A Joint Phrasal and Dependency Model for Paraphrase Alignment

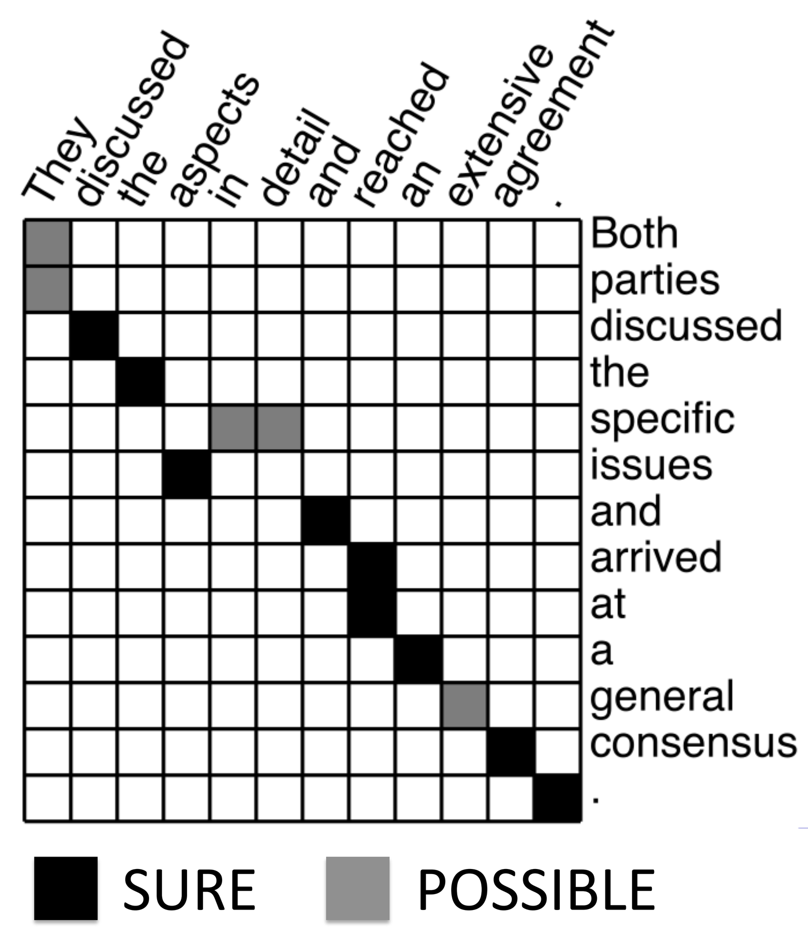
Kapil Thadani, Scott Martin and Michael White
kapil@cs.columbia.edu, {scott,mwhite}@ling.ohio-state.com

COLUMBIA UNIVERSITY

THE OHIO STATE UNIVERSITY

Monolingual Alignment

- **Goal:** identify semantically equivalent words and phrases across pairs of text segments
- Monolingual variant of MT alignment
- Useful for paraphrasing, entailment, sentence fusion, QA matching, plagiarism detection etc

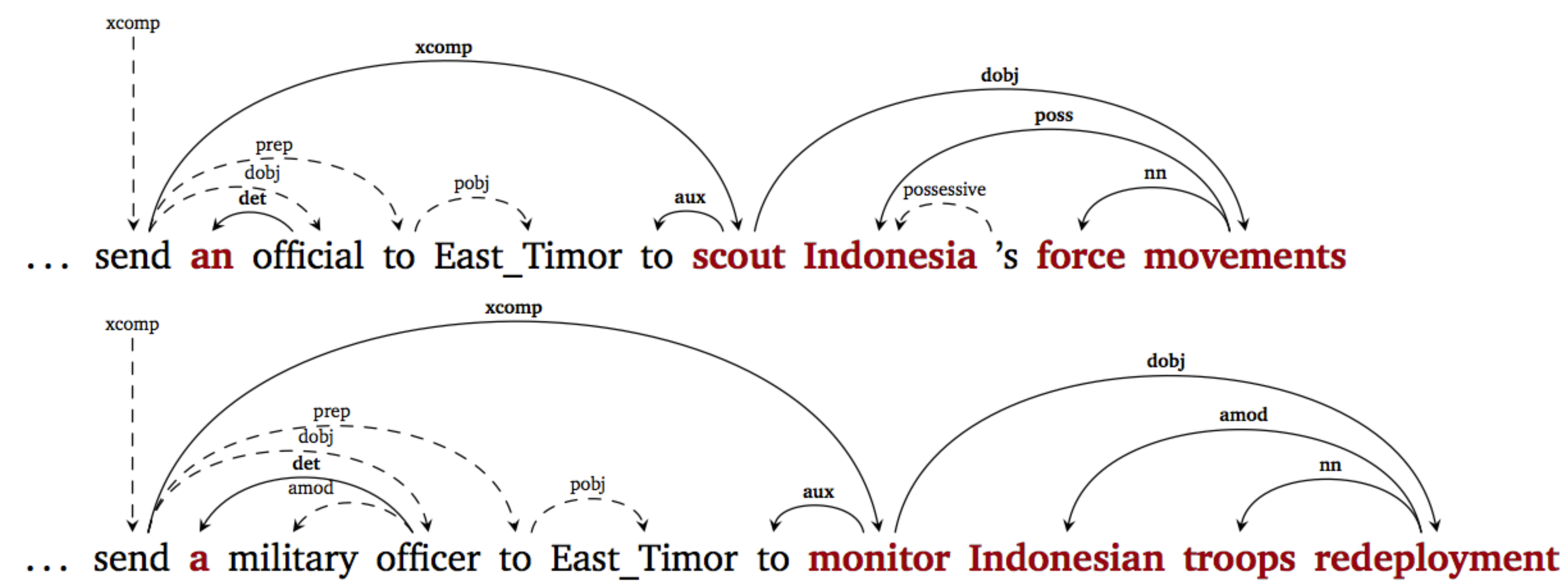


Challenges

- Phrase-based representations are natural but paraphrase recall is problematic
- ~65% of token alignments in Edinburgh training corpus supported by dependencies

Can we improve alignment using syntax?

Example: token alignments with syntactic support **missed** by the Meteor aligner (Denkowski & Lavie, 2011)

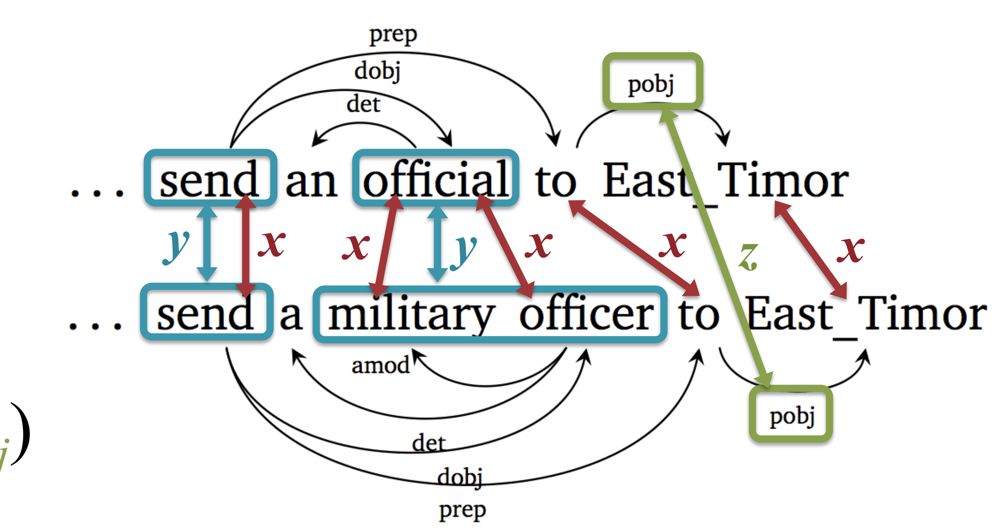


Approach

- Supervised structured prediction for pairwise text alignment
- Joint inference to **simultaneously** align phrases and dependency arcs
- Assumes the score of an alignment factors into scores for **phrase edits** and **arc edits**

Joint inference via ILP

Indicator variables for phrase edits y, arc edits z and token pair alignments x



Objective: $\max \sum_i y_i \cdot \mathbf{w}_{phr}^T \Phi(y_i) + \sum_j z_j \cdot \mathbf{w}_{arc}^T \Phi(z_j)$

subject to constraints:

- Only one active y per token, one active z per dependency
- Each active x participates in exactly one active y
- Each z activates two x to pair off its heads and dependents respectively and, conversely, is activated if these x are active

Learning

- Structured perceptron with averaging (Collins, 2002) to learn weights
- **Phrase edit** features $\Phi(y)$ from MANLI (MacCartney et al., 2008): phrase sizes, lexical + contextual similarity, relative positions, etc
- **Arc edit** features $\Phi(z)$ note the label category (e.g., *subj*) of the dependencies

Edinburgh Paraphrase Corpus

- Human-aligned paraphrase corpus (Cohn et al., 2008) from three sources:
  1. Multiple Translation Chinese corpus
  2. Multiple Jules Verne translations
  3. MSR paraphrase corpus
- Retokenized, truecased, NEs collapsed
- 715 training + 305 test instances (70:30 split)

http://www.ling.ohio-state.edu/~mwhite/data/coling12/

Baselines

1. **Meteor** (Denkowski & Lavie, 2011)
   - MT evaluation metric
   - best configuration over training set (*max-accuracy*)
2. **Phrase-based MANLI** (Thadani & McKeown, 2011)
   - Supervised aligner with ILP inference
   - Improvement on MANLI (MacCartney et al., 2008) which outperforms GIZA++ (Och & Ney 2003), HMMs (Liang et al., 2006) and Stanford's RTE aligner (Chambers et al., 2007)

Results (F_1)

Tokens	**Meteor**	**Phrase-based**	**Phrase+Arc**
SURE	75.49%	77.85%	**79.20**%
SURE + POSSIBLE	72.32%	75.37%	**77.57**%
Dependencies	**Meteor**	**Phrase-based**	**Phrase+Arc**
SURE	65.60%	75.10%	**76.30**%
SURE + POSSIBLE	62.57%	78.79%	**80.92**%

All gains significant under Wilcoxon's signed rank test

Funded by: IARPA, via DoI/NBC contract number D11PC20153
AFRL, subcontract to FA8750-09-C-0179