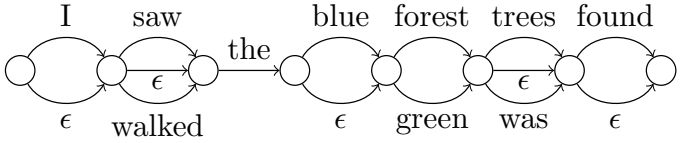
Machine Translation System Combination by Confusion Forest

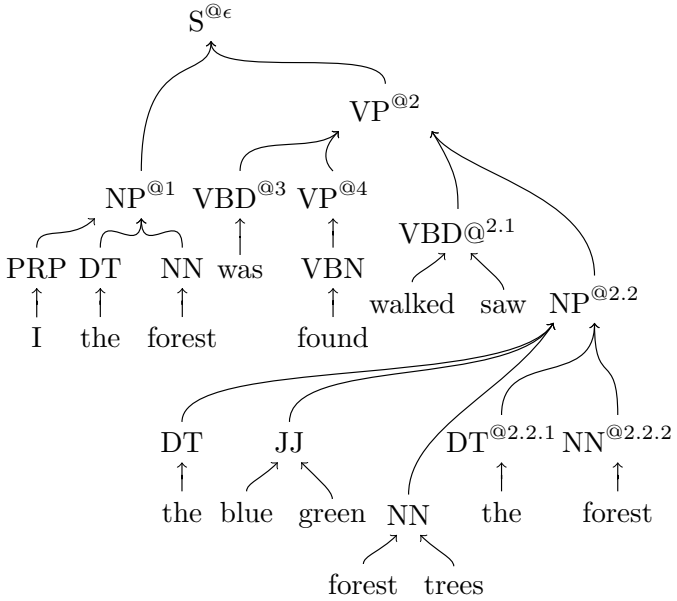
Taro Watanabe and Eiichiro Sumita
National Institute of Information and Communication Technology (NICT)
{taro.watanabe, eiichiro.sumita}@nict.go.jp

Confusion Network



- Multiple hypotheses encoded as a lattice structure (Bangalore et al., 2001)
- The best path = the best translation
- Problems:
  - Handling syntactically different translations.
  - Spurious repetitions/insertions due to alignment error.
  - Partial Solution: incremental network construction + multiple skeletons (Rosti et al., 2007)

Confusion Forest



- Multiple hypotheses encoded as a forest structure, or hypergraph (Billot and Lang, 1989, Mi et al., 2008)
- The best derivation = the best translation
- Represents a syntactic consensus by sharing hyperedges.

Forest Construction

- A grammar based approach:
  1. Parse system hypotheses.
  2. Learn a CFG by extracting a set of rules that constitute the parsed trees.
  3. Forest generation from the CFG using the Earley's algorithm through non-terminal rewrites.

Scan:

$$\frac{[\mathrm{X} \to \alpha \bullet x\beta, h] : u}{[\mathrm{X} \to \alpha x \bullet \beta, h] : u}$$

Predict:

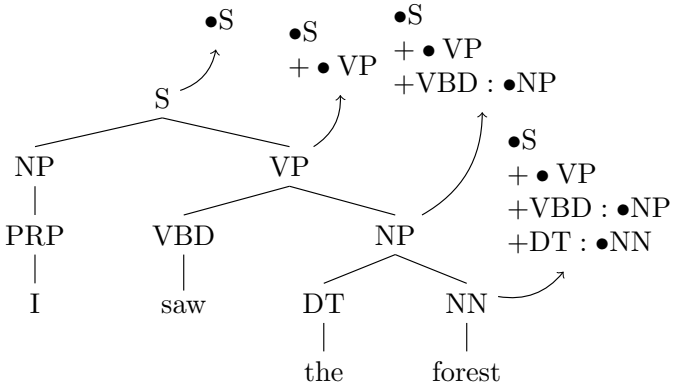
$$\frac{[\mathrm{X} \to \alpha \bullet \mathrm{Y}\beta, h]}{[\mathrm{Y} \to \bullet\gamma, h+1] : u} \quad \mathrm{Y} \overset{u}{\to} \gamma \in \mathcal{G}, h < H$$

Complete:

$$\frac{[\mathrm{X} \to \alpha \bullet \mathrm{Y}\beta, h] : u \quad [\mathrm{Y} \to \gamma\bullet, h+1] : v}{[\mathrm{X} \to \alpha\mathrm{Y} \bullet \beta, h] : u \otimes v}$$

Ambiguous Grammar

- Replace each non-terminal symbol in the parsed tree by the state representation of Earley's algorithm, then extract a CFG.
- Limit the labels by its vertical/horizontal orders for better generalizations (Klein and Manning, 2003).



Experiments

- WMT10 shared task: {cz,de,es,fr}-to-en
- Implemented in "cicada": a hypergraph toolkit based on a semiring parsing framework
- CN: transformed into a forest by lattice parsing with a monotone grammar
- CF: Stanford parser followed by the rule extraction with v={3,4,∞} and h={1,2, ∞}
- The same feature set for CN and CF

BLEU

language	cz-en	de-en	es-en	fr-en
system min	14.09	15.62	21.79	16.79
max	23.44	24.10	29.97	**29.17**
CN	23.70	24.09	**30.45**	**29.15**
$\mathrm{CF}_{v=\infty,h=\infty}$	**24.13**	24.18	**30.41**	**29.57**
$\mathrm{CF}_{v=\infty,h=2}$	**24.14**	**24.58**	**30.52**	28.84
$\mathrm{CF}_{v=\infty,h=1}$	**24.01**	23.91	**30.46**	**29.32**
$\mathrm{CF}_{v=4,h=\infty}$	**23.93**	23.57	29.88	28.71
$\mathrm{CF}_{v=4,h=2}$	**23.82**	22.68	29.92	28.83
$\mathrm{CF}_{v=4,h=1}$	**23.77**	21.42	30.10	28.32
$\mathrm{CF}_{v=3,h=\infty}$	23.38	23.34	29.81	27.34
$\mathrm{CF}_{v=3,h=2}$	23.30	23.95	30.02	28.19
$\mathrm{CF}_{v=3,h=1}$	23.23	21.43	29.27	26.53

Oracle BLEU

language	cz-en	de-en	es-en	fr-en
rerank	29.40	32.32	36.83	36.59
CN	38.52	34.97	47.65	46.37
$\mathrm{CF}_{v=\infty,h=1}$	31.09	34.65	39.27	39.51
$\mathrm{CF}_{v=4,h=1}$	31.44	34.62	39.69	39.90
$\mathrm{CF}_{v=3,h=1}$	31.55	34.60	39.72	39.97

Hypergraph size

language	cz-en	de-en	es-en	fr-en
CN	2,222.68	47,231.20	2,932.24	11,969.40
$\mathrm{CF}_{v=\infty,h=1}$	230.08	540.03	262.30	386.79
$\mathrm{CF}_{v=4,h=1}$	254.45	651.10	302.01	477.51
$\mathrm{CF}_{v=3,h=1}$	286.01	802.79	349.21	575.17