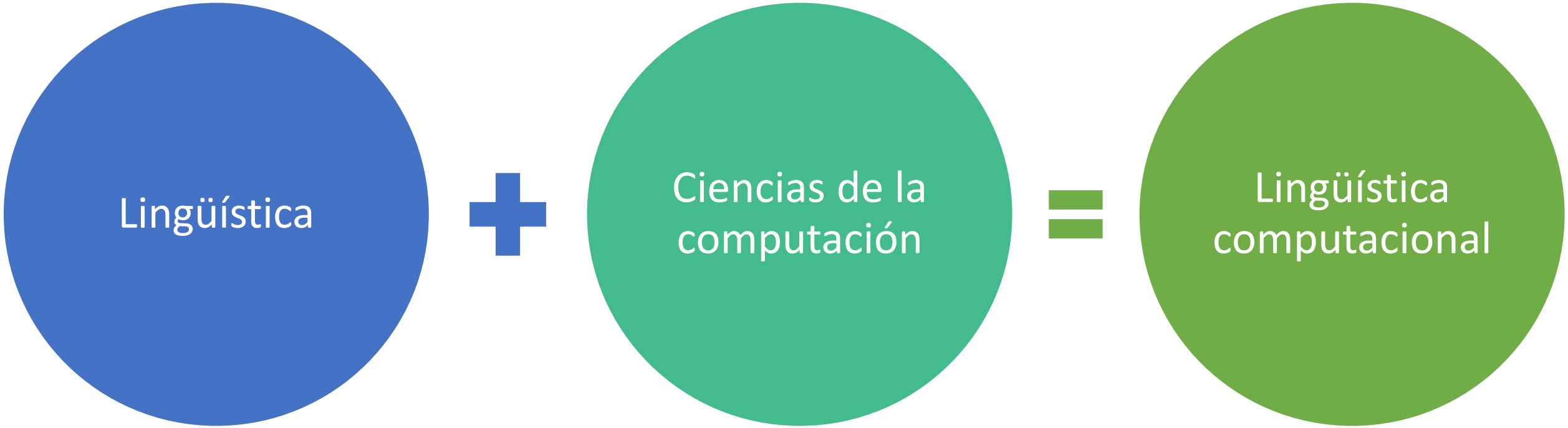
Lingüística computacional

Definición y alcance

Escuela Nacional de Antropología e Historia (ENAH)
Agosto – diciembre de 2015

Definición y alcance



Definición y alcance

- Estudio del lenguaje desde una perspectiva computacional
- Busca crear modelos computacionales para distintos tipos de fenómenos lingüísticos
- Estos modelos pueden ser hechos manualmente (basados en conocimiento) o mediante aproximaciones empíricas (basados en datos)
- Tiene un objetivo teórico: explicación computacional de un fenómeno lingüístico
- Tiene un objetivo práctico: desarrollar tecnología que trabaje con lenguaje natural

Definición y alcance

- La complejidad del sistema lingüístico no puede ser abordada desde una sola disciplinaria.
- La combinación de herramientas metodológicas de distintas disciplinas puede mejorar el entendimiento de las lenguas humanas.
- La computación ha incrementado el entendimiento de muchos fenómenos distintos

Definición y alcance

Disciplinas
Lingüística computacional
Procesamiento de lenguaje natural
Minería de textos
Ingeniería lingüística

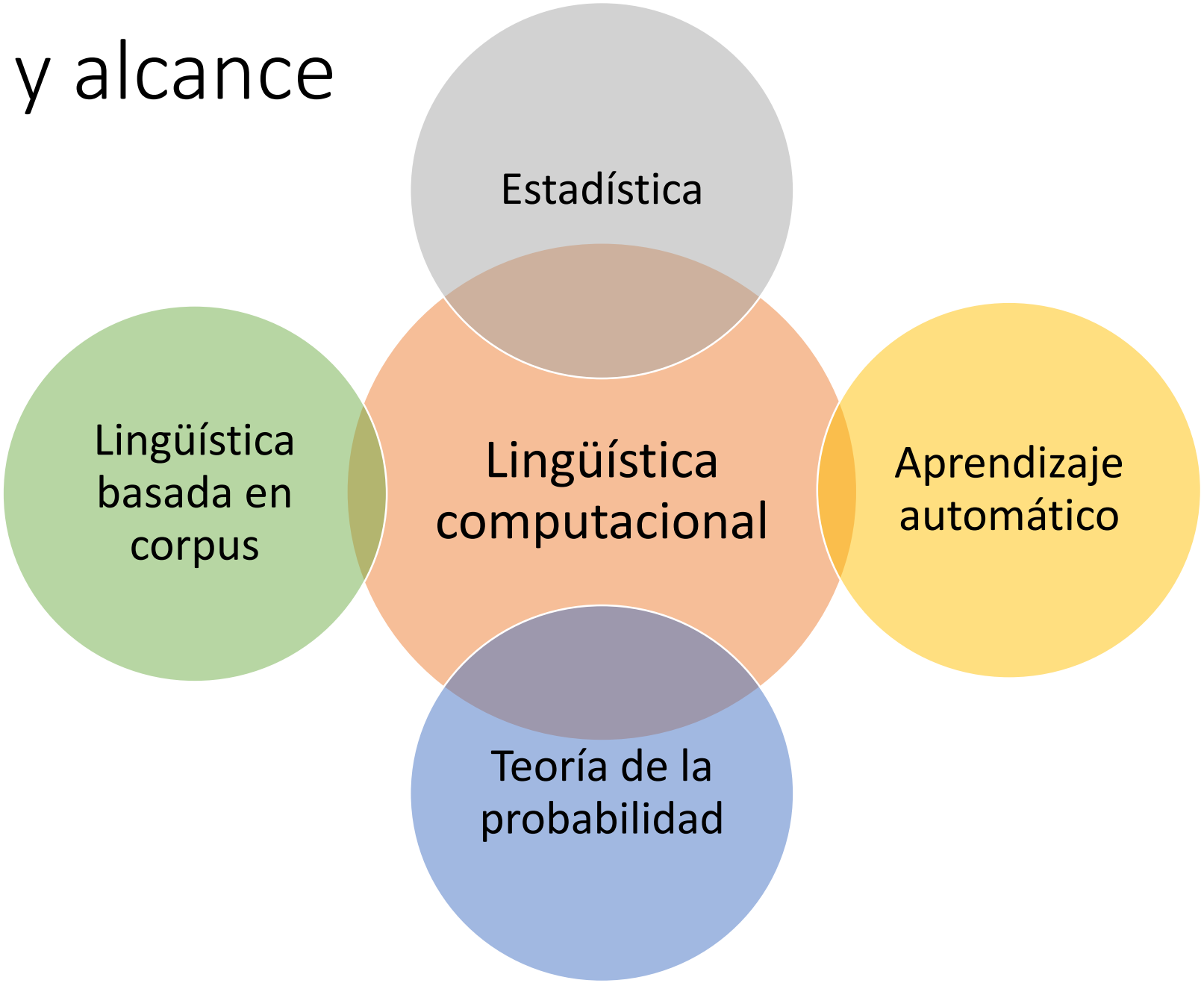
- Teórica
- Lingüística

- Aplicada
- Computacional

Definición y alcance

Disciplina	Objetivo
Lingüística computacional	Estudiar el lenguaje a través de modelos para computadora
Procesamiento de lenguaje natural	Desarrollar programas y herramientas de análisis del lenguaje
Minería de textos	Analizar colecciones de documentos (lenguaje) para descubrir patrones de asociación
Ingeniería lingüística	Desarrollar software que analice lenguaje

Definición y alcance



Y otras…

Definición y alcance

- Gramaticalidad de una oración

vs

- ¿Cuáles son los patrones en el uso del lenguaje?

Definición y alcance

Racionalismo (1960 - 1985)

- Gran parte del conocimiento humano está determinada de manera genética (herencia), no se deriva de los sentidos.
- ¿Cómo los niños aprenden en corto tiempo y con pocos estímulos algo tan complejo como el lenguaje? La parte principal del lenguaje es innata.

Empirismo (1920 – 1960 y hoy)

- Existen estructuras iniciales (generales) en el cerebro.
- Operaciones iniciales: asociaciones, reconocimiento de patrones y generalización.

Definición y alcance

Lingüística computacional estadística

- Principal herramienta es contar cosas (estadística)
- Teoría de la probabilidad

Definición y alcance

- Corpus
  - Colección de materiales hablados o escritos
- Corpus lingüísticos
  - Colección de materiales hablados o escritos recopilados bajo ciertos criterios de investigación para análisis lingüísticos.
- Estudios basados en corpus (Lingüística de corpus)
- Corpus (en lingüística computacional)
  - Colección de documentos o grabaciones

Definición y alcance

Algoritmo

- Serie de pasos finitos y ordenados para resolver un problema.

¿Quién hace el algoritmo?

- Una persona
- Una computadora (aprendizaje automático)

Definición y alcance

¿Por qué dejar que una máquina haga el algoritmo?

- Demasiados datos.
- Problemas muy complejos.
- Demasiadas variables.
- Le toma menos tiempo.

Definición y alcance

Aprendizaje automático

- Estudio y desarrollo de algoritmos que pueden “aprender” a partir de datos y hacer predicciones sobre ellos.
- Esos algoritmos construyen un modelo a partir de los datos para hacer predicciones o “tomar decisiones”.

Definición y alcance

Aprendizaje automático

- Supervisado: la computadora recibe ejemplos de entrada a partir de los cuales “aprende” reglas para predecir ejemplos de salida.
- No supervisado: la computadora NO recibe ejemplos de entrada y tiene que “descubrir” una estructura en los datos.

Definición y alcance

Aprendizaje automático

- Supervisado: clasificación.
- No supervisado: agrupamiento.

Definición y alcance

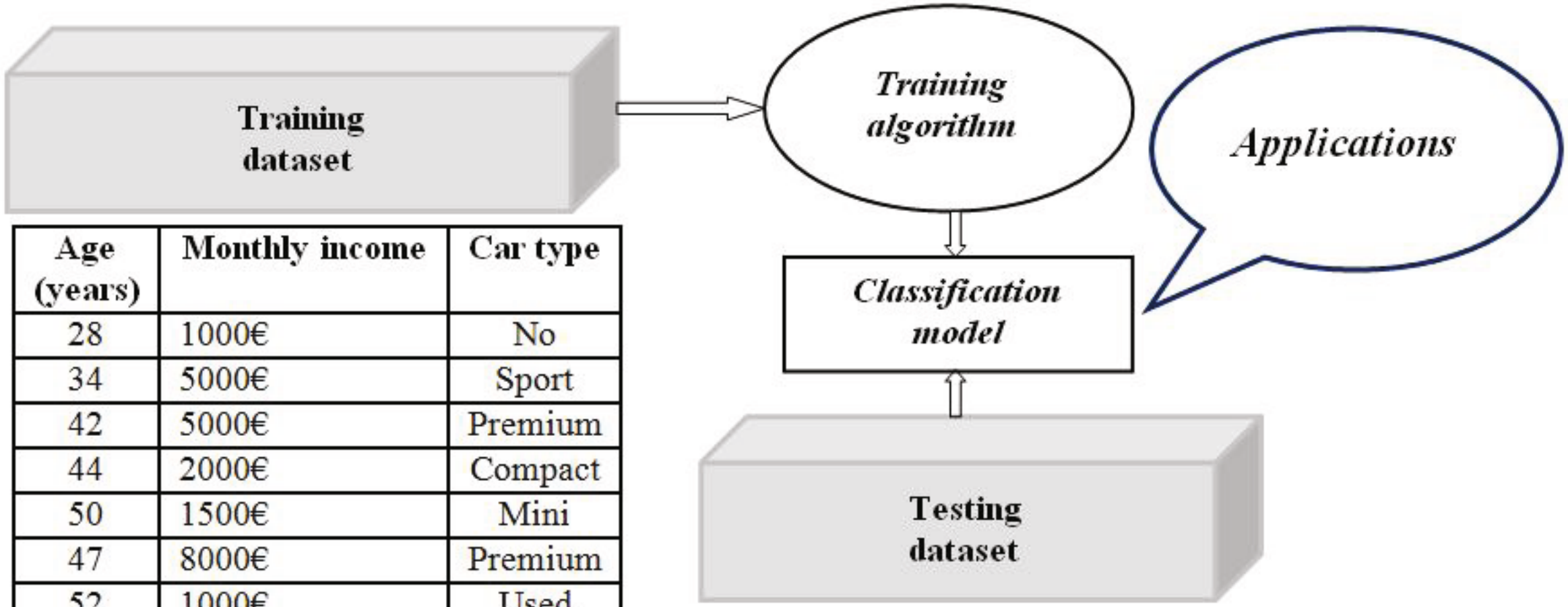
Clasificación automática

- Clase: etiqueta asignada a cada ejemplo (objeto) clasificado
- Características: atributos de los ejemplos (objetos) que son usados para clasificarlos
- Conjunto de datos de entrenamiento: conjunto de ejemplos (objetos) previamente clasificados a partir de los cuales se obtiene un modelo clasificador
- Conjunto de datos de evaluación: conjunto de nuevos ejemplos que serán clasificados usando el modelo clasificador

Definición y alcance

Clasificación automática

- Modelo clasificador: una función que recibe los valores de las características de un ejemplo y regresa su correspondiente clase (=predicción)
- Ejemplo = objeto, vector, instancia, registro



Age (years)	Monthly income	Car type
28	1000€	No
34	5000€	Sport
42	5000€	Premium
44	2000€	Compact
50	1500€	Mini
47	8000€	Premium
52	1000€	Used
55	500€	Used
45	1500€	Compact
63	800€	No
41	1000€	Mini
46	1200€	Used

Age (years)	Monthly income	Car type
51	1100€	? (*Used*)
38	1800€	? (*Compact*)
27	1000€	? (*No*)

Definición y alcance

Clasificación automática

- Identificación de perfiles de clientes
- Identificación de movimientos fraudulentos en tarjetas de crédito
- Clasificación de especies
- Clasificación de galaxias

Definición y alcance

Clasificación automática

- Clasificación de estudiantes por interés lingüístico

Definición y alcance

Clasificación automática

- Ejemplos de clasificación en lingüística
- Formule una estrategia de clasificación para cada ejemplo

Definición y alcance

Agrupamiento (*clustering*)

- Dividir datos (objetos) en grupos (*clusters*) de acuerdo a su similitud
- Los datos NO están agrupados previamente
- Los datos NO tienen una clase asignada previamente
- NO hay conjunto de datos de entrenamiento
- Descubrir las clases
- Descubrir la estructura de los datos

Definición y alcance

Agrupamiento (*clustering*)

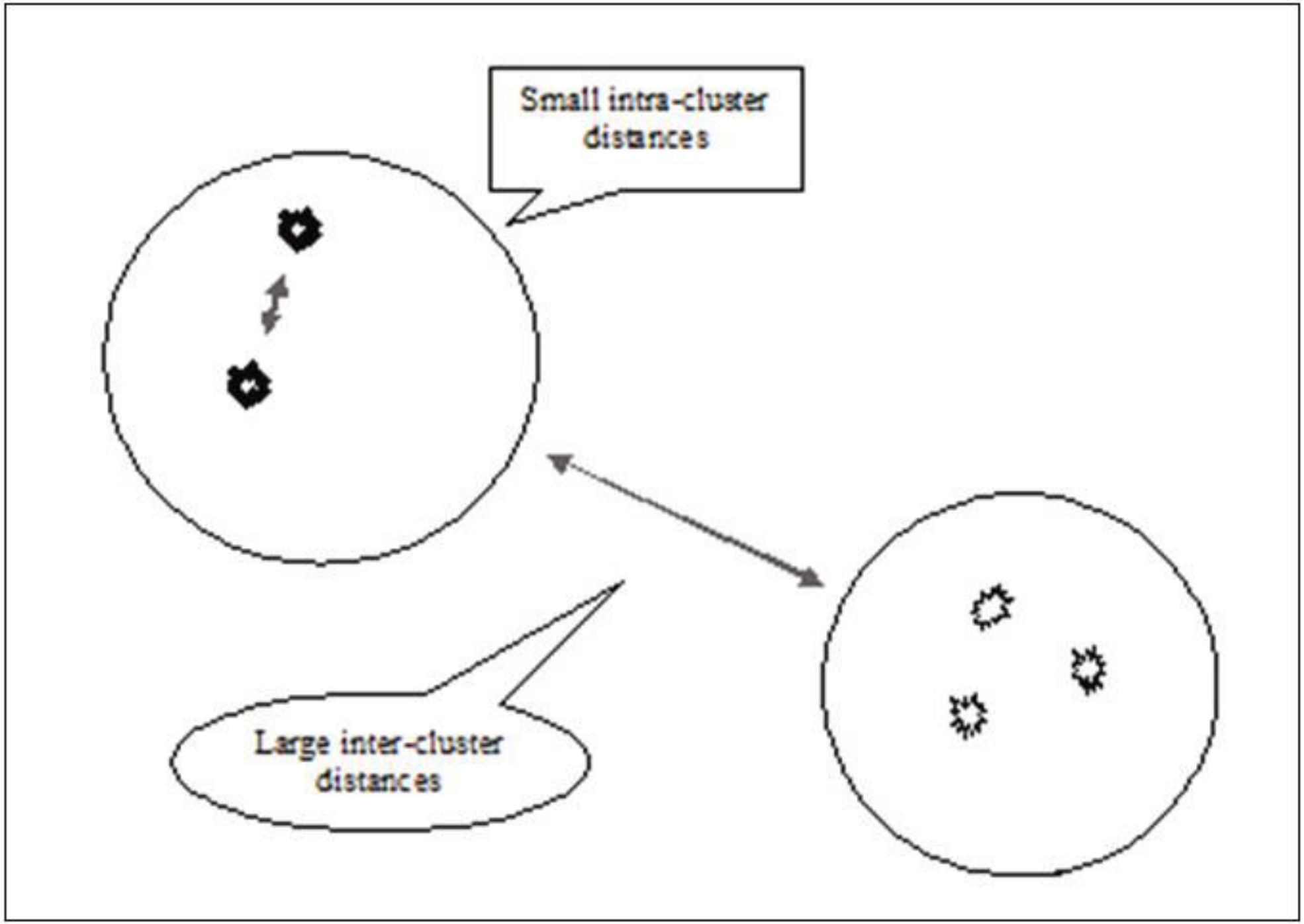
- Dividir tal que
- 1) Objetos de un mismo cluster son más similares entre sí y
- 2) Objetos de distintos clusters son menos similares entre sí

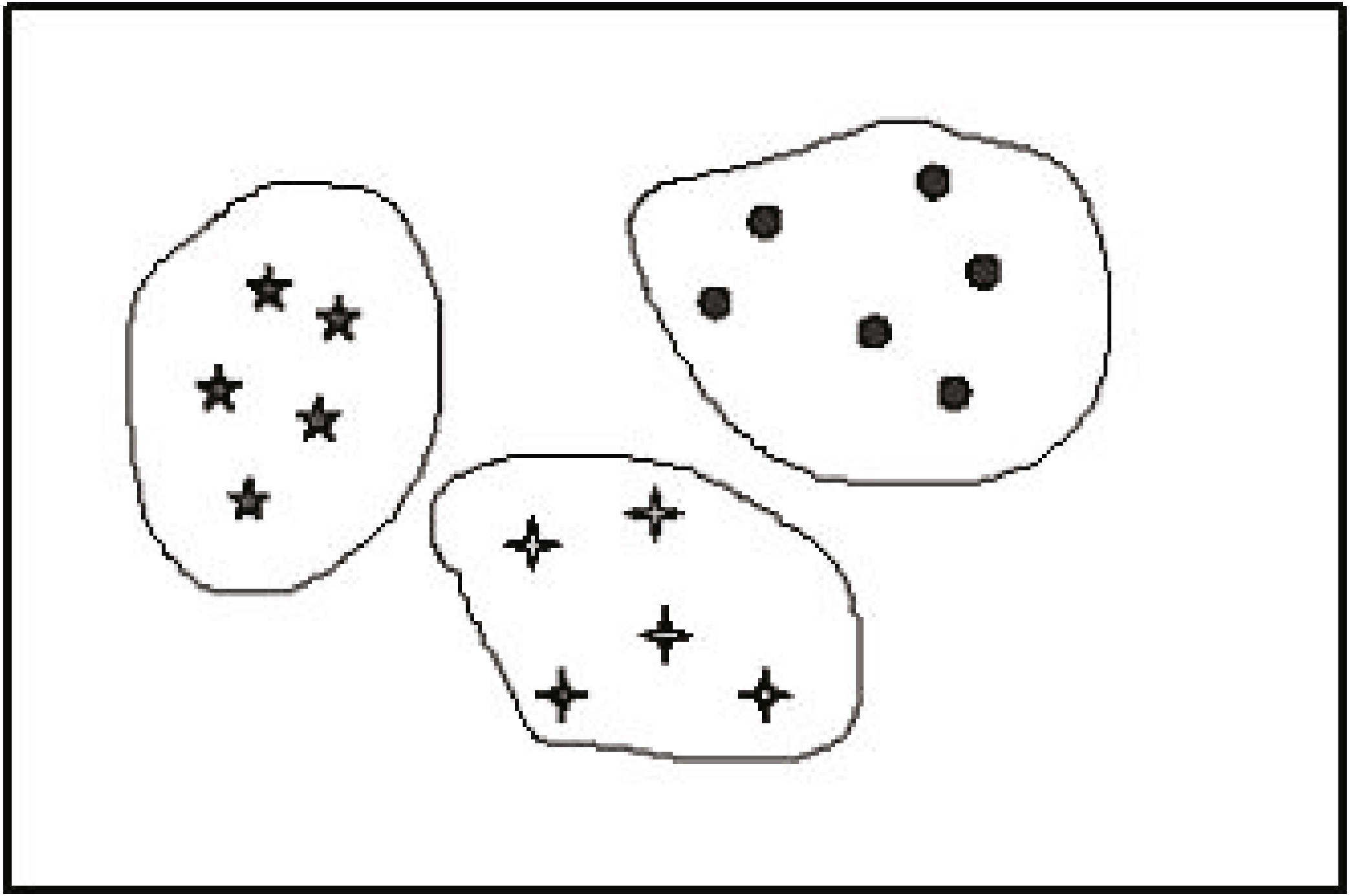
- 1) = similitud intraclusters
- 2) = desemejanza interclusters

Definición y alcance

Agrupamiento (*clustering*)

- Medida de similitud o distancia
- A + distancia, - similitud


Small intra-cluster distances
Large inter-cluster distances



Definición y alcance

Agrupamiento (*clustering*)

- Segmentación de mercado = separar clientes en distintos grupos
- Agrupamiento de documentos = separar documentos en temas
- Descubrimiento de familias de genes

Definición y alcance

Agrupamiento (*clustering*)

- Agrupamiento de estudiantes por perfil

Definición y alcance

Agrupamiento (*clustering*)

- Ejemplos de agrupamiento en lingüística
- Formule una estrategia de agrupamiento para cada ejemplo

Fin