

Traducción Automática

Trabajo 2

Ramon Ruiz Dolz

January 2019

1 Tree to string translation

- 1.1 (***) Define a small model: a parsing tree, a r-table, an n-table and a t-table. Then, apply the estimation algorithm described in slide 15. Keep in mind that the models should be small enough to make the computations easy. Therefore the number of posible transformation should keep very small.

Para la realización de este ejercicio se ha considerado una traducción de español a japonés. Concretamente se ha trabajado con la frase "Me gusta ver películas" que en japonés se traduciría como "Watashi wa eiga wo mirukoto ga suki desu". Para ello partimos de la siguiente gramática,

A ->B C D	0.95
A ->B B D	0.05
D ->E F	0.9
D ->C F	0.1
B ->"me"	0.8
B ->"gusta"	0.2
C ->"gusta"	0.7
C ->"ver"	0.3
E ->"ver"	1
F ->"películas"	1

Table 1: Gramática empleada con sus probabilidades asociadas

También se han definido las siguientes tablas de reordenamiento, inserción y traducción.

B C D	B C D	0.0375
	B D C	0.8125
	C B D	0.0325
	C D B	0.0425
	D B C	0.035
	D C B	0.04
BBD	BBD	0.12
	BDB	0.83
	DBB	0.05
E F	E F	0.08
	F E	0.92
CF	CF	0.11
	FC	0.89

Table 2: r-table

parent	TOP	A	A	A	D	D
node	A	B	C	D	E	F
P(none)	0.8	0.1	0.03	0.87	0.15	0.2
P(left)	0.05	0.15	0.07	0.07	0.05	0.15
P(right)	0.15	0.75	0.9	0.06	0.8	0.65

Table 3: n-table (nodes)

word	P(ins)
wa	0.15
wo	0.1
ga	0.45
desu	0.2
to	0.07
ni	0.03

Table 4: n-table (words)

me	gusta	ver	películas
watashi (0.82)	suki (0.84)	mirukoto (0.92)	eiga (0.87)
anata (0.06)	daisuki (0.16)	kiku (0.08)	hon (0.1)
boku (0.12)			NULL (0.03)

Table 5: t-table

Con la gramática, los reordenamientos, inserciones y traducciones definidas se llegan a generar cuatro árboles distintos, como podemos observar en las Figure 1, Figure 2, Figure 3, Figure 4.

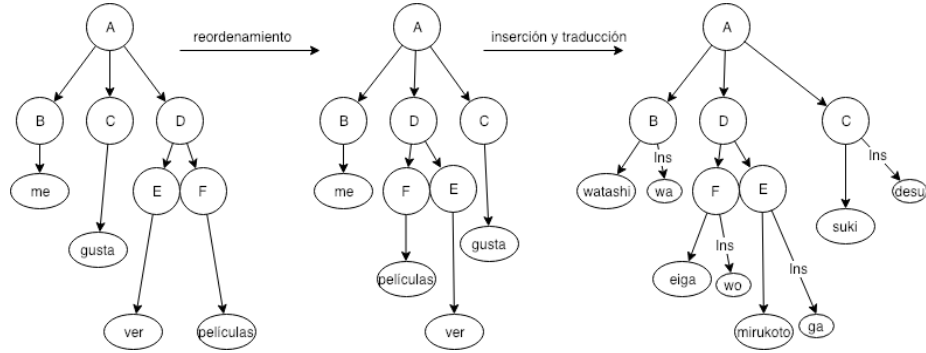


Figure 1: Árbol 1 generado por la gramática

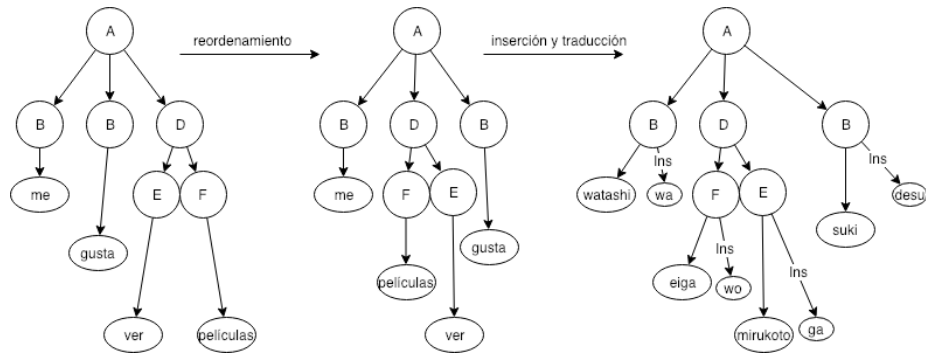


Figure 2: Árbol 2 generado por la gramática

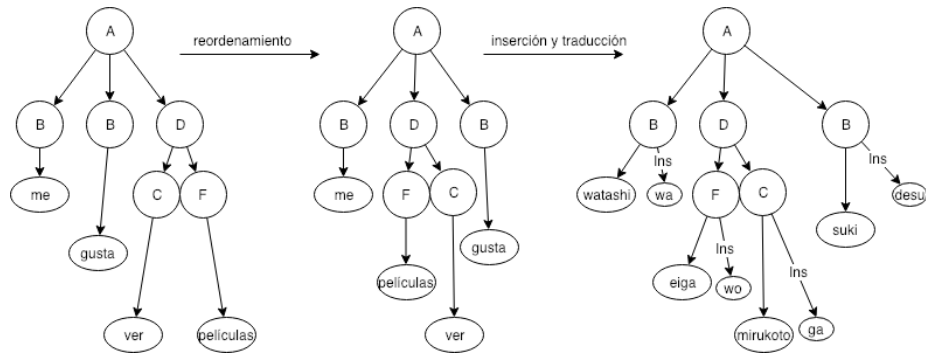


Figure 3: Árbol 3 generado por la gramática

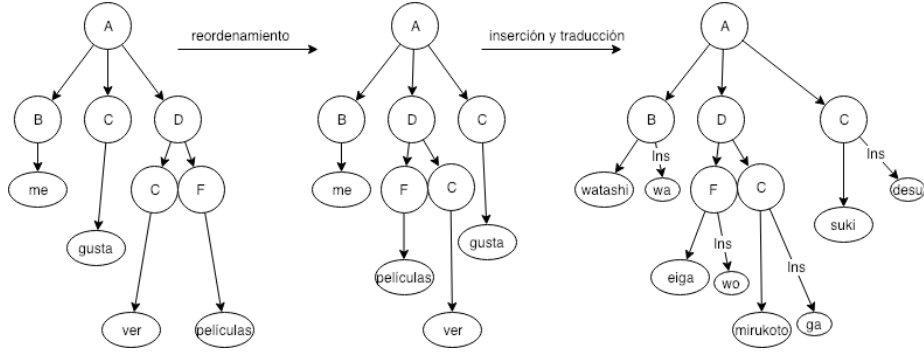


Figure 4: Árbol 4 generado por la gramática

Finalmente procederemos a realizar la estimación del parámetro correspondiente con la regla (C -> "gusta"). Para ello se hace uso de la siguiente fórmula (diapositiva 15),

$$\bar{\theta} = \frac{\sum_{f,E:L(E)=f} \frac{1}{P(f,E)} \sum_{\theta:L(\theta(E)=f} N(n_i, \theta) \prod_{i=1}^n n_i r_i t_i}{\sum_{f,E:L(E)=f} \frac{1}{P(f,E)} \sum_{\theta:L(\theta(E)=f} \sum_i N(n_i, \theta) \prod_{j=1}^n n_i r_i t_i} \quad (1)$$

Para hacer más fácil la estimación del nuevo valor del parámetro, se han precalculado las probabilidades de los cuatro árboles por separado. Las probabilidades obtenidas se muestran a continuación,

$$\begin{aligned} P_1 &= 0.8125 * 0.92 * (0.75 * 0.15) * (0.65 * 0.1) * (0.8 * 0.45) * (0.9 * 0.2) \\ &\quad * 0.82 * 0.84 * 0.92 * 0.87 * 0.95 * 0.9 * 0.8 * 0.7 * 1 * 1 \\ &= 7.79 * 10^{-5} \end{aligned} \quad (2)$$

$$\begin{aligned} P_2 &= 0.83 * 0.92 * (0.75 * 0.15) * (0.65 * 0.1) * (0.8 * 0.45) * (0.75 * 0.2) \\ &\quad * 0.82 * 0.84 * 0.92 * 0.87 * 0.05 * 0.9 * 0.8 * 0.2 * 1 * 1 \\ &= 3.98 * 10^{-8} \end{aligned} \quad (3)$$

$$\begin{aligned} P_3 &= 0.83 * 0.89 * (0.75 * 0.15) * (0.65 * 0.1) * (0.9 * 0.45) * (0.75 * 0.2) \\ &\quad * 0.82 * 0.84 * 0.92 * 0.87 * 0.05 * 0.1 * 0.8 * 0.2 * 0.3 * 1 \\ &= 4.34 * 10^{-8} \end{aligned} \quad (4)$$

$$\begin{aligned} P_4 &= 0.8125 * 0.89 * (0.75 * 0.15) * (0.65 * 0.1) * (0.9 * 0.45) * (0.75 * 0.2) \\ &\quad * 0.82 * 0.84 * 0.92 * 0.87 * 0.95 * 0.1 * 0.8 * 0.7 * 0.3 * 1 \\ &= 2.82 * 10^{-6} \end{aligned} \quad (5)$$

Por lo tanto la reestimación del parámetro asociado a (C -> "gusta") es,

$$\bar{\theta}(C - >"gusta") = \frac{P_1 + P_4}{P_1 + P_2 + P_3 + P_4} = 0.9989 \quad (6)$$

Entonces, el parámetro que modela esta regla, quedaría modificado con este nuevo valor calculado.

2 Hierarchical machine translation

2.1 (*) Write an example of an alignment between two sentences and the rules that can be obtained with that alignment.

En este segundo ejercicio se ha considerado una traducción español a inglés, concretamente de la frase "Me gustaría comer una pizza en la trattoria". Siguiendo el modelo de alineaciones se alcanza la siguiente tabla,

trattoria									X	X
la										
en								X		
pizza							X			
una						X				
comer			X		X					
gustaría		X								
Me	X									
	I	would	like	to	eat	a	pizza	in	the	trattoria

A raíz de estos alineamientos es posible deducir las siguientes reglas dónde se puede observar como se han remplazado algunas parejas de *phrases* por símbolos no terminales.

S \rightarrow Me gustaría comer A en B, I would like to eat A in B

A \rightarrow una pizza, a pizza

B - > la trattoría, the trattoria