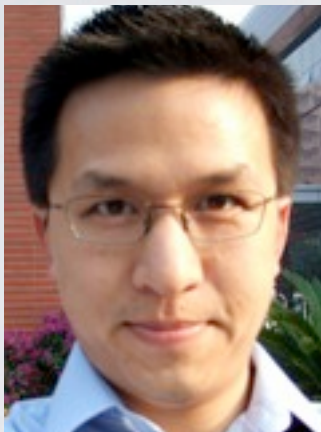


CSCI 599

MACHINE TRANSLATION

11-1-11 11:00

INSTRUCTORS



David Chiang 蔣偉

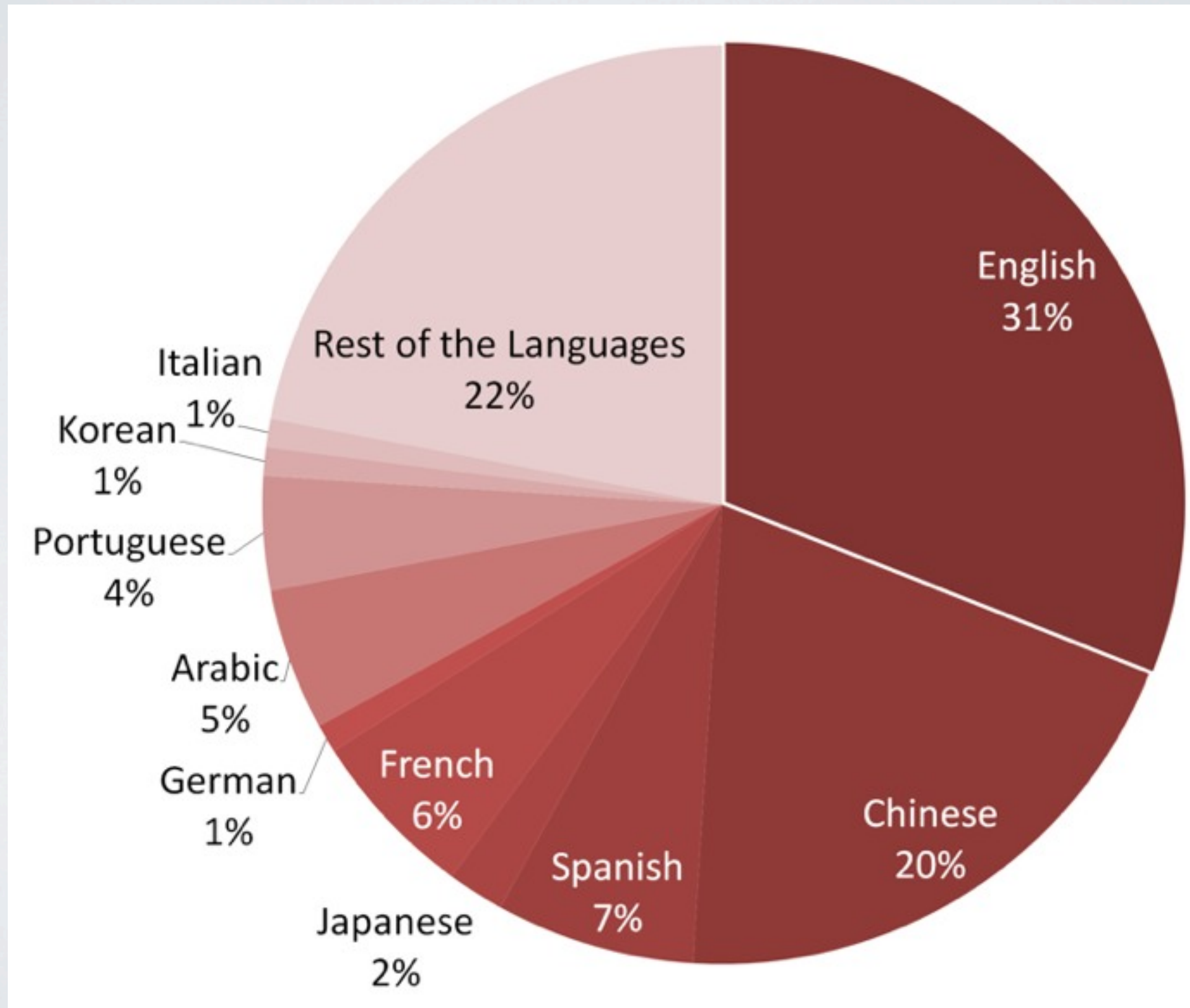


Liang Huang 黃亮

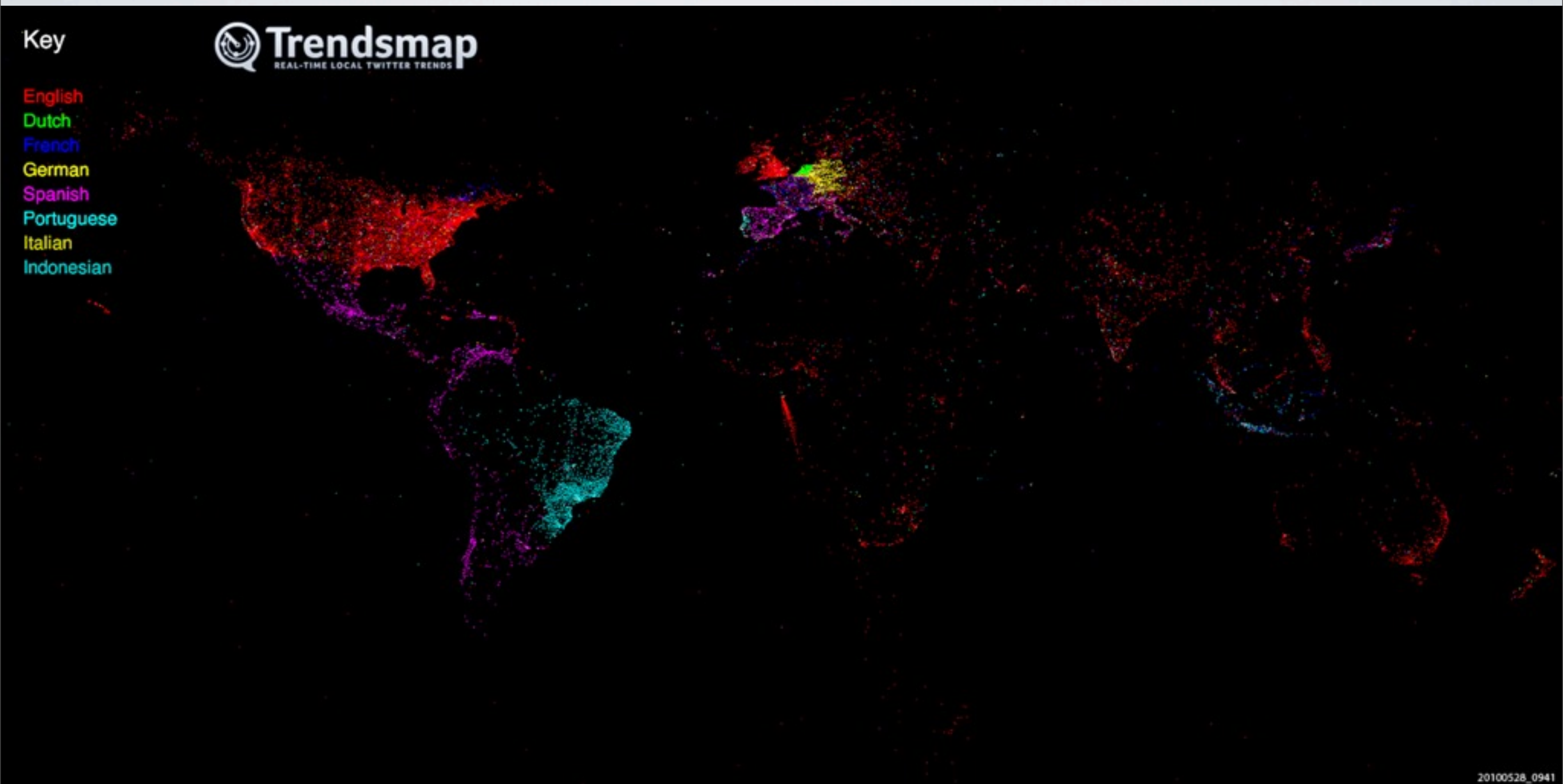


Kevin Knight 武凱文

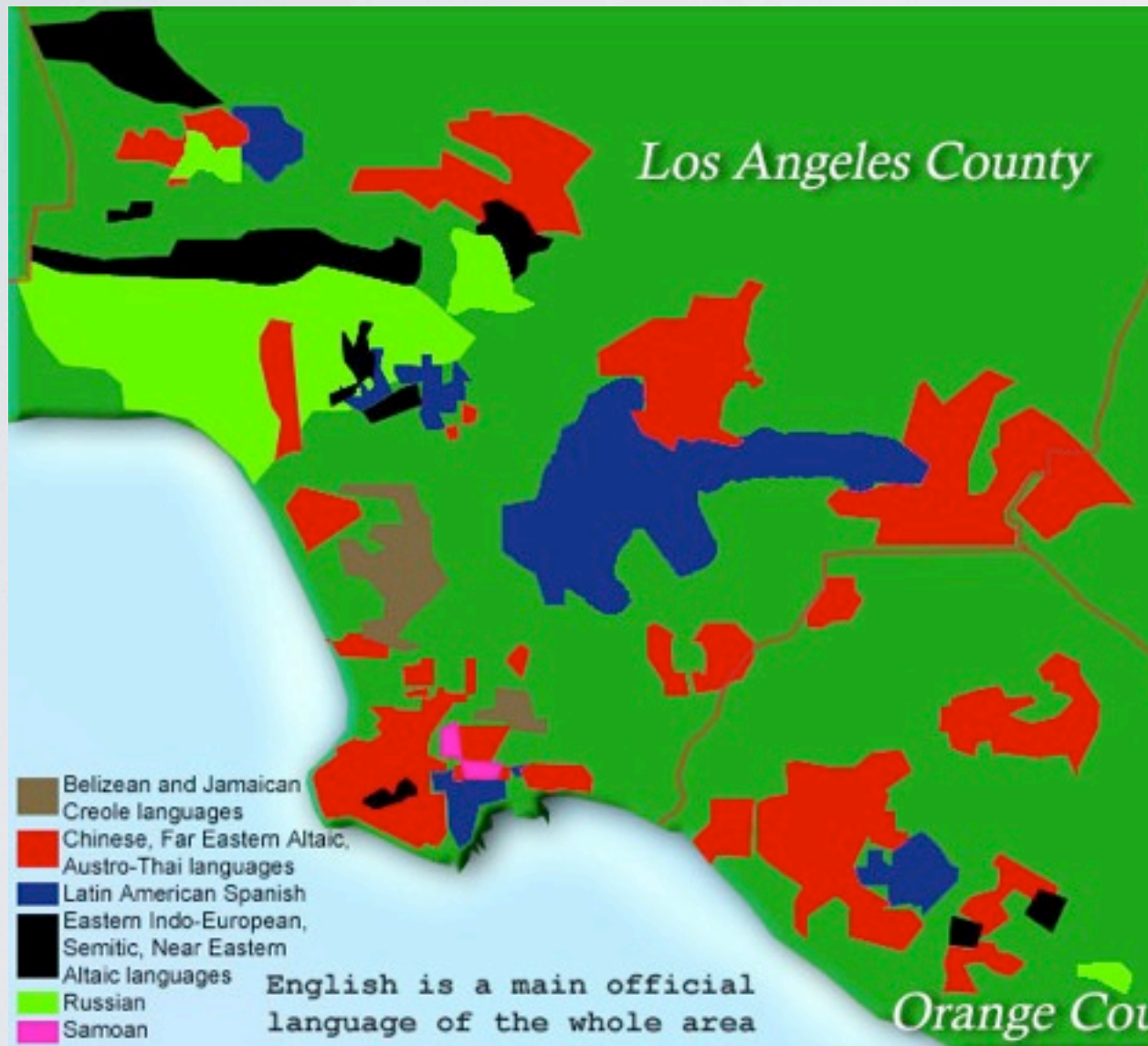
LANGUAGES ON THE WEB



LANGUAGES ON TWITTER



LANGUAGES IN LA



WHY DO WE NEED MT?



WHY DO WE NEED MT?



WHY DO WE NEED MT?



ADVERTENCIA

**SI USTED TIENE UN
MARCAPASO O CUALQUIER
OTRO IMPLANTE SIMILAR,
NO PASE SOBRE ESTE AVISO!
POR FAVOR PIDA ASISTENCIA
Y LAS MEDIDAS
APPROPIADAS SERAN TOMADAS.**



ADVERTENCIA

**NO HAY
SALVA VIDAS
PRESENTE**

LOS NIÑOS MENORES DE 14
AÑOS NO DEBEN NADAR NI USAR
EL PISCINA O JACUZZI SIN LA
PRESENCIA DE UN ADULTO.

ADULTOS NO DEBEN NADAR SOLOS.



Código de Ordenanza de la Ciudad
de Minneapolis Título 11, Capítulo 234

**ES PROHIBIDO FUMAR EN
ESTE ESTABLECIMIENTO**

Llame al 612-673-2424 con preguntas o quejas.

Honesdale

Days Inn Tannersville
 ●●●●○ 42 críticas
 Última crítica 6 ene 2011
 No. 1 de 3 hoteles en Tannersville

Settlers Inn
 ●●●●○ 42 críticas
 Última crítica 5 ene 2011
 No. 1 de 3 B&Bs / hostales en Hawley

Villa Roma Resort and Conference Center
 ●●●●○ 178 críticas
 Última crítica 5 ene 2011
 No. 1 de 1 hoteles en Callicoon

Harrington House
 ●●●●○ 13 críticas
 Última crítica 4 ene 2011
 No. 1 de 4 en Milford

Hampton Inn Clarks Summit
 ●●●●○ 12 críticas
 Última crítica 31 dic 2010
 No. 1 de 7 hoteles en Clarks Summit

Ramada Clarks Summit Hotel
 ●●●●○ 22 críticas
 Última crítica 31 dic 2010
 No. 4 de 7 hoteles en Clarks Summit

Days Inn Scranton
 ●●●●○ 32 críticas
 Última crítica 22 dic 2010
 No. 8 de 9 hoteles en Scranton

Hotel Wayne
 ●●●●○ 15 críticas
 Última crítica 21 dic 2010
 No. 2 de 2 hoteles en Honesdale

Scranton East Hotel & Convention Center
 ●●●●○ 46 críticas
 Última crítica 29 oct 2010
 No. 3 de 4 hoteles en Dunmore

Comfort Inn Pocono Lakes Region
 ●●●●○ 24 críticas
 Última crítica 21 oct 2010
 No. 1 de 1 hoteles en Lake Ariel

“ mucha sabor de los "beets" y sueños del muerte ”

tiburondelnorte ☒ 12 contribuciones
 whitewater colorado

30 nov 2007

estubimos perdidos cuando encontramos este lugar. estuve enteresante, pero estuve algunas cosas que son, eh, misteriosos. mi familia y yo teniamos un cuarto cerca del "cuarto de irrigation." mucho agua circulando por vientecuarto horas por dia. despues algun tiempo, fue tranquilo. los vinos del este hotel fueron... no era malo, pero eran muy curioso. tienen sabores de la tierra, y... [más](#)

☒ Traducción automática ?
☐ Original en inglés

Traducido por:
 Language Weaver

“ súper! ”

Sword85 ☒ 1 contribución

12 nov 2010

Me encanta este lugar! Me pasé dos semanas en esta granja y me he realmente disfruté de mi estancia! Yo sugiero a todo el mundo [más](#)

☒ Traducción automática ?
☐ Original en inglés

Traducido por:
 Language Weaver

“ interesante ”

yiga ☒ 1 contribución
 Surrey, Canadá

8 nov 2010

La ubicación es estupenda aunque direcciones estaban extraño verdad 156 pasos desde el buzón tinto ligero, ve a la izquierda". Cuando me dieron allí un hombre muy extraño persiguieron mi coche en el Lane pero el dueño me saludó y me ofreció zumo de la remolacha que era degustación interesante. Las únicas quejas que tengo era los pilas de con... [más](#)

han estado en Honesdale.

[Conéctate con Facebook](#)

Guías gratuitas de Honesdale

Consigue tu guía rápida de los mejores hoteles, restaurantes y cosas que hacer.



Hotels



Restaurants



Things to Do

[¡Listo para llevar!](#)



[VER MAPA DE SCHRUTE FARMS](#)

Por la zona de Schrute Farms**Alojamientos por la zona con mejor puntuación**

HOTEL	MXN*	DIST.	PUNTUACIÓN
The Harvest Inn	N/D	0.3 km	●●●●○
Double W Ranch Bed and Breakfast	N/D	5.1 km	●●●●○
James Manning House	N/D	5.3 km	●●●●○
Ponderosa Pines Campgrounds	N/D	12.3 km	●●●●○

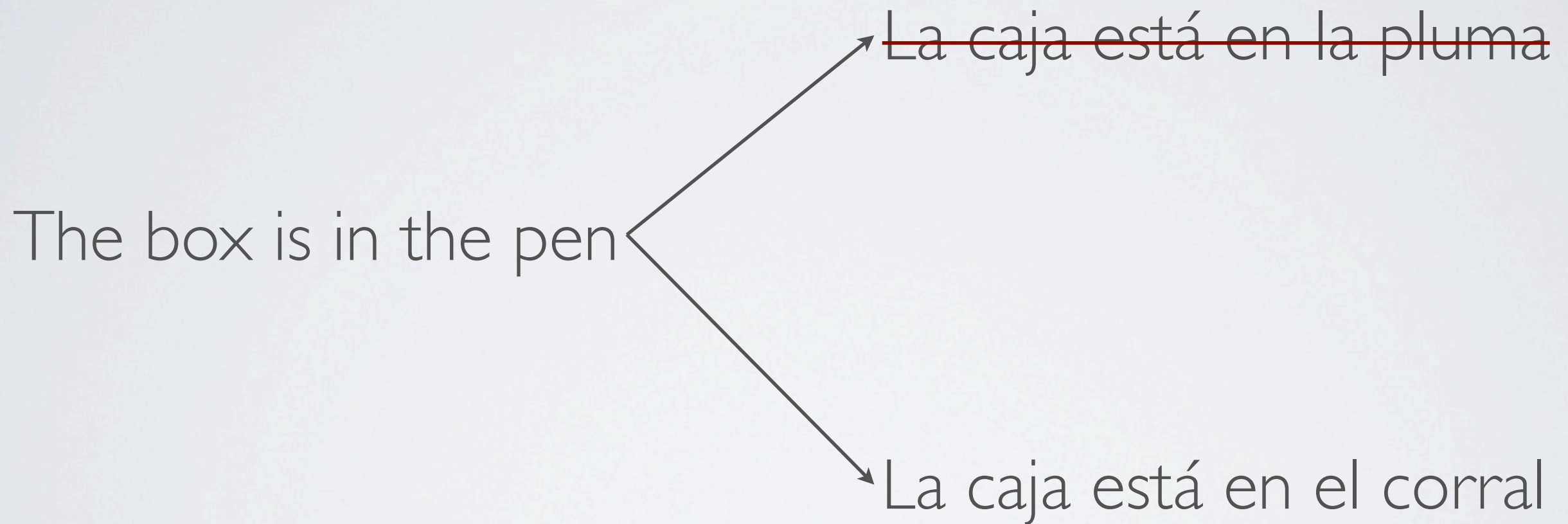
WHY IS MT HARD?

Coverage

chiliagon —————> ?

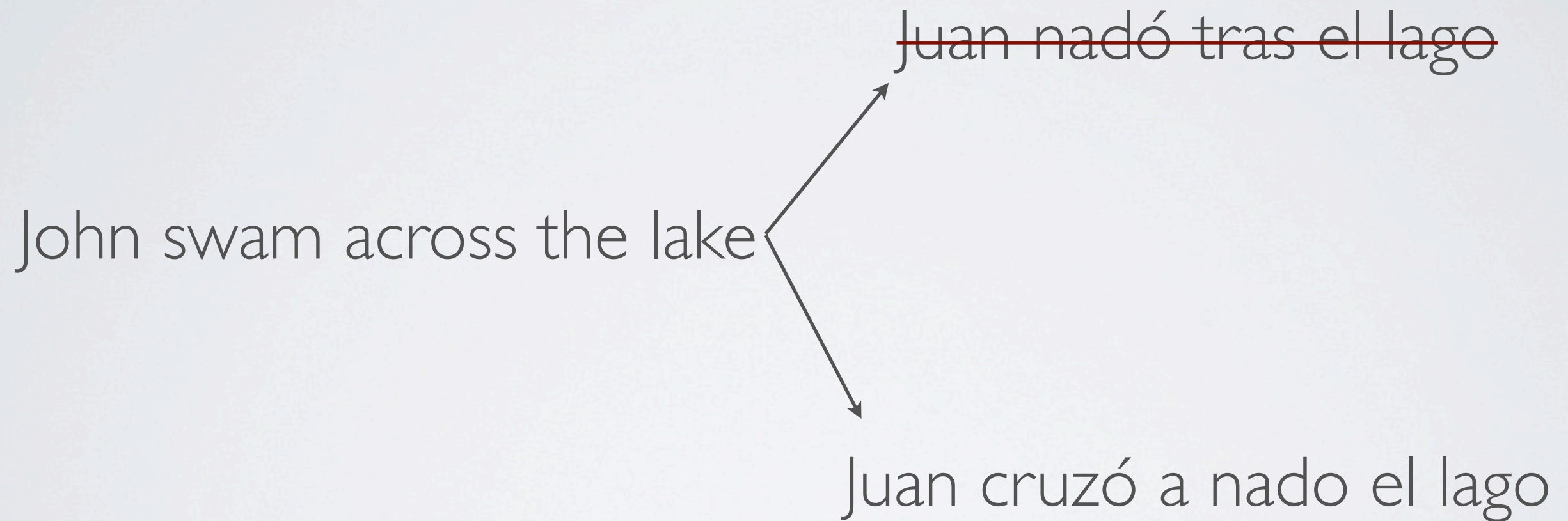
WHY IS MT HARD?

Ambiguity



WHY IS MT HARD?

Divergence



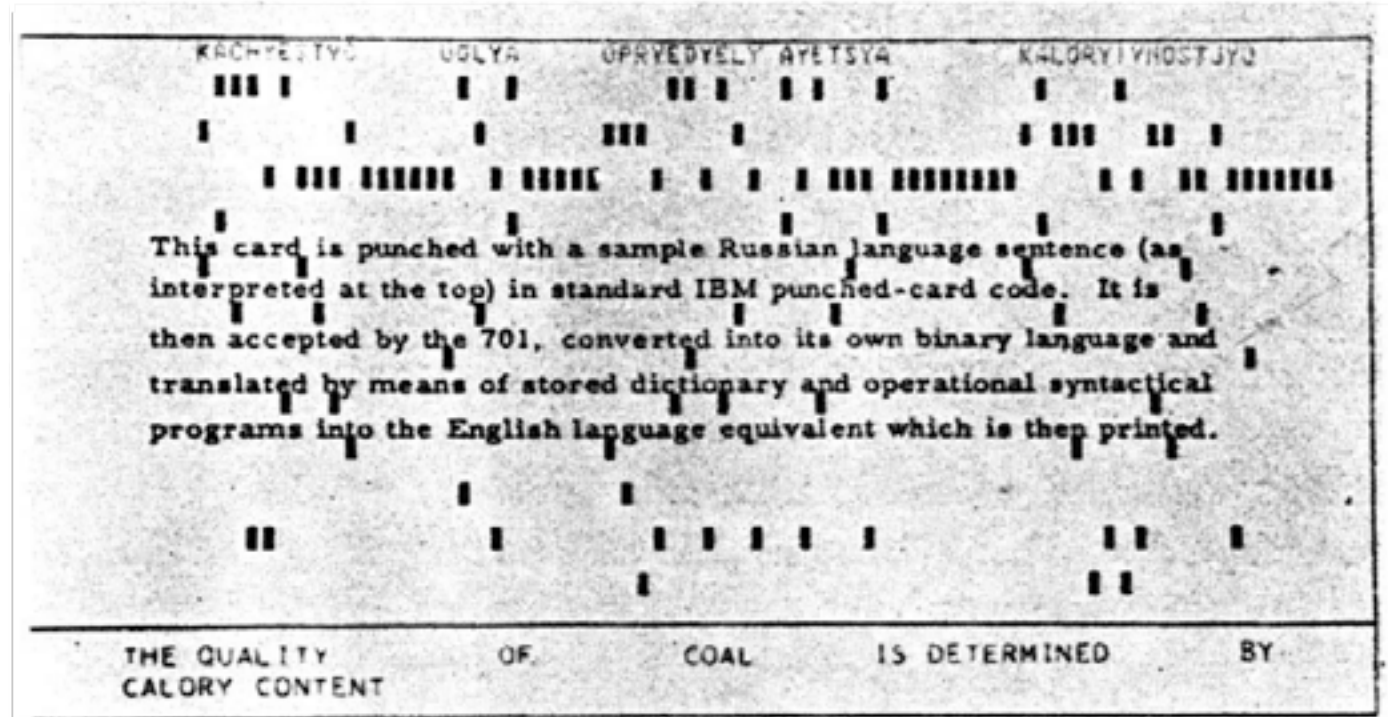
IN THE BEGINNING



One naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say: “This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.”

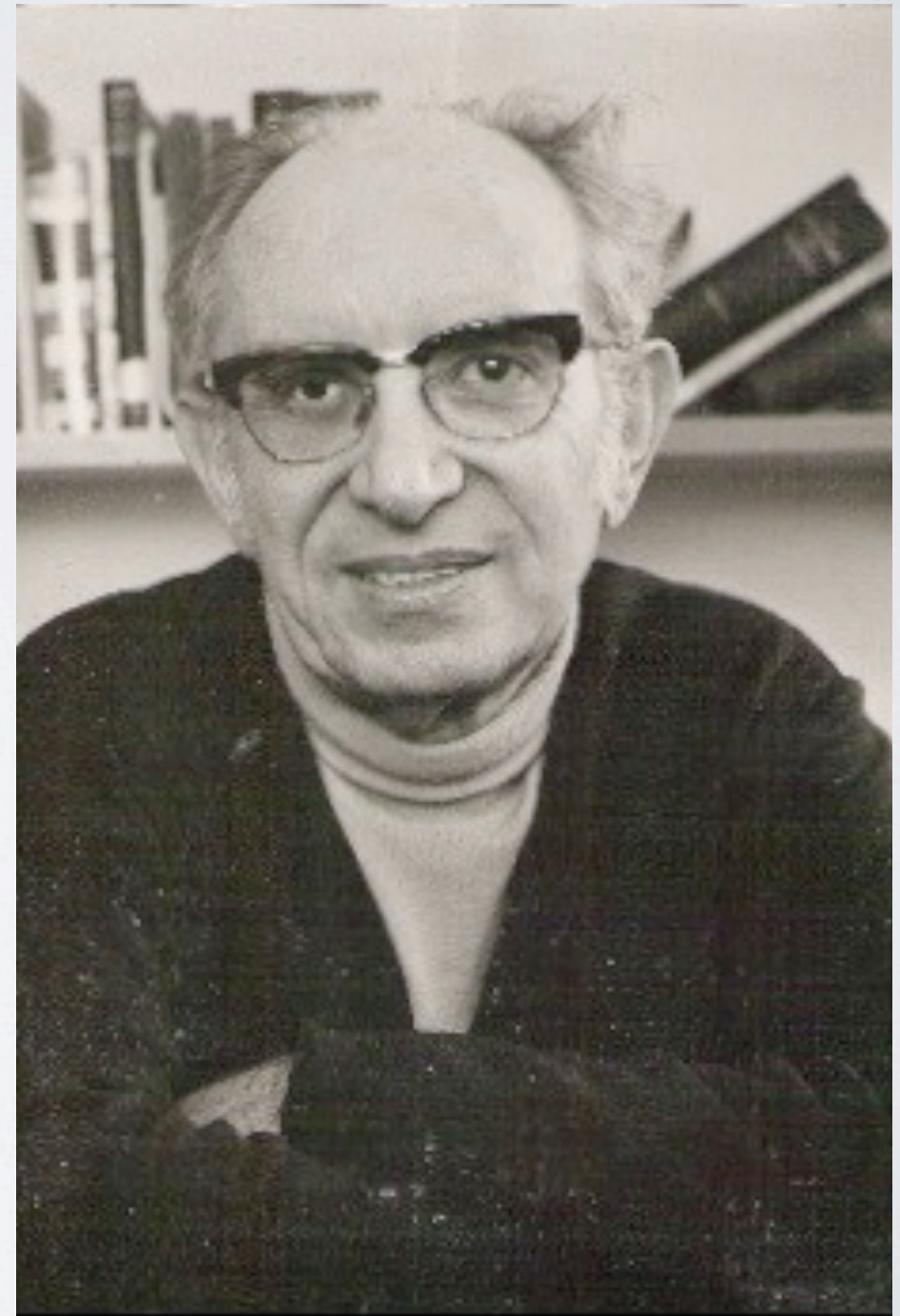
Warren Weaver, 1947

IBM-GEORGETOWN



BAR-HILLEL

- Syntactic transfer
- “Semantic barrier”: *The box was in the pen*



ALPAC REPORT

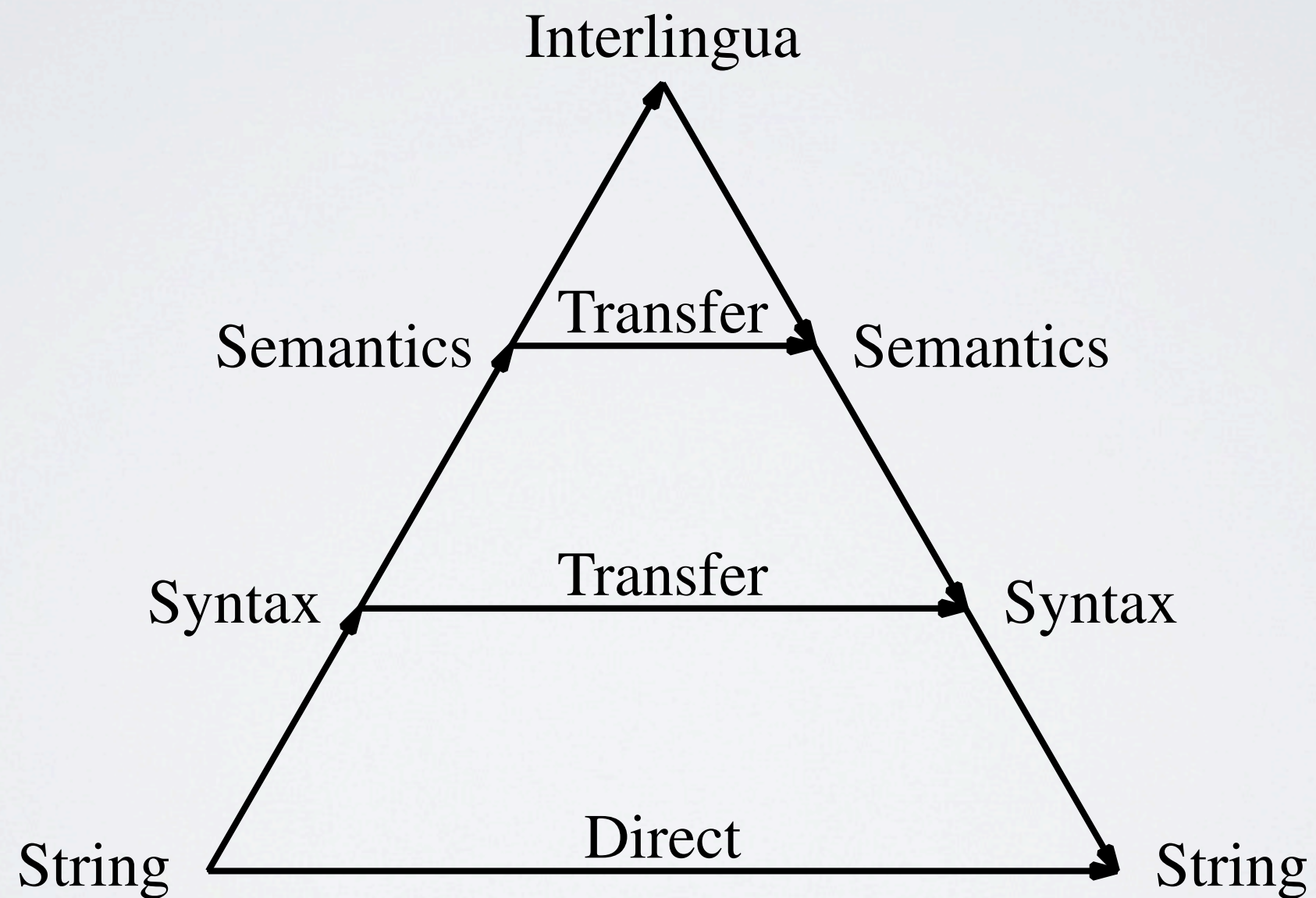
LANGUAGE AND MACHINES

COMPUTERS IN TRANSLATION AND LINGUISTICS

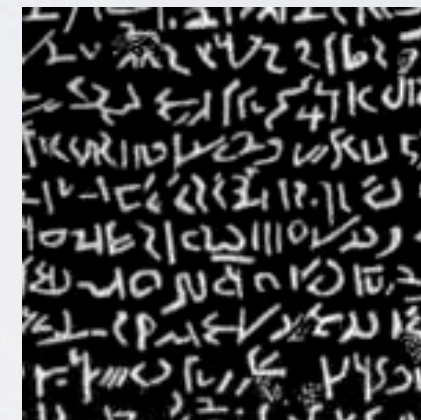
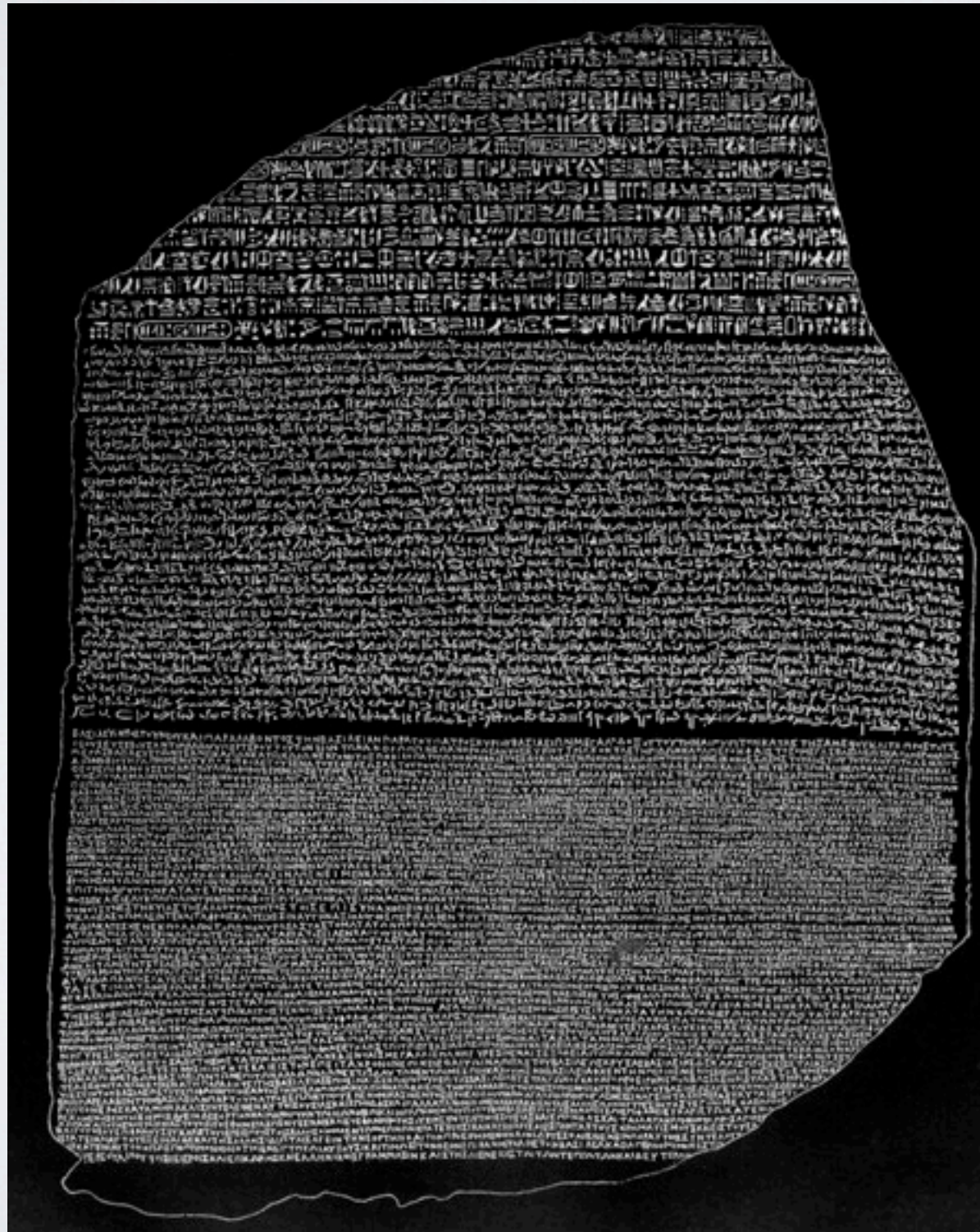
A Report by the
Automatic Language Processing Advisory Committee
Division of Behavioral Sciences
National Academy of Sciences
National Research Council

Publication 1416
National Academy of Sciences National Research Council
Washington, D. C. 1966

RULE-BASED MT



STATISTICAL MT



A SCI-FI EXAMPLE

(KNIGHT, 1997)

Your assignment: translate this
Centauri sentence into Arcturan

farok crrrok hihok yorok klok kantok ok-yurp

farok crrrok hihok yorok klok kantok ok-yurp

1c. ok-voon ororok sprok . 1a. at-voon bichat dat .	7c. lalok farok ororok lalok sprok izok enemok . 7a. wat jjat bichat wat dat vat eneat .
2c. ok-drubel ok-voon anak plok sprok . 2a. at-drubel at-voon pippat rrat dat .	8c. lalok brok anak plok nok . 8a. iat lat pippat rrat nnat .
3c. erok sprok izok hihok ghrok . 3a. totat dat arrat vat hilat .	9c. wiwok nok izok kantok ok-yurp . 9a. totat nnat quat oloat at-yurp .
4c. ok-voon anak drok brok jok . 4a. at-voon krat pippat sat lat .	10c. lalok mok nok yorok ghrok klok . 10a. wat nnat gat mat bat hilat .
5c. wiwok farok izok stok . 5a. totat jjat quat cat .	11c. lalok nok crrrok hihok yorok zanzanok . 11a. wat nnat arrat mat zanzanat .
6c. lalok sprok izok jok stok . 6a. wat dat krat quat cat .	12c. lalok rarok nok izok hihok mok . 12a. wat nnat forat arrat vat gat .

(Knight, 1997)

farok crrrok hihok yorok klok kantok ok-yurp

1c. ok-voon ororok sprok . 1a. at-voon bichat dat .	7c. lalok farok ororok lalok sprok izok enemok . 7a. wat jjat bichat wat dat vat eneat .
2c. ok-drubel ok-voon anak plok sprok . 2a. at-drubel at-voon pippat rrat dat .	8c. lalok brok anak plok nok . 8a. iat lat pippat rrat nnat .
3c. erok sprok izok hihok ghrok . 3a. totat dat arrat vat hilat .	9c. wiwok nok izok kantok ok-yurp . 9a. totat nnat quat oloat at-yurp .
4c. ok-voon anak drok brok jok . 4a. at-voon krat pippat sat lat .	10c. lalok mok nok yorok ghrok klok . 10a. wat nnat gat mat bat hilat .
5c. wiwok farok izok stok . 5a. totat jjat quat cat .	11c. lalok nok crrrok hihok yorok zanzanok . 11a. wat nnat arrat mat zanzanat .
6c. lalok sprok izok jok stok . 6a. wat dat krat quat cat .	12c. lalok rarok nok izok hihok mok . 12a. wat nnat forat arrat vat gat .

(Knight, 1997)

farok **crrrok** hihok yorok klok kantok ok-yurp

1c. ok-voon ororok sprok . 1a. at-voon bichat dat .	7c. lalok farok ororok lalok sprok izok enemok . 7a. wat jjat bichat wat dat vat eneat .
2c. ok-drubel ok-voon anak plok sprok . 2a. at-drubel at-voon pippat rrat dat .	8c. lalok brok anak plok nok . 8a. iat lat pippat rrat nnat .
3c. erok sprok izok hihok ghrok . 3a. totat dat arrat vat hilat .	9c. wiwok nok izok kantok ok-yurp . 9a. totat nnat quat oloat at-yurp .
4c. ok-voon anak drok brok jok . 4a. at-voon krat pippat sat lat .	10c. lalok mok nok yorok ghrok klok . 10a. wat nnat gat mat bat hilat .
5c. wiwok farok izok stok . 5a. totat jjat quat cat .	11c. lalok nok crrrok hihok yorok zanzanok . 11a. wat nnat arrat mat zanzanat .
6c. lalok sprok izok jok stok . 6a. wat dat krat quat cat .	12c. lalok rarok nok izok hihok mok . 12a. wat nnat forat arrat vat gat .

(Knight, 1997)

farok crrrok **hihok** yorok klok kantok ok-yurp

1c. ok-voon ororok sprok . 1a. at-voon bichat dat .	7c. lalok farok ororok lalok sprok izok enemok . 7a. wat jjat bichat wat dat vat eneat .
2c. ok-drubel ok-voon anak plok sprok . 2a. at-drubel at-voon pippat rrat dat .	8c. lalok brok anak plok nok . 8a. iat lat pippat rrat nnat .
3c. erok sprok izok hihok ghirok . 3a. totat dat arrat vat hilat .	9c. wiwok nok izok kantok ok-yurp . 9a. totat nnat quat oloat at-yurp .
4c. ok-voon anak drok brok jok . 4a. at-voon krat pippat sat lat .	10c. lalok mok nok yorok ghirok klok . 10a. wat nnat gat mat bat hilat .
5c. wiwok farok izok stok . 5a. totat jjat quat cat .	11c. lalok nok crrrok hihok yorok zanzanok . 11a. wat nnat arrat mat zanzanat .
6c. lalok sprok izok jok stok . 6a. wat dat krat quat cat .	12c. lalok rarok nok izok hihok mok . 12a. wat nnat forat arrat vat gat .

(Knight, 1997)

farok crrrok hihok **yorok** klok kantok ok-yurp

1c. ok-voon ororok sprok . 1a. at-voon bichat dat .	7c. lalok farok ororok lalok sprok izok enemok . 7a. wat jjat bichat wat dat vat eneat .
2c. ok-drubel ok-voon anak plok sprok . 2a. at-drubel at-voon pippat rrat dat .	8c. lalok brok anak plok nok . 8a. iat lat pippat rrat nnat .
3c. erok sprok izok hihok ghrok . 3a. totat dat arrat vat hilat .	9c. wiwok nok izok kantok ok-yurp . 9a. totat nnat quat oloat at-yurp .
4c. ok-voon anak drok brok jok . 4a. at-voon krat pippat sat lat .	10c. lalok mok nok yorok ghrok klok . 10a. wat nnat gat mat bat hilat .
5c. wiwok farok izok stok . 5a. totat jjat quat cat .	11c. lalok nok crrrok hihok yorok zanzanok . 11a. wat nnat arrat mat zanzanat .
6c. lalok sprok izok jok stok . 6a. wat dat krat quat cat .	12c. lalok rarok nok izok hihok mok . 12a. wat nnat forat arrat vat gat .

(Knight, 1997)

farok crrrok hihok yorok **clock** kantok ok-yurp

<p>1c. ok-voon ororok sprok .</p> <p>1a. at-voon bichat dat .</p>	<p>7c. lalok farok ororok lalok sprok izok enemok .</p> <p>7a. wat jjat bichat wat dat vat eneant .</p>
<p>2c. ok-drubel ok-voon anak plok sprok .</p> <p>2a. at-drubel at-voon pippat rrat dat .</p>	<p>8c. lalok brok anak plok nok .</p> <p>8a. iat lat pippat rrat nnat .</p>
<p>3c. erok sprok izok hihok ghirok .</p> <p>3a. totat dat arrat vat hilat .</p>	<p>9c. wiwok nok izok kantok ok-yurp .</p> <p>9a. totat nnat quat oloat at-yurp .</p>
<p>4c. ok-voon anak drok brok jok .</p> <p>4a. at-voon krat pippat sat lat .</p>	<p>10c. lalok mok nok yorok ghirok clock .</p> <p>10a. wat nnat gat mat bat hilat .</p>
<p>5c. wiwok farok izok stok .</p> <p>5a. totat jjat quat cat .</p>	<p>11c. lalok nok crrrok hihok yorok zanzanok .</p> <p>11a. wat nnat arrat mat zanzanat .</p>
<p>6c. lalok sprok izok jok stok .</p> <p>6a. wat dat krat quat cat .</p>	<p>12c. lalok rarok nok izok hihok mok .</p> <p>12a. wat nnat forat arrat vat gat .</p>

(Knight, 1997)

farok crrrok hihok yorok klok kantok ok-yurp

<p>1c. ok-voon ororok sprok .</p> <p>1a. at-voon bichat dat .</p>	<p>7c. lalok farok ororok lalok sprok izok enemok .</p> <p>7a. wat jjat bichat wat dat vat eneat .</p>
<p>2c. ok-drubel ok-voon anak plok sprok .</p> <p>2a. at-drubel at-voon pippat rrat dat .</p>	<p>8c. lalok brok anak plok nok .</p> <p>8a. iat lat pippat rrat nnat .</p>
<p>3c. erok sprok izok hihok ghirok .</p> <p>3a. totat dat arrat vat hilat .</p>	<p>9c. wiwok nok izok kantok ok-yurp .</p> <p>9a. totat nnat quat oloat at-yurp .</p>
<p>4c. ok-voon anak drok brok jok .</p> <p>4a. at-voon krat pippat sat lat .</p>	<p>10c. lalok mok nok yorok ghirok klok .</p> <p>10a. wat nnat gat mat bat hilat .</p>
<p>5c. wiwok farok izok stok .</p> <p>5a. totat jjat quat cat .</p>	<p>11c. lalok nok crrrok hihok yorok zanzanok .</p> <p>11a. wat nnat arrat mat zanzanat .</p>
<p>6c. lalok sprok izok jok stok .</p> <p>6a. wat dat krat quat cat .</p>	<p>12c. lalok rarok nok izok hihok mok .</p> <p>12a. wat nnat forat arrat vat gat .</p>

(Knight, 1997)

A SCI-FI EXAMPLE

(KNIGHT, 1997)

Your assignment: translate this
Centauri sentence into Arcturan

farok crrrok hihok yorok klok kantok ok-yurp
jjat arrat mat bat oloat at-yurp

Next: put the Arcturan words in Arcturan order

Clients do not sell pharmaceuticals in Europe .

<p>1e. Garcia and associates . 1s. Garcia y asociados .</p>	<p>7e. the clients and the associates are enemies . 7s. los clients y los asociados son enemigos .</p>
<p>2e. Carlos Garcia has three associates . 2s. Carlos Garcia tiene tres asociados .</p>	<p>8e. the company has three groups . 8s. la empresa tiene tres grupos .</p>
<p>3e. his associates are not strong . 3s. sus asociados no son fuertes .</p>	<p>9e. its groups are in Europe . 9s. sus grupos estan en Europa .</p>
<p>4e. Garcia has a company also . 4s. Garcia tambien tiene una empresa .</p>	<p>10e. the modern groups sell strong pharmaceuticals . 10s. los grupos modernos venden medicinas fuertes .</p>
<p>5e. its clients are angry . 5s. sus clientes estan enfadados .</p>	<p>11e. the groups do not sell zenzanine . 11s. los grupos no venden zanzanina .</p>
<p>6e. the associates are also angry . 6s. los asociados tambien estan enfadados .</p>	<p>12e. the small groups are not modern . 12s. los grupos pequenos no son modernos .</p>

(Knight, 1997)

TRANSLATION AS DECODING



$$\begin{aligned}\arg \max_e p(e \mid f) &= \arg \max_e p(e, f) \\ &= \arg \max_e p(e)p(f \mid e)\end{aligned}$$

OVERVIEW

- Word-based alignment and translation
 - Language models
 - Evaluation
- Phrase-based translation
 - Discriminative training
- Interlude: Subword translation
- Syntax-based translation