Compositions of Tree-to-Tree Statistical Machine Translation Models

Andreas Maletti

Institute for Natural Language Processing
University of Stuttgart, Germany

Montréal — July 28, 2016

Statistical machine translation

Required resource

- Parallel corpus (containing sentences in both languages)

Additional resources

- Language model data (text in the target language)
- Word alignments
- Parse trees
- Various (morphological analyzers, parsers)

Statistical machine translation

Assume translation direction  $\rightarrow$ 
(English to Catalan)

but no parallel corpus for this pair

Pivot approach

but parallel corpus for  $\rightarrow$  $\rightarrow$ 
(parallel corpus for English-Spanish and Spanish-Catalan)

(similar motivation for non-deterministic pre- or post-processing steps)

Statistical machine translation

Consequences

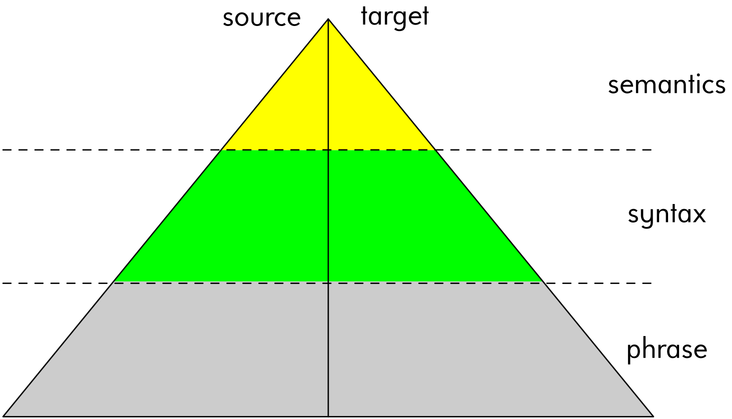
- 2 translation models (operating in sequence)
- Inefficiencies in sequential operation (due to sequential pruning)
- Theoretical guarantees missing for many operations (e.g., tuning)

Remedies

- Partial Evaluation [MayKniVog10]
- Composition

Statistical machine translation

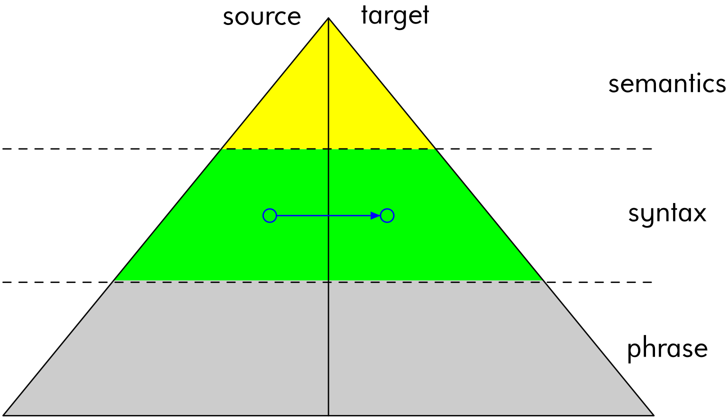
Vauquois triangle:



Translation model:

Statistical machine translation

Vauquois triangle:



Translation model: tree-to-tree

Overview

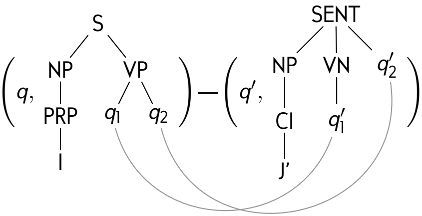
1. Background
2. Theoretical Model
3. Unweighted Compositions
4. Weighted Compositions

Extended tree transducers

Definition

Transducer $M = (Q, \Sigma, (q_1, q_2), R, \mathrm{wt})$ with

- finite set Q of states
- finite set Σ of symbols
- source initial state q_1 and target initial state q_2
- finite set R of rules (see below)
- weight assignment $\mathrm{wt} \colon R \to A$ (into a semiring)

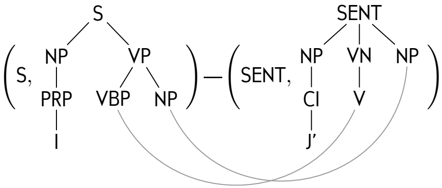


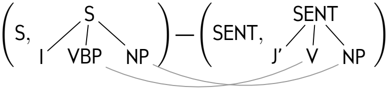
Extended tree transducers

Restrictions

Transducer $M = (Q, \Sigma, (q_1, q_2), R, \mathrm{wt})$ is

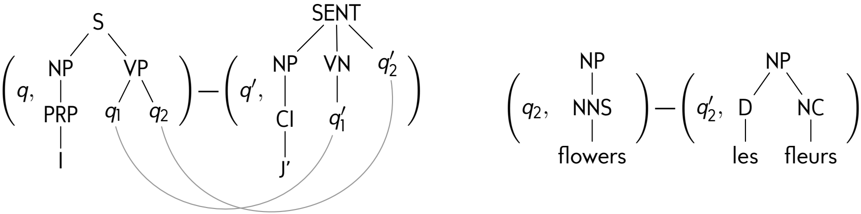
- STSG if $Q = \Sigma$ and state = root label
- SCFG if STSG and all rules are shallow
- ε-free if no left-hand side is in Q
- strict if no right-hand side is in Q
- simple if ε-free and strict



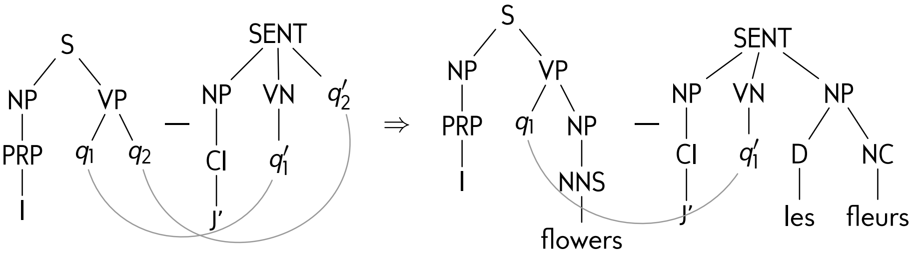


Extended tree transducers

Rules:



Use in a derivation step:



Extended tree transducers

Initial sentential form:

$$q_1 \quad \text{—} \quad q_2$$

Final sentential form: No more linked states

Weight

- of a derivation: product of the weights of the used rules
- of a tree pair: sum of all derivations for the pair

Extended tree transducers

Composition

Given two weighted relations $\tau \colon T_\Sigma \times T_\Delta \to A$ and $\tau' \colon T_\Delta \times T_\Gamma \to A$

$$(\tau \,;\, \tau')(t, s) = \sum_{u \in T_\Delta} \tau(t, u) \cdot \tau'(u, s)$$

Overview

1. Background
2. Theoretical Model
3. Unweighted Compositions
4. Weighted Compositions

Known results

$$\mathrm{SCFG} \subsetneq \mathrm{STSG} \subsetneq \mathrm{TT}$$

[Eis03] and folklore

Model	Composition closure	Reference
top-down transducer	1	[Eng75]
simple transducer	2	[ArnDau82]
other transducer	∞	[EngFulMal16]

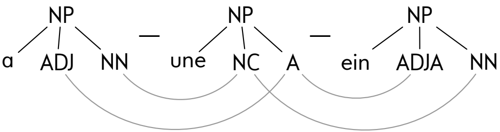
(top-down tree transducer = all left-hand sides shallow)

First result

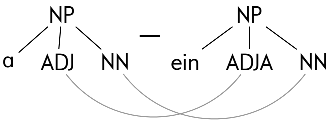
Theorem

SCFG are closed under composition

Matching original rules

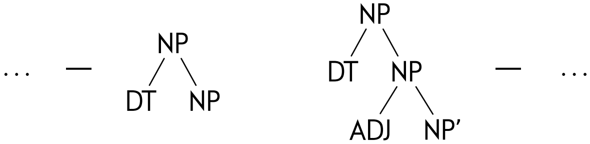


Newly constructed rule

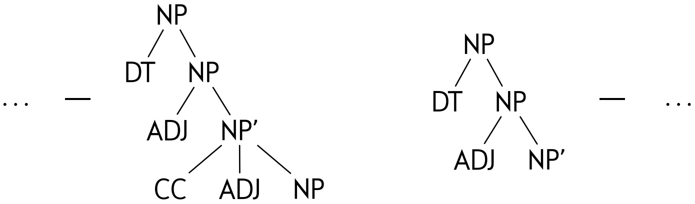


Second result

Difficulty for STSGs:



Extending the rules does not help:

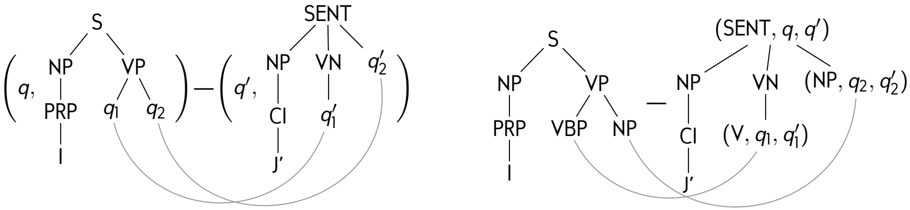


Second result

Theorem

Compositions of $n \geq 2$ simple STSGs are as expressive as compositions of n simple transducers

Proof idea: Encode finite states in the intermediate tree(s)



Second result

Corollary

The composition closure for simple STSGs is obtained at level 2

Proof idea:

$$\text{s-STSG} \subsetneq \text{s-TT} \subsetneq \text{s-TT}^2 = \text{s-TT}^3 = \text{s-STSG}^3 = \text{s-STSG}^2$$

Third result

Theorem

The composition hierarchy for the remaining STSGs is infinite

Proof idea: Inspect the corresponding proof for transducers and realize that the counterexamples can be generated by STSGs

Overview

1 Background

2 Theoretical Model

3 Unweighted Compositions

4 Weighted Compositions

Known results and main lemma

$\text{s-wTT} ; \text{wREL} \subseteq \text{s-wTT}$ $\qquad$ $\text{wREL} ; \text{s-wTT} \subseteq \text{s-wTT}$

follows from [FulMalVog11] and [Kui99]

Lemma

$$\text{s-wTT} \subseteq \text{su-TT}_{\text{inj}} ; \text{wREL}$$

Proof idea: Simply annotate rule identifier of applied rule to output and apply weights using the relabeling

($\text{su-TT}_{\text{inj}}$ = injective relations computed by simple unamb. transducers)

First weighted result

Theorem

The composition closure of weighted simple STSGs is obtained at level 2

Proof idea:

$$
\begin{aligned}
&\text{s-wTT} ; \text{s-wTT}^2 && \subseteq \text{su-TT}_{\text{inj}} ; \text{wREL} ; \text{s-wTT}^2 \\
\subseteq\ &\text{su-TT}_{\text{inj}} ; \text{s-wTT}^2 && \subseteq \text{su-TT}^2_{\text{inj}} ; \text{wREL} ; \text{s-wTT} \\
\subseteq\ &\text{su-TT}^2_{\text{inj}} ; \text{s-wTT} && \subseteq \text{su-TT}^3_{\text{inj}} ; \text{wREL} \\
\subseteq\ &\underbrace{\text{s-TT}^2}_{\text{injective}} ; \text{wREL} && \subseteq \text{su-TT}^2_{\text{inj}} ; \text{wREL} \subseteq \text{s-wTT}^2
\end{aligned}
$$

Second weighted result

Theorem

The composition hierarchy for the remaining weighted STSGs is infinite

Proof idea: Utilize linking technique of [Mal15] to lift the corresponding unweighted result

Summary

Model	Composition closure
(weighted) SCFGs	1
(weighted) simple STSGs	2
(weighted) other STSGs	∞

Literature

- May, Knight, Vogler: Efficient inference through cascades of weighted tree transducers. ACL 2010
- Eisner: Learning non-isomorphic tree mappings for machine translation. ACL 2003
- Engelfriet: Bottom-up and top-down tree transformations: a comparison. Math. Syst. Theor., 1975
- Arnold, Dauchet: Morphismes et bimorphismes d'arbres. Theor. Comput. Sci. 1982
- Engelfriet, Fülöp, M.: Composition closure of linear extended top-down tree transducers. Theor. Comput. Syst. 2016
- Fülöp, M., Vogler: Weighted extended tree transducers. Fundam. Informaticae 2011
- Kuich: Full abstract families of tree series I. In: Jewels are Forever, Springer 1999
- M.: The power of weighted regularity-preserving multi bottom-up tree transducers. Int. J. Found. Comput. Sci. 2015