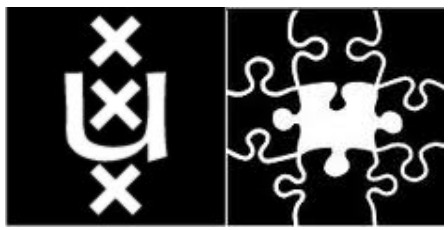
Modelling the Adjunct/Argument Distinction in Hierarchical Phrase-Based Translation

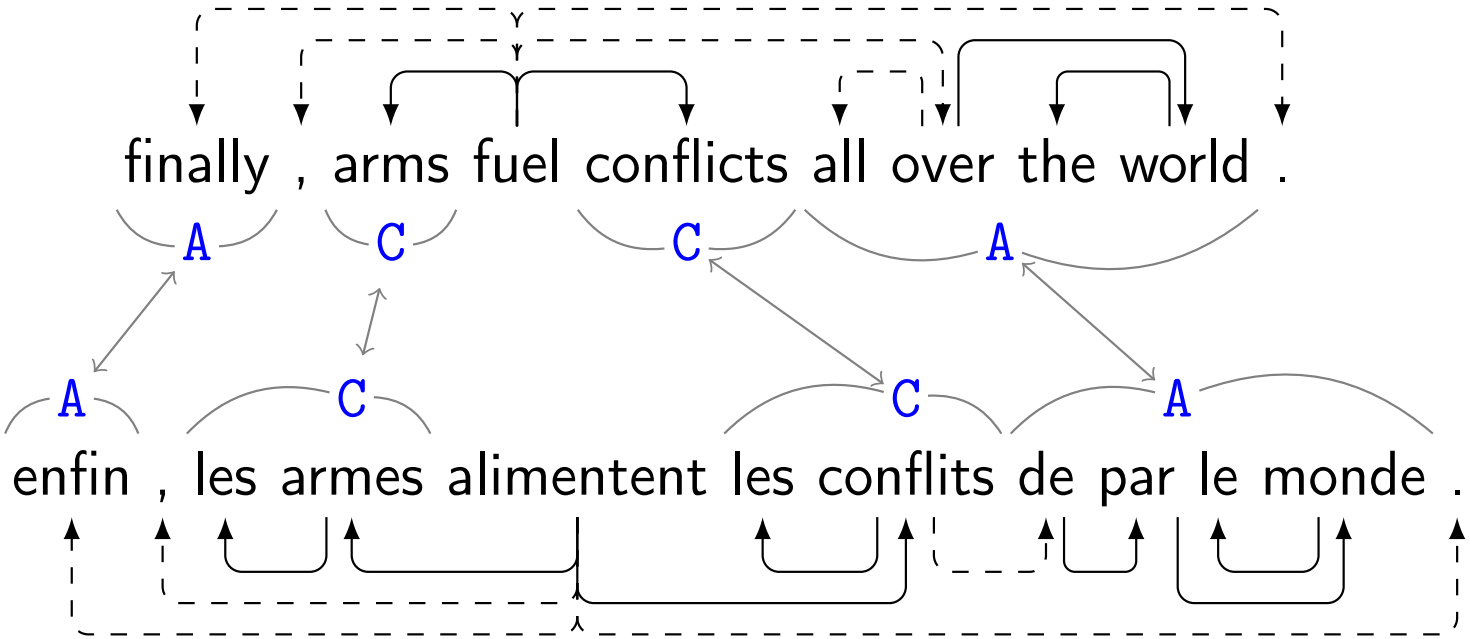
Sophie Arnoult and Khalil Sima'an



Institute for Logic, Language and Computation
University of Amsterdam

Deep Machine Translation Workshop, September 4, 2015

The Adjunct/Argument Distinction for Hiero



Minimally explain recursion in Hiero

- distinction is semantically driven
- adjunction is a central device for recursion

Interpretation of the Adjunct/Argument Distinction

A restrictive interpretation of the adjunct/argument distinction

- not modelling selectional preferences as in STAG
- adjuncts and arguments are interpreted as types in SCFG

Interpretation of adjuncts

- adjuncts as modifiers
- not only in semantic frames

Model

- Syntax-Augmented Machine Translation (SAMT)
  - labelled Hiero model
  - phrase labels derived from syntactic annotations through combinatory rules
- unlike SAMT
  - minimal labels
  - bilingual source/target annotations
- phrase-length constraint (10 tokens)
  - no labelled reordering at sentence level

Labelling procedure

Procedure

- adjunct/argument labels
- combinatory rules for phrase labels
- (bilingual) phrase-pair labels

Adjunct/Argument labels

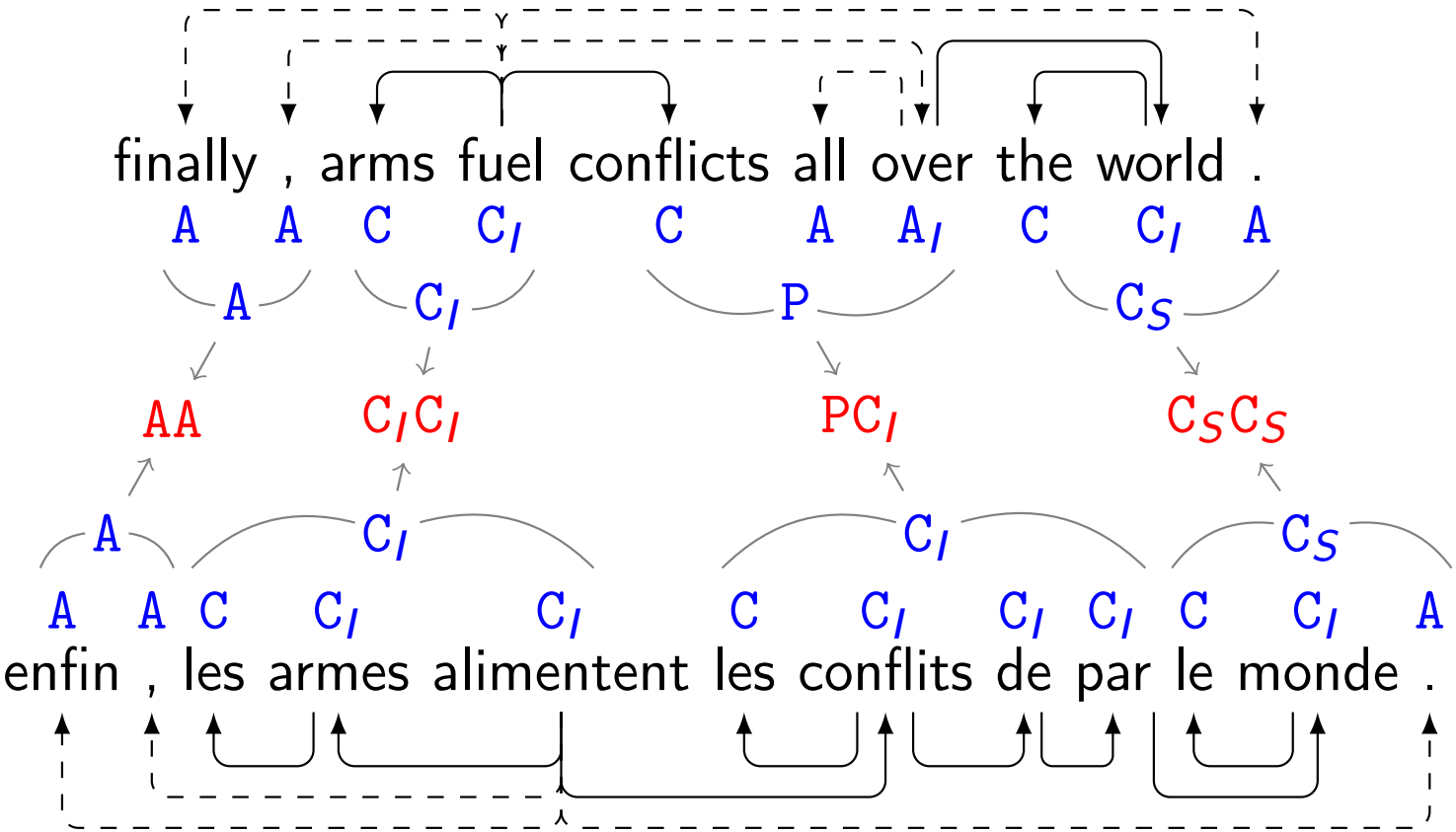
- use dependency annotations
- map *modifier* and *punctuation* labels to adjuncts

Combinatory Rules for Phrase Labels

- derive phrase labels from adjunct (A) and argument (C) labels
- SAMT-like combinatory rules
- extension is minimal and reflects characteristics of adjunction

	phrase type	*resulting label*
if	constituent	A or C
else if	constituent sequence	
	if all adjuncts	A
	else	C_S
else if	const. less subconstituents	
	if all adjuncts	A or C
	else	A_I or C_I
else		P

Labelled Models



First Results

- French-English Europarl
- in-domain LM data, dev/test sets
- training with 200k sentence pairs

	labels	BLEU		METEOR		TER	
		dev	test	dev	test	dev	test
Hiero	1	32.1	**31.8**	34.9	**34.8**	52.9	**53.3**
AA-Src	6	31.9$^{\nabla\nabla}$	31.3$^{\nabla\nabla}$	34.8$^{\nabla}$	34.7$^{\nabla\nabla}$	53.0	53.5$^{\nabla\nabla}$
AA-Trg	6	32.0$^{\nabla}$	31.6$^{\nabla\nabla}$	**34.9**	34.7$^{\nabla}$	**52.9**	53.5$^{\nabla\nabla}$
AA-Bi	36	31.9$^{\nabla}$	31.5$^{\nabla\nabla}$	34.8	34.7$^{\nabla\nabla}$	53.0	53.5$^{\nabla}$

Relabelling by Clustering

- compare labels according to their *lhs/rhs* behaviour
- two-component distance
  - lhs distance

$$d_{LHS} = \Sigma_{RHS} |\Delta_{LHS} P(rhs|lhs)|$$

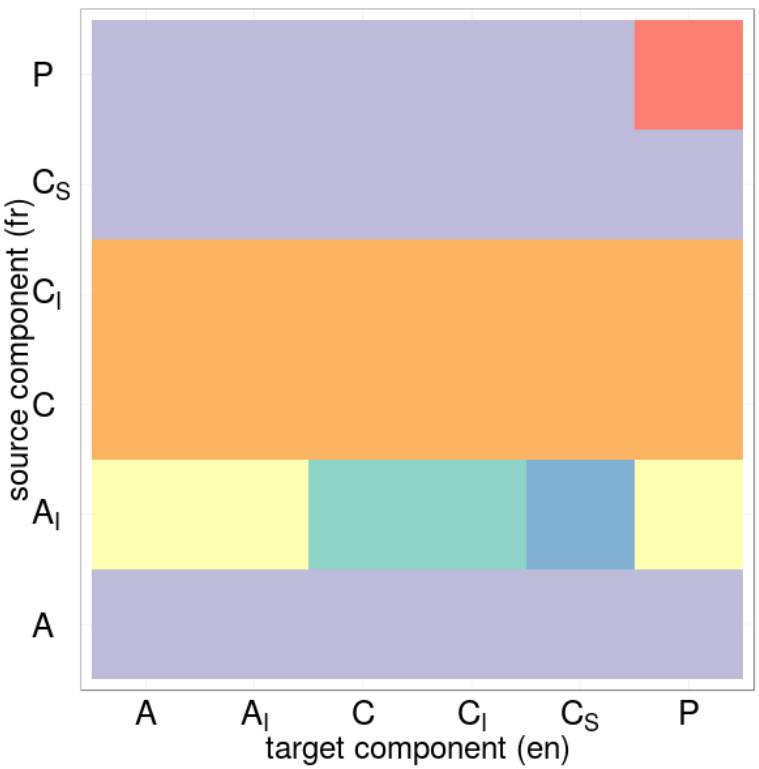
  - rhs distance

$$d^{cond}_{RHS} = \Sigma_{LHS} |\Delta_{RHS} P(lhs|rhs)|$$
$$d^{joint}_{RHS} = \Sigma_{LHS} |\Delta_{RHS} P(lhs, rhs)|$$

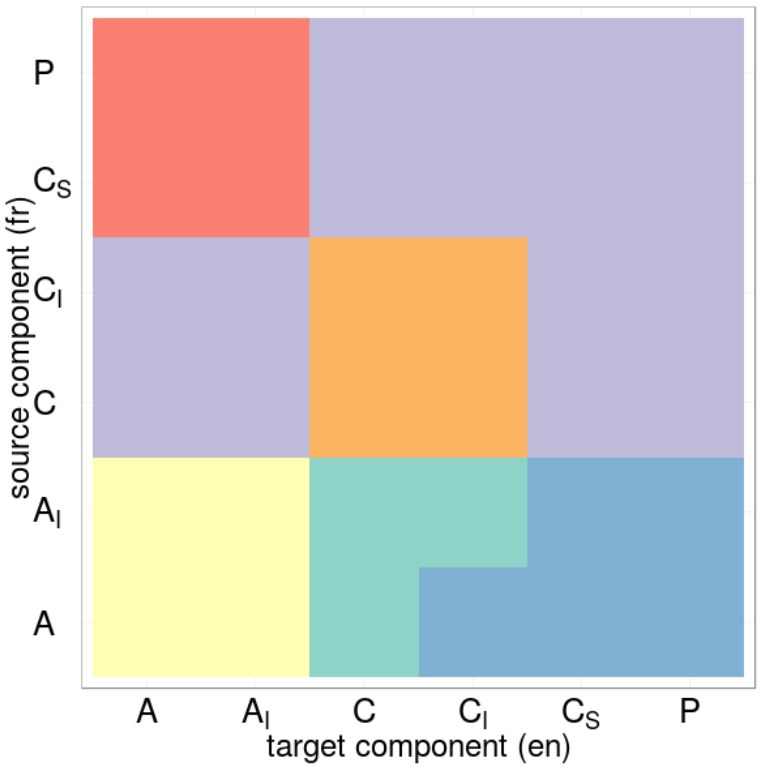
- probabilities estimated from the dev-set AA-Bi grammar
- clustering stops at six clusters

Label Clusters

$d_{LHS} + d^{cond}_{RHS}$



$d_{LHS} + d^{joint}_{RHS}$



Results with Clustered Labels (1)

	labels	BLEU		METEOR		TER	
		dev	test	dev	test	dev	test
AA-Bi	36	**31.9**	31.5	34.8	34.7	**53.0**	53.5
Cl-cond	6	31.8$^{\blacktriangledown}$	31.4	34.8	34.7	53.1	53.6
Cl-joint	6	**31.9**	**31.8**$^{\blacktriangle\blacktriangle}$	**34.9**	**34.8**$^{\blacktriangle\blacktriangle}$	**53.0**	**53.3**$^{\blacktriangle\blacktriangle}$

Results with Clustered Labels (2)

	labels	BLEU		METEOR		TER	
		dev	test	dev	test	dev	test
Hiero	1	32.1	**31.8**	34.9	**34.8**	52.9	**53.3**
Cl-cond	6	31.8$^{\triangledown\triangledown}$	31.4$^{\triangledown\triangledown}$	34.8	34.7$^{\triangledown}$	53.1$^{\triangledown\triangledown}$	53.6$^{\triangledown\triangledown}$
Cl-joint	6	31.9$^{\triangledown\triangledown}$	**31.8**	34.9	**34.8**	53.0	**53.3**

Future Work

- better method to reshape the bilingual-label set
  - clustering works, but only allows merging
- lift phrase-length constraint
  - reordering rules
  - swap for recursion constraint
- extend experimental set-up
  - other language pairs

Thank you.