

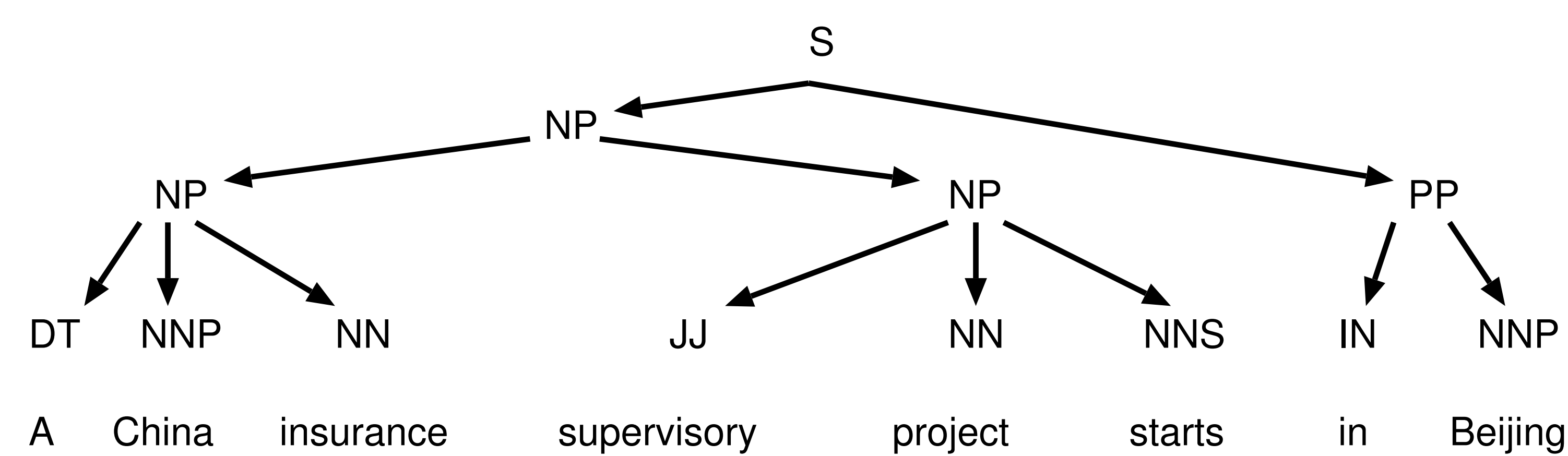
Introduction

- Using syntactic knowledge in hierarchical machine translation is reported to improve translation quality
- However, typically only one method is analyzed in each paper
- Here we compare and combine three different methods
 - Parse Matching (Vilar et al., 2008)
 - Soft Syntactic Labels (Venugopal et al., 2009)
 - String-to-Dependency (Shen et al. 2008)

Parse Matching (par)

up to +0.3 BLEU

- A lexical phrase extracted from a sentence is syntactically valid if it corresponds to the yield of a parse tree node
- Syntactically invalid phrases are penalized

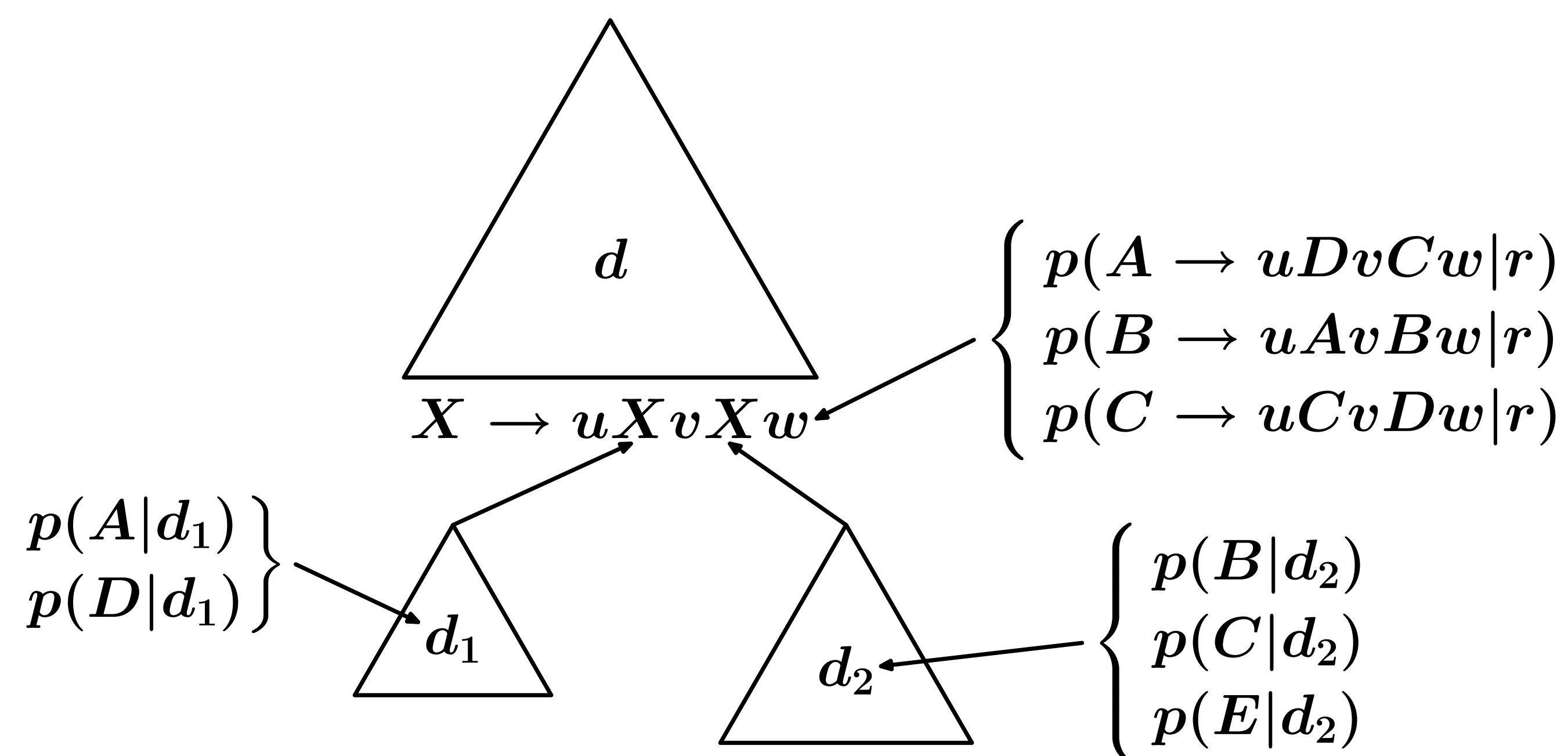


(Vilar et al., 2008)

Soft Syntactic Labels (syn)

up to +0.9 BLEU

- Labels of the parse trees are appended as additional information to the phrases
- A probability measuring the well-formedness of the translation parse tree is computed

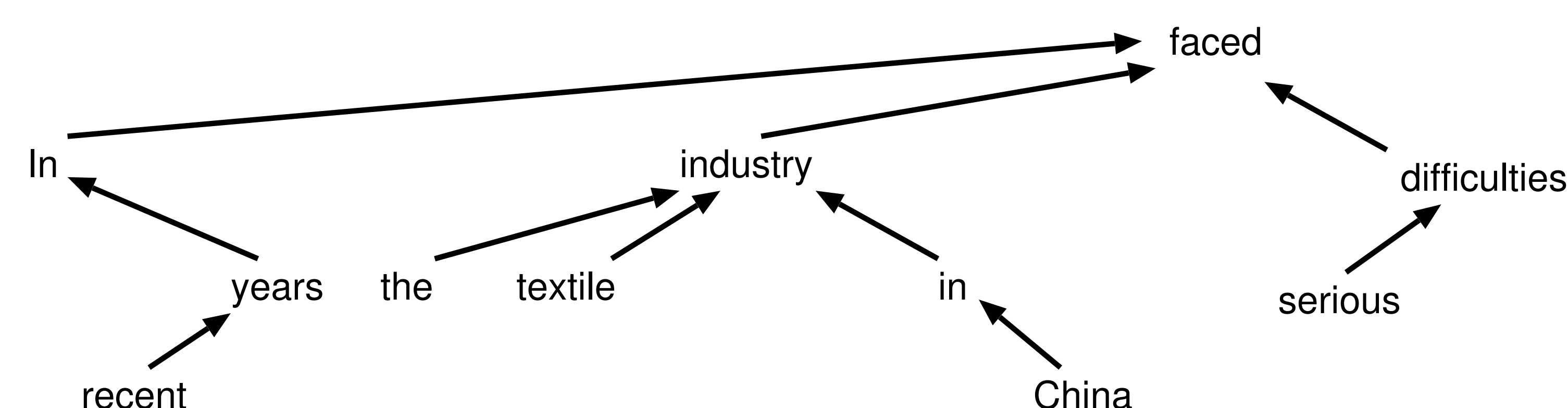


(Venugopal et al., 2009)

String-to-Dependency (dep)

up to +0.6 BLEU

- The dependency tree for the translation is reconstructed during search
- The well-formedness of the tree is measured applying dependency-level language models as rescoring



(Shen et al. 2008)

Data Used

- Medium-sized (3M sent.) training set from NIST Chinese-English
- NIST 2006 evaluation set as a development corpus
- NIST 2008 evaluation set as the blind test corpus

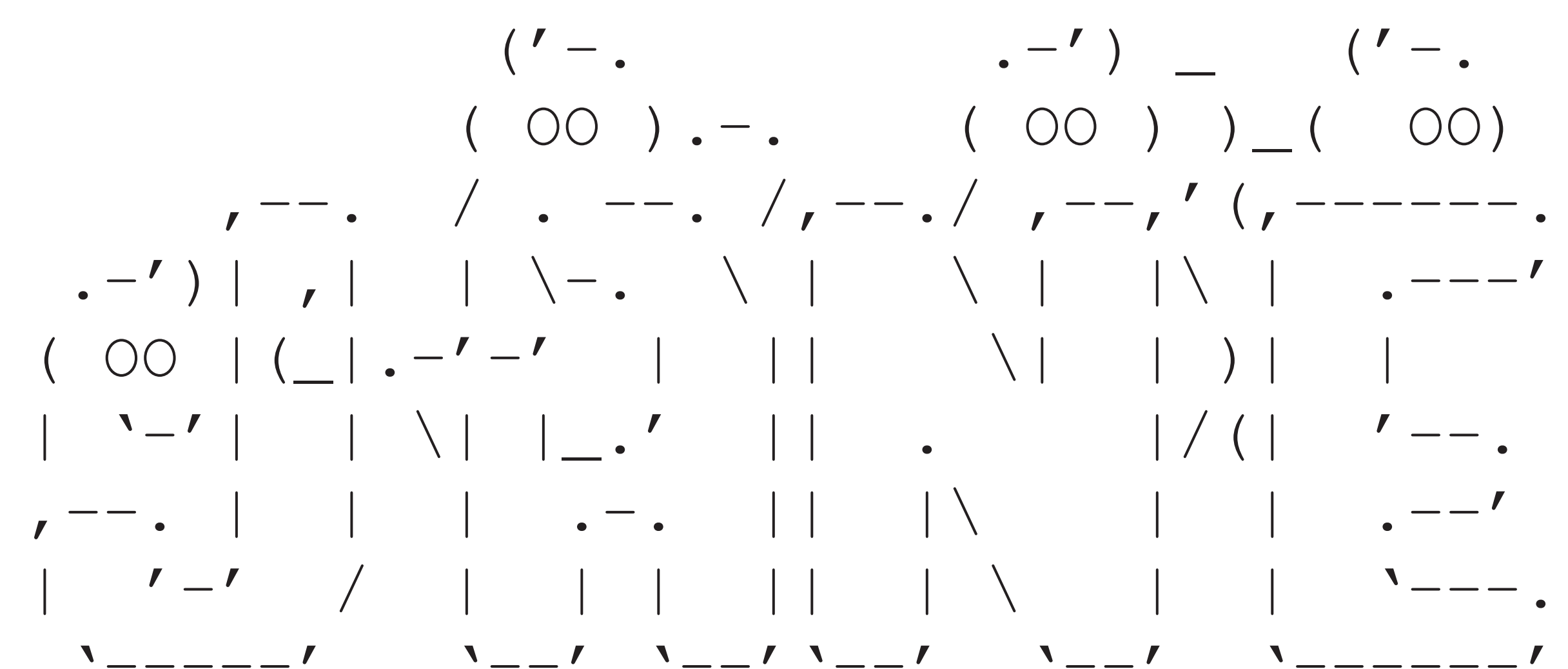
Experimental Results

	NIST '06 (dev)		NIST '08 (test)	
	BLEU[%]	TER[%]	BLEU[%]	TER[%]
baseline hierarchical	31.37	63.16	24.04	68.37
parsematch	31.37	63.09	24.36	67.89
dependency	32.23	61.93	24.61	66.69
syntax labels	32.22	62.12	24.99	67.15
par + dep	32.03	62.45	24.61	67.57
syn + par	32.36	62.27	25.31	67.28
syn + dep	32.90	61.39	25.39	66.66
all additional models	32.89	60.99	25.11	66.39

Translation Examples

reference	I hope that all of these worries are needless.
baseline	I hope that all these are worried that is.
par	We hope that all concerned that this is groundless.
dep	We hope that all this worry is groundless.
syn	I hope that all this concern is that.
par + dep	We hope that all these are worried about is the.
syn + par	I hope that all this worry is groundless.
syn + dep	I hope that all this worry is unnecessary.
all	I hope that all this worry is superfluous.

Commercial Break



- RWTH's hierarchical phrase-based translation toolkit
- Released as open source software
- Implemented in C++
- Applied in different MT evaluations (WMT, Gale, IWSLT, ...)
- <http://www.hltpr.rwth-aachen.de/jane>

(Vilar et al., 2010)

Conclusion

- The above mentioned methods seem to complement each other
- Their combination achieves an increase of 1% in BLEU and a reduction of nearly 2% in TER
- All methods are released as open-source within the Jane framework