

## Processament del Llenguatge Humà 6. Desambiguació de sentits



UNIVERSITAT POLITÈCNICA DE CATALUNYA  
BARCELONATECH

Facultat d'Informàtica de Barcelona



# Índice

Desambiguación  
de sentidos  
(WSD)

Métodos de  
WSD

- 1 Desambiguación de sentidos (WSD)
  - Objetivo y motivación
  - Recursos
- 2 Métodos de WSD
  - Tipos de algoritmos
  - Basados en corpus: supervisados
  - Basados en conocimiento

# Índice

Desambiguación  
de sentidos  
(WSD)

Objetivo y  
motivación

Métodos de  
WSD

- 1 Desambiguación de sentidos (WSD)
  - Objetivo y motivación
    - Recursos
- 2 Métodos de WSD

# Objetivo

- Los recursos léxico semánticos nos aportan los diferentes sentidos de cada palabra (polisemia)

<b>lema</b>	<b>PoS</b>	<b>sentido</b>
perro	NN	1. (m,f) mamífero domestico
		2. (m,f) persona despreciable
		3. (m) mal o daño ocasionado por engaño
		...
	JJ	1. (f) dicho de una cosa: mala o indigna
		...

# Objetivo

- Los recursos léxico semánticos nos aportan los diferentes sentidos de cada palabra (polisemia)

lema	PoS	sentido
perro	NN	1. (m,f) mamífero domestico
		2. (m,f) persona despreciable
		3. (m) mal o daño ocasionado por engaño
		...
	JJ	1. (f) dicho de una cosa: mala o indigna
		...

- **Objetivo:** seleccionar el sentido correcto de una palabra en una oración (NN, JJ, VB y quizás ADV)

# Motivación

Desambiguación  
de sentidos  
(WSD)

Objetivo y  
motivación

Métodos de  
WSD

- WSD es potencialmente útil en muchas tareas de PLN:
  - Adquisición automática de conocimiento léxico
  - Análisis semántico de la oración
  - Análisis del sentimiento
  - Extracción de información
  - Question answering
  - Traducción automática

# Motivación

- P.e.: Análisis semántico: se necesita seleccionar el sentido correcto de cada palabra para construir el significado de la oración

sentido	glosa de WordNet 1.5
age_N_1	the length of time something (or someone) has existed
age_N_2	a historic period

He was mad about stars at the age of nine .

# Motivación

- P.e.: TA: traducir la palabra *dog* en una oración en inglés al español

NOUN

1. (animal)

a. **el perro (m), la perra (f)**

My dog is a German Shepherd. — Mi perro es un pastor alemán.



2. (colloquial) (wicked person)

a. **el bribón (m), la bribona (f)**

My coworker is a lazy dog; I'm always having to do his work. — Mi colega es un bribón perezoso; siempre le tengo que estar haciendo el trabajo.

b. **el canalla (m), la canalla (f) (colloquial)**

That dog started cheating on his girlfriend almost as soon as they started going out. — Ese canalla le pegó cuernos a su novia prácticamente tan pronto empezaron a salir.

3. (pejorative) (unattractive woman)

TRANSITIVE VERB

4. (to follow)

a. **seguir**

The neighborhood bullies dogged him all the way to his house. — Los matones del vecindario lo siguieron el camino entero hasta llegar a su casa.

5. (to plague)

a. **perseguir**

He has been dogged by scandal his entire career. — El escándalo lo ha perseguido durante su carrera entera.

Source: <http://www.spanishdict.com>.

Desambiguación  
de sentidos  
(WSD)

Objetivo y  
motivación

Métodos de  
WSD

# Índice

Desambiguación  
de sentidos  
(WSD)

Recursos

Métodos de  
WSD

## 1 Desambiguación de sentidos (WSD)

- Objetivo y motivación

- Recursos

## 2 Métodos de WSD

# Recursos

Desambiguación  
de sentidos  
(WSD)

Recursos

Métodos de  
WSD

Recursos léxicos que suelen usarse:

- basados en conocimiento: WordNet, BabelNet (redes semánticas léxicas)
- basados en corpus: word embeddings, sense embeddings

# Recursos

## Corpora y marcos de evaluación:

- Todas las ocurrencias de **una palabra** anotadas con su sentido correcto (una muestra por palabra)
  - SemEval Lexical Sample Task (training/Test corpus)
  - aprendizaje supervisado de modelos de desambiguación

800004

Mr Purves is tight-lipped about what happens then.  
He vexed rumour-mongers, who <tag '520051'>bet</> on a bid  
for Midlan sooner.

800005

Mr Jones loses his <tag '519914'>bet</>:1,000 people attended  
Cowley pools last year.

# Recursos

## Corpora y marcos de evaluación:

- Todas las ocurrencias de **una palabra** anotadas con su sentido correcto (una muestra por palabra)
  - SemEval Lexical Sample Task (training/Test corpus)
  - aprendizaje supervisado de modelos de desambiguación

800004

Mr Purves is tight-lipped about what happens then.  
He vexed rumour-mongers, who `<tag '520051'>bet</>` on a bid  
for Midlan sooner.

800005

Mr Jones loses his `<tag '519914'>bet</>`:1,000 people attended  
Cowley pools last year.

- Todas las ocurrencias de **todas las palabras** anotadas con su sentido correcto
  - Semcor, SemEval All Words Task (Test corpus)
  - métodos no supervisado

# Índice

Desambiguación  
de sentidos  
(WSD)

Métodos de  
WSD

- 1 Desambiguación de sentidos (WSD)
  - Objetivo y motivación
  - Recursos
  
- 2 Métodos de WSD
  - Tipos de algoritmos
  - Basados en corpus: supervisados
  - Basados en conocimiento

# Índice

Desambiguación  
de sentidos  
(WSD)

Métodos de  
WSD

Tipos de algoritmos

## 1 Desambiguación de sentidos (WSD)

## 2 Métodos de WSD

- Tipos de algoritmos
  - Basados en corpus: supervisados
  - Basados en conocimiento

# Tipos de algoritmos para WSD

- **Basados en corpus:**

- **Supervisados:**

- Ocurrencias de una palabra anotadas con sus sentidos correctos
    - P.e.: Naïve Bayes, kNN, SVM, redes neuronales

# Tipos de algoritmos para WSD

- **Basados en corpus:**

- **Supervisados:**

- Ocurrencias de una palabra anotadas con sus sentidos correctos
    - P.e.: Naïve Bayes, kNN, SVM, redes neuronales

- **Semisupervisados:**

- Unas pocas ocurrencias anotadas de una palabra. Muchas no anotadas
    - P.e.: Algoritmo de Yarowsky (Boostrapping)

# Tipos de algoritmos para WSD

Desambiguación  
de sentidos  
(WSD)

Métodos de  
WSD

Tipos de algoritmos

- **Basados en corpus:**
  - **Supervisados:**
    - Ocurrencias de una palabra anotadas con sus sentidos correctos
    - P.e.: Naïve Bayes, kNN, SVM, redes neuronales
  - **Semisupervisados:**
    - Unas pocas ocurrencias anotadas de una palabra. Muchas no anotadas
    - P.e.: Algoritmo de Yarowsky (Boostrapping)
- **Basados en el conocimiento:**
  - **No supervisados**
  - P.e.: Algoritmo de Lesk (accesible via NLTK), UKB (accesible via TextServer)

# Índice

Desambiguación  
de sentidos  
(WSD)

Métodos de  
WSD

Basados en corpus:  
supervisados

## 1 Desambiguación de sentidos (WSD)

## 2 Métodos de WSD

- Tipos de algoritmos
- Basados en corpus: supervisados
- Basados en conocimiento

# Basados en corpus: supervisados

WSD como problema de clasificación: aprender **un modelo** para desambiguar menciones de **una palabra** en un texto

Desambiguación  
de sentidos  
(WSD)

Métodos de  
WSD

Basados en corpus:  
supervisados

## Basados en corpus: supervisados

WSD como problema de clasificación: aprender **un modelo** para desambiguar menciones de **una palabra** en un texto

Aplicar **cada modelo** a una nueva oración previamente lematizada

veo un banco de peces desde el banco

Desambiguación  
de sentidos  
(WSD)

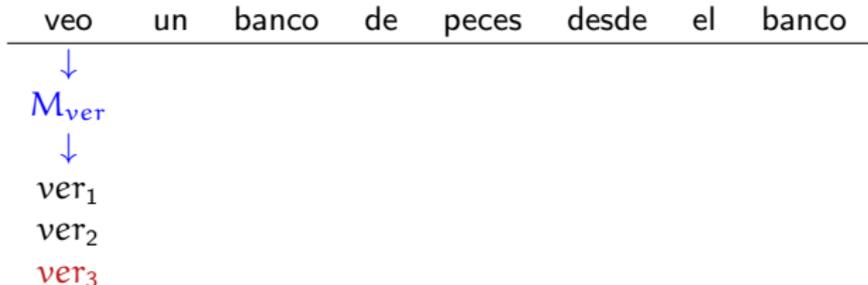
Métodos de  
WSD

Basados en corpus:  
supervisados

# Basados en corpus: supervisados

WSD como problema de clasificación: aprender **un modelo** para desambiguar menciones de **una palabra** en un texto

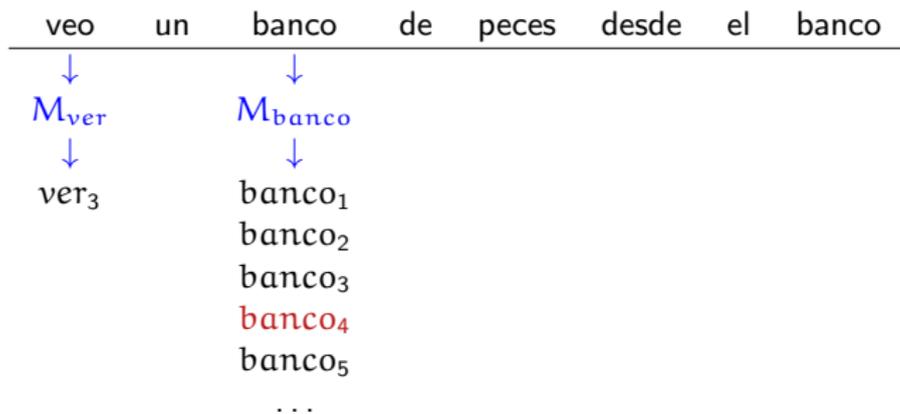
Aplicar **cada modelo** a una nueva oración previamente lematizada



# Basados en corpus: supervisados

WSD como problema de clasificación: aprender **un modelo** para desambiguar menciones de **una palabra** en un texto

Aplicar **cada modelo** a una nueva oración previamente lematizada



Desambiguación  
de sentidos  
(WSD)

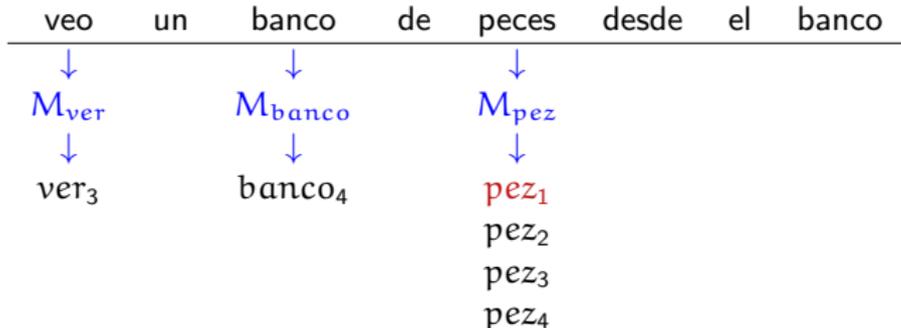
Métodos de  
WSD

Basados en corpus:  
supervisados

## Basados en corpus: supervisados

WSD como problema de clasificación: aprender **un modelo** para desambiguar menciones de **una palabra** en un texto

Aplicar **cada modelo** a una nueva oración previamente lematizada





## Basados en corpus: supervisados

WSD como problema de clasificación: aprender **un modelo** para desambiguar menciones de **una palabra** en un texto

- **Conjunto de clases:**

$\{\text{sentido}_1 \dots, \text{sentido}_k\}$

P.e.:

44 sentidos diferentes de la palabra bajo (NN, JJ, VB)

## Basados en corpus: supervisados

WSD como problema de clasificación: aprender **un modelo** para desambiguar menciones de **una palabra** en un texto

Desambiguación  
de sentidos  
(WSD)

Métodos de  
WSD

Basados en corpus:  
supervisados

- **Conjunto de clases:**

$\{\text{sentido}_1 \dots, \text{sentido}_k\}$

P.e.:

44 sentidos diferentes de la palabra bajo (NN, JJ, VB)

- **Corpus anotado:**

$\{ \langle \text{mención}_i, \text{contexto}_i, \text{sentido\_correcto}_i \rangle \}$

P.e.:

texto:	el	niño	$e_1^+$ bajo	toca	el	bajo
POS:	DT	NN	JJ	VB	DT	NN
			01206474-a			02803349-n

# Basados en corpus: supervisados

WSD como problema de clasificación: aprender **un modelo** para desambiguar menciones de **una palabra** en un texto

Desambiguación  
de sentidos  
(WSD)

Métodos de  
WSD

Basados en corpus:  
supervisados

- **Conjunto de clases:**

$\{\text{sentido}_1 \dots, \text{sentido}_k\}$

P.e.:

44 sentidos diferentes de la palabra bajo (NN, JJ, VB)

- **Corpus anotado:**

$\{\langle \text{mención}_i, \text{contexto}_i, \text{sentido\_correcto}_i \rangle\}$

P.e.:

<b>texto:</b>	el	niño	bajo	toca	el	
<b>POS:</b>	DT	NN	JJ	VB	DT	

01206474-a

$e_2^+$   
bajo  
NN  
02803349-n

# Basados en corpus: supervisados

WSD como problema de clasificación: aprender **un modelo** para desambiguar menciones de **una palabra** en un texto

- Ejemplos de aprendizaje:

$\{e^+\}$ :  $\{ \langle \text{mención}_i, \text{contexto}_i, \text{sentido\_correcto}_i \rangle \}$

$\{e^-\}$ :  $\{ \langle \text{mención}_i, \text{contexto}_i, \text{sentido\_incorrecto}_i \rangle \}$

# Basados en corpus: supervisados

WSD como problema de clasificación: aprender **un modelo** para desambiguar menciones de **una palabra** en un texto

- Ejemplos de aprendizaje:

$\{e^+\}$ :  $\{ \langle \text{mención}_i, \text{contexto}_i, \text{sensido\_correcto}_i \rangle \}$

$\{e^-\}$ :  $\{ \langle \text{mención}_i, \text{contexto}_i, \text{sensido\_incorrecto}_i \rangle \}$

- Su representación en aprendizaje basado en atributos:

- de contexto local: palabra+posición, lema+posición, POS+posición

p.e.: *come up with* → w+1\_up, w+2\_with

Desambiguación  
de sentidos  
(WSD)

Métodos de  
WSD

Basados en corpus:  
supervisados

# Basados en corpus: supervisados

WSD como problema de clasificación: aprender **un modelo** para desambiguar menciones de **una palabra** en un texto

- Ejemplos de aprendizaje:

$\{e^+\}$ :  $\{ \langle \text{mención}_i, \text{contexto}_i, \text{sentido\_correcto}_i \rangle \}$

$\{e^-\}$ :  $\{ \langle \text{mención}_i, \text{contexto}_i, \text{sentido\_incorrecto}_i \rangle \}$

- Su representación en aprendizaje basado en atributos:

- de contexto local: palabra+posición, lema+posición, POS+posición

p.e.: *come up with* → w+1\_up, w+2\_with

- de contexto amplio: bolsa de palabras, lemas, bigramas, ...

p.e.: *I was **studying** at U.P.C. in Barcelona for 2 years* → l+\_year, co+\_U.P.C.\_Barcelona.

# Basados en corpus: supervisados

WSD como problema de clasificación: aprender **un modelo** para desambiguar menciones de **una palabra** en un texto

- Ejemplos de aprendizaje:

$\{e^+\}$ :  $\{ \langle \text{mención}_i, \text{contexto}_i, \text{sentido\_correcto}_i \rangle \}$

$\{e^-\}$ :  $\{ \langle \text{mención}_i, \text{contexto}_i, \text{sentido\_incorrecto}_i \rangle \}$

- Su representación en aprendizaje basado en atributos:

- de contexto local: palabra+posición, lema+posición, POS+posición

p.e.: *come up with* → w+1\_up, w+2\_with

- de contexto amplio: bolsa de palabras, lemas, bigramas, ...

p.e.: *I was **studying** at U.P.C. in Barcelona for 2 years* → l+\_year, co+\_U.P.C.\_Barcelona.

- de sintaxis: funciones sintácticas

p.e.: *cats **eat** fish.* → subj\_cat, obj\_fish

# Basados en corpus: supervisados

WSD como problema de clasificación: aprender **un modelo** para desambiguar menciones de **una palabra** en un texto

- Ejemplos de aprendizaje:

$\{e^+\}$ :  $\{ \langle \text{mención}_i, \text{contexto}_i, \text{sentido\_correcto}_i \rangle \}$

$\{e^-\}$ :  $\{ \langle \text{mención}_i, \text{contexto}_i, \text{sentido\_incorrecto}_i \rangle \}$

- Su representación en aprendizaje basado en atributos:

- de contexto local: palabra+posición, lema+posición, POS+posición

p.e.: *come up with* → w+1\_up, w+2\_with

- de contexto amplio: bolsa de palabras, lemas, bigramas, ...

p.e.: *I was **studying** at U.P.C. in Barcelona for 2 years* → l+\_year, co+\_U.P.C.\_Barcelona.

- de sintaxis: funciones sintácticas

p.e.: *cats **eat** fish.* → subj\_cat, obj\_fish

- de semántica: sentidos de palabras previas, word embeddings

# Basados en corpus: supervisados

WSD como problema de clasificación: aprender **un modelo** para desambiguar menciones de **una palabra** en un texto

- Ejemplos de aprendizaje:

$\{e^+\}$ :  $\{ \langle \text{mención}_i, \text{contexto}_i, \text{sentido\_correcto}_i \rangle \}$

$\{e^-\}$ :  $\{ \langle \text{mención}_i, \text{contexto}_i, \text{sentido\_incorrecto}_i \rangle \}$

- Su representación en aprendizaje basado en atributos:

- de contexto local: palabra+posición, lema+posición, POS+posición

p.e.: *come up with* → w+1\_up, w+2\_with

- de contexto amplio: bolsa de palabras, lemas, bigramas, ...

p.e.: *I was **studying** at U.P.C. in Barcelona for 2 years* → l+\_year, co+\_U.P.C.\_Barcelona.

- de sintaxis: funciones sintácticas

p.e.: *cats **eat** fish.* → subj\_cat, obj\_fish

- de semántica: sentidos de palabras previas, word embeddings

- Su representación en redes neuronales suele ser via word embeddings

# Ejercicio

Representar la siguiente oración via atributos de contexto local y amplio para ser usados por un algoritmo de AA

**Ejemplo** He was mad about stars at the **age** of nine .  
**age.n.01**

+ PoS ('He', 'PRP'), ('was', 'VBD'), ('mad', 'JJ'),  
('about', 'IN'), ('stars', 'NNS'), ('at', 'IN'),  
('the', 'DT'), ('age', 'NN'), ('of', 'IN'),  
('nine', 'CD'), (',', ',')

- 1 Aporta la bolsa de palabras del contexto
- 2 Aporta atributos sobre formas en una ventana de  $\pm 2$  palabras.
- 3 Aporta dos atributos más.

# Basados en corpus: supervisados

WSD como problema de clasificación: aprender **un modelo** para desambiguar menciones de **una palabra** en un texto

Problemáticas:

- Problema de la existencia de modelos para todas las palabras de un idioma
- Problema de la adquisición de corpus anotado

Desambiguación  
de sentidos  
(WSD)

Métodos de  
WSD

Basados en corpus:  
supervisados

# Basados en corpus: supervisados

WSD como problema de clasificación: aprender **un modelo** para desambiguar menciones de **una palabra** en un texto

Desambiguación  
de sentidos  
(WSD)

Métodos de  
WSD

Basados en corpus:  
supervisados

Problemáticas:

- Problema de la existencia de modelos para todas las palabras de un idioma
- Problema de la adquisición de corpus anotado

Alternativas:

- métodos semisupervisados (requieren pocos ejemplos anotados y muchos no anotados -p.e.: bootstrapping-)
- métodos basados en conocimiento

# Índice

Desambiguación  
de sentidos  
(WSD)

Métodos de  
WSD

Basados en  
conocimiento

## 1 Desambiguación de sentidos (WSD)

## 2 Métodos de WSD

- Tipos de algoritmos
- Basados en corpus: supervisados
- Basados en conocimiento
  - Algoritmo de Lesk
  - UKB

## Basados en conocimiento: Lesk

$$\text{Lesk}(w) = \underset{s_i \in S(\{w\})}{\operatorname{argmax}} \forall_{s_j \in S(C(w))} |\text{Def}(s_i) \cap \text{Def}(s_j)|$$

Desambiguación  
de sentidos  
(WSD)

Métodos de  
WSD

Algoritmo de Lesk

$w$ : palabra de categoría abierta

$S(X)$ : conjunto de los sentidos de todos los lemas en  $X$

$C(w)$ : conjunto de los lemas del contexto de la palabra  $w$ .

$\text{Def}(s)$ : conjunto de lemas de la definición del sentido  $s$ .

- Desambigua una sola palabra en su contexto
- Útil para desambiguar palabras en oraciones cortas o en una ventana centrada de  $K$  palabras contextuales.

# Basados en conocimiento: Lesk

## Ejemplo

**Input:** "pine cone"

### **PINE**

1. kinds of evergreen tree with needle-shaped leaves
2. waste away through sorrow or illness

### **CONE**

1. solid body which narrows to a point
2. something of this shape whether solid or hollow
3. fruit of certain evergreen trees

# Basados en conocimiento: Lesk

## Ejemplo

Input: "pine cone"

### PINE

1. kinds of evergreen tree with needle-shaped leaves
2. waste away through sorrow or illness

### CONE

1. solid body which narrows to a point
2. something of this shape whether solid or hollow
3. fruit of certain evergreen trees

**Solution** (sin contar las *stopwords*)

Mejor intersección:  $\text{Pine\#1} \cap \text{Cone\#3} = 2.$

sentido para "pine": Pine#1

# Basados en conocimiento: Lesk

## Ejemplo

Input: "pine cone"

### PINE

1. kinds of evergreen tree with needle-shaped leaves
2. waste away through sorrow or illness

### CONE

1. solid body which narrows to a point
2. something of this shape whether solid or hollow
3. fruit of certain evergreen trees

**Solution** (sin contar las *stopwords*)

Mejor intersección: Pine#1  $\cap$  Cone#3 = 2.

sentido para "cone": Cone#3

# Basados en conocimiento: Lesk

## Algoritmo de Lesk simplificado

$$\text{Lesk}(w) = \operatorname{argmax}_{s_i \in S(\{w\})} |\text{Def}(s_i) \cap C(w)|$$

$w$ : palabra de categoría abierta

$S(X)$ : conjunto de los sentidos de todos los lemas en  $X$

$C(w)$ : conjunto de los lemas del contexto de la palabra  $w$ .

$\text{Def}(s)$ : conjunto de lemas de la definición del sentido  $s$ .

- En general, mejores resultados que el algoritmo de Lesk original

# Ejercicio

Dada la siguiente oración:

I went to the bank to deposit money.

y las siguientes definiciones de los 2 sentidos de la palabra **bank**:

- 1 sloping land (especially the slope beside a body of water)
- 2 a financial institution that accepts deposits and channels the money into lending activities

aplica el algoritmo de Lesk simplificado para encontrar el sentido correcto

# Basados en conocimiento: Lesk

## Extensiones para paliar la baja cobertura:

- Cambio de función de similaridad (p.e., coseno)
- Enriquecimiento con WordNet (Adapted/Extended Lesk) (Banerjee and Pederson, 2002/2003)
  - Uso de ejemplos de synsets de WordNet
  - Uso de datos de hiperónimos y/o hipónimos
- Filtrado de sentidos del contexto via POS-tagging y ponderación de sentidos candidatos via frecuencia de uso (WordNet) (Vasilei et al 2004)
- Enriquecimiento con WordNet y Wikipedia (Enhanced Lesk) (Basile et al. 2014)
- Representar palabras via word embeddings o sense embeddings (Oele y Gertjan, 2019)
- ...

# Basados en conocimiento: UKB

- Alternativas para desambiguar una o **todas las palabras del contexto a la vez**
- Basado en el algoritmo de **PageRank** de Google
  - input:** red de páginas web definida por sus enlaces
  - output:** relevancia de cada página web incluida en la red

# Basados en conocimiento: UKB

Desambiguación  
de sentidos  
(WSD)

Métodos de  
WSD

UKB

- Alternativas para desambiguar una o **todas las palabras del contexto a la vez**
- Basado en el algoritmo de **PageRank** de Google
  - input:** **red de páginas web** definida por sus enlaces
  - output:** relevancia de cada **página web** incluida en la red
- Analogía:
  - input:** texto a desambiguar y **grafo de sentidos de palabras** definido por sus relaciones (p.e. WordNet)
  - output:** relevancia de cada **sentido de palabra** incluida en el texto

# Basados en conocimiento: UKB

## 1. PRELIMINAR ¿Cómo funciona **PageRank**?

- \* Relevancia de cada página = prob. de ser visitada via enlaces

Desambiguación  
de sentidos  
(WSD)

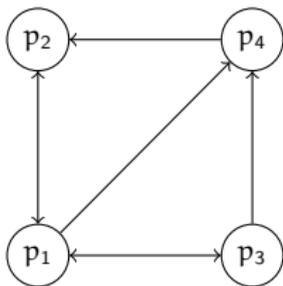
Métodos de  
WSD

UKB

# Basados en conocimiento: UKB

## 1. PRELIMINAR ¿Cómo funciona PageRank?

- \* Relevancia de cada página = prob. de ser visitada via enlaces



matriz de transición

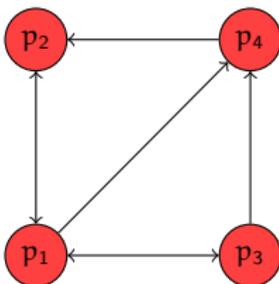
$$H = \begin{bmatrix} 0 & 1 & 1/2 & 0 \\ 1/3 & 0 & 0 & 1 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 1/2 & 0 \end{bmatrix}$$

# Basados en conocimiento: UKB

## 1. PRELIMINAR ¿Cómo funciona PageRank?

- \* Relevancia de cada página = prob. de ser visitada via enlaces
- \* Encontrar la distribución estacionaria

$$v_{(i+1)} = H \cdot v_i \quad v_0 = [1/n]_n$$



matriz de transición

$$H = \begin{bmatrix} 0 & 1 & 1/2 & 0 \\ 1/3 & 0 & 0 & 1 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 1/2 & 0 \end{bmatrix}$$

vector inicial de relevancias

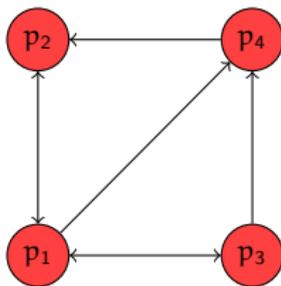
$$v_0 = \begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}$$

# Basados en conocimiento: UKB

## 1. PRELIMINAR ¿Cómo funciona PageRank?

- \* Relevancia de cada página = prob. de ser visitada via enlaces
- \* Encontrar la distribución estacionaria

$$v_{(i+1)} = H \cdot v_i \quad v_0 = [1/n]_n$$



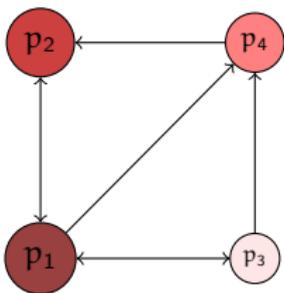
$$v_1 = \begin{bmatrix} - \\ - \\ - \\ - \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1/2 & 0 \\ 1/3 & 0 & 0 & 1 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 1/2 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}$$

# Basados en conocimiento: UKB

## 1. PRELIMINAR ¿Cómo funciona PageRank?

- \* Relevancia de cada página = prob. de ser visitada via enlaces
- \* Encontrar la distribución estacionaria

$$v_{(i+1)} = H \cdot v_i \quad v_0 = [1/n]_n$$



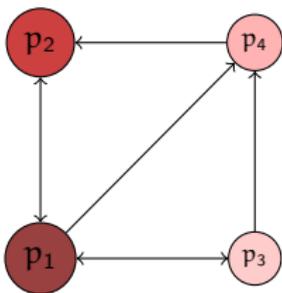
$$v_1 = \begin{bmatrix} 0.375 \\ 0.333 \\ 0.083 \\ 0.208 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1/2 & 0 \\ 1/3 & 0 & 0 & 1 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 1/2 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}$$

# Basados en conocimiento: UKB

## 1. PRELIMINAR ¿Cómo funciona PageRank?

- \* Relevancia de cada página = prob. de ser visitada via enlaces
- \* Encontrar la distribución estacionaria

$$v_{(i+1)} = H \cdot v_i \quad v_0 = [1/n]_n$$



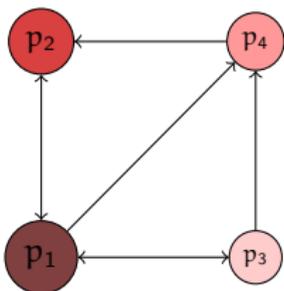
$$v_2 = \begin{bmatrix} 0.374 \\ 0.333 \\ 0.125 \\ 0.166 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1/2 & 0 \\ 1/3 & 0 & 0 & 1 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 1/2 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0.375 \\ 0.333 \\ 0.083 \\ 0.208 \end{bmatrix}$$

# Basados en conocimiento: UKB

## 1. PRELIMINAR ¿Cómo funciona PageRank?

- \* Relevancia de cada página = prob. de ser visitada via enlaces
- \* Encontrar la distribución estacionaria

$$v_{(i+1)} = H \cdot v_i \quad v_0 = [1/n]_n$$



$$v_3 = \begin{bmatrix} 0.395 \\ 0.291 \\ 0.125 \\ 0.187 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1/2 & 0 \\ 1/3 & 0 & 0 & 1 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 1/2 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0.374 \\ 0.333 \\ 0.125 \\ 0.166 \end{bmatrix}$$

# Basados en conocimiento: UKB

Desambiguación  
de sentidos  
(WSD)

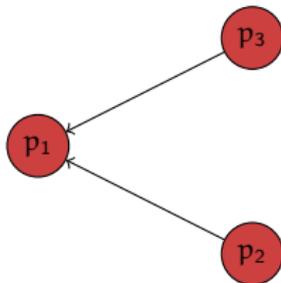
Métodos de  
WSD

UKB

## 1. PRELIMINAR ¿Cómo funciona PageRank?

**PROBLEMA:** páginas sin enlaces salientes y grafos desconexos

$$v_{(i+1)} = H \cdot v_i \quad v_0 = [1/n]_n$$



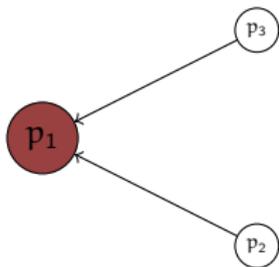
$$v_1 = \begin{bmatrix} - \\ - \\ - \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

# Basados en conocimiento: UKB

## 1. PRELIMINAR ¿Cómo funciona PageRank?

**PROBLEMA:** páginas sin enlaces salientes y grafos desconexos

$$v_{(i+1)} = H \cdot v_i \quad v_0 = [1/n]_n$$



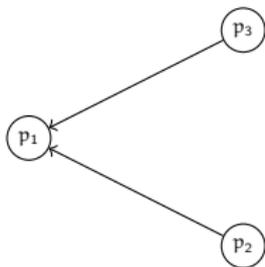
$$v_1 = \begin{bmatrix} 0.66 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

# Basados en conocimiento: UKB

## 1. PRELIMINAR ¿Cómo funciona PageRank?

**PROBLEMA:** páginas sin enlaces salientes y grafos desconexos

$$v_{(i+1)} = H \cdot v_i \quad v_0 = [1/n]_n$$



$$v_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0.66 \\ 0 \\ 0 \end{bmatrix}$$

# Basados en conocimiento: UKB

## 1. PRELIMINAR ¿Cómo funciona **PageRank**?

**PROBLEMA:** páginas sin enlaces salientes y grafos desconexos

**SOLUCIÓN:** selección aleatoria de una página

$$v_{(i+1)} = H \cdot v_i \quad v_0 = [1/n]_n$$

# Basados en conocimiento: UKB

## 1. PRELIMINAR ¿Cómo funciona PageRank?

PROBLEMA: páginas sin enlaces salientes y grafos desconexos

**SOLUCIÓN:** selección aleatoria de una página

$$v_{(i+1)} = M \cdot v_i \quad v_0 = [1/n]_n$$

$$M = (1 - \alpha) \cdot H + \alpha \cdot B$$

M: matriz de PageRank

H: matriz de transición

$\alpha$ : probabilidad de selección aleatoria (por defecto 0.15)

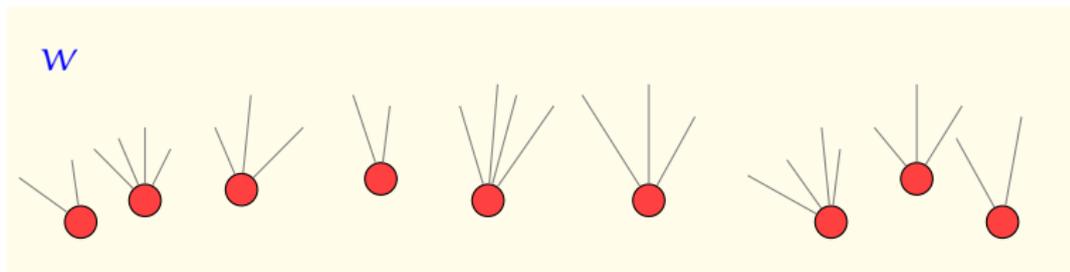
B: matriz  $[1/n]_n^n$

# Basados en conocimiento: UKB

## 2. WSD usando PageRank

- \* Usa WordNet como grafo

$$v_{(i+1)} = M_W \cdot v_i \quad v_0 = [1/n]_n$$



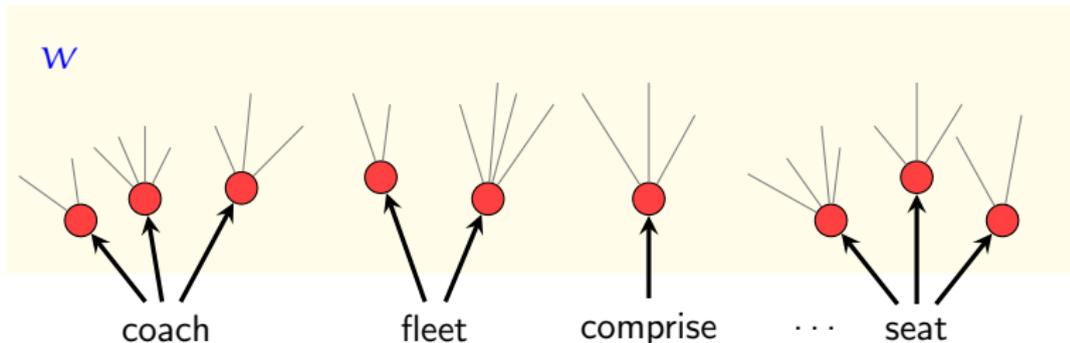
# Basados en conocimiento: UKB

## 2. WSD usando PageRank

- \* Usa WordNet como grafo

$$v_{(i+1)} = M_W \cdot v_i \quad v_0 = [1/n]_n$$

- \* Se enfoca en los synsets de las palabras del texto



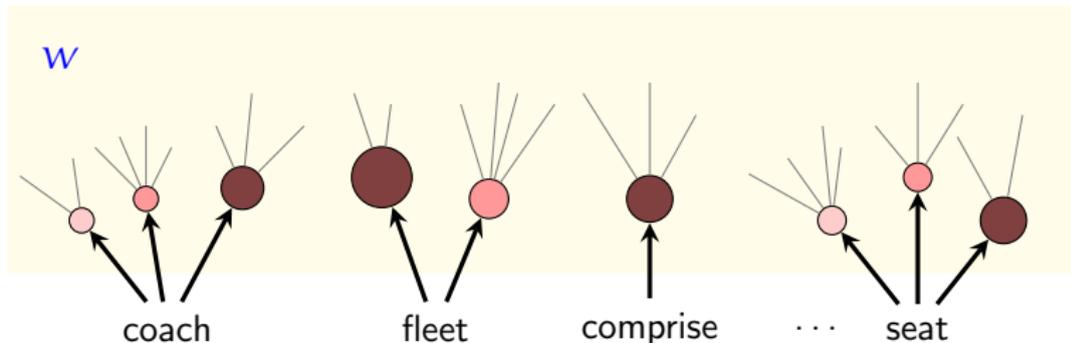
# Basados en conocimiento: UKB

## 2. WSD usando PageRank

- \* Usa WordNet como grafo

$$v_{(i+1)} = M_W \cdot v_i \quad v_0 = [1/n]_n$$

- \* Se enfoca en los synsets de las palabras del texto



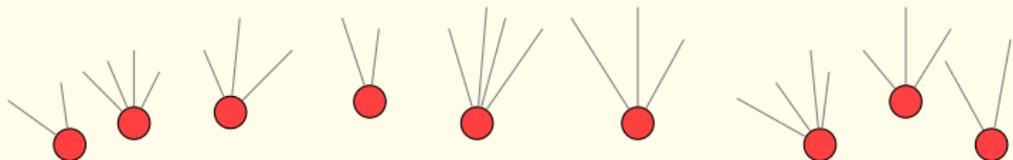
# Basados en conocimiento: UKB

## 2. WSD usando PageRank

¿Cómo se enfoca en los synsets de las k palabras del texto?

$$v_{(i+1)} = M_W \cdot v_i \quad v_0 = [1/n]_n$$

W



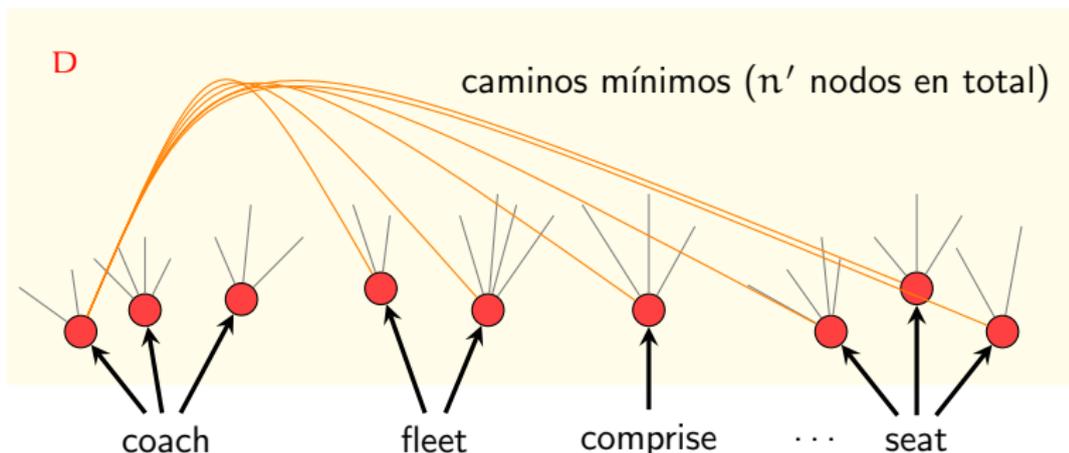
# Basados en conocimiento: UKB

## 2. WSD usando PageRank

¿Cómo se enfoca en los synsets de las  $k$  palabras del texto?

**OPCIÓN 1.** Restringir  $W$  al subgrafo de desambiguación

$$v_{(i+1)} = M_D \cdot v_i \quad v_0 = [1/n']_{n'}$$



# Basados en conocimiento: UKB

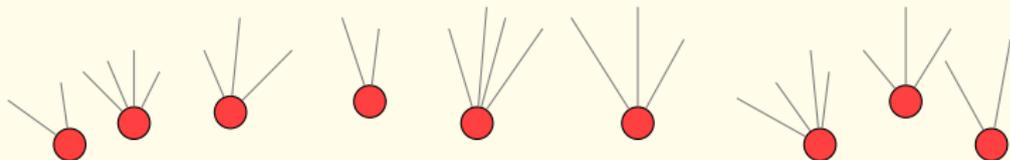
## 2. WSD usando PageRank

¿Cómo se enfoca en los synsets de las  $k$  palabras del texto?

**OPCIÓN 2.** Personalizar  $B$  a las  $k$  palabras del texto

$$v_{(i+1)} = M_W \cdot v_i \quad v_0 = [1/n]_n$$

$W$



# Basados en conocimiento: UKB

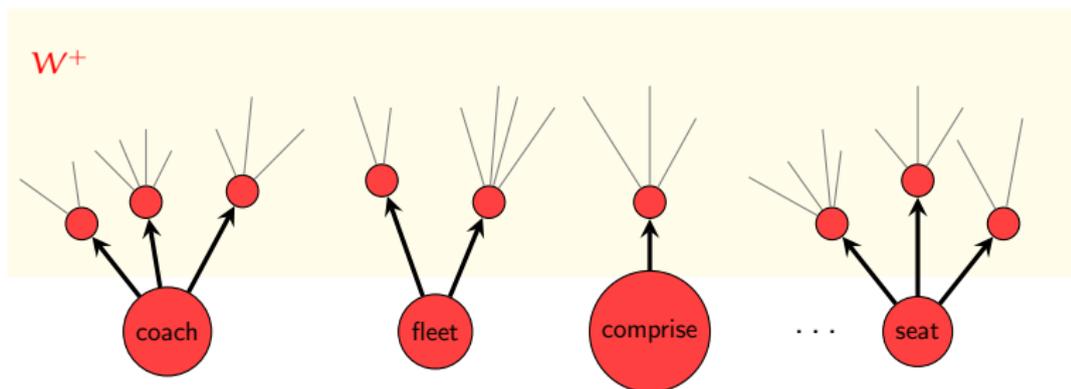
## 2. WSD usando PageRank

¿Cómo se enfoca en los synsets de las  $k$  palabras del texto?

**OPCIÓN 2.** Personalizar  $B$  a las  $k$  palabras del texto

$$v_{(i+1)} = M_{W^+} \cdot v_i \quad v_0 = [1/n + k]_{n+k}$$

- Insertamos las palabras como nuevos nodos ( $|W^+| = n' = n + k$ )



# Basados en conocimiento: UKB

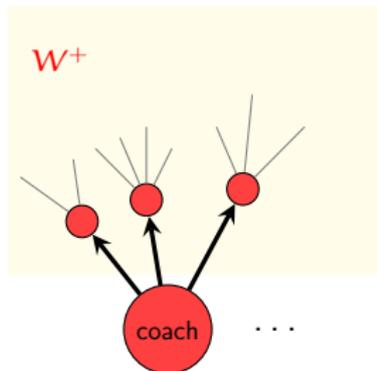
## 2. WSD usando PageRank

¿Cómo se enfoca en los synsets de las  $k$  palabras del texto?

**OPCIÓN 2.** Personalizar  $B$  a las  $k$  palabras del texto

$$M_{W^+} = (1 - \alpha) \cdot H_{W^+} + \alpha \cdot B_{W^+}$$

- Insertamos las palabras como nuevos nodos ( $|W^+| = n' = n + k$ )



$$H_{W^+} = \begin{bmatrix} \overbrace{\quad\quad\quad}^n & \overbrace{\quad\quad\quad}^k \\ \begin{matrix} H_W & \begin{matrix} 0 & \dots \\ 1/3 & \dots \\ 1/3 & \dots \\ \dots & \dots \\ 0 & \dots \\ 0 & \dots \\ 1/3 & \dots \end{matrix} \end{matrix} \\ \hline \begin{matrix} 0 & 0 \end{matrix} \end{bmatrix} \begin{matrix} \left. \vphantom{\begin{matrix} 0 \\ 1/3 \\ 1/3 \\ \dots \\ 0 \\ 0 \\ 1/3 \end{matrix}} \right\} n \\ \left. \vphantom{\begin{matrix} 0 \\ 0 \end{matrix}} \right\} k \end{matrix}$$

# Basados en conocimiento: UKB

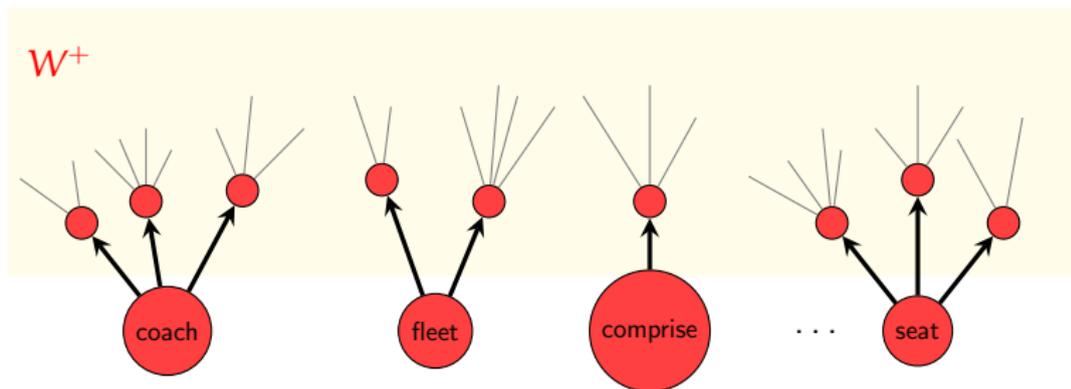
## 2. WSD usando PageRank

¿Cómo se enfoca en los synsets de las  $k$  palabras del texto?

**OPCIÓN 2.** Personalizar  $B$  a las  $k$  palabras del texto

$$M_{W^+} = (1 - \alpha) \cdot H_{W^+} + \alpha \cdot B_{W^+}$$

- Concentramos en ellos la probabilidad de salto aleatorio



# Basados en conocimiento: UKB

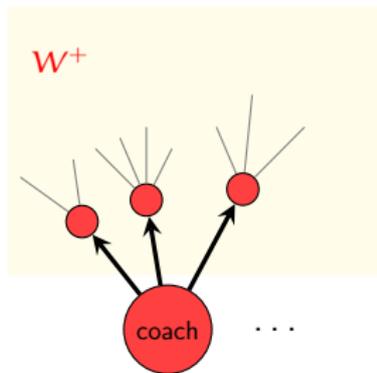
## 2. WSD usando PageRank

¿Cómo se enfoca en los synsets de las  $k$  palabras del texto?

**OPCIÓN 2.** Personalizar  $B$  a las  $k$  palabras del texto

$$M_{W^+} = (1 - \alpha) \cdot H_{W^+} + \alpha \cdot B_{W^+}$$

- Concentramos en ellos la probabilidad de salto aleatorio



$$B_{W^+} = \begin{bmatrix} \overbrace{\quad\quad\quad}^n & \overbrace{\quad\quad\quad}^k \\ \quad\quad\quad 0 & \quad\quad\quad 0 \\ \hline 1/n' & \dots & \dots & 1/n' \\ \dots & \dots & \dots & \dots \\ 1/n' & \dots & \dots & 1/n' \end{bmatrix} \begin{matrix} \left. \vphantom{\begin{matrix} 0 \\ 0 \end{matrix}} \right\} n \\ \left. \vphantom{\begin{matrix} 1/n' \\ \dots \\ 1/n' \end{matrix}} \right\} k \end{matrix}$$