

Inteligencia Artificial

Introducción a la inteligencia artificial

Primavera 2007

profesor: Luigi Ceccaroni



¿Qué es la inteligencia artificial?

- No hay una sola definición, sino varias aproximaciones, que Russell-Norvig resumen en cuatro principales.
- Las aproximaciones siguen diferentes puntos de vista
- Sus influencias son diversas (Filosofía, Matemática, Psicología, Biología ...)
- Sus campos de aplicación son amplios e interrelacionados

Bases de la Inteligencia Artificial

- Filosofía
- Matemáticas
- Economía
- Neurociencia
- Psicología
- Ingeniería computacional
- Teoría de control y cibernética
- Lingüística

Bases de la Inteligencia Artificial

Filosofía

- Debate sobre la posibilidad de una inteligencia mecánica
 - Platón: leyes que gobiernan el pensamiento (silogismos, lógica)
 - Descartes, Leibniz: la mente está ligada al mundo físico
 - John Locke: en el principio fue la Mente (1690)
 - Hume (1779), Russell: el conocimiento es fruto de la percepción y se adquiere por la experiencia (**inducción**). El conocimiento está representado por teorías lógicas.
 - Darwin (1857): destrucción del *ex nihilo nihil fit* a través de la teoría de la evolución por selección natural

Bases de la Inteligencia Artificial

Matemáticas

- Las bases filosóficas necesitan reglas formales
- Boole, Frege: fundamentos de la lógica matemática
- Gödel, Turing: límites de lo computable (teorema de incompletitud)
- Fermat, Bernoulli, Bayes: probabilidad, razonamiento probabilístico

Bases de la Inteligencia Artificial

Psicología

- Psicología cognitiva: teorías sobre la conducta, bases del comportamiento racional
 - Representación de los estímulos externos
 - Manipulación consciente de la representación
 - Actuación consecuente

Bases de la Inteligencia Artificial

Ingeniería computacional

- Para la existencia de la IA es necesario un mecanismo para soportarlo (hardware)
- También son necesarias herramientas para desarrollar programas de IA

Lingüística

- Chomsky: representación del conocimiento, gramática de la lengua
- Lingüística computacional

Definición de IA

- **Sistemas que actúan como humanos**

El estudio de cómo lograr que los ordenadores realicen tareas que, de momento, la gente hace mejor (Rich y Knight, 1991)

- **Sistemas que piensan como humanos**

El esfuerzo de hacer que los ordenadores piensen ... *máquinas con mentes* en el más amplio sentido literal (Haugeland, 1985)

- **Sistemas que piensan racionalmente**

El estudio de las facultades mentales a través del estudio de modelos computacionales (Charniak y McDermott, 1985)

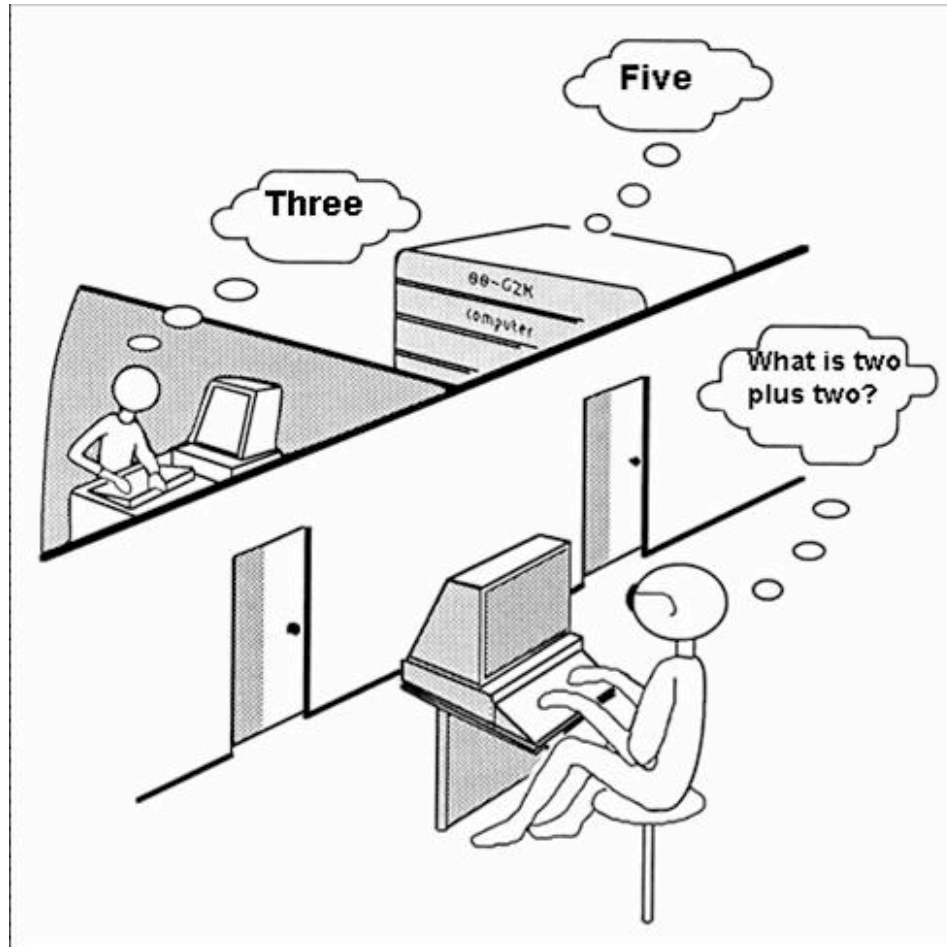
- **Sistemas que actúan racionalmente**

El estudio que busca explicar y emular el comportamiento inteligente en términos de procesos computacionales (Shalkoff, 1990)

Sistemas que actúan como humanos

- El modelo es el hombre; el objetivo es construir un sistema que pase por humano
- Prueba de Turing: si un sistema la pasa es inteligente
- Capacidades necesarias: Procesamiento del Lenguaje Natural, Representación del Conocimiento, Razonamiento, Aprendizaje
- Pasar la Prueba no es el objetivo primordial de la IA
- La interacción de programas con personas hace que sea importante que éstos puedan actuar como humanos

La prueba de Turing



Sistemas que piensan como humanos

- El modelo es el funcionamiento de la mente humana.
- Se intenta establecer una teoría sobre el funcionamiento de la mente (experimentación psicológica).
- A partir de la teoría se pueden establecer modelos computacionales.
- Influencia de las Ciencias Cognitivas

El misterio de la conciencia

- La influencia de las neurociencias y de las ciencias cognitivas
- ¿Qué es la conciencia?
- Si identificamos el núcleo cognitivo de la conciencia, ¿podemos incorporarlo en una máquina?
- Cuestiones, antes objeto solo de especulaciones teológicas, ahora estudiadas por la neurociencia cognitiva
- Las mayores religiones localizan la conciencia en un ánima que sobrevive a la muerte del cuerpo.
- Para nosotros la conciencia es la vida misma.

El misterio de la conciencia

- Woody Allen: “No quiero alcanzar la inmortalidad a través de mi obra. Quiero alcanzarla no muriéndome”.
- Conciencia como base para la moralidad
- La conciencia no depende del lenguaje.
- La conciencia no es el simple conocimiento de nuestra existencia.

Problemas fáciles y difíciles

- David Chalmers trata el tema de la conciencia en términos de dos problemas:
 - Problema Fácil: distinguir entre pensamiento consciente e inconsciente (Freud):
 - superficies que tenemos delante
 - planes para el día
 - sueños con ojos abierto durante las clases
 - control del latido del corazón
 - reglas que ordenan las palabras cuando hablamos
 - secuencias de contracciones musculares
 - Problema Difícil: explicar cómo puede nacer la experiencia subjetiva de la computación neuronal

El cerebro como máquina

- La “hipótesis asombrosa” de Francis Crick:
 - Nuestros pensamientos, sensaciones, alegrías, dolores consisten enteramente de actividad fisiológica en los tejidos del cerebro.
- La conciencia es un producto biológico natural, tan vacía de elementos sobrenaturales como la digestión o la circulación de la sangre.
- Robots inteligentes y el sentido del Problema Difícil

Sistemas que piensan racionalmente

- Las leyes del pensamiento racional se fundamentan en la **lógica** (silogismos de Aristóteles)
- La lógica formal está en la base de los programas inteligentes (logicismo)
- Se presentan dos obstáculos:
 - Es muy difícil formalizar el conocimiento
 - Hay un gran salto entre la capacidad teórica de la lógica y su realización práctica

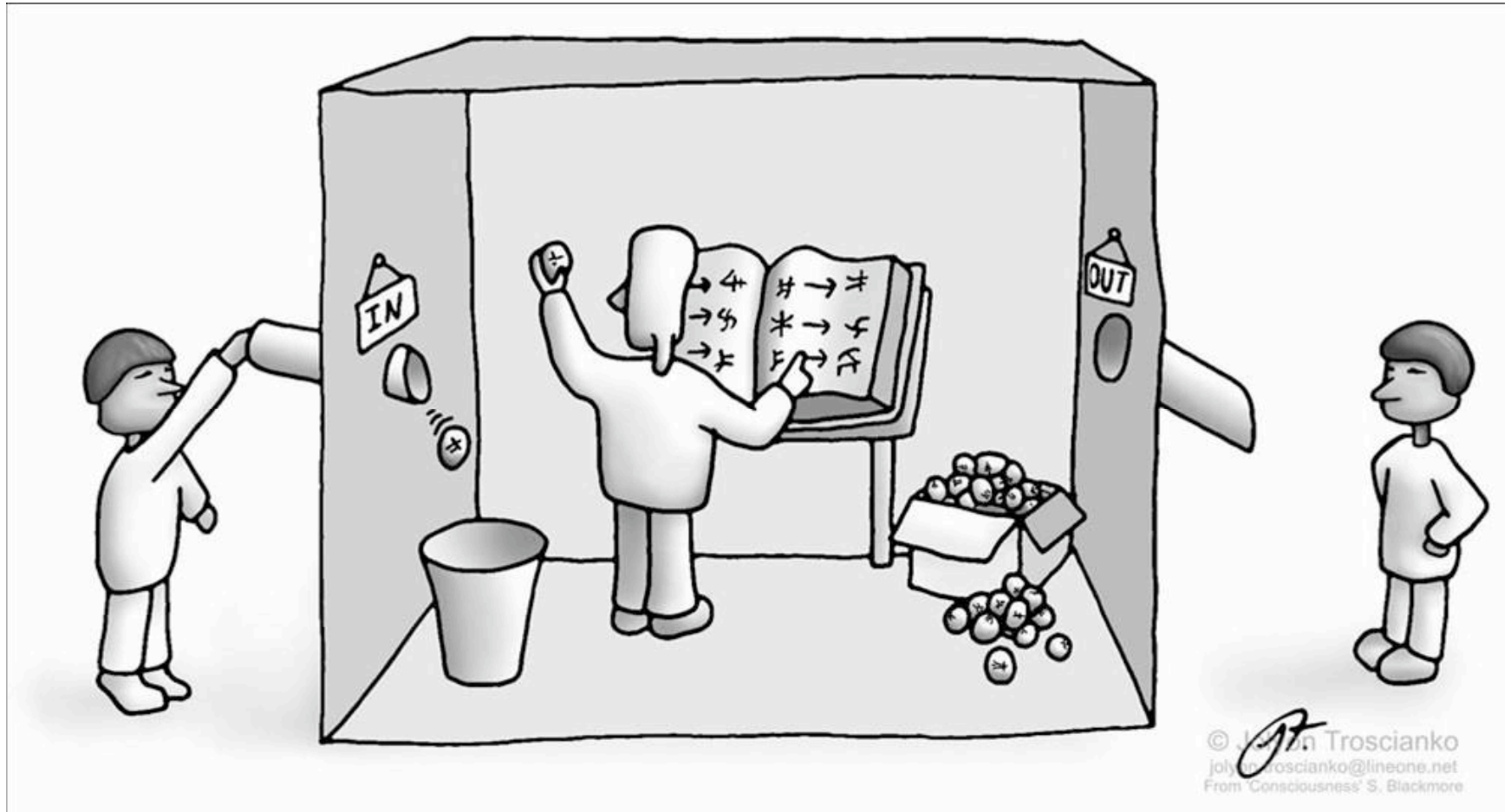
Sistemas que actúan racionalmente

- Actuar racionalmente significa conseguir unos objetivos dadas unas creencias.
- El paradigma es el **agente racional**, que se aplica, por ejemplo, a muchos sistemas robóticos.
- Un agente *percibe y actúa*, siempre teniendo en cuenta el entorno en el que está situado.
- Las capacidades necesarias:
 - percepción
 - procesamiento del lenguaje natural
 - representación del conocimiento
 - razonamiento
 - aprendizaje automático
- Visión de la actuación general y no centrada en el modelo humano
- Ejemplos:
 - Stanley
 - Grand challenge 2005
 - Asimo

¿Es la IA posible?

- La posibilidad de la inteligencia artificial plantea problemas filosóficos complejos
- ¿Las máquinas pensantes poseen consciencia?
- **La habitación china** (Searle, 1980)
- ¿Es la inteligencia una propiedad emergente de los elementos biológicos que la producen?
- No hay una conclusión definitiva.

La habitación china



Historia de la IA

Los inicios (1943-1956)

- McCulloch y Pitts: primer modelo de neurona, demostración de que cualquier función es computable por una red de neuronas, aprendizaje
- Hebb: regla de aprendizaje Hebbiano o de Hebb
- Minsky, Edmonds: primer simulador de red neuronal (40 neuronas)
- Taller de Dartmouth (bautizo de la IA): McCarthy, Minsky, Shannon, Rochester, More, Samuel, Solomonoff, Selfridge, Newell, Simon
- Newell, Simon: *Teórico Lógico (TL)*

Historia de la IA

Entusiasmo inicial, grandes expectativas (1952-1969)

- SRGP: sistema de resolución general de problemas, imitación de las demostraciones humanas
- Samuel: programa jugador de damas, aprendizaje
- Sistemas planificadores, representación interna del mundo
- Aplicaciones en dominios reducidos: integración simbólica, mundo de bloques
- Redes neuronales: perceptrones, aprendizaje

Historia de la IA

La dura realidad (1966-1973)

- Muchos sistemas de IA actuaban de manera prometedora en problemas sencillos y fallaban ante problemas reales
- Muchos problemas eran intratables (NP); el tamaño del problema es relevante
- La búsqueda exhaustiva es impracticable
- Minsky, Papert: en *Perceptrons* se demuestra la escasa capacidad de representación de los perceptrones

ELIZA (1966)

- **Conversación con ELIZA**

E> Hello. I am ELIZA. How can I help you?

P> I am feeling sad.

E> I am sorry to hear you are sad.

P> I had some nightmares tonight.

E> You say you had some nightmares tonight.

P> Yes, my dreams are worrying me.

E> What does that dream suggest to you?

P> Thrill, scare.

E> I am not sure I understand you fully.

P> My dreams annoy me.

E> Do you dream often?

P> Yes, daily.

Traducción automática

- Traducción automática, mecanismos simples

inglés → ruso (o castellano) → inglés

“The spirit is willing, but the flesh is weak” →

→ Дух охотно готов, но плоть слаба →

→ “The vodka is good, but the meat is rotten”

(Alternativas: “Spirit is willingly ready, but flesh is weak” o
“The alcohol is arranged, but the meat is weak”)

Historia de la IA

Los sistemas basados en el conocimiento (1969-1979)

- Los mecanismos generales de resolución de problemas no son practicables (métodos débiles)
- DENDRAL: el conocimiento del dominio del problema ayuda a resolverlo mejor
- **Sistemas Expertos** (MYCIN, PROSPECTOR):
incertidumbre, factores de certeza
- Representación del conocimiento y razonamiento (redes semánticas, marcos, Prolog)

Historia de la IA

La IA se industrializa (1980 hasta el presente)

- El éxito de los sistemas expertos lleva a su uso comercial: R1, XCON (sistema de configuración en Digital Equipment Corporation)
- Japoneses: quinta generación de ordenadores
- Estadounidenses: MCC
- Software para el desarrollo de la IA
- Máquinas de arquitectura dedicada

Historia de la IA

La actualidad

- El retorno de las redes neuronales (física, psicología, modelos conexionistas)
- Uso de modelos bien fundamentados, comprensión de las características de los problemas, minería de datos, redes de Bayes
- Construcción de *agentes totales* o *arquitecturas de agentes completas* (SOAR)

Áreas de trabajo de la IA

- **Áreas Básicas**

- Representación del conocimiento
- Resolución de problemas mediante búsqueda

- **La IA hoy en día está principalmente relacionada con actividades en los siguientes campos:**

- representación del conocimiento
- planificación autónoma
- juegos
- control autónomo
- diagnosis
- planificación logística
- robótica
- procesamiento de lenguaje
- resolución de problemas

Áreas que se trabajan en la FIB

Representación del conocimiento

- Esquemas de representación sintácticos: lógica clásica, sistemas de producción
- Esquemas de representación semánticos: redes semánticas, marcos, ontologías, FOAF
- Lógicas no clásicas: creencias, incertidumbre, sustancias temporales y espaciales

Complejidad del conocimiento

- Creencias

Juan cree que “A” y Pedro cree que “no A”

- Incertidumbre, vaguedad

La temperatura es alta

Es muy posible que la fiebre alta signifique una infección primaria

- Tiempo

El suceso A fue antes que B y simultáneamente a C

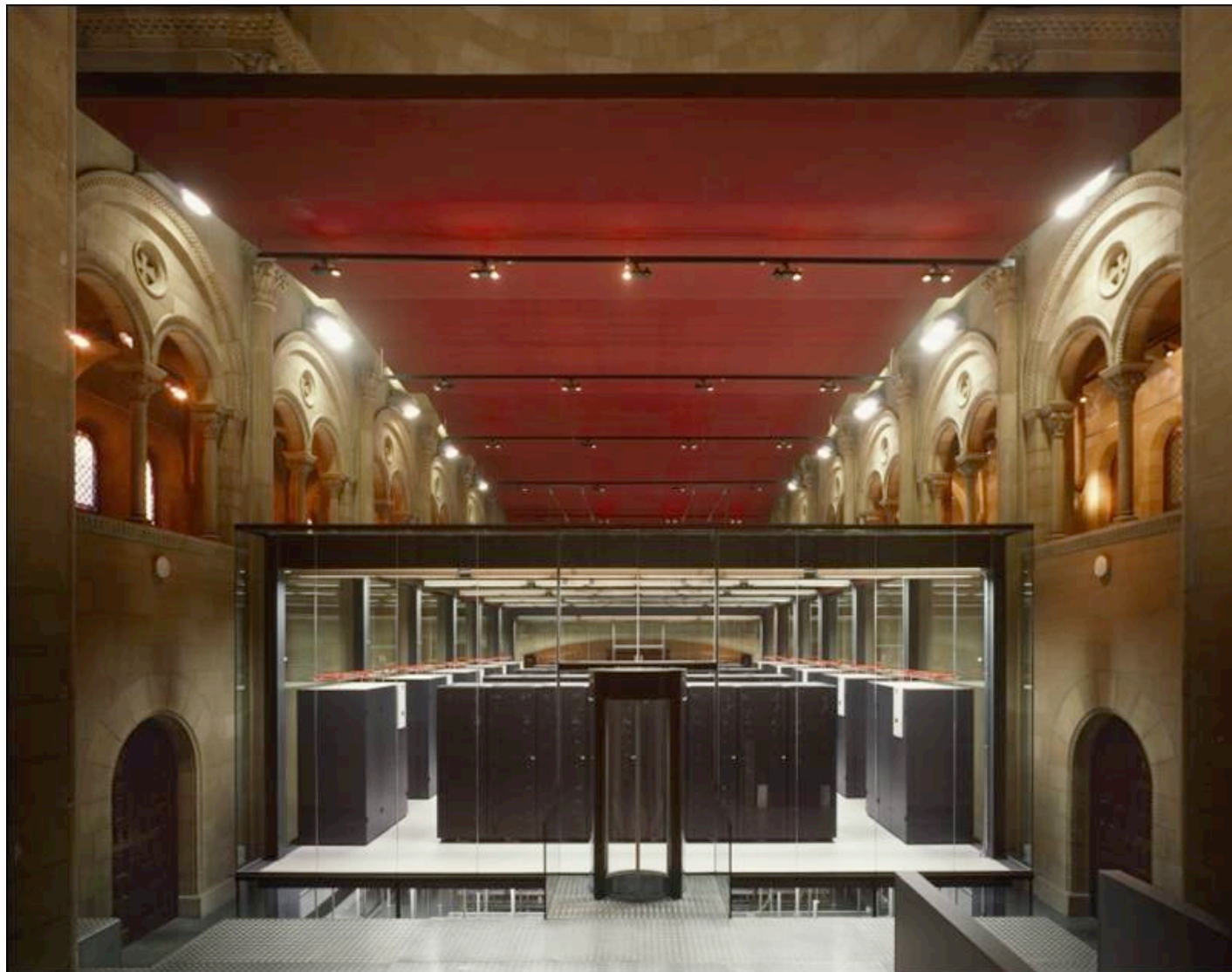
Resolución de problemas

- Búsqueda Heurística (ascensión de colinas, A^* , IDA*)
- Juegos (MINIMAX, poda alfa-beta)
- Satisfacción de restricciones (búsqueda con vuelta atrás para PSR, comprobación hacia adelante)

Resolución de problemas - Juegos



Resolución de problemas - Supercomputación



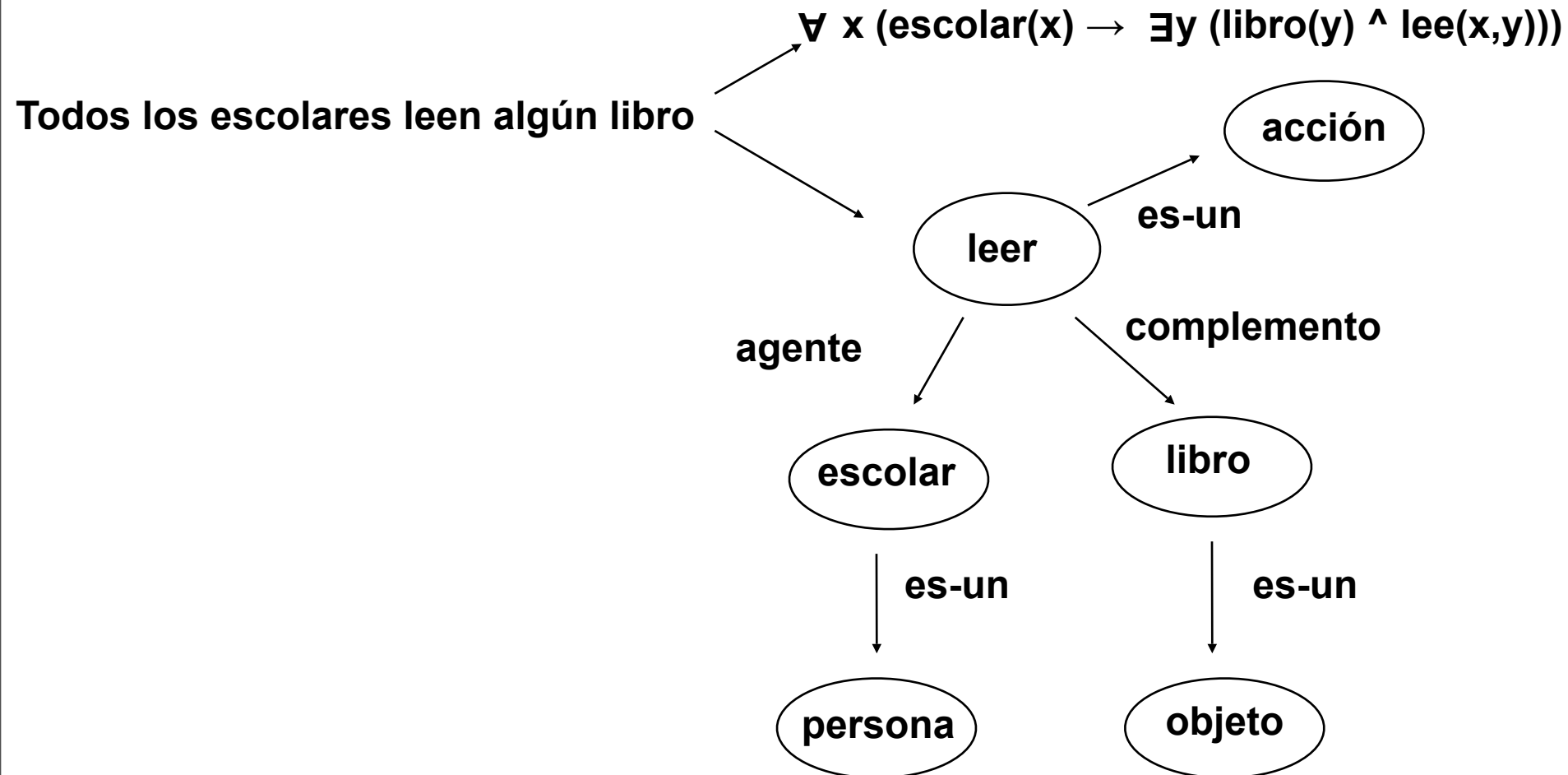
Procesamiento del lenguaje natural o lingüística computacional (1)

- Analizadores: léxicos, sintácticos, semánticos
- Generación automática de diccionarios
- Etiquetadores Sintácticos
- Traducción Automática
- Diálogos
- Generación del lenguaje

Procesamiento del Lenguaje Natural (2)

- Requerimientos para el entendimiento del lenguaje
 - Entendimiento de la estructura de las sentencias
 - Comprensión de la materia bajo estudio
 - Comprensión del contexto
- Representación del conocimiento
(representación pensada específicamente para que un ordenador pueda razonar a partir de ella)

Procesamiento del Lenguaje Natural (3)



Razonamiento Automático

- Lógica de primer orden, resolución
- Razonamiento basado en casos
- Lógicas no monótonas
- Lógicas probabilísticas

Sistemas Basados en el Conocimiento (SBCs)

- Ingeniería del Conocimiento
- Ontologías
- Explicación de razonamientos
- Validación de sistemas

SBCs - Aplicaciones

- Diagnóstico
- Predicción
- Supervisión, Monitorización
- Detección / Corrección de fallos
- Control
- Diseño

Aprendizaje automático

- Aprendizaje inductivo
- Aprendizaje deductivo
- Algoritmos genéticos

Áreas que no se trabajan en la FIB

Planificación de Tareas

- Planificación lineal (STRIPS)
- Planificación parcialmente ordenada (NONLIN)
- Planificación jerárquica
- Planificación en tiempo real

Planificación de Tareas: actualidad

- Tweak (1987): demostraciones de completitud y intratabilidad
- SNLP (1991): implementación de un planificador de **orden parcial** completo
- UCPOP (1992): primer planificador para Lenguaje de Descripción de Acciones
- Graphplan (1997): heurísticas extraídas de un grafo de planificación
- FastForward (2000): buscador de **espacio de estados** especialmente exitoso
- LPG (2002): buscador de **grafos de planificación** especialmente exitoso

Percepción

- Visión
- Reconocimiento del habla
- Robótica (planificación de trayectorias)

Aprendizaje automático

- Analogía
- Redes neuronales

Lecturas recomendadas y otros recursos

Libros

- Russell, S., Norvig, P., *Artificial Intelligence: A Modern Approach*, Prentice Hall (2003)
- Russell, S., Norvig, P., *Inteligencia Artificial: Un Enfoque Moderno*, Pearson (2004)

En Internet

- Google, Yahoo
- <http://www.lsi.upc.es/~bejar/ia/ia.html>
- <http://www.lsi.upc.es/~luigi>

Bibliografía complementaria

Darwin's Dangerous Idea [La peligrosa idea de Darwin] de **Daniel Dennett**

Libro fundamental a favor del darwinismo materialista

Víctimas: Noam Chomsky, Roger Penrose, John Searle y, especialmente, Stephen Jay Gould