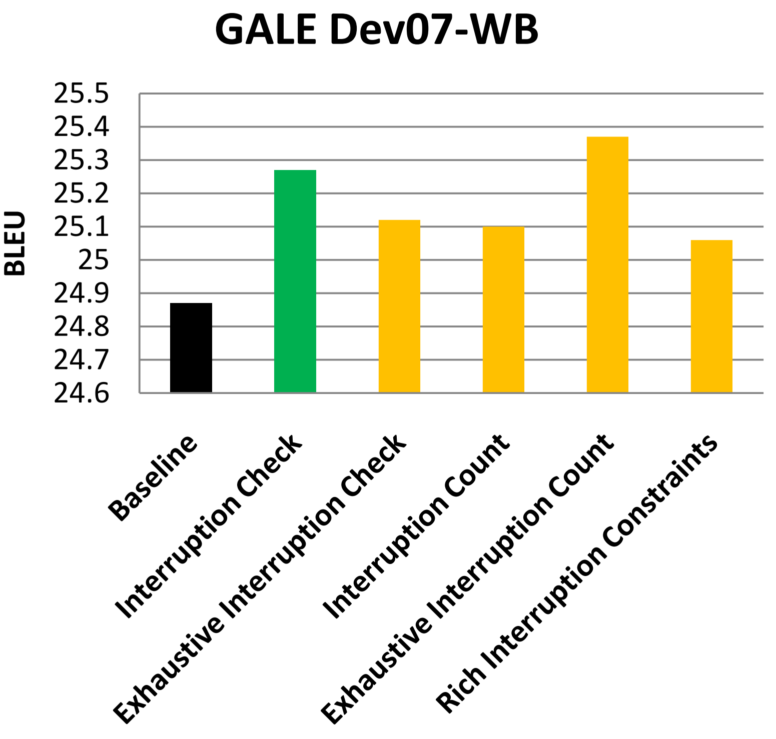
- Train 2 MALT dependency parser models: M1 with 10% of treebank and M2 with all treebank.
- Performance on CoNLL-07 dependency test set
  - **M1**: 19.41%
  - **M2**: 86.21%
- Apply to MT
  - M2 is better than M1

- Are the improvements subsumed by a strong reordering model and system scale?

- What if we translate from X->English?

GALE Chinese-English





Cohesive constraints obtained improvements even with **large scale system** and **strong reordering models**

Outline

- Cohesion Decoding Approach
- Experiments
- **Conclusions & Future Work**

Conclusions & Future Work

- Conclusions
  - Cohesive constraints are helpful
  - The effectiveness was shown when using with a strong reordering model
  - Obtained improvements with 3 language pairs and also covered a wide range of training corpus sizes, ranging from 500K up to 11M sentence pairs
- Future work
  - A source side dependency reordering model: Learning reordering events of the phrases based on source subtree movements
  - A hierarchical source side dependency reordering model: extend Galley&Manning (2008).

Questions

?