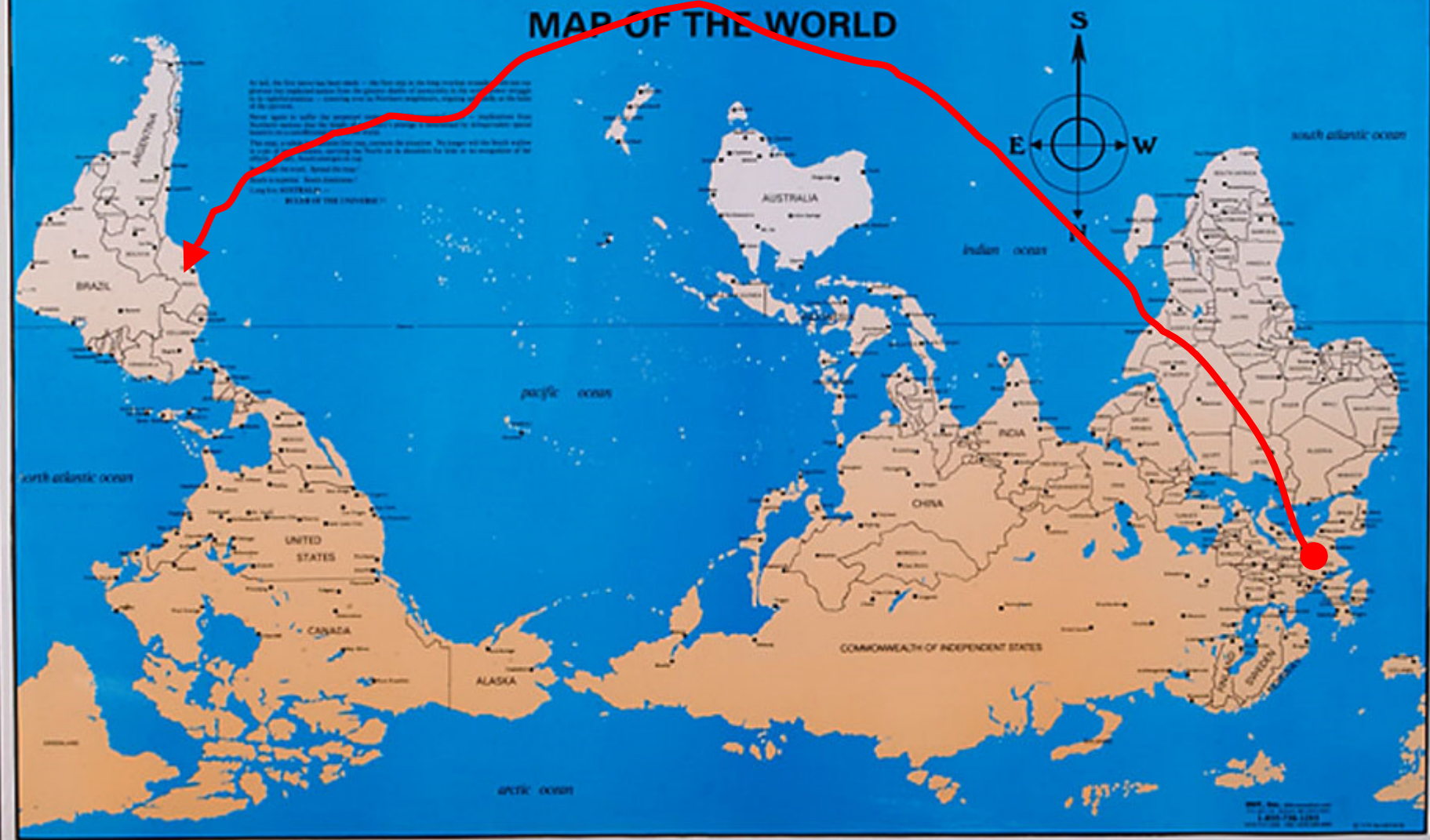
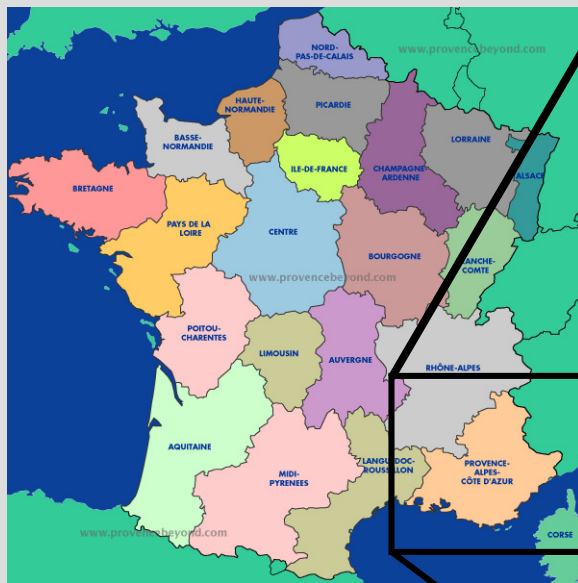


McARTHUR'S UNIVERSAL CORRECTIVE MAP OF THE WORLD

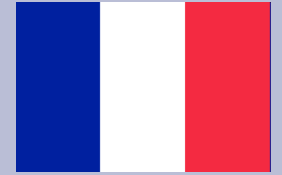




Region Provence-Alpes-Côte d'Azur





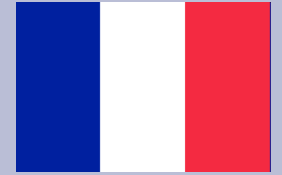


Avignon, Francia

Laboratoire Informatique d'Avignon (LIA)



LABORATOIRE
INFORMATIQUE D'AVIGNON
Université d'Avignon et des Pays de Vaucluse

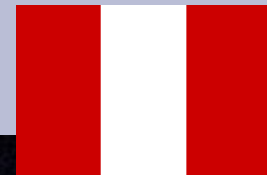


¿ *Avignon* o *Aviñón* ?
¿ *France* o *Francia* ?

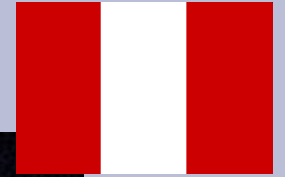
Laboratorio Informático de Aviñón (LIA)



LABORATOIRE
INFORMATIQUE D'AVIGNON
Université d'Avignon et des Pays de Vaucluse

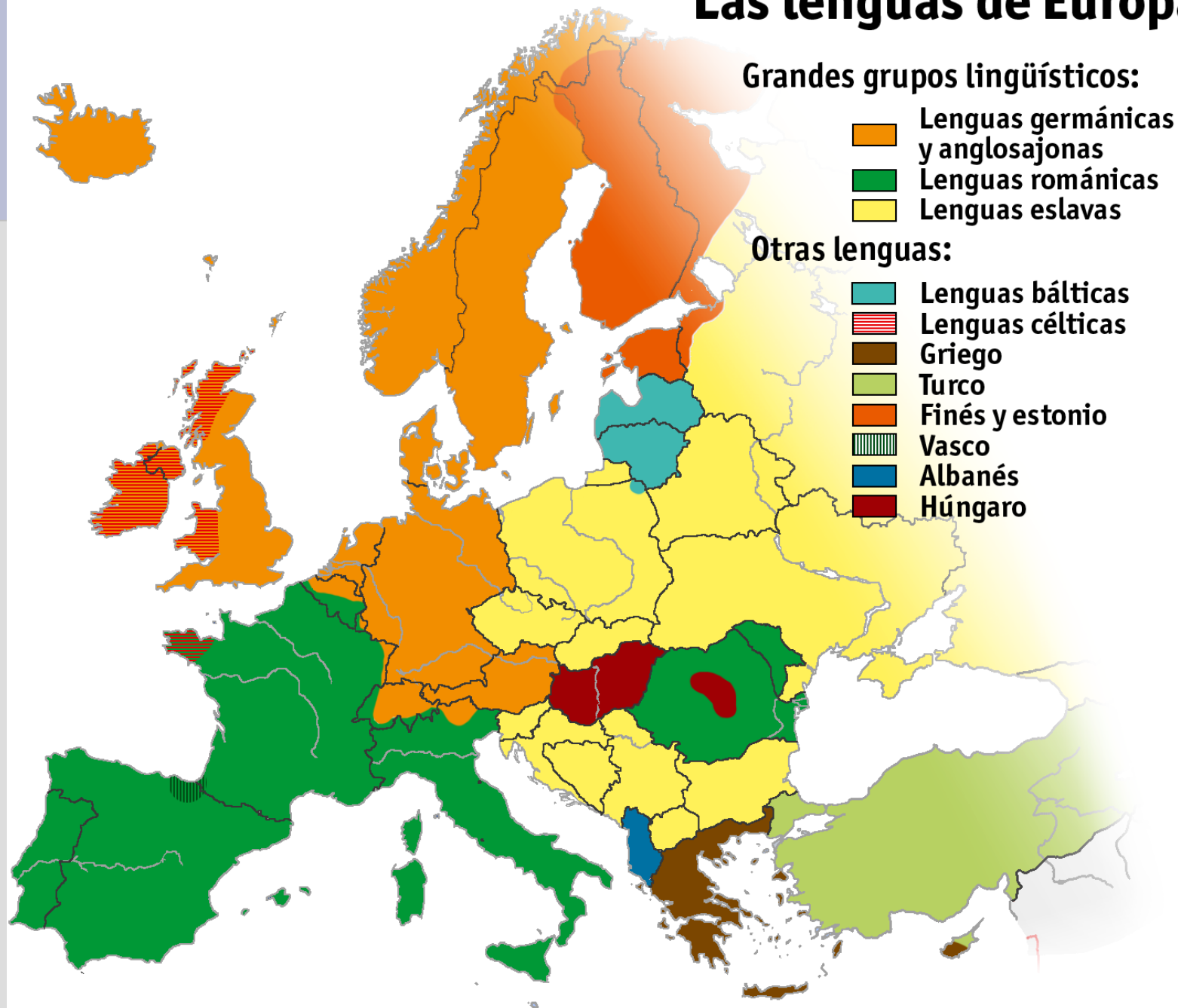


¿ Perú o *Pérou* ?



¿ Plaza de Armas? o *Place d'Armes* ?

Las lenguas de Europa

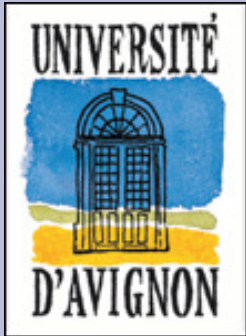


Mapa lingüístico de Europa



Gran diversidad : gran problema...





UNIVERSITÉ D'AVIGNON
ET DES PAYS DE VAUCLUSE



Universidad Privada Antenor Orrego



UPAO

...una gran Universidad!

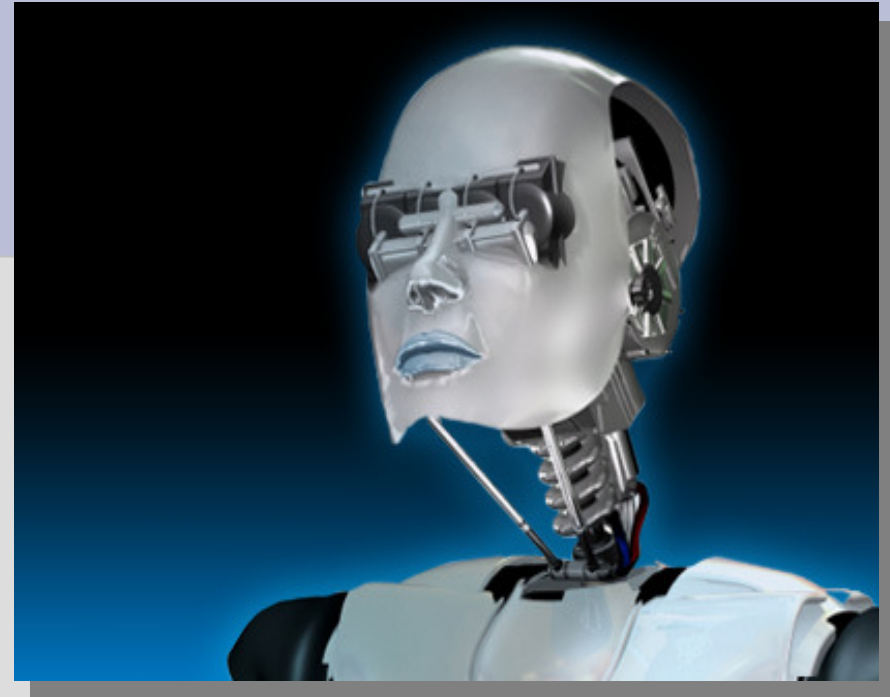
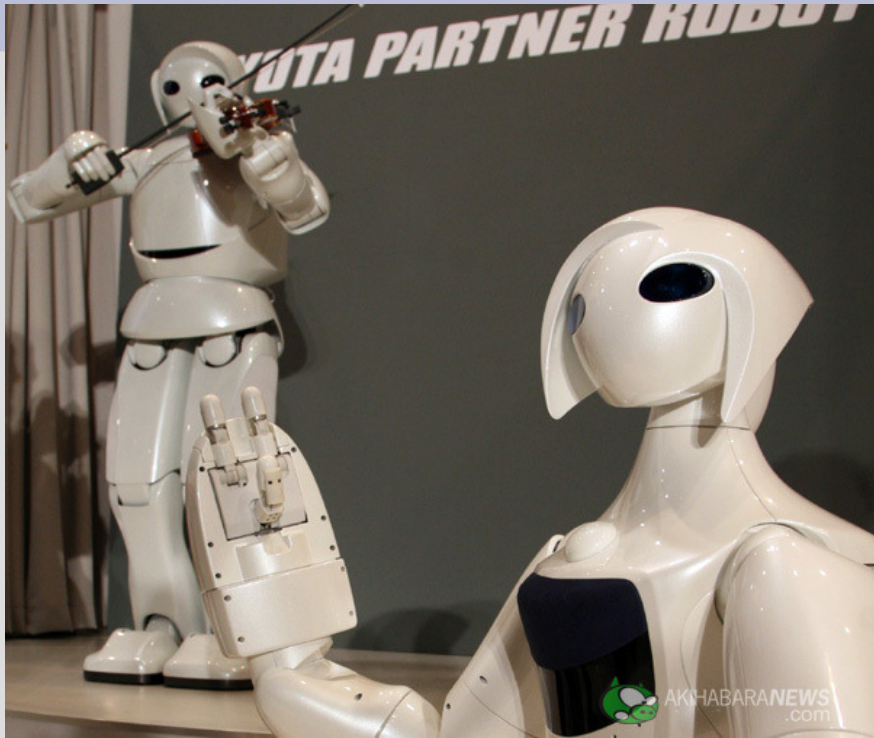
¿Qué es el Procesamiento de Lenguaje Natural?

Juan-Manuel Torres Moreno

`juan-manuel.torres@univ-avignon.fr`

Laboratoire Informatique d'Avignon (France)

**V Congreso Internacional de
Ingeniería de Computación y Sistemas
TRUJILLO**



Inteligencia aunque sea Artificial...

¿Puede imaginarse una plática donde humanos y robots discuten sobre un tema del entorno en que se desenvuelven?



¿Puede imaginarse una plática donde humanos y robots discuten sobre un tema del entorno en que se desenvuelven?

- ¡Eso es sólo ficción!
- ¡O mas aún: ciencia ficción!
- Pero aun siendo ficción, se trata de un objetivo ***atractivo*** que está mereciendo atención especial en ***Inteligencia Artificial***

Procesamiento de Lenguaje Natural (PLN)

- Procesamiento *automático* de lenguajes naturales: español, francés, inglés, chino, alemán, árabe...
- Se hace referencia como **Lingüística computacional**
- Nace por motivos militares :
 - Descifrar mensajes del ejército enemigo
 - Resolver problemas de **traducción automática** (Turing 1954)

PLN

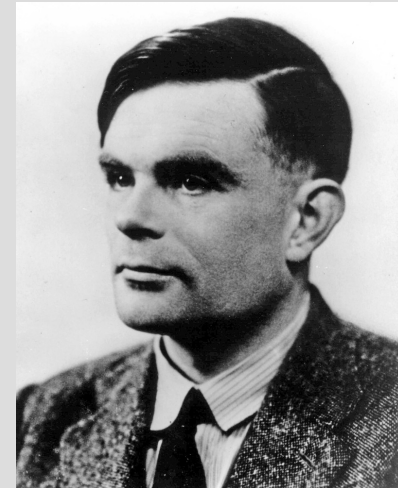
- Problema **multidisciplinario** del que se ocupa:
 - la lingüística
 - la lógica
 - la informática
 - la psicología cognitiva
 - la ingeniería
 - las matemáticas
 - la epistemología...

¿Qué hace el PLN ?

- Investigación en continuo desarrollo
- Aplicaciones:
 - Traducción automática
 - Recuperación/Busqueda de información
 - Elaboración automática de resúmenes
 - Interfaces hombre/máquina
 - Generación de texto
 - Detección de opiniones, emociones
 - Diálogo...

¿Por qué es difícil el PLN ?

- Un sueño desde hace mucho tiempo ¿Por qué tantas dificultades ?
 - Cantidad de lenguas humanas
 - Complejidad de las lenguas
 - Ambigüedad
 - Incertidumbre
 - Fenómenos culturales
 - ...



Alan Turing

¿Qué tipo de textos trata el PLN?

- ¿Textos literarios...?

Bergère ô tour Eiffel le troupeau des ponts bêle ce matin

(Pastor oh Torre Eiffel el rebaño de puentes bala esta mañana)

Apollinaire, Alcools 1913

À la nue accablante tu

S. Mallarmé

(A la pesadisima nube/desnuda tu)

- ¿Documentos (científicos, periodísticos)?

- vocabulario técnico
- normalizado
- formas sintácticas suficientemente simples



¿Qué tipo de textos trata el PLN?

- **Texto literario**

- Expresa voluntaria y *ambiguamente* un mensaje *complejo*.
- No puede resumirse o reformularse
- Soporta mal la traducción, los sinónimos...

- **Documento** (científico o periodístico)

- El texto se desvanece frente a su contenido
- El documento envejece rápidamente: textos recientes que deben ser procesados rápidamente para difundir su contenido

¿Cómo hacer artefactos tecnológicos de PLN *reales*?

Ideas intuitivas:

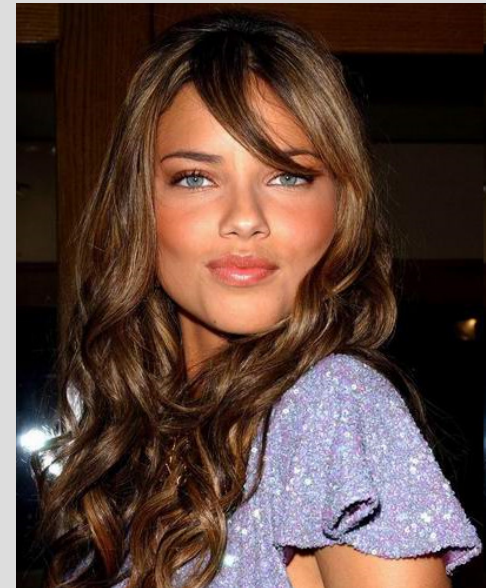
- Si el problema es la lengua... ¡hagamos uso de la **lingüística**!
- La lingüística debe estar en el **centro de la problemática** de PLN
- La lingüística ***debe aportar soluciones*** a los problemas del PLN
- La lingüística *tiene un **marco teórico*** adecuado
- El **análisis** lingüístico es **fino**...

La lingüística estudia las lenguas humanas

- **Paradigma de la lingüística**
 - Una lengua humana puede ser representada por una gramática formal
- **Gramática formal**
 - Conjunto de reglas que deciden si una frase pertenece (*gramaticalmente correcta*) o no (*agramatical*) a una lengua

¿La lingüística resuelve los problemas del PLN?

- **Vi al hombre con el telescopio**
 - ¿Usé un telescopio para ver al hombre?
 - ¿El hombre tenía un telescopio ?
- **Me gusta acariciar los gatos. A tu novia también**
 - Ah bon ??



J'amerais descendre un avocat

Quisiera (*descendre = bajar/matar*) un (*avocat = palta/abogado*)

?

J'amerais descendre un avocat

Quisiera (*descendre* = *bajar/matar*) un (*avocat* = *palta/abogado*)



?



Mmmh ¿Tienes problemas ? No importa... ¿Cuánto pagas?

¿La lingüística resuelve problemas del PLN?

- **Los humanos son muy buenos para**
 - **Resolver la ambigüedad:** elegir el sentido pertinente de una frase
 - **Robustez:** detectan y no aceptan pequeñas desviaciones de las frases
 - **Desempeño:** son capaces de procesar eficazmente frases complejas: ***ellas*** representan desafíos enormes a la lingüística

Ambigüedad

Significación: *llama* (camélido) vs *llama* (fuego) vs *llama* (verbo llamar)



?



Ambigüedad

Etiquetas gramaticales (Análisis sintáctico):

La belle ferme le voile

(a) **La bella granja lo oculta**

(b) **La bella cierra el velo**

(a) **Art Adj Nom Pro Ver**

(b) **Art Nom Ver Art Nom**

?

Ambigüedad

- Las lenguas humanas son ambiguas e inciertas... ¡eso hace su riqueza!
- Dificultad para ser analizadas automáticamente
- Las gramáticas formales producen un número de análisis de frases ***exponencial*** respecto al número de palabras de una frase

Ambigüedad

- **Un sistema PLN basado en lingüística debe resolver entre un gran número de posibles caminos**
- Para un humano es diferente: poseemos vastos recursos extra lingüísticos:
 - Conocimiento del mundo
 - Preferencias culturales
 - Experiencia...

Ambigüedad

- Una gramática formal no tiene acceso a conocimientos extra-lingüísticos
 - *En Peru los ministros pueden ser chuponeados !* 
- Lingüística incapaz de procesar la ambigüedad : no puede resolver la incertidumbre
- Determinista

Robustez

- Frases no-gramaticales

- *John not home* (falta el verbo)

- No gramaticales y ambiguas

- *L'artiste peins la nuit* (Mal acuerdo 3a persona:
*L'artiste **peint** la nuit* ; “pinta la noche” o “pinta
de noche” ?)

- Uso no comun de la lengua



- *La bica es poder la trica la sabiduría*
 - *Quien no tiene de inga tiene de mandinga...*
 - *Coma... sin-vergüenza*

Desempeño

- *Remarcablemente, el **linguista** (computacional o no) **pueden** llegar a pensar que el **modulo comun** de todas las aplicaciones de lenguaje natural debe ser un modelo de procesamiento **linguistico**; y más **aun**, que **deberia** jugar un rol central en la **interaccion** (vocal o escrita) hombre-**maquina**. Sin embargo la **pratica** en vigor para **esas tareas** no incluye absolutamente los modelos reales del **prosesamiento linguistico**.*

¿La lingüística resuelve problemas del PLN?

- Frases largas... estilos
- Uso de sinónimos y antónimos
- Gramática compleja
- Sintaxis pobre: **palavras mal escritaz o mal acéntuadas**

¿Los modelos lingüísticos pueden aun jugar un rol en las tecnologías del lenguaje natural?

La semántica...

- ¿Se necesita comprender un texto para *procesarlo* adecuadamente?
- El caso de **FRUMP**: un sistema que genera resúmenes automáticos por *comprensión* del texto
 - Análisis lingüístico
 - Interpretación usando conocimientos
 - Generación de texto: activación de *scripts* por palabras claves « comprendidas »

FRUMP

a small earthquake shook several Southern Illinois counties Monday night, the National Earthquake Information Service in Golden, Colo., reported. Spokesman Don Finley said the quake measured 3.2 on the Richter scale, "probably not enough to do any damage or cause any injuries." The quake occurred about 7:48 p.m. CST and was centered about 30 miles east of Mount Vernon, Finley said. It was felt in Richland, Clay, Jasper, Effington, and Marion Counties.

(TEXT0) ↵

There was an earthquake in Illinois with a 3.2 richter scale. (RESUMEN) ↵

¡RESUMEN REMARCABLE ↵!

FRUMP : resultados fuera de su “comprensión”

- ¿Los 50 scripts de FRUMP son suficientes para interpretar el mundo?
- Conocimientos *codificados* manualmente
¿Es fácil hacer nuevos scripts para otros temas ?

Ciudad del Vaticano. La noticia de la muerte del Papa sacude al mundo. Murió el martes pasado de forma misteriosa... (TEXTO)

Sismo en el Vaticano : un muerto. (RESUMEN)

¿La lingüística resuelve problemas del PLN?

- *Las gramáticas formales **limitan** las fronteras del lenguaje*
 - *las frases no-gramaticales son clasificadas como no pertenecientes a la lengua*
- *Las frases extrañas, mal escritas o con sintáxis pobre (chat, SMS, e-mail...) no son un problema insoluble para un humano*
 - *La interpretación es elegida en función de su **uso corriente***

¿La lingüística puede procesar ésto?

from <juan.manuel.torres@univ-avignon.fr>

to <carlos.raymundo@upao.edu.pe>

Bonjour!!! el vuelo af estuvo uffftffff!! (+ de 13) grcias por el eticket eh? :-)

A 1/2 noche no dieron de comer un monton y casi nos da indi gestion... Vi n+1 pelis, ya ni se. La n++ je rien compris y hizo que me pusiera super >-(

confirmo @ pasado: doy 1 conf. Cong.upao y 1 +tecnico en el WShop Salu2!

JM

66% de palabras corregidas

from <juan.manuel.torres@univ-avignon.fr>

to <carlos.raymundo@upao.edu.pe>

¡Hola! El vuelo de Air France estuvo ¡cansadísimo!
(Fueron más de 13 horas). Gracias por el boleto
electrónico. Me dio mucho gusto.

A media noche nos dieron de comer muchísimo y casi
nos indigestamos... Vi muchas películas. No sé
cuántas. La última no la entendí e hizo que me
pusiera muy enojado.

Confirmando tu correo electrónico pasado: doy una
conferencia en el congreso de UPAO y un seminario
mas técnico en el Workshop. ¡Saludos!

Juan Manuel

¿La lingüística resuelve problemas del PLN?

- *El desempeño sale de **su** campo*
- *La robustez también*
- *Las gramáticas formales son **inadecuadas** para resolver la ambigüedad porque no pueden resolver la incertidumbre*

¿Qué hacemos entonces?

¿Qué queda?

¿Como atacar el problema de la lengua natural?

Enfoque pragmatico:

- Intuiciones
- Matematicas
- Inteligencia Artificial
- Ingenieria



¡Los métodos probabilistas!

- Abordan la incertitud formalizando un **modelo de lenguaje** como una *distribución de probabilidad*
- Ambigüedad: cada análisis tiene asociado un valor $0 \leq p \leq 1$: indica el **grado de pertenencia** de la frase a la lengua
- Los valores son más que números: probabilidades fundadas en la **Teoría de Probabilidades**

Teoría de probabilidades

- Modelización adecuada de la noción intuitiva de la probabilidad de eventos (independientes o no)
- Camino empírico para estimar las probabilidades : la **estadística**
 - Interpretación de la teoría de probabilidades

Ejemplo 1: traducción automática

- *La belle ferme le voile*

(a) *La bella cierra el velo*

(b) *La bella granja lo oculta*



Ejemplo 1: traducción automática

- ***La belle ferme le voile***
 - (a) *La bella cierra el velo*
 - (b) *La bella granja lo oculta*



Ejemplo 1: traducción automática

- *La belle ferme le voile*

(a) *La bella cierra el velo*

(b) *La bella granja lo oculta*

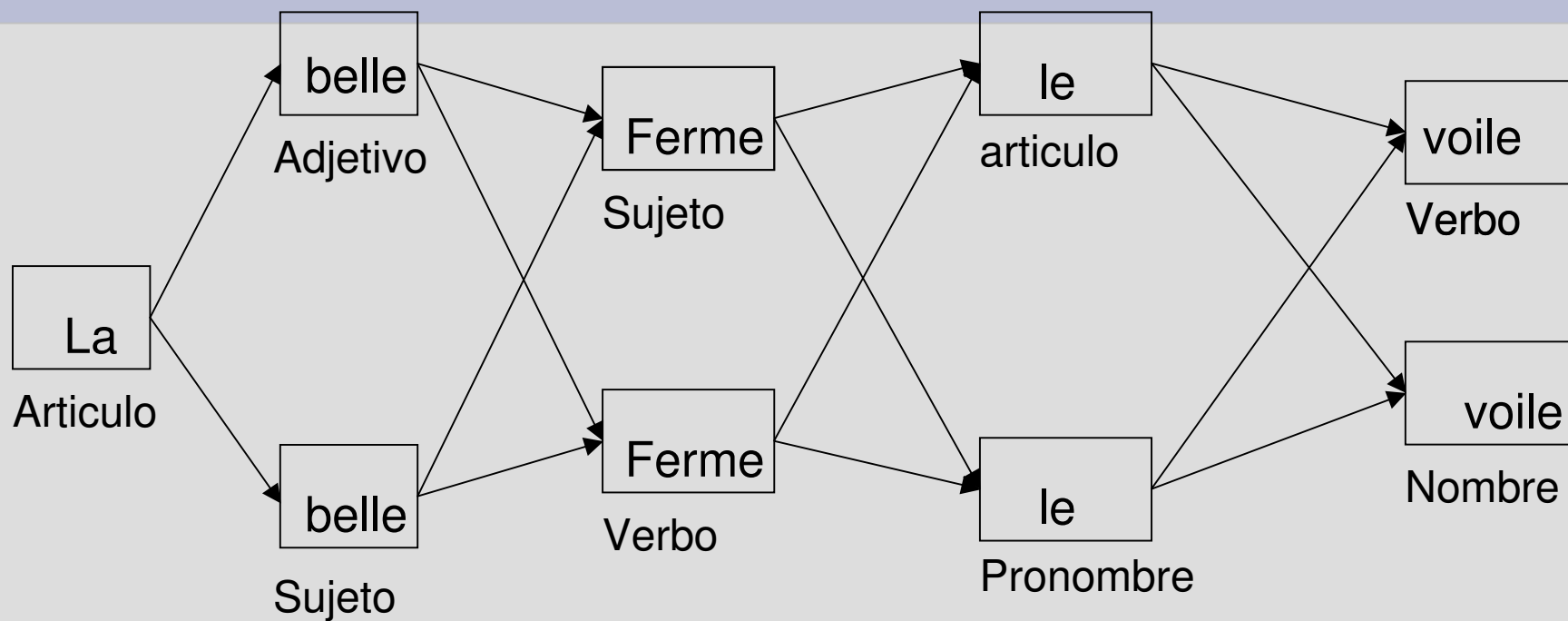
(a) Art Nom Ver Art Nom ? o bien :

(b) Art Adj Nom Pro Ver ?



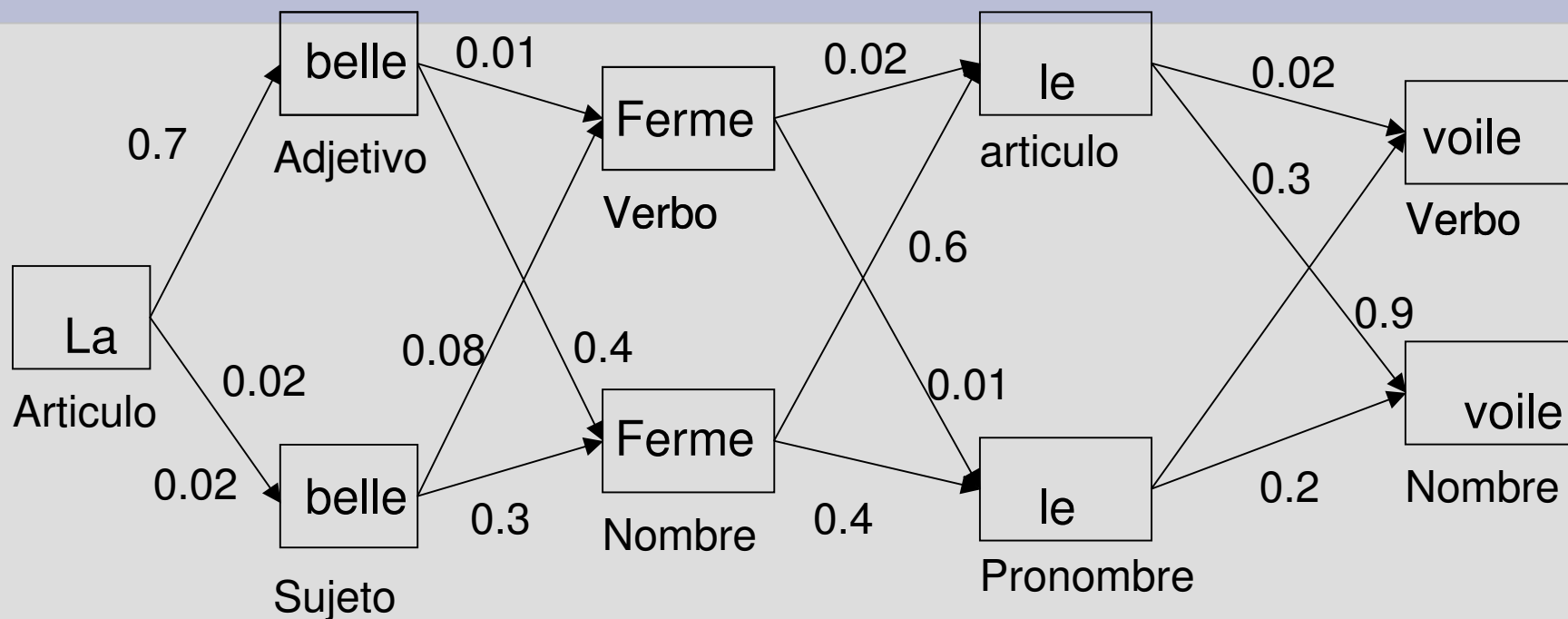
La linguística no puede resolver esta ambigüedad...

Etiquetas gramaticales

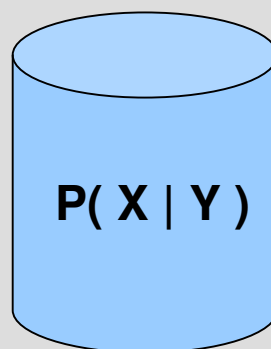


¿Qué camino tomar ?

Calcular las etiquetas gramaticales

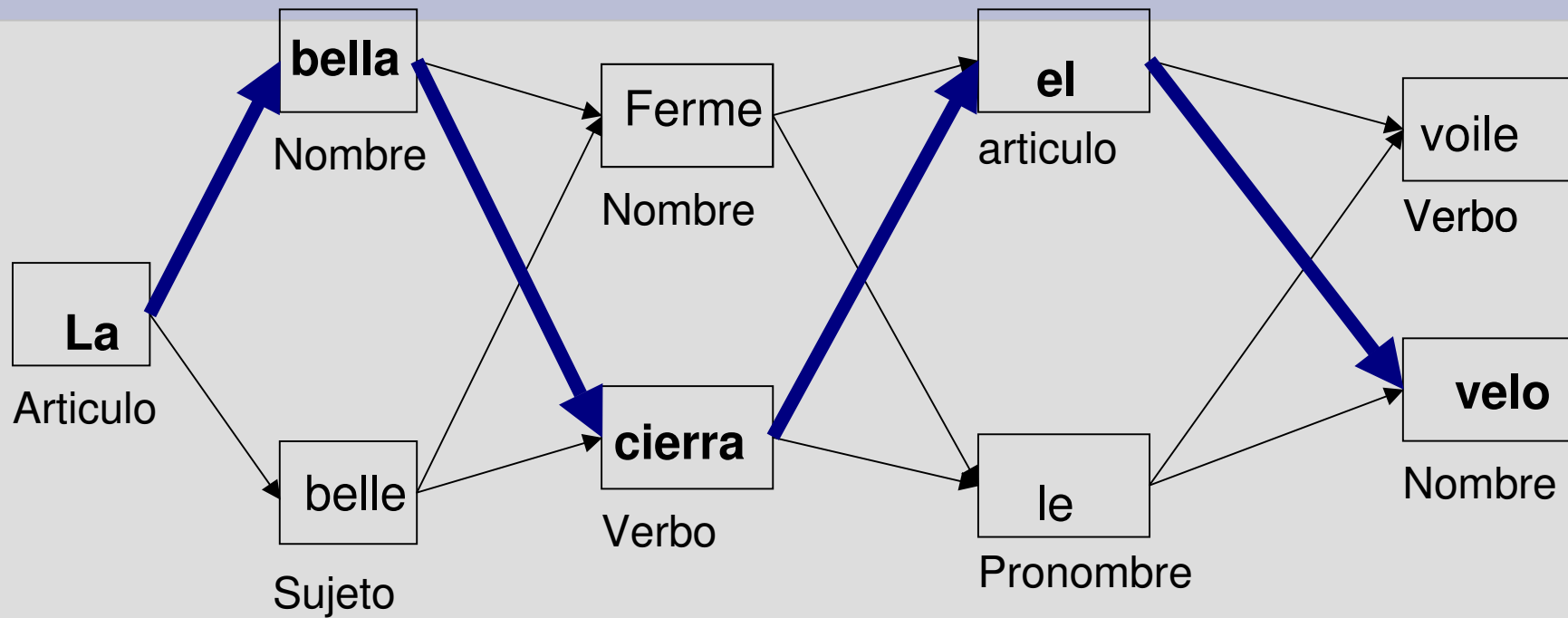


Probabilidades a partir
de un Corpus (miles de
palabras)



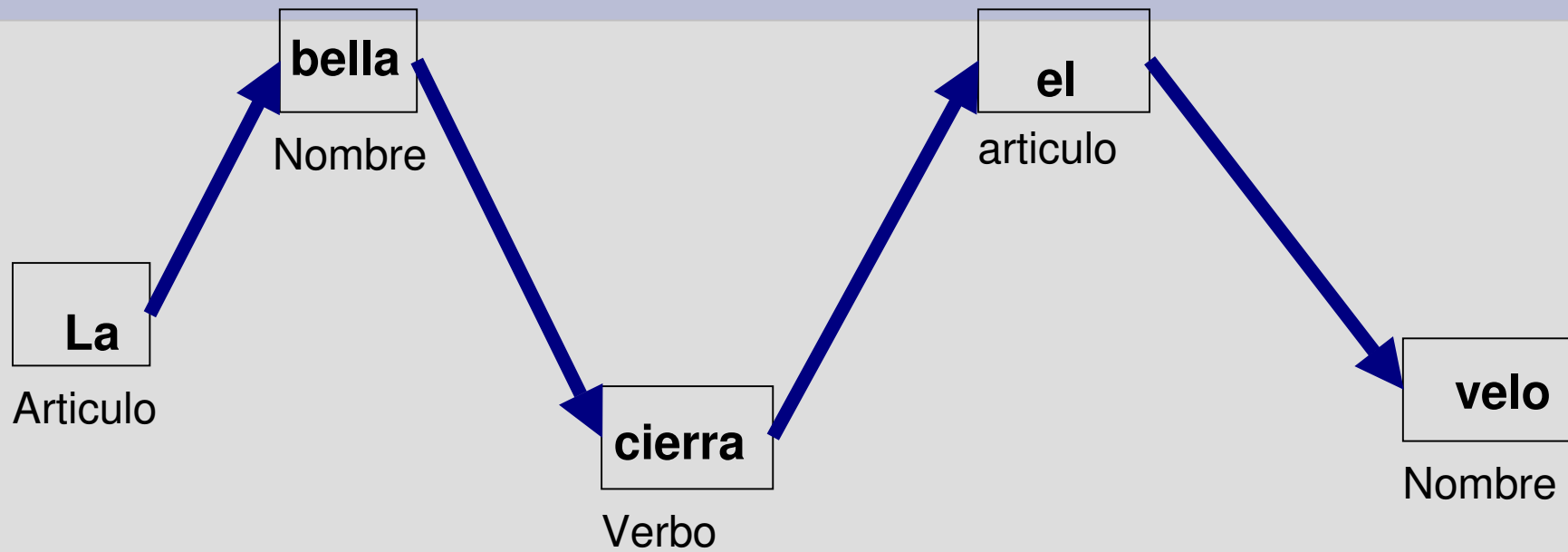
INTERNET

Traduccion final



LA / Art BELLA / Nom CIERRA / Ver EL / Art VELO / Nom

Traduccion final



Ejemplo 2: resúmen por métodos estadísticos

Ciudad del Vaticano. La noticia de la muerte del Papa sacude al mundo. Murió el martes pasado de forma misteriosa... (TEXTO)

La noticia de la muerte del Papa sacude al mundo. (RESUMEN)

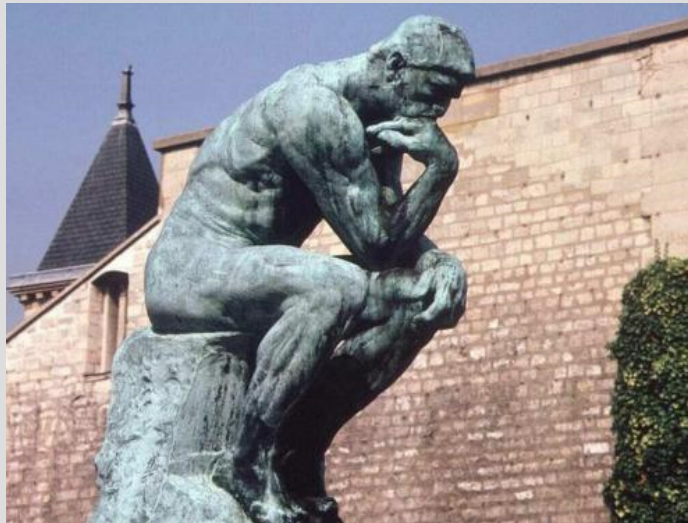
NO HAY COMPRENSION. NO HAY ANALISIS LEXICO NI SEMANTICO. NO NECESITA SCRIPTS NI RECURSOS EXTRA LINGUISTICOS... PERO EL RESUMEN ES *ACCEPTABLE*

Ventajas

- Extiende naturalmente la teoría de conjuntos para **tratar la incertidumbre** (implicaciones directas en ambigüedad, robustez y desempeño)
- **Interpretación directa y empírica** a partir de la estadística
- Enlace directo con la **teoría del aprendizaje automático**
- Ventajas metodológicas importantes: la **optimización** y la **modularización**

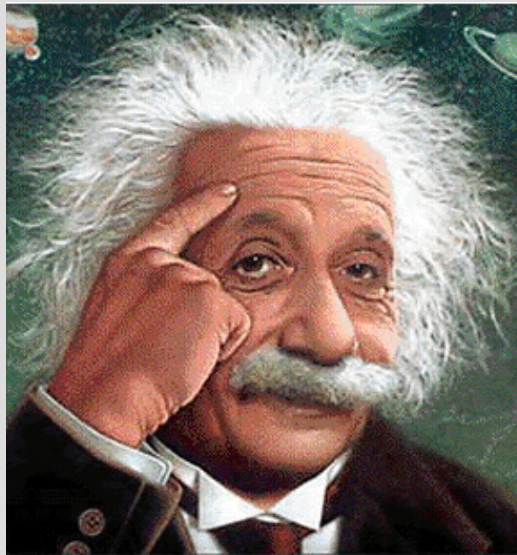
Abordar el problema pragmáticamente

- Todo mundo puede decir que piensa, pero nadie puede decir **cómo** esta pensando...
- Es difícil explicar **cómo** se llega a una conclusión a partir de una lectura **entre líneas**



Abordar el problema pragmáticamente

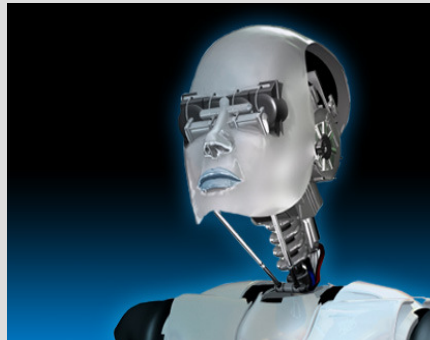
- Todavía no sabemos escribir programas que comprendan el texto como lo hace un humano



- La lengua es demasiado compleja para modelarla con reglas... **¡que la máquina las aprenda!**

Reflexiones

- Un programa no puede reproducir la manera en que las personas leen y producen documentos
- Caminos muy diferentes: un abismo de experiencias, de percepciones, de emociones...
- El camino de la máquina es **inhumano**... ahí reside su fuerza



Artefactos PLN

Google Traduction - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils ?

http://translate.google.com/translate_t#es|fr|Trujillo es la ciudad de la eterna primavera,%0ASin embargo hoy ha llovido un poco fuerte...%0AVallejo ha

Google

Les plus visités À la une

Aide

Google Traduction BETA

Texte et sites Web Traduction des résultats Dictionnaire Outils

Texte à traduire

Texte original :

Trujillo es la ciudad de la eterna primavera.
Sin embargo hoy ha llovido un poco fuerte...
Vallejo habría pensado que se hallaba en París,
muriendo con aguacero.

Espagnol » Français Traduire

Traduction : Espagnol » Français

Trujillo est la ville du printemps éternel.
Mais, aujourd'hui, a plu un peu trop ...
Vallejo aurait pu penser qu'il était à Paris, en
train de mourir à l'averse.

+ Proposer une meilleure traduction

Page Web à traduire

http://

Espagnol » Français Traduire

Traduccion automatica

Terminé

Démarrer UPAO2008 Yahoo! Messenger Pollita (patricia_velazquez) Microsoft PowerPoint - [...] Google Traduction - ...

FR 21:44

The Minimum Number of Errors in the N -Parity and its Solution with an Incremental Neural Network

J. MANUEL TORRES-MORENO^{1*}, JULIO C. AGUILAR²,
and MIRTA B. GORDON³

¹École Polytechnique de Montréal, Département de Génie informatique, CP 6079 Succ.
Centre-ville, H3C3A7 Montréal (Québec) Canada. e-mail: juan-manuel.torres@polymtl.ca

²Laboratorio Nacional de Informática Avanzada (LANIA), Rébsamen 80-91090 Xalapa,
México

³Laboratoire Leibniz – IMAG (CNRS), 46, Avenue Félix Viallet, 38031 Grenoble Cedex,
France

Abstract. The N -dimensional parity problem is frequently a difficult classification task for Neural Networks. We found an expression for the minimum number of errors v_f as function of N for this problem, performed by a perceptron. We verified this quantity experimentally for $N = 1, \dots, 15$ using an optimal train perceptron. With a constructive approach we solved the full N -dimensional parity problem using a minimal feedforward neural network with a single hidden layer of $h = N$ units.

Key words. classification tasks, minimerror, monoplan, parity problem, perceptrons, supervised learning

1. Introduction

The Neural Networks Community has studied the N -dimensional parity problem for a long time. In their celebrated book, Minsky and Papert [1] elegantly demonstrated that perceptrons are unable to solve non linearly separable problems, such as the parity of 2 inputs (or their equivalent one, the or-exclusive problem, XOR). The capacity of a simple perceptron is limited, since it is unable to solve problems that are linearly separable [6, 7]. The problem still becomes difficult in small dimensions: $N \leq 5$ and it is increased exponentially in function of the number of available patterns. This problem has been attacked by several methods such as Gradient Backpropagation (BP) and its variations. These methods have difficulties even in small dimensions due to the problem of the local minima in those in which the minimization of the cost function may fall. An alternative approach is to use Incremental Neural Networks methods that add units while learning errors exist, following a suitable heuristic, as we show it in Section 4.

The N -dimensional parity problem can be formulated as a supervised learning problem with a learning set of P patterns with N binary inputs $\xi_i = \pm 1$, $i = 1, \dots, N$, and a binary output $\tau = \pm 1$:

$$C = \{(\xi^{\mu}, \tau^{\mu}), \mu = 1, \dots, P\} \quad (1)$$

Resumenes automáticos

gradient search algorithms to escape from the multiple local minima. This problem has become a classic benchmark for classification algorithms [1], given that it is a highly not linearly separable problem (non LS). The classifier should learn how to discriminate against it if a given pattern belongs to the positive class, $\tau = +1$, or the negative one, $\tau = -1$. The problem is considered exhaustive learning, because all the $P = 2^N$ different patterns examples must be learned. The N -parity's problem becomes quickly complicated due to the neighboring states in input's space (its Hamming distances are $d_H = 1$) with opposite outputs. A solution with binary poid perceptrons has been should in [22].

The input's space for the N -dimensional parity and the separating hyperplanes w are represented for values of $N = 2, 3, 4$, in Figures 1(a), 1(b) and 2.

2. Finding the Minimum Number of Errors

Following a constructive approach, an incremental Neural Network adds hidden units one to one, until it is capable of eliminating learning errors. In N -dimensional parity, the problem is quite difficult for the first unit because it should find the

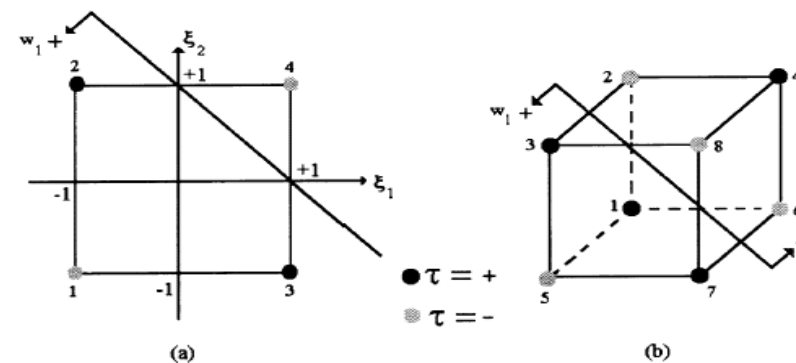


Figure 1. N -dimensional parity: (a) $N = 2$, (b) $N = 3$.

Resume automatique - Recherche Google - Mozilla Firefox

Echier Edition Affichage Aller à Marque-pages Outils ?

http://www.google.fr/search?hs=jv5&hl=fr&client=firefox-a&rls=org.mozilla%3Afr%3Aofficial&q=Resume+automatique&btnG=Rechercher&meta=

TALN Nautitia Gestion des... Web Term Documen... SENSUS demo System Quirk UNED Group in Natu... Departamento de In... Program 1 SELFHTML: CGI/Perl... UNIVERSIA-MIT-OP...

Google Web Images Groupes Annuaire Actualités plus »

Resume automatique Rechercher Recherche avancée Préférences

Rechercher dans : Web Pages francophones Pages : France

Web Résultats 1 - 10 sur un total d'environ 1 250 000 pour Resume automatique. (0,25 secondes)

Pertinence Mining - Traitement de l'information textuelle
... ne nécessite aucune inscription et ne sert que de DEMO de sensibilisation au couplage de deux technologies : **Résumé automatique** + Moteur de recherche ...
www.pertinence.net/ - 15k - [En cache](#) - [Pages similaires](#)

RALI: Résumé automatique
Les activités du RALI dans le domaine du **résumé automatique** s'articulent autour de trois projets principaux: SumUM pour les articles scientifiques ...
rali.iro.umontreal.ca/Resume/index.fr.html - 7k - [En cache](#) - [Pages similaires](#)

Résumé automatique de textes
Le **résumé automatique** de textes : solutions et perspectives, J. Minel, - Narrative Summarization, I. Mani, - **Résumé** de textes juridiques par identification ...
www.atala.org/article.php?id_article=24 - 18k - [En cache](#) - [Pages similaires](#)

Le résumé automatique des textes scientifiques et techniques ...
Le **résumé** de texte **automatique** s'inscrit dans le cadre de l'ingénierie ...
To produce the RAFI system (**Résumé Automatique** à Fragments Indicateurs' ...
www.atala.org/article.php?id_article=208 - 21k - [En cache](#) - [Pages similaires](#)
[[Autres résultats domaine www.atala.org](#)]

Appel a participation : Filtrage et resume automatique de l ...
Évaluation, rectification et pertinence du **résumé automatique** de texte pour une utilisation en réseau, Abderrafih LEHMAM. ...
www.lri.fr/Services/Archives-majordomo/interactif/msg00030.html - 16k - [En cache](#) - [Pages similaires](#)

NetEconomie :: Pertinence.net veut démocratiser le résumé automatique
Pertinence.net veut démocratiser le **résumé automatique**. vedette La Jeune pousse
Pertinence.net propose un service en ligne de **résumé automatique** des ...
www.neteconomie.com/perl/navig.pl/neteconomie/infos/article/20011004152944 - 25k - 14 sep 2005 - [En cache](#) - [Pages similaires](#)

sujet stage résumé automatique
Résumé automatique par filtrage sélectif Contexte La tâche de **résumé** est avant tout une tâche de compréhension de texte : dans une première étape il s'agit ...
www.limsi.fr/Individu/mhp/stageDEA/NH-05.html - 3k - 13 sep 2005 - [En cache](#) - [Pages similaires](#)

Outils Froids - Resume Automatique
Outils de **résumé automatique**. De La Cartographie Pour Pertinence Mining : Fourre Tout Numero Quatre : Pertinence **Resume** Les Resultats De Google ...
www.outilsfroids.net/texts/ResumeAutomatique.shtml - 19k - [En cache](#) - [Pages similaires](#)

GFII - Séminaire "Le résumé automatique : état de l'art technique ...
MINEL Jean-Luc, Filtrage sémantique : du **résumé automatique** à la fouille de ...
Pertinence Summarizer : un outil de **résumé automatique** au service de la ...
www.gfii.fr/040543_seminaire.htm - 7k - [En cache](#) - [Pages similaires](#)

Terminé

Démarrer 3 Explorateur ... 3 OpenOffice.o... 9 Firefox ~/Projets/Versio... PERL - Microsoft ... Courrier entrant ... Microsoft PowerP... 09:37 a.m.

Búsqueda de información

...Artefactos PLN

- Generadores de texto (meteo en francés/inglés...)
- Correctores de ortografía/sintaxis/gramática
- Comprensión de la palabra
- Síntesis del habla
- Diálogo hombre/máquina
- ...

Conclusion

- Se necesita únicamente escribir programas que *razonablemente* procesen masas de documentos en lugar de las personas... y que lo hagan bien y rápidamente
- En tanto que el enfoque sea **eficaz**, poco importa si es inhumano... siempre y cuando los resultados sean satisfactorios

Conclusion...

- El PLN numérico es *comprensible*: los números dicen cosas (a quien los sabe escuchar):
 - Se puede saber por qué el sistema siguió un camino y no otro
- Probablemente no necesitamos escribir programas que verdaderamente comprendan el texto
 - Ejemplo: resumidores automáticos

**Los ingenieros y científicos en computación
están construyendo un mundo nuevo.
Un mundo muy diferente del que ahora
vivimos**





Voilà!

Avez-vous des questions?



Juan-Manuel Torres

`juan-manuel.torres@univ-avignon.fr`

Laboratoire Informatique d'Avignon (France)