

Traducción Automática.

Igor Boguslavsky

Universidad Politécnica de
Madrid,

Academia de Ciencias de Rusia

¿Para que sirve el procesamiento automático de lenguaje natural?

- Ayuda en preparación de textos
- Recuperación de información
- Extracción de información
- Respuesta a las preguntas
- Interfaces en lenguaje natural
- Procesamiento de voz
- Generación de texto
- Traducción automática
- ...

Extracción de información: nombramientos o despido de los gerentes

El vicepresidente segundo y ministro de Economía, Pedro Solbes, ha elegido a Miguel Ángel Fernández Ordóñez para gobernador del Banco de España. Ocupará el puesto de Julio Segura, que pasará a ser consejero de la Comisión Nacional del Mercado de Valores

PERSONA: *Miguel Ángel Fernández Ordóñez*

PUESTO: *gobernador*

ORGANIZACIÓN: *el Banco de España*

EMPEZAR / DEJAR: *empezar*

PERSONA: *Julio Segura*

PUESTO: *gobernador*

ORGANIZACIÓN: *el Banco de España*

EMPEZAR / DEJAR: *dejar*

PERSONA: *Julio Segura*

PUESTO: *consejero*

ORGANIZACIÓN: *Comisión Nacional del Mercado ...*

EMPEZAR / DEJAR: *empezar*

Universal Networking Language: ¿de dónde?

- Barrera lingüística en la sociedad de información
- Un problema global de la humanidad: Universidad de ONU, Tokyo.
- Hiroshi Uchida, el autor del sistema TA ATLAS
- 15 grupos de varios países

UNL vs. Traducción automática

- Sistemas TA existentes no pueden resolver el problema
 - Pocas lenguas
 - Baja calidad
- ¿Por qué UNL es mejor?
 - Estrategia de interlingua
 - Análisis semiautomático

Estrategias de traducción automática



Traducción directa

What's the time? →

Traducción directa

What's the time? → *Qué*

Traducción directa

What's the time? → *Qué es*

Traducción directa

What's the time? → **Qué es el tiempo?*

Traducción directa

What's the time? → **Qué es el tiempo?*

Qué hora es? →

Traducción directa

What's the time? → **Qué es el tiempo?*

Qué hora es? → *What*

Traducción directa

What's the time? → **Qué es el tiempo?*

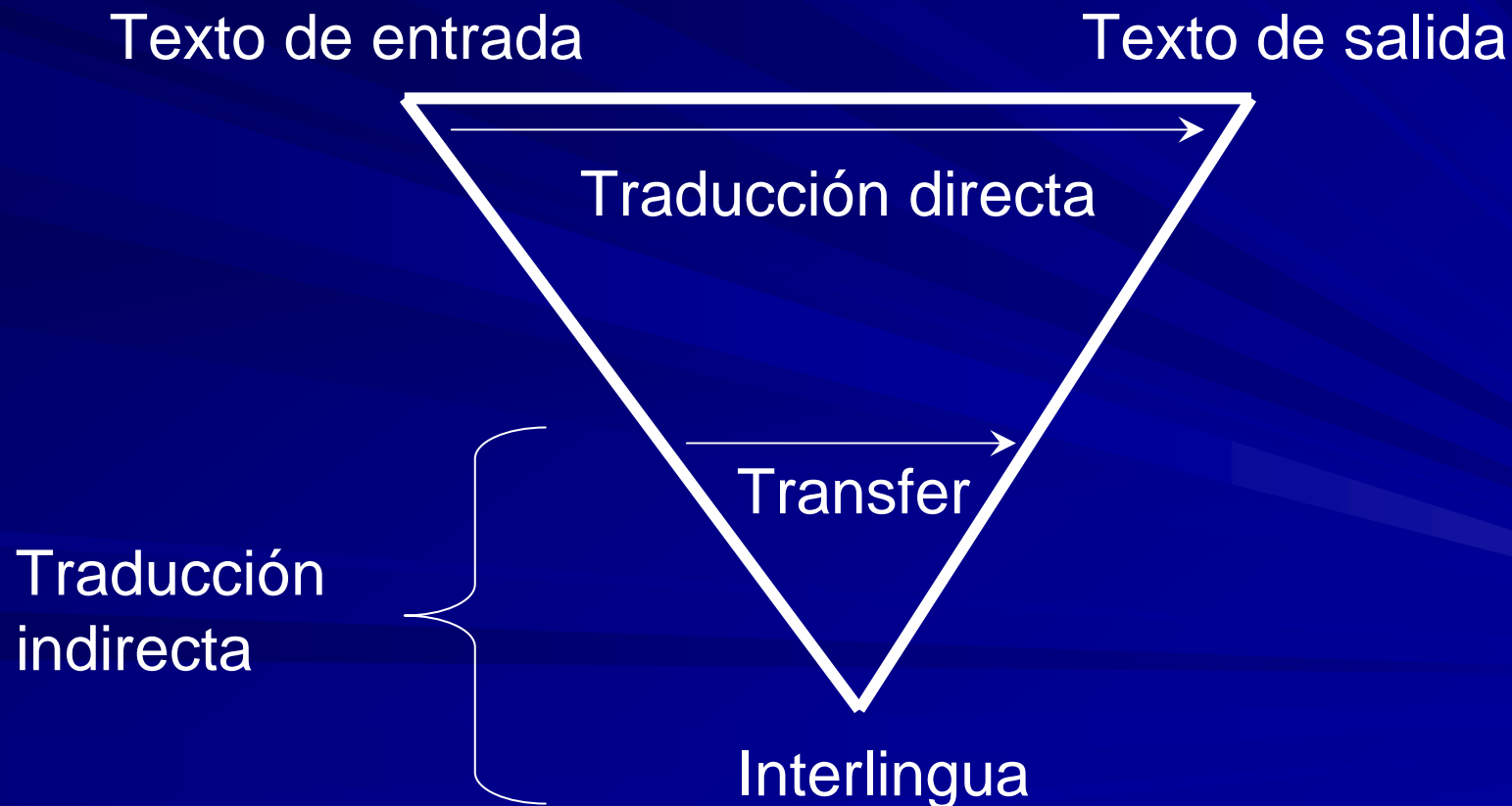
Qué hora es? → *What hour*

Traducción directa

What's the time? → **Qué es el tiempo?*

Qué hora es? → **What hour is?*

Estrategias de traducción automática



Una demo

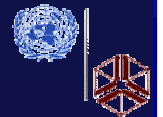
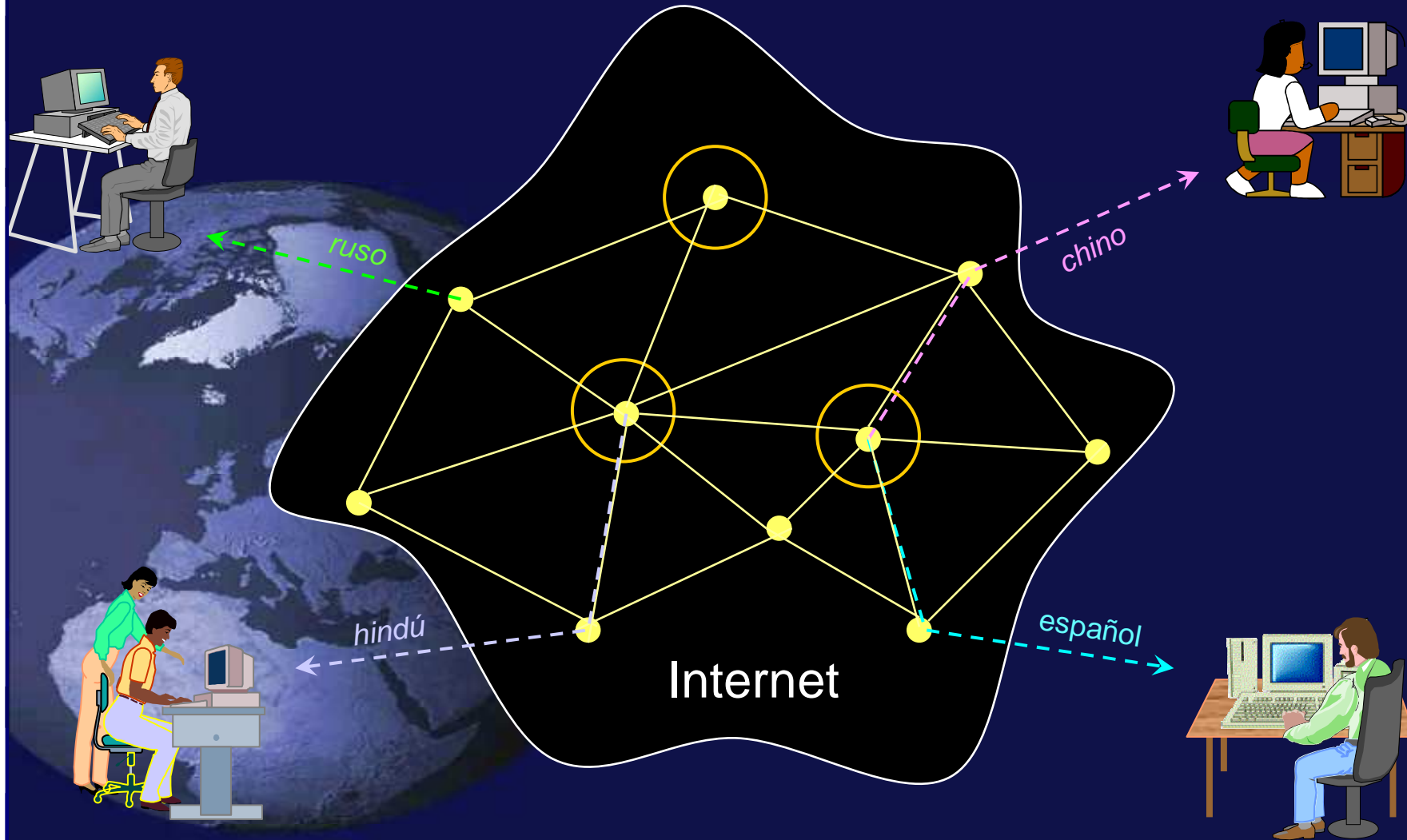
- *We made a general move.*
 - general1 (adjetivo) - general2 (sustantivo)
 - move1 (verbo) - move2 (sustantivo)

Universal Networking Language

- Meaning representation language
- Deconverter: UNL \Rightarrow NL
- Enconverter: NL \Rightarrow UNL
- Deconverter + Enconverter = Language Server
- El sistema UNL = una red de los servidores de lenguas



Sistema UNL



¿Como construir una interlingua?

- Palabras
- Sintaxis

Universal Words

Dos enfoques al diseño de las palabras para una lengua semántica:

- Hay una correspondencia uno a uno entre las unidades semánticas y sentidos léxicos.
- Unidades semánticas componen un pequeño grupo de los primitivos conceptuales que se combinan para formar sentidos mas grandes de las palabras.

Primer enfoque

- Ventaja: correspondencia trivial entre unidades semánticas y palabras del lenguaje natural
- Defecto: sentidos léxicos de varias lenguas no coinciden
 - *corner* --> CORNER
 - *rincón* (inner corner) vs. *esquina* (outside corner)
 - INSIDE_CORNER + OUTSIDE_CORNER

Más ejemplos (1)

- Japonés:

‘hermano mayor’ vs. ‘hermano menor’

Ruso:

zhenit'sja ‘casarse (de un varón)’ vs.
vyxodit' zamuzh ‘casarse (de una mujer)’

Más ejemplos (2)

- Sueco:

mormor 'madre de la madre' vs. *farmor*
'madre del padre'

- Malayo:

'arroz en el campo (non recogido)' vs.
'arroz non cocido' vs. 'arroz cocido' vs.
'cocido como papilla' vs. 'arroz preparado
como puré' vs. 'arroz glutinoso'

Más ejemplos (3)

- Japonés:

‘llevar o ponerse (en general)’ vs. ‘abrigo o cazadora’ vs. ‘calzados o pantalones’ vs. ‘sombrero’ vs. ‘anillo o guantes’ vs. ‘cinturón, corbata o tapabocas’ vs. ‘gafas o collar vs. ‘brocha’

¿Como construir una lengua semántica según el enfoque 1?

- Elegir una lengua para servir de base.
- Introducir palabras nuevas según las distinciones encontradas en otras lenguas:
 - 2 palabras para *corner* debido al español,
 - 2 palabras para *el río* debido al francés,
 - 2 palabras para *casarse* debido al ruso,
 - 6 palabras para *arroz* debido al malayo,
 - 8 palabras para *llevar* debido al japonés.

Distinción excesiva

Inglés

rice

Japonés

‘arroz en el
campo’
‘arroz non
cocido’
‘arroz cocido’
‘cocido como
papilla’
‘arroz preparado
como puré’
‘arroz glutinoso’

Español

arroz

Segundo enfoque: primitivos semánticos

- *comer* = INGERIR + SÓLIDO
- *beber* = INGERIR + LIQUIDO
- *essen* = INGERIR + SÓLIDO + HUMANO
- *fressen* = INGERIR + SÓLIDO + ANIMAL

¡insuficiente!: la representación debe ser estructurada.

¿Quién es HUMANO en *essen* - él que come o él que está comido?

Ejemplo

- inglés: *He drove into the town*
- español: *Entró en la ciudad en coche*

- *drive* = 'TYPE: move; INSTR: car'
into = 'DIRECTION: into'
- *entrar* = 'TYPE: move; DIRECTION: into'
en coche = 'INSTR: car'

Desventaja

Demasiado complicado!

- Encontrar el conjunto de nociones simples y suficientes para representar todas las distinciones de todas lenguas
- Descomponer los sentidos de todas las palabras por medio de estas nociones

¿Qué se hace en UNL?

- Los sentidos de palabras de todas las lenguas se representan con las palabras de UNL – Palabras Universales (UW).
- UWs deben ser suficientemente sencillas para ser comprensibles para los hablantes de varias lenguas y portadores de varias culturas.
- UWs deben ser suficientemente complejas para poder expresar los sentidos específicos de varias lenguas.

Estos requisitos están garantizados debido a:

- UWs se construyen sobre la base de las palabras de una lengua natural (inglés).
- UNL cuenta con ciertos medios para modificar los sentidos de palabras inglesas y hacerlos más cercanos a los de palabras locales.
- Eso se asegura gracias a las restricciones que acompañan palabras inglesas (headwords).

UW

- Una UW corresponde a un sentido de una palabra de un idioma
- El diccionario UNL es el conjunto de todos sentidos expresados por palabras de los lenguajes naturales

¿Cómo se representa la ambigüedad léxica? *COACH*

coach1.1 (= autobús)

coach1.2 (= vagón)

coach1.3 (= carruaje)

coach2.1 (= entrenar a los deportistas)

coach2.2 (= preparar a un alumno para un examen)

coach3.1 (= entrenador)

coach3.2 (= repetidor)

coach(icl>public_transport) = *coach1.1*

coach(icl>car) = *coach1.2*

coach(icl>carriage) = *coach1.3*

coach(icl>trainer) = *coach3.1*

coach(icl>teacher) = *coach3.2*

coach(icl>do,obj>sportsman) = *coach2.1*,

coach(icl>do,obj>pupil) = *coach2.2*.

Conceptos ausentes en inglés

Encontrar la palabra inglesa más cercana con el sentido más ancho y estrecharlo con las restricciones:

- *pez* ⇒ fish(icl>living thing);
- pescado* ⇒ fish(icl>food)
- *zhenitsja* ⇒ marry(agt>male)
- vyjodit zamuzh* ⇒ marry(agt>female)

come

priletet' – 'llegar volando'

priplyt' – 'llegar nadando o
navegando'

pripolzti – 'llegar arrastrandose'

pribezhat' – 'llegar corriendo'

come(met>run)

come(met>plane)

Estrategia UNL

- Sentidos léxicos de cada lengua se representan por medio de las UWs.
- Si el sentido léxico no tiene equivalente en inglés, hay que encontrar la palabra inglesa mas cercana con el sentido mas general y restringir su sentido.

Soluciones del problema de deferencias léxicas

- Despreciar la restricción:
grandmother(com>father) => *abuela*
grandmother(com>mother) => *abuela*
- Exteriorizar la restricción:
come(met>swim) => *llegar nadando*
- Elegir la palabra más cercana por medio de la Base de Conocimientos UNL

Base de Conocimientos UNL

casarse

marry(agt>person)

zhenit'sja

marry(agt>male)

vyxodit' zamuzh

marry(agt>female)



Expresión UNL

- Expresión UNL es un hipergrafo orientado
- Nodos - Universal Words (UW)
- Arcos – relaciones semánticas que conectan dos UWs
- UW pueden ser dotadas con los atributos que den información suplementaria

Hipernodos

- Hay dos clases de nodos: simples y compuestos (hipernodos).
- El hipernodo es un subgrafo que funciona como un nodo único:
 - puede recibir y emitir relaciones
 - puede tener sus propios atributos
- El hipernodo puede contener nodos simples y otros hipernodos.

Hipernodos

la convergencia de la tecnología y cultura moderna

- (tecnología y cultura) moderna
- tecnología y (cultura moderna)

(tecnología y cultura) moderna

