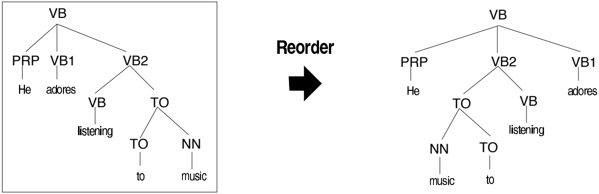
A Syntax-based Statistical Translation Model

Kenji Yamada Kevin Knight

February 7, 2006

Principal purpose of the natural language which recently machine translation and processes met to the success which is limited. Because most machine translations require pre-edit and the second volume collection stage, inclusion of the human of type is included.
– Uncyclopedia.com

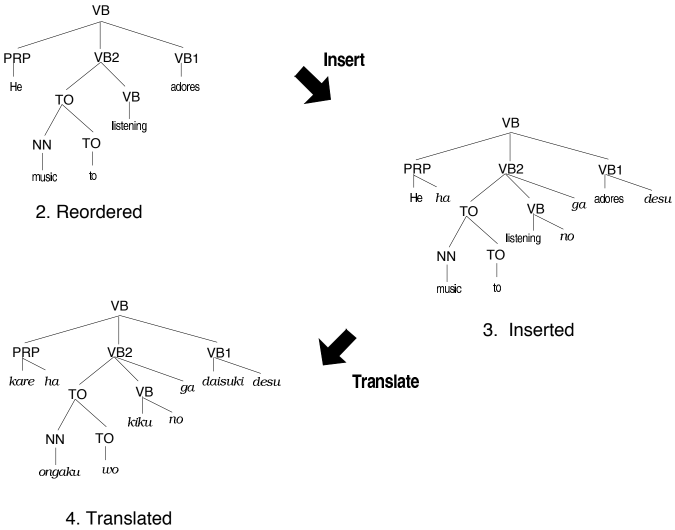
From Tree to Tree′



1. Channel Input

2. Reordered

From Tree′ to Pulp

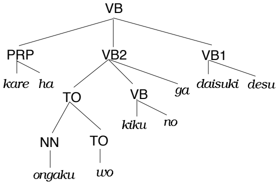


From Pulp to Text

Reading off Leaves

kare ha ongaku wo kiku no ga daisuki desu

5. Channel Output



4. Translated

Definitions

$\boldsymbol{\mathcal{E}}$ = English Tree **f** = French String

- Operations
  N = Insertion **R** = Reorder **T** = Translation
- Operation Variables
  ν = Insertion ρ = Reorder τ = Translation
- Features
  $\mathcal{N}$ = Insertion $\mathcal{R}$ = Reorder $\mathcal{T}$ = Translation
- Feature Variables
  $\mathcal{N}$ = Insertion $\mathcal{R}$ = Reorder $\mathcal{T}$ = Translation
- Probablities
  $n(\nu|\mathcal{N})$ = Insertion $r(\rho|\mathcal{R})$ = Reorder $t(\tau|\mathcal{T})$ = Translation

$\theta =< \nu, \rho, \tau >$ is a set of values of $< \mathbf{N}, \mathbf{R}, \mathbf{T} >$.
$\boldsymbol{\theta} = \theta_1, \theta_2, \ldots, \theta_n$ for a parse tree $\boldsymbol{\mathcal{E}} = \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$.

Formal Description

- Propability of a French sentence given an English parse tree:

$$\mathrm{P}(\mathbf{f}|\boldsymbol{\mathcal{E}}) = \sum_{\boldsymbol{\theta}:\mathrm{Str}(\boldsymbol{\theta}(\boldsymbol{\mathcal{E}}))=\mathbf{f}} \mathrm{P}(\boldsymbol{\theta}|\boldsymbol{\mathcal{E}})$$

- Probablity of a particular sequence of operations:

$$\begin{aligned}\mathrm{P}(\boldsymbol{\theta}|\boldsymbol{\mathcal{E}}) &= \mathrm{P}(\theta_1, \theta_2, \ldots, \theta_n|\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) \\ &= \prod_{i=1}^{n} \mathrm{P}(\theta_i|\theta_1, \theta_2, \ldots, \theta_{i-1}, \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)\end{aligned}$$

- Assuming each operation is independent yields:

$$\begin{aligned}\mathrm{P}(\boldsymbol{\theta}|\boldsymbol{\mathcal{E}}) &= \mathrm{P}(\theta_1, \theta_2, \ldots, \theta_n|\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) \\ &= \prod_{i=1}^{n} \mathrm{P}(\theta_i|\varepsilon_i)\end{aligned}$$

Formal Description

- Assuming the operations are independent, we get:

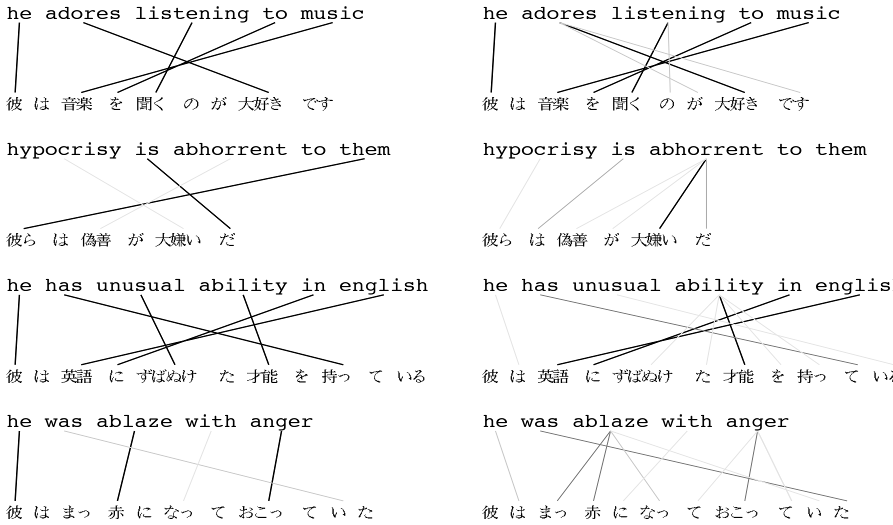
$$\begin{aligned}
\mathrm{P}(\theta_i|\varepsilon_i) &= \mathrm{P}(\nu_i, \rho_i, \tau_i|\varepsilon_i) \\
&= \mathrm{P}(\nu_i|\varepsilon_i)\mathrm{P}(\rho_i|\varepsilon_i)\mathrm{P}(\tau_i|\varepsilon_i) \\
&= \mathrm{P}(\nu_i|\mathcal{N}(\varepsilon_i))\mathrm{P}(\rho_i|\mathcal{R}(\varepsilon_i))\mathrm{P}(\tau_i|\mathcal{T}(\varepsilon_i)) \\
&= n(\nu_i|\mathcal{N}(\varepsilon_i))r(\rho_i|\mathcal{R}(\varepsilon_i))t(\tau_i|\mathcal{T}(\varepsilon_i))
\end{aligned}$$

- In Summary:

$$\mathrm{P}(\mathbf{f}|\boldsymbol{\mathcal{E}}) = \sum_{\boldsymbol{\theta}:\mathrm{Str}(\boldsymbol{\theta}(\boldsymbol{\mathcal{E}}))=\mathbf{f}} \mathrm{P}(\boldsymbol{\theta}|\boldsymbol{\mathcal{E}})$$

$$= \sum_{\boldsymbol{\theta}:\mathrm{Str}(\boldsymbol{\theta}(\boldsymbol{\mathcal{E}}))=\mathbf{f}} \prod_{i=1}^{n} n(\nu_i|\mathcal{N}(\varepsilon_i))r(\rho_i|\mathcal{R}(\varepsilon_i))t(\tau_i|\mathcal{T}(\varepsilon_i))$$

Comparison to IBM Model 5



Comparison to IBM Model 5

	Align. score	Perfect sents	Perplexity
YK Model	0.582	10	15.79
IBM Model 5	0.431	0	9.84