UNIVERSITAT POLITÈCNICA DE VALÈNCIA

MIARFID
Máster en Inteligencia Artificial, Reconocimiento de Formas e Imagen Digital

DSIC
DEPARTAMENTO DE SISTEMAS INFORMÁTICOS Y COMPUTACIÓN

Máster en Inteligencia Artificial, Reconocimiento de Formas e Imagen Digital
Universitat Politècnica de València

Traducción basada en modelos sintácticos

Traducción Automática

Autor: Juan Antonio López Ramírez

Curso 2019-2020

Índice general

CAPÍTULO 1
Introducción

El objetivo de este trabajo es realizar los ejercicios propuestos en los temas de traducción automática basada en modelos sintácticos. Concretamente, se han escogido los temas de *Hierarchical machine translation* y *Tree to string translation*. Los ejercicios, y su puntuación, son:

- Ejercicio 1 (*) de *Hierarchical machine translation*. Escribir un ejemplo de una alineación entre dos frases y las reglas que pueden ser obtenidas con dicha alineación.
- Ejercicio 3 (***) de *Tree to string translation*. Definir un modelo pequeño: un árbol de análisis, una tabla de reordenamiento, una tabla de inserción (tanto para nodos como para palabras) y una tabla de traducción. Luego, aplicar el algoritmo de estimación descrito en la diapositiva 15 de las transparencias, teniendo en cuenta que el número de posibles transformaciones debe mantenerse muy pequeño para facilitar los cálculos.

CAPÍTULO 2

Ejercicio 1. Hierarchical machine translation.

En este primer ejercicio hemos tenido en cuenta una traducción del español al inglés. La frase ha sido *Me gustaría comer un kebab en el Tigris*. Hemos seguido el modelo de alineaciones y hemos alcanzado la tabla que se puede apreciar en la figura 2.1.

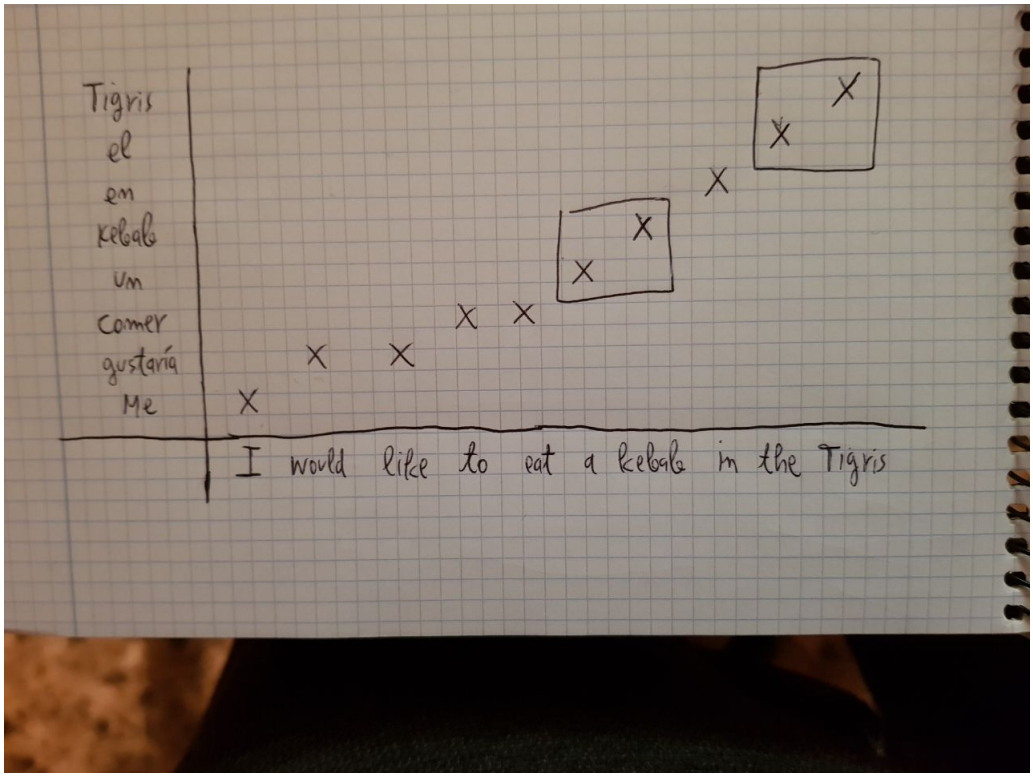

Figura 2.1: Tabla que muestra los alinemientos.

A partir de estos alineamientos, deducimos las siguientes reglas, en las que se puede apreciar que algunos pares de *phrases* han sido sustituidos por símbolos terminales:

A —>Me gustaría comer B en C, I would like to eat B in C
B —>un kebab, a kebab
C —>el Tigris, the Tigris

CAPÍTULO 3

Ejercicio 3. Tree to string translation.

Para este ejercicio hemos tenido en cuenta una traducción del español al japonés. La frase a traducir ha sido *Me gusta escuchar música*, que en japonés sería algo así como *Watashi wa ongaku wo kiku no ga suki desu*. El punto de partida ha sido la siguiente gramática:

0.95 $A \rightarrow B\ C\ D$
0,05 $A \rightarrow B\ B\ D$
0,95 $D \rightarrow E\ F$
0,05 $D \rightarrow C\ F$
0,90 $B \rightarrow 'me'$
0,10 $B \rightarrow 'gusta'$
0,80 $C \rightarrow 'gusta'$
0,20 $C \rightarrow 'escuchar'$
1,00 $E \rightarrow 'escuchar'$
1,00 $F \rightarrow 'musica'$

Luego, hemos tenido que definir las tablas de reordenamiento, inserción (nodos y palabras) y traducción, que se pueden observar en la figuras 3.1.

Tabla de reordenamiento

B C D	B C D	0,038
	B D C	0,813
	C B D	0,033
	C D B	0,043
	D B C	0,035
	D C B	0,038
B B D	B B D	0,1
	B D B	0,85
	D B B	0,05
E F	E F	0,1
	F E	0,9
C F	C F	0,1
	F C	0,9

Tabla de inserción (nodos)

padre	TOP	A	A	A	D	D	D
nodo	A	B	C	D	C	E	F
P(none)	0,8	0,1	0,03	0,87	0,05	0,15	0,2
P(izquierda)	0,05	0,15	0,07	0,07	0,05	0,05	0,15
P(derecha)	0,15	0,75	0,9	0,06	0,9	0,8	0,65

Tabla de inserción (palabras)

palabra	P(inserción)
wa	0,1
wo	0,1
ga	0,295
desu	0,2
no	0,295
ni	0,01

Tabla de traducción

me	gusta	escuchar	música
watashi (0,82)	suki (0,84)	kiku (0,92)	ongaku (0,87)
anata (0,06)	daisuki (0,16)	mirukoto (0,08)	NULL (0,13)
boku (0,12)			

Figura 3.1: Tablas de reordenamiento, inserción (nodos y palabras) y traducción.

Una vez hecho todo esto, se generan cuatro árboles diferentes, que se pueden apreciar en la figura 3.2.

Por último, hemos realizado la estimación del parámetro que se corresponde con la regla:

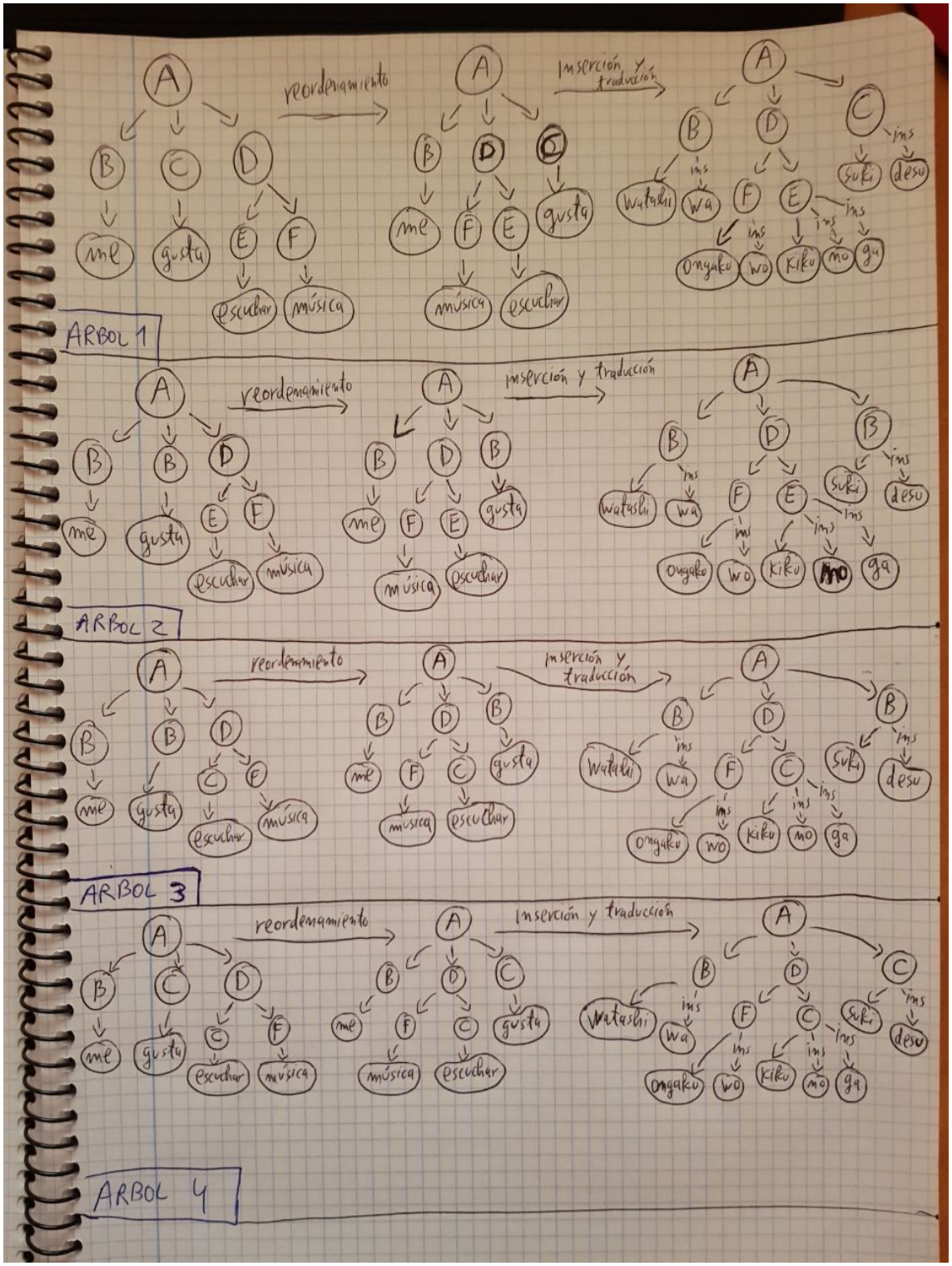

Figura 3.2: Árboles generados teniendo en cuenta la gramática y las tablas.

0.70 $C \rightarrow$ *'gusta'*

Para hacer esto, hemos aplicado la fórmula de la transparencia 15 de las diapositivas. Antes de eso, hemos calculado las probabilidades de los cuatro árboles por separado. Las probabilidades obtenidas se muestran en la figura 3.3.

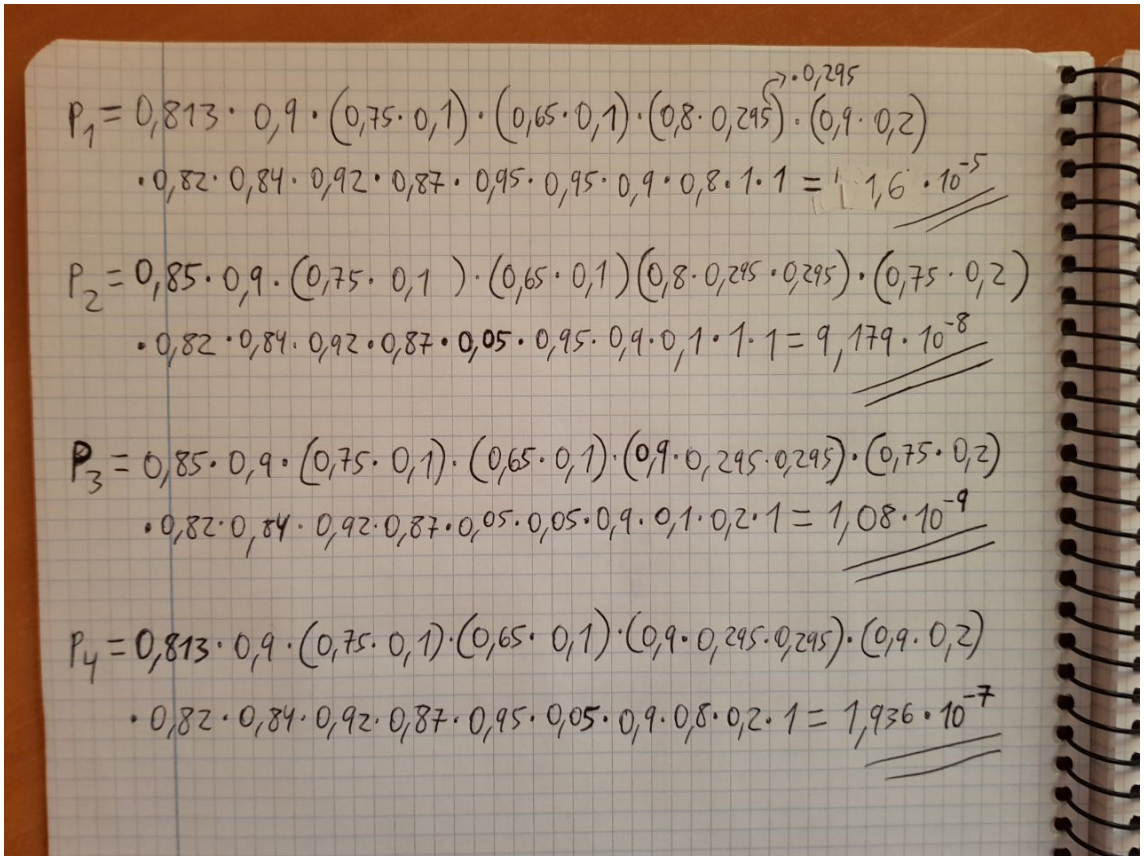

Figura 3.3: Probabilidades obtenidas de cada árbol.

Entonces, la reestimación del parámetro asociado a la regla es la que se observa en la figura 3.4, por lo que el parámetro quedaría ajustado con este nuevo valor.

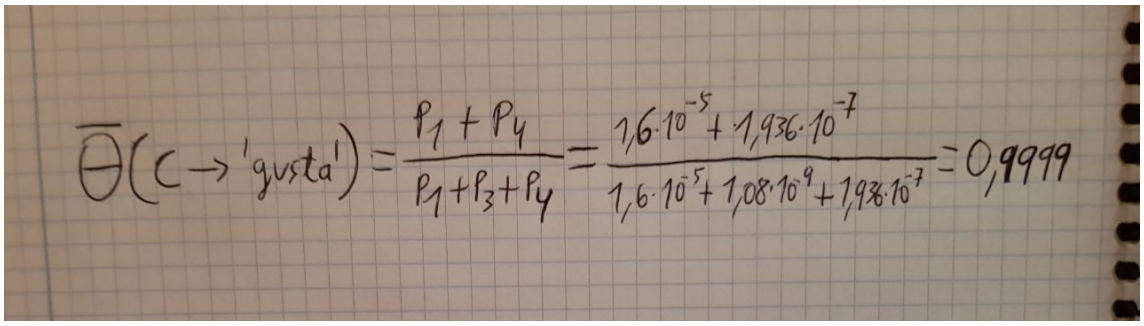

Figura 3.4: Reestimación obtenida.