PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

FACULTAD DE LETRAS

Curso de traducción automática de lenguas naturales

César Antonio Aguilar

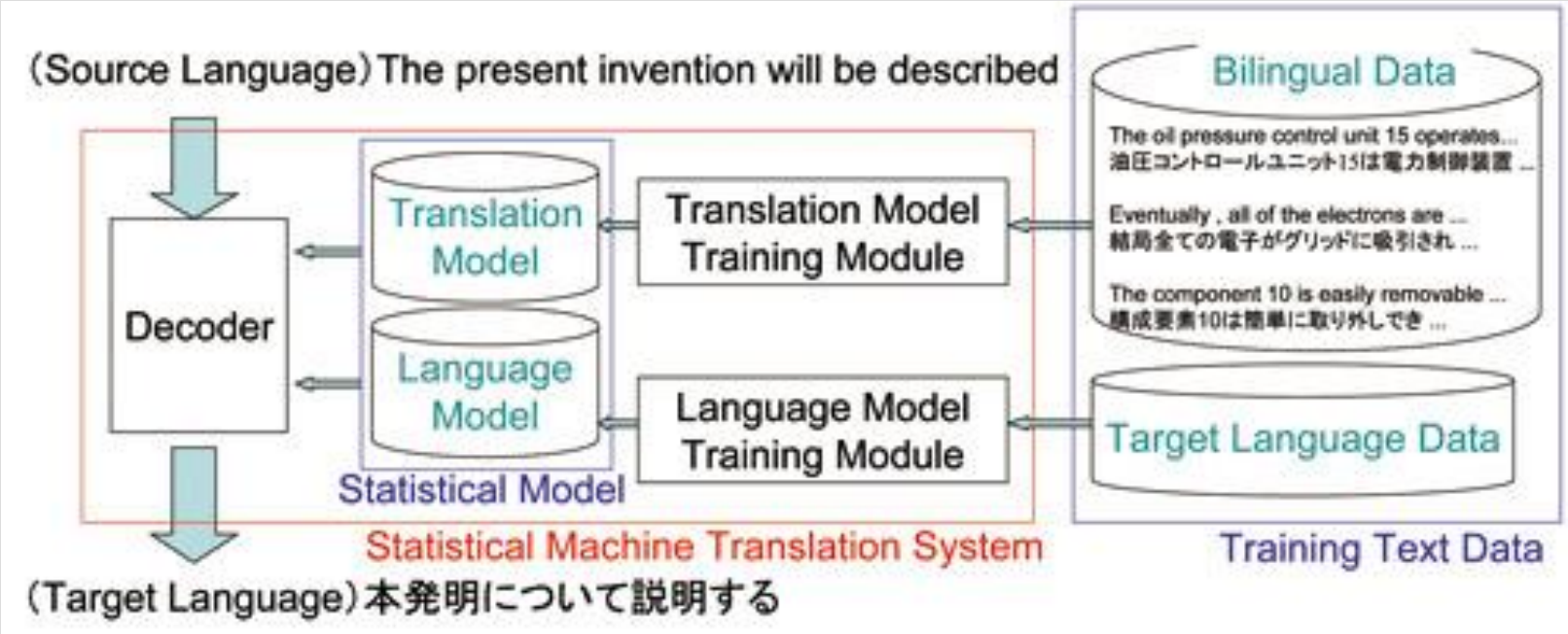
Facultad de Lenguas y Letras

28/08/2015

Cesar.Aguilar72@gmail.com

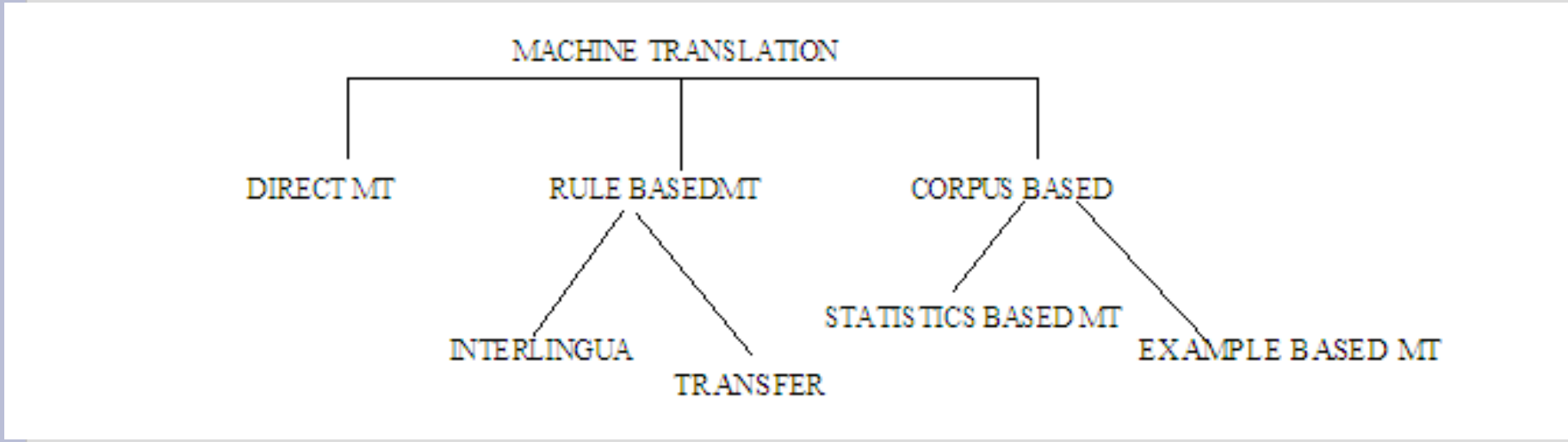
Síntesis de la clase pasada (1)

Como vimos en nuestra primera clase, la traducción automática es una línea de investigación de larga historia, la cual ha pasado por varias etapas: una génesis basada en el trabajo criptográfico desarrollado en los años 40, que dio lugar a una etapa de experimentación con modelos lingüísticos, y finalmente, esto se ha complementado con el uso de métodos estadísticos.



Síntesis de la clase pasada (2)

Tomando en cuenta tales etapas, se han desarrollado distintos métodos para realizar una traducción vía el empleo de computadores:



De manera resumida, describimos 4 en la clase pasada: (i) la traducción directa (palabra por palabra), (ii) la traducción basada en análisis gramaticales, (iii) la traducción basada en aproximaciones semánticas, y (iv), la traducción basada en una interlengua.

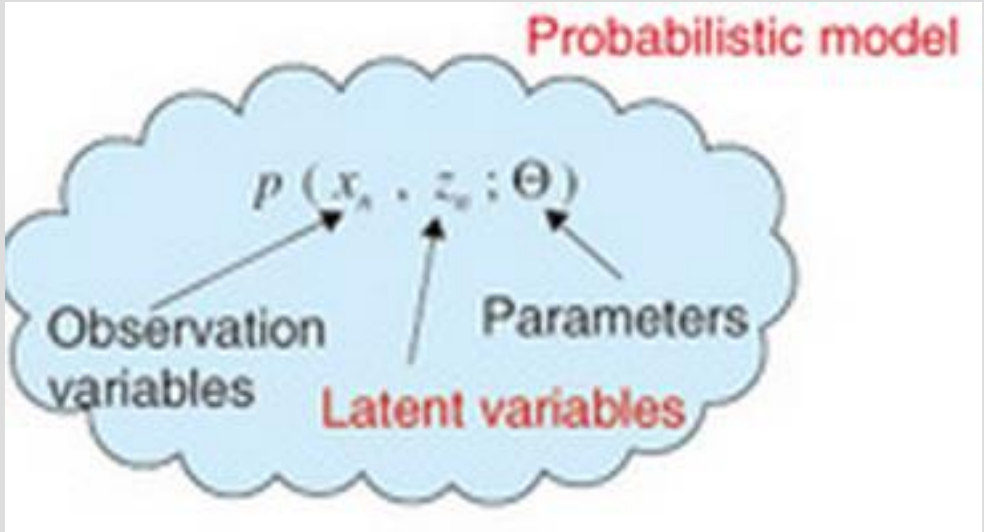
Traducción basada en estadísticas (1)

Este método de traducción (o mejor dicho, se trata de un paradigma tanto teórico como metodológico) fue introducido en 1949 por **Warren Weawer** (de quien ya vimos algo la clase pasada), basándose en la teoría de la información planteado por su colega **Claude Shannon** (1916-2001).

Traducción basada en estadísticas (2)

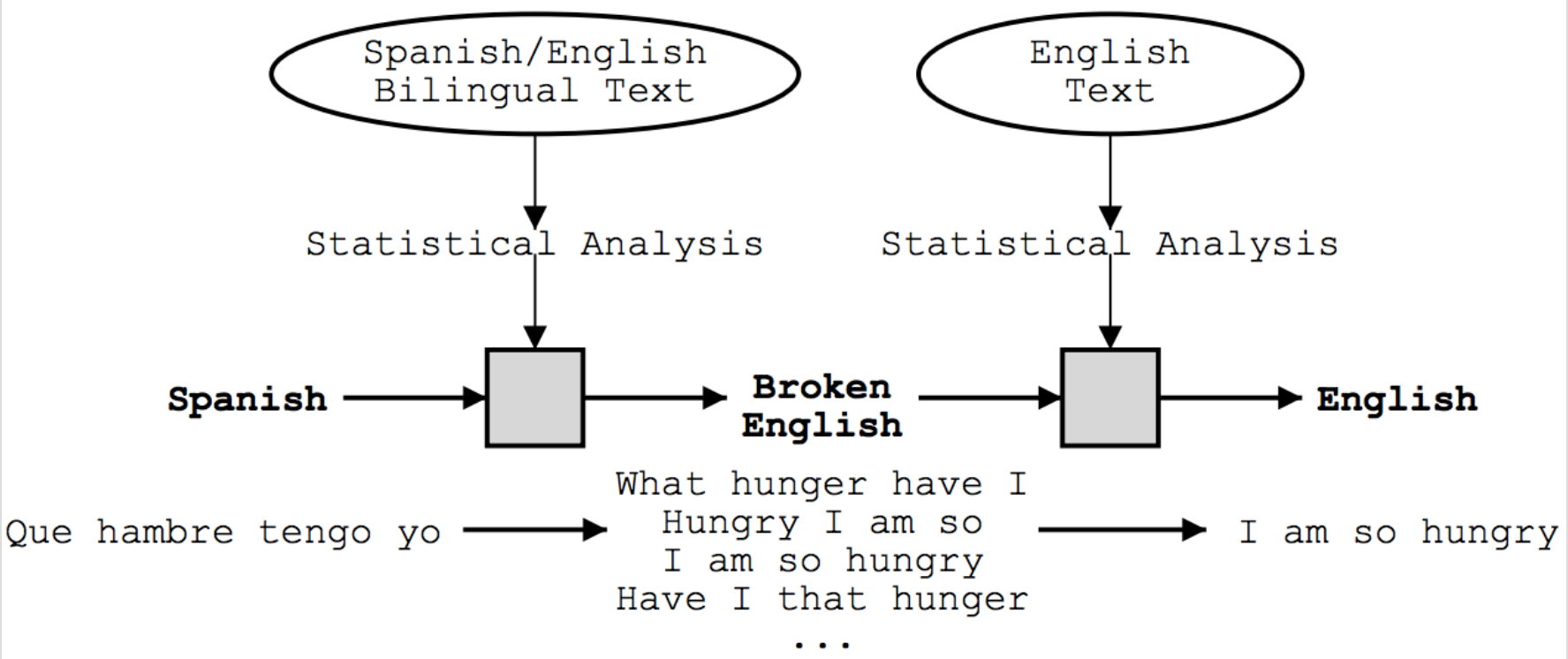
A grandes rasgos, una traducción de este tipo supone que es imposible lograr una traducción perfecta, ya sea que la realice una computadora o un traductor.

Dado esto, una forma de lograr una traducción aceptable es realizar una estimación estadística, la cual estaría guiada por dos variables concretas: (i) **su grado de fiabilidad** (*¿es correcta o no?*), (ii) **su grado de naturalidad** (*¿suena bien tal traducción en la lengua de llegada?*).



Traducción basada en estadísticas (3)

Veamos un ejemplo concreto: si tuvieran que traducir al inglés una oración con un estilo muy madrileño como: *¡Pues mira…, qué hambre tengo yo!* ¿Cómo lo harían? ¿Y al hacerlo, qué ponderan: el lograr una copia fidedigna en inglés (esto es, su *grado de fidelidad*); o buscan más bien que sea entendible (es decir, su grado de naturalidad)?



Traducción basada en estadísticas (4)

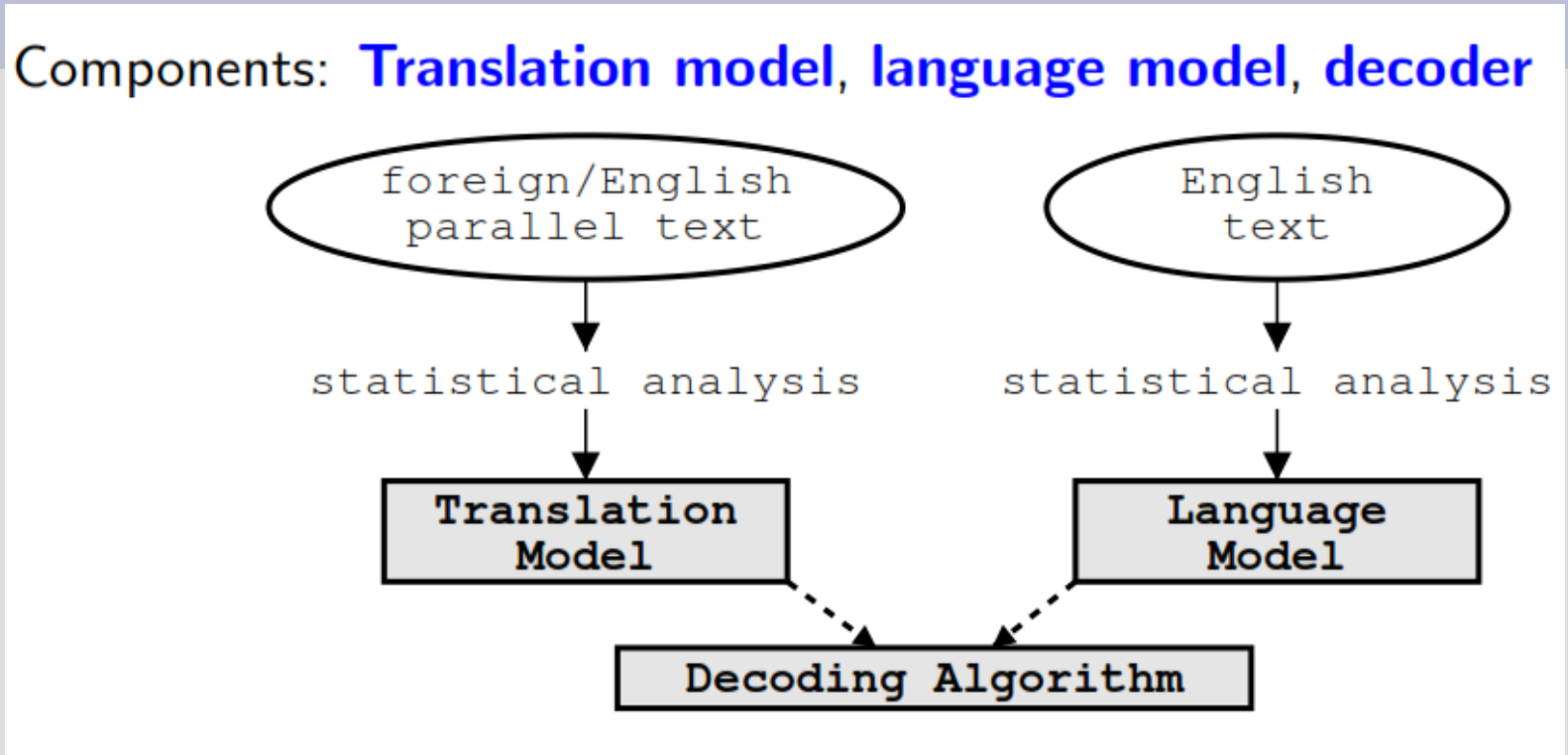
Dada la noción anterior, pensemos entonces que existe la *probabilidad* de que hay al menos una frase en inglés (*ing*) es equivalente a nuestra frase en español (*esp*), tomando en cuenta las variables anteriores: fidelidad y naturalidad:

$$P(ing|esp)$$

Si estamos de acuerdo con esta probabilidad, entonces lo que requerimos para dar con la mejor traducción es *contar todas las posibles frases equivalentes que tenemos en inglés, y luego escoger cuál es la más confiable y natural a nuestra frase en español.*

Traducción basada en estadísticas (5)

Esta noción la podemos entender mejor viendo el siguiente gráfico:



En nuestro caso, nuestro **modelo de traducción** es el español, en tanto que nuestro **modelo del lenguaje** es el inglés (recuerden, lo que queremos es traducir de español a inglés).

Traducción basada en estadísticas (6)

Y ahora, consideraremos un algoritmo de traducción que esté basado en la siguiente fórmula:

$$\hat{E} = \underset{E \in English}{\operatorname{argmax}} \; P(E|S)$$

$$= \underset{E \in English}{\operatorname{argmax}} \; \frac{P(S|E)P(E)}{P(S)}$$

$$= \underset{E \in English}{\operatorname{argmax}} \; \underbrace{P(S|E)}_{\text{Modelo de traducción}}\underbrace{P(E)}_{\text{Modelo de lenguaje}}$$

Traducción basada en estadísticas (6)

Lo que quiere decir:

La mejor traducción resultante ($\hat{E}$) es equivalente a reconocer a la expresión (E) óptima (esto es lo que representa la abreviatura *argmax*), que sea una frase realmente en inglés ($E \in English$), tomando en cuenta la probabilidad de que existe una frase en inglés adecuada para la frase a traducir del español ($P(E|S)$).

Lo mismo, pero de manera abreviada:

$$\tilde{e} = \arg\max_{e \in e^*} p(e|f) = \arg\max_{e \in e^*} p(f|e)p(e)$$

Cálculo de fluidez (1)

¿Qué es lo que estamos calculando con estas fórmulas? De acuerdo con **Daniel Jurafsky** (Stanford University), los elementos de esta fórmula nos ayudan a calcular precisamente los grados de fidelidad y naturalidad que buscamos en nuestra traducción.

Empecemos por la naturalidad o **fluidez** (ing.: *Fluency*). Digamos que en inglés podemos expresar un evento del tipo: *un auto estuvo a punto de chocar conmigo*. Así, lo que podríamos hacer es establecer un rango entre lo que suene más descriptivo, hasta lo que suene más vago, p. e.:

We need a metric that ranks this sentence

That car almost crash to me

as less fluent than this one:

That car almost hit me

Cálculo de fluidez (2)

El cálculo de la fluidez es sencillo e interesante. Primero, lo formulamos del siguiente modo:

$$P(E)$$

Lo cual sólo indica que hay una probabilidad *X* de que encontremos una frase que describa el evento anterior. Ahora, si lo atendemos con más detalle, esperamos que tal frase se ubique dentro del siguiente rango:

$$P(me|hit) > P(to|crash)$$

Esto es hacer un cálculo respecto al conocimiento que tenemos de nuestra propia lengua: ¿puedo solamente expresar que *algo casi me pega*; o puedo entrar en detalles y decir que *un carro estuvo a punto de chocar con el auto que yo conducía*?

Cálculo de fidelidad (1)

Y ahora vamos a ver la fidelidad. Si queremos traducir al inglés una frase en español como:

Maria no dió una bofetada a la bruja verde

¿Cuál de las siguientes opciones les parece la más adecuada?:

Mary didn't slap the green witch
Mary not give a slap to the witch green
Maria gave no a slap to the green witch
The green witch was not slapped by Mary

Para determinar la probabilidad de nuestras opciones, usamos esta fórmula, en donde *S* representa *Spanish,* y *E* representa *English*:

$$P(S|E)$$

Cálculo de fidelidad (2)

Relacionando ambos cálculos, lo que estamos suponiendo para decidir cuál será nuestra mejor traducción es:

La formula *P(S|E)* plantea que nuestro **modelo de traducción** es una especie de **bolsa de palabras**, que nos permite buscar y contrastar cuál es la opción más confiable para traducir lo que queremos desde la lengua de origen.

Por otro lado, la formula *P(E)* nos ofrece un **modelo de lengua**, que nos permite ir generando todas las opciones posibles de nuestra lengua de llegada, de tal modo que priorizamos cuál de ellas es la más fluida, es decir, la que suena *más natural*.

Interludio: algo sobre probabilidades (1)

Para entender mejor cómo trabajan los métodos estadísticos en traducción, vamos a hacer una mínima revisión sobre un concepto clave aquí: la noción de **probabilidad**. Tratemos de definirla, realizando el siguiente experimento:

Primer experimento: supongamos que tenemos un dado y nos preguntamos, ¿cuántas veces puede caer el número 4 en 6 tiradas?

Delimitemos:

1. Un dado tiene 6 caras.
2. Cada cara representa un número del 1 al 6.
3. En teoría, en 6 tiradas me puede salir al menos un número del 1 al 6.

Interludio: algo sobre probabilidades (1)

Lo anterior se puede representar en una tabla del tipo:

Tiradas						
1	1/6	1/6	1/6	1/6	1/6	1/6
2						
3						
4						
5						
6	1	2	3	4	5	6

Caras del dado

Interludio: algo sobre probabilidades (3)

Algunas definiciones:

- A nuestra serie de tiradas del dado (esto es, 6 las veces que lo arrojamos para que nos salga el número 6, conforman un *espacio simple* (o *Es*).

- Cada tirada puede ser vista como un *fenómeno* o *hecho* en donde podemos confirmar o no si sale el número 6.

- Si lanzamos 6 veces nuestro dado, y pensamos que por lo menos 1 vez va a caer el número 6, eso quiere decir que las posibilidades de que salga este número es de 1/6, o lo que es lo mismo, 0.16.

Interludio: algo sobre probabilidades (4)

Segundo experimento: supongamos que ahora jugamos con una moneda, y queremos ver si en tres lanzamientos nos sale dos veces Sol.

Condiciones:

1. Nuestra moneda tiene dos lados: Águila y Sol.
2. Nuestra moneda no está “calada”.
3. En teoría, la posibilidad de que se cumpla el hecho que de caiga Águila o Sol es de un 50% para cada uno, o sea 0.5.

Interludio: algo sobre probabilidades (5)

Para decirlo de una manera más “elegante”:

- Hay dos hechos posibles, Águila (Ag) y Sol (So)
- Espacio simple *Es* es {H,T}
- Al lanzar una moneda al aire, suponemos que hay las mismas probalidades de que caiga Ag o So.
- La suma de estas dos posibilidades nos da 1.
- P({Ag}) = 0.5
- P({So}) = 0.5
- P({Ag,So}) = P({Ag})+P({So}) = 1.0

Interludio: algo sobre probabilidades (6)

- Lanzemos 3 veces nuestra moneda.
- Las combinaciones de lanzamientos conforma nuestro *Es*, y éste es representable como una cadena.

- *Es* ={AgAgAg; AgAgSo; AgSoAg; SoAgAg; AgSoSo; SoAgSo; SoSoAg; SoSoSo}

- Asumimos que cada hecho es *equiprobable*.
- **Pregunta:** ¿cuántas probabilidades tenemos de que caiga 2 veces Sol?
- A = {SoSoAg; SoAgSo; AgSoSo}
- P(A) = P({SoSoAg}) + P({SoAgSo}) + P({SoAgSo})
- = 1/8 + 1/8 + 1/8
- =3/8

Interludio: algo sobre probabilidades (7)

- Una duda: ¿y si nuestra modena está trucada y no permite que caiga Sol más que una sola vez?

- Nuestras opciones de combinación se reducen, por ejemplo:

- *Es* ={AgAgAg; AgAgSo; AgSoAg; SoAgAg; AgSoSo; SoAgSo; SoSoAg; SoSoSo}

- **Pregunta:** ¿cuántas probabilidades tenemos de que caiga Sol por lo menos una sola vez?
- B = {AgAgSo; AgSoAg; SoAgAg}
- P(B) = P({AgAgSo}) + P({AgSoAg}) + P({SoAgAg})
- = 2/8 + 2/8 + 2/8
- =6/8

Interludio: algo sobre probabilidades (8)

Tiradas

Ag								
Ag								
Ag								
Ag			2/8					
Ag		2/8						
Ag	2/8							
Ag								
Ag	So	So	So	So	So	So	So	So

Caras del dado

Probabilidades independientes y condicionales (1)

Para nuestros análisis, vamos a tomar en cuenta dos tipos de probabilidades: una en donde suponemos que no hay una relación de ningún tipo entre dos (o más variables) para que se dé un fenómeno. Decimos que tales fenómenos están regidos por el azar (y de manera absoluta).

Algo tan simple como decir: *me duelen las muelas del juicio, seguro es que va a llover hoy*, es postular justo un fenómeno que, si se llegara a cumplir, no veríamos una relación clara entre las variables que intervienen en él. La fórmula para representar esto es:

$$P(A \text{ and } B) = P(A) \times P(B)$$

Probabilidades independientes y condicionales (2)

En cambio, si hay una relación subyacente entre dos variables, y ésta condiciona la manifestación de un fenómeno, entonces hablamos de **probabilidades condicionales**. En otras palabras, lo que queremos señalar es que cada vez que se presenta una asociación entre dos variables (A y B), observamos una alta probabilidad de que ocurra un fenómeno. Así, las dos variables se influyen mutuamente. La fórmula para presentar esto es:

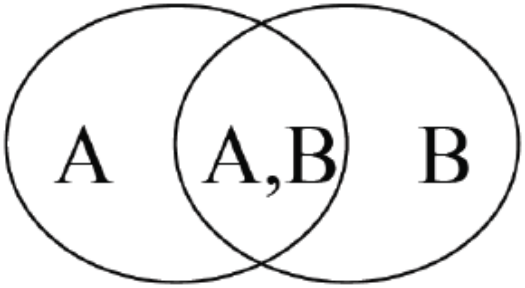
$$p(B|A) = p(A \cap B) / p(A)$$

Volviendo a nuestro juego de cartas, podríamos decir que conforme resolvamos quién tiene las Reinas que nos faltan para nuestro póker, eso condiciona que podamos formar la secuencia ganadora, en la menor cantidad de manos posibles.

Probabilidades independientes y condicionales (3)

Si nos ponemos más formales, una probabilidad condicional equivale a decir que:

$$P(A \mid B) = \frac{P(A,B)}{P(B)}$$



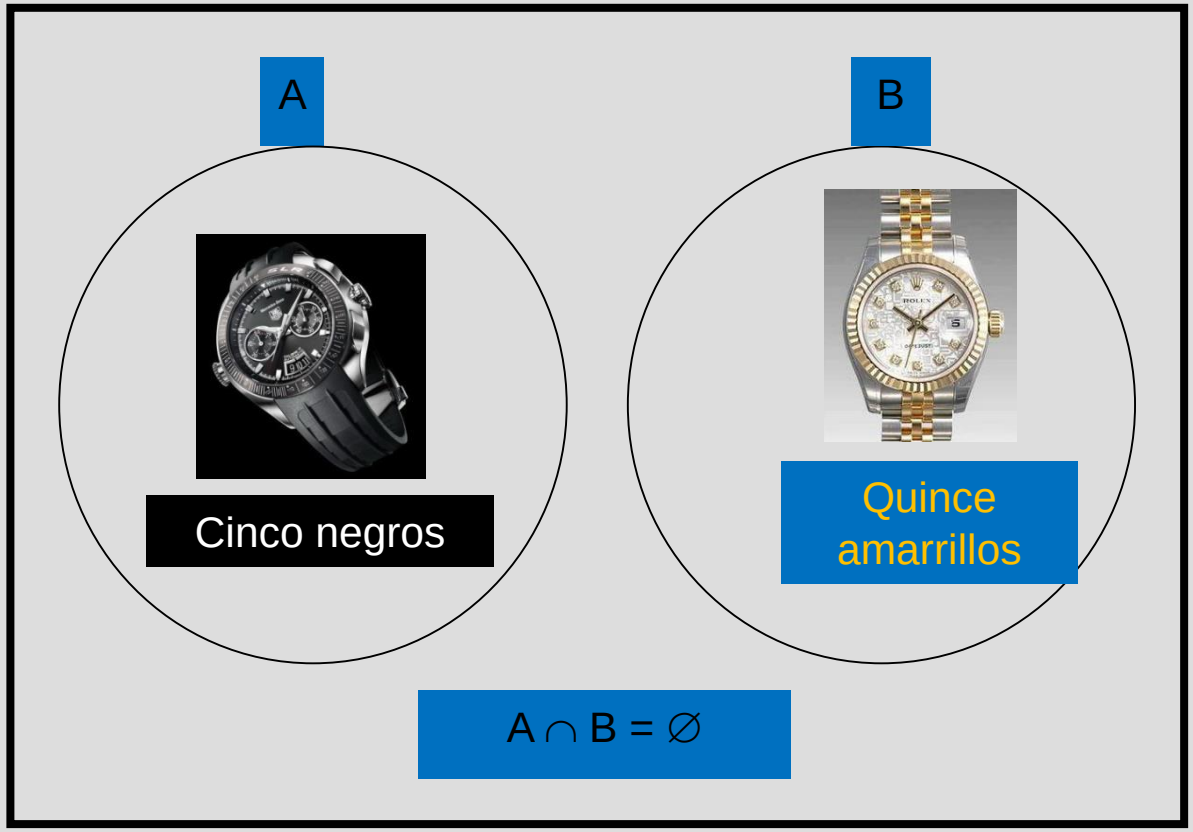
Recordando las matemáticas de la secundaria, esto se puede representar con un diagrama de Venn:

Hay que notar que: $P(A,B)=P(A|B) \cdot P(B)$

También: $P(A,B) = P(B,A)$

Probabilidades independientes y condicionales (4)

Para entender mejor estas probabilidades, veamos el siguiente ejemplo: supongamos que tenemos un conjunto de relojes de la siguiente forma:



Probabilidades independientes y condicionales (5)

Si calculamos su probabilidad independiente, decimos que:

$$P(A \cup B) = P(A) + P(B)$$

Esto quiere decir que *la probabilidad la unión de los conjuntos A y B es igual a la suma de la probabilidad de A y la probabilidad de B*.

Si esto es así, lo que estamos suponiendo es que entre los conjuntos A y B no tienen elementos en común, aunque forman parte de un mismo universo. En otras palabras, nuestro conjunto Universal son 20 relojes, pero los separamos a partir de un rasgo: su color, lo que nos permite generar dos conjuntos específicos.

Probabilidades independientes y condicionales (6)

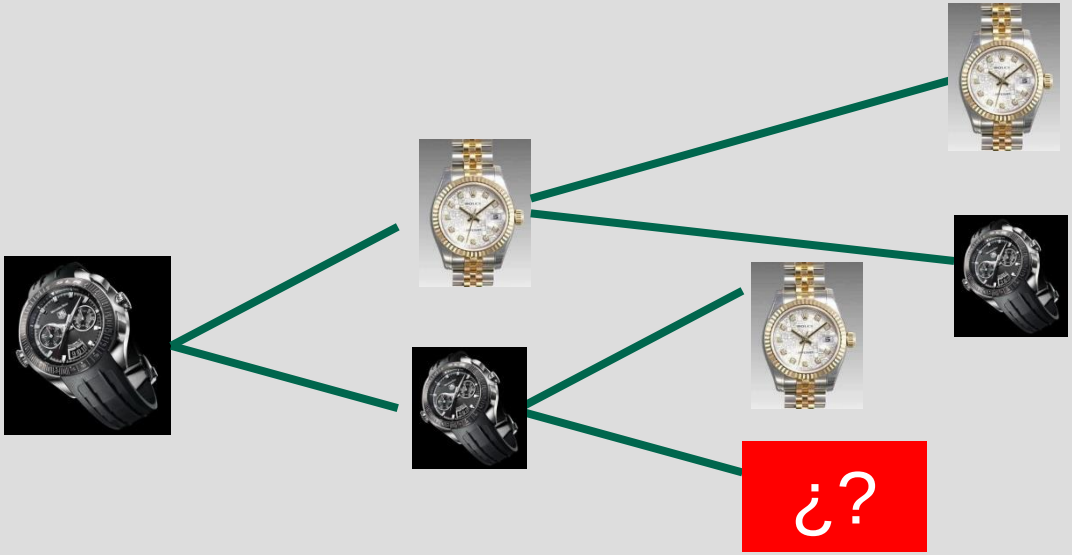
Veamos ahora el caso contrario: supongamos que los eventos A y B tienen un punto de cruce entre sí, de tal manera que forman una intersección.

Si quisiera calcular de probabilidad dos eventos no disjuntos A y B, debo excluir su punto de intersección, esto es:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Probabilidades independientes y condicionales (7)

Volviendo a la extracción de relojes, si sacara un reloj negro del conjunto A, y me quedo con 4, ¿conservo las mismas probabilidades para realizar mi experimento, o éstas se ven afectadas por tal decisión?



Probabilidades independientes y condicionales (7)

Cuando realizamos este experimento, suponiendo que tenemos las mismas condiciones entre la primera y la segunda extracción, los escenarios que tenemos son:

Negros	Amarillos	P(negros)	P(amarillos)	# Experimento
5	15	5/20	15/20	Exp. 1 = 1N + 2 A
4	13	4/17	13/17	Exp. 2 = 1N + 2 A
3	11	3/14	11/14	Exp. 3 = 1N + 2 A

A este hecho de forzar las probabilidades de los resultados esperados en cada experimento es lo que llamamos condicionalidad, lo que da pie a las llamadas probabilidades condicionales.

Probabilidades independientes y condicionales (8)

Una forma de definir las probabilidades condicionales es la siguiente:

- Sean A y B dos eventos estadísticamente dependientes, entonces, la probabilidad condicional de A dado B, denotado por P(A / B), es la probabilidad de que suceda A dado que se sabe que el evento B ocurrió.

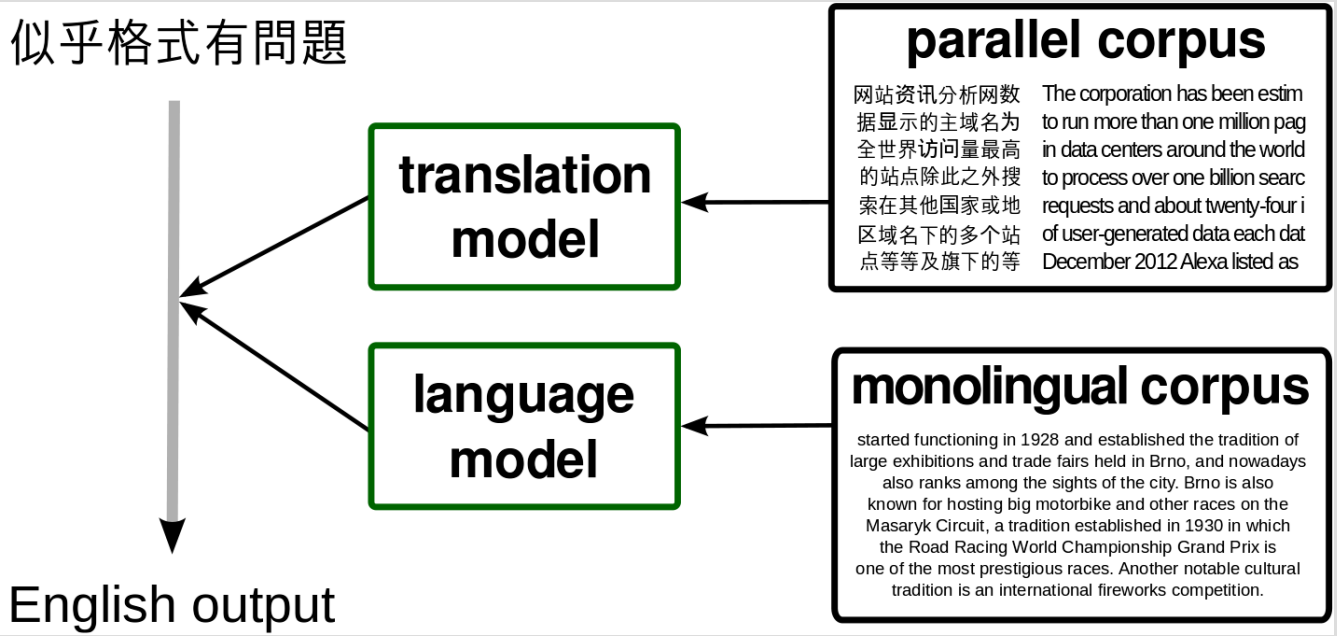
$$P(A / B) = \frac{P(A \cap B)}{P(B)}$$

P(A/B) = probabilidad de que ocurra el evento A, dado que el evento B ya ocurrió.
$P(A \cap B)$ = probabilidad de que ocurra A y B
P(B) = probabilidad de que ocurra el primer evento B.

Uso de corpus en traducción (1)

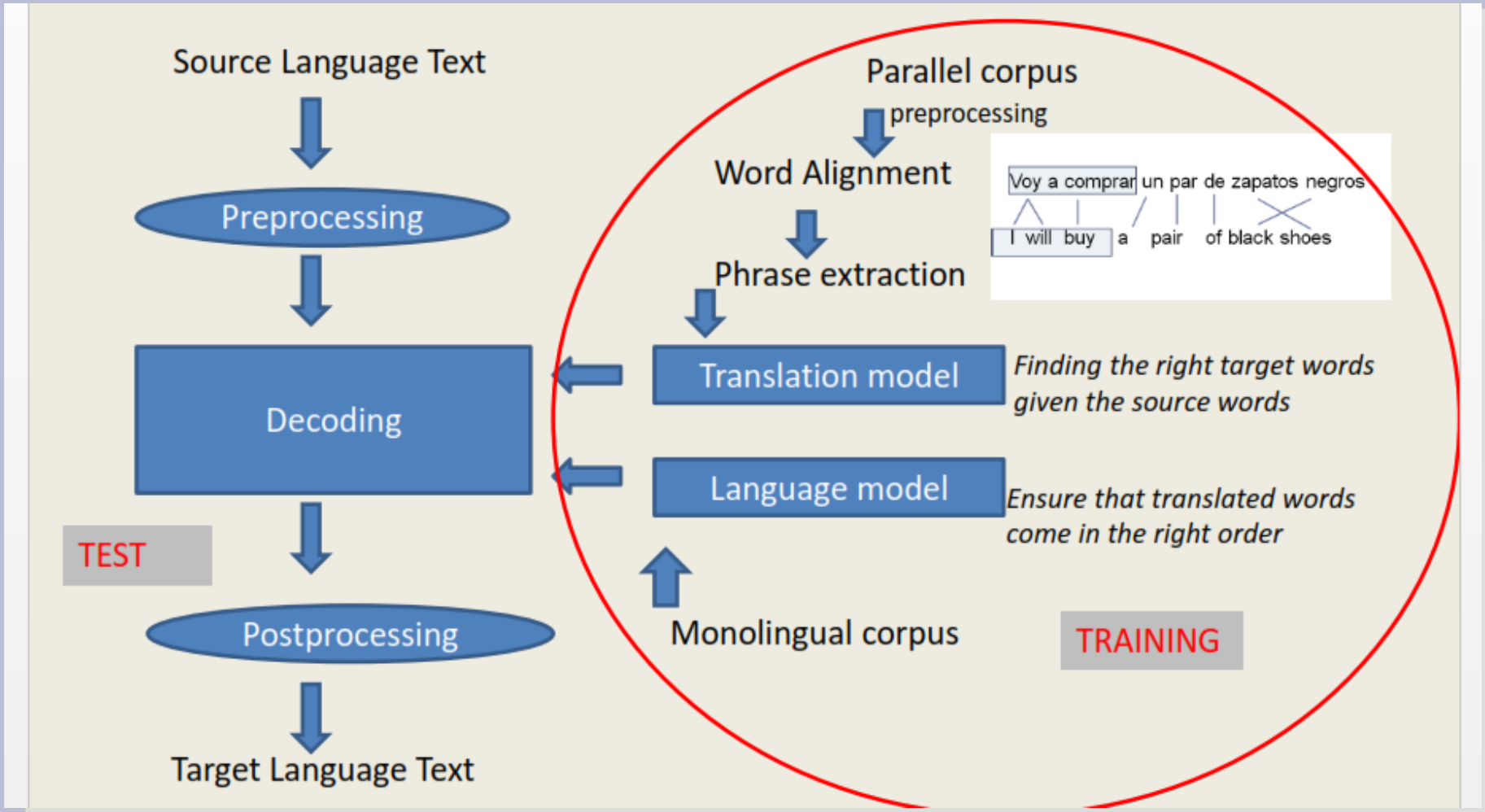
Considerando entonces el paradigma estadístico de la traducción, resulta necesario contar con amplias colecciones de datos lingüísticos que nos den idea sobre cómo funciona un modelo del lenguaje determinado, e igualmente necesitamos datos venidos de una lengua extranjera que nos ayuden a evaluar si es posible encontrar expresiones que puedan ser equivalentes a las de nuestro idioma.

En pocas palabras, necesitamos trabajar con **corpus paralelos** para aprovechar al máximo este enfoque estadístico.



Uso de corpus en traducción (1)

¿Para qué nos sirven los corpus paralelos? Revisando el gráfico, parece que para muchas cosas:



Gracias por su atención

Blog del curso:
http://cesaraguilar.weebly.com/curso-de-procesamiento-del-lenguaje-natural.html