

# Indexation et recherche d'images par fusion d'informations textuelles et visuelles

Sabrina TOLLARI

Directeurs de thèse : Hervé GLOTIN et Jacques LE MAITRE

Laboratoire LSIS

Université du Sud Toulon-Var

Soutenance de thèse de doctorat, le 24 octobre 2006











# Plan

- Motivation : la recherche d'images
- Problématique :
  - Trouver des méthodes de fusion d'informations textuelles et visuelles efficaces et efficientes
- État de l'art
- Méthodologie et corpus
- Contributions
  - DIMATEX : un système rapide d'auto-annotation
  - Recherche des caractéristiques visuelles d'un mot
  - Sélection des dimensions sur une base d'images mal annotées
- Conclusion

# Systemes de recherche d'images sur le Web

[Google](#) [Web](#) [Images](#) [Groupes](#) [Annuaire](#) [Actualités](#)  
  [Images - Recherche avancée](#) [Préférences](#)

**Images** Résultats **1 - 20** sur un total d'environ **520** pour **house water man filetype:jpg -lukas**. (0,23 secondes)  
 Afficher: **Toutes les tailles** - [Grandes](#) - [Moyennes](#) - [Petites](#)

			
<p><b>Water Gap House.jpg</b> 795 x 535 pixels - 45 ko <a href="http://www.dutotmuseum.com/.../Water%20Gap%20House.jpg">www.dutotmuseum.com/.../Water%20Gap%20House.jpg</a></p>	<p><b>Clear WindowPlants HouseB...</b> 640 x 480 pixels - 69 ko <a href="http://www.briannelsonconsulting.com/images-clearwin...">www.briannelsonconsulting.com/images-clearwin...</a></p>	<p><b>indxrvsn.jpg</b> 448 x 260 pixels - 23 ko <a href="http://www.usbr.gov/mp/watershare/resources/catalog.cfm">www.usbr.gov/mp/watershare/resources/catalog.cfm</a></p>	<p><b>tsunami01.jpg</b> 742 x 557 pixels - 63 ko <a href="http://www.hanneketravels.net/alaska/tsunami01.jpg">www.hanneketravels.net/alaska/tsunami01.jpg</a></p>
			
<p><b>snow_l.jpg</b> 600 x 400 pixels - 49 ko <a href="http://www.software-e-commerce.com/.../don+t++its+over">www.software-e-commerce.com/.../don+t++its+over</a></p>	<p><b>Bevan5.jpg</b> 434 x 314 pixels - 19 ko <a href="http://www.lutsk.ukrpack.net/~scm/Pic/Bevan5.jpg">www.lutsk.ukrpack.net/~scm/Pic/Bevan5.jpg</a></p>	<p><b>LoyalWedellBulldozerTarps...</b> 640 x 480 pixels - 87 ko <a href="http://www.surpluscitysales.com/partytencitycustome...">www.surpluscitysales.com/partytencitycustome...</a></p>	<p><b>11683.jpg</b> 300 x 462 pixels - 49 ko <a href="http://www.kamat.com/.../architecture/common/11683.htm">www.kamat.com/.../architecture/common/11683.htm</a></p>

# Indexation d'images

## ■ Indexation textuelle

- Manuelle : coûteuse, subjective
- Automatique à partir du nom, de la légende ou du texte entourant l'image
- Ne décrit pas le contenu de l'image, beaucoup d'erreurs d'indexation, mais apporte des informations sémantiques

## ■ Indexation visuelle

- Couleurs, formes, textures
- Localisation, régions d'intérêt, segmentation
- Décrit le contenu visuel de l'image, mais extraction de la sémantique difficile !

## Indexation visuelle et fossé sémantique



(a)



(b)



(c)

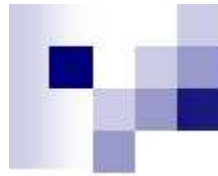
Les images (a) et (b) ont des descripteurs de couleurs similaires, mais un sens différent.

Les images (b) et (c) ont des descripteurs de couleurs différents, mais un sens similaire.

« The semantic gap is the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation » (Smeulders et al., 2000).



# Problématique



## Problématique générale de la thèse

- Trouver des méthodes efficaces de fusion des informations textuelles et visuelles
  - pour améliorer les systèmes de recherche d'images
  - à partir d'une base d'images généralistes annotées pour lesquelles les descripteurs visuels sont connus d'avance
- Difficulté à prendre en compte :
  - le passage à l'échelle
    - Les techniques classiques de recherche et d'apprentissage ne sont pas forcément efficaces et efficientes sur de grandes bases d'images



## Contributions

- Les contributions de ma thèse sont :
  - Un système d'auto-annotation rapide même sur un grand corpus d'images
  - Une méthode pour caractériser visuellement un mot-clé
  - Une méthode de sélection des dimensions sur des données mal-annotées



# État de l'art



## Systemes de recherche d'images

	Systeme visuel seul	Systeme visuel ou textuel, mais pas de fusion	Systeme visuel, textuel et fusion
1993		QBIC	
1995		Chabot	Piction
1996	VisualSeek	Photobook	FourEyes
1997	MARS		ImageRover
1999	NeTra, Blobword		
2000			PicHunter
2001	SIMPLIcity, IKONA		
2002	PicSOM		
2003			a-LIP

# L'auto-annotation d'images

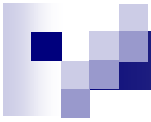
- Définition : l'annotation automatique (ou auto-annotation) consiste à associer un groupe de mots à une image uniquement à partir du contenu visuel de cette image au moyen d'un système informatique.
- Les systèmes d'auto-annotation d'images peuvent être vus comme des sous-modules d'un système de RI combinant texte et visuel.

# Auto-annotation d'images à partir du contenu visuel

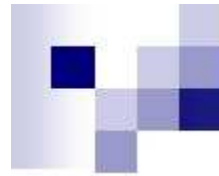
- Grand nombre de modèles :
  - Modèles basés sur l'analyse de la sémantique latente (LSA, PLSA...)
  - Modèles basés sur la distribution de Dirichlet (MoM-LDA, Corr-LDA...)
  - Modèles de traduction de langues...
- Différentes façons de combiner informations textuelles et visuelles pour l'auto-annotation
  - Fusion précoce des espaces textuel et visuel :
    - LSA, PLSA...
  - Combinaisons indépendantes des espaces:
    - MoM-LDA, GM-LDA, MoM-HAM I-2...
  - Combinaisons dépendantes des espaces :
    - Corr-LDA, MoM-HAM D-2...
  - Combinaisons à différents niveaux d'une hiérarchie :
    - MoM-HAM, MoM-LDA, Mix-Hier...

## Lacunes des systèmes d'auto-annotation d'images

- Les systèmes d'auto-annotation sont de plus en plus efficaces
- Mais peu de systèmes sont construits pour prendre en compte :
  - le temps de calcul
  - le nombre de paramètres
  - la complexité du modèle
- Or, pour permettre le passage à l'échelle, il est important que le système soit efficient



# Méthodologie et corpus



## Méthodologie générale

- Pour pouvoir mesurer la capacité de nos systèmes à fusionner informations textuelles et visuelles, nous proposons de les évaluer pour différentes tâches :
  - Pour la tâche d'auto-annotation d'images à partir du contenu visuel
  - Ou pour la tâche de classification d'images
- Nous utiliserons dans les deux cas le même ensemble d'images généralistes annotées : le corpus COREL utilisé par de nombreuses équipes de recherche.

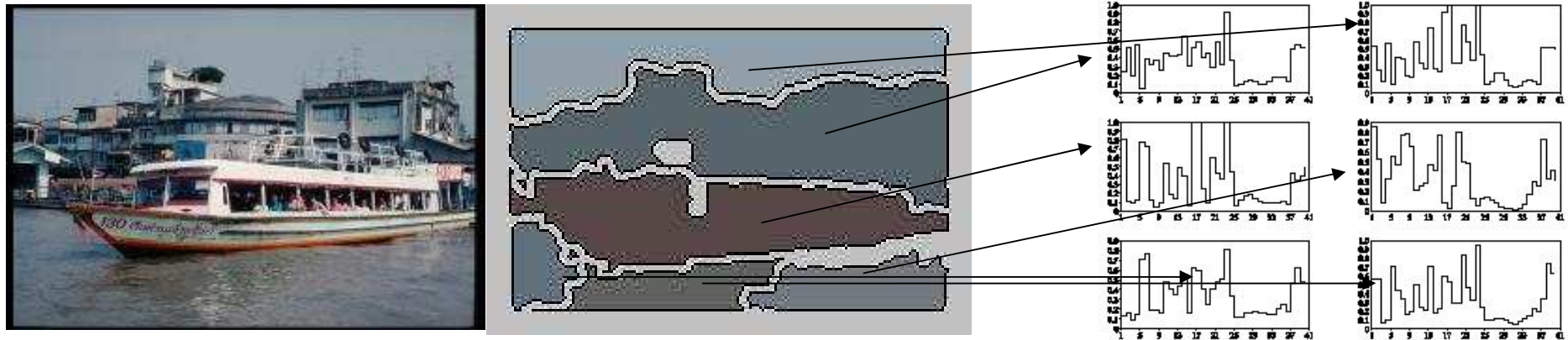
## Le corpus COREL (1/3)

- 10 000 images généralistes fournies par James Z. Wang  
<http://wang.ist.psu.edu>
- Chaque image est :
  - Annotée par de 1 à 5 mots-clés choisis manuellement dans un lexique de 250 mots-clés environ
  - Segmentée en 10 régions maximum
  - Chaque région (appelée « blob ») est décrite par un vecteur de 40 composantes visuelles
- Les annotations, les segmentations et les vecteurs visuels proviennent des données utilisées dans l'article :  
Kobus Barnard, P. Duygulu, N. de Freitas, D. Forsyth, D. Blei, and M. I. Jordan, « *Matching Words and Pictures* », Journal of Machine Learning Research, Vol 3, pp 1107-1135, 2003.

## Le corpus COREL (2/3)

- Algorithme de segmentation utilisée par K. Barnard et al. :

J. Shi, J. Malik, « Normalized Cuts and Image Segmentation »,  
 IEEE Patterns Analysis and Machine Intelligence, vol.22, n°8, 2000



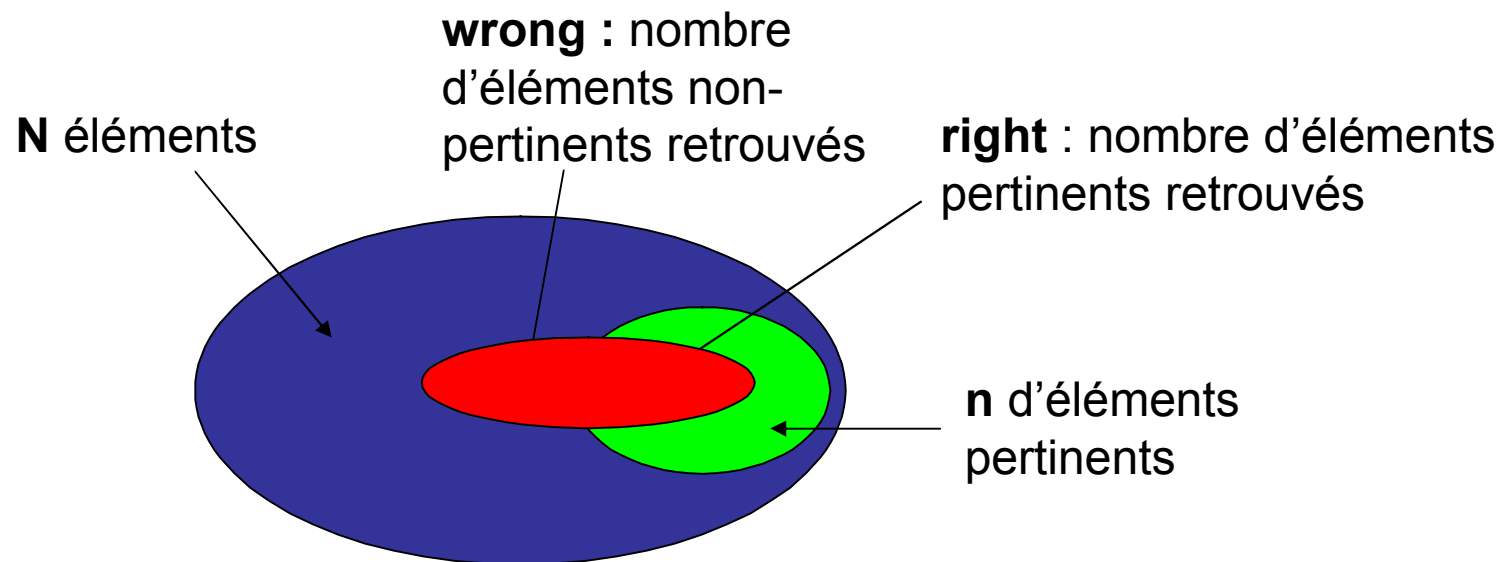
*water boat harbor building*

## Le corpus COREL (3/3)

- Chaque blob de l'image est décrit par un vecteur visuel de 40 composantes extraites par K. Barnard et al. :
  - 6 dimensions de formes (aire du blob...)
  - 18 dimensions de couleurs (RVB, rvS, Lab),
  - 16 dimensions de textures (filtres gaussiens...).
  
- Nous avons normalisé le corpus :
  - par estimation MLE de distributions Gamma des vecteurs visuels pour la génération de distributions de probabilités et supprimer les artefacts. Les valeurs sont comprises entre 0 et 1.
  
- **Nous appelons par la suite cet espace à 40 dimensions l'espace U**

## Le score normalisé (NS)

$$-1 \leq \text{NS} = \underbrace{\text{right}/n}_{\text{sensibilité}} - \underbrace{\text{wrong}/(N-n)}_{1\text{-spécificité}} \leq 1$$




Les éléments peuvent être :

- Les mots prédits pour chaque image dans le cas de l'auto-annotation
- Les images dans le cas de la classification




# Plan

- Motivation : la recherche d'images
- Problématique :
  - Trouver des méthodes de fusions d'informations textuelles et visuelles efficaces et efficientes
- État de l'art
- Méthodologie et corpus
- **Contributions**
  - **DIMATEX : un système rapide d'auto-annotation**
  - **Recherche des caractéristiques visuelles d'un mot**
  - **Sélection des dimensions sur une base d'images mal annotées**
- Conclusion



# DIMATEX : un système rapide d'auto-annotation d'images



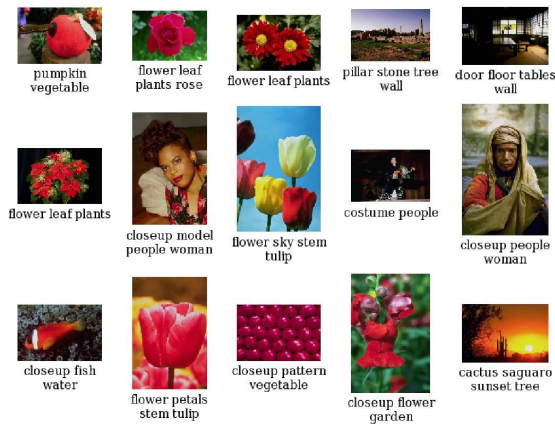
# DIMATEX : un système rapide d'auto-annotation d'images

## ■ Principe :

- Construction d'une table de distributions jointes entre informations textuelles et visuelles à partir des données d'apprentissage à l'aide d'une technique issue des bases de données (VA-Files)
- Ajout d'un modèle probabiliste simple afin de prédire une distribution de mots pour une nouvelle image

# Construction de la table de distributions jointes

- Principe des VA-Files (Weber et al., 1998):
  - Chaque dimension de l'espace visuel est séparée en deux segments
  - L'espace est partitionné en  $2^n$  clusters
  - Chaque vecteur visuel de l'ensemble d'apprentissage est codé en une séquence de bits de longueur  $n$



Cluster 0000110110001



Cluster 0111010110000



Cluster 1111111111110

## Construction de la table de distributions jointes

- La table de distributions jointes est estimée ainsi :

- pour tout mot  $w$  et pour tout cluster  $C_k$  :

$$P(w, C_k | \mathcal{A}) = \sum_{J \in \mathcal{A}} P(J | \mathcal{A}) \sum_{b \in J} P(w | C_k, b, J, \mathcal{A}) P(C_k | b, J, \mathcal{A}) P(b | J, \mathcal{A})$$

- où l'on peut supposer que :

- $P(J | \mathcal{A})$  suit une distribution uniforme
- $P(w | C_k, b, J, \mathcal{A}) = 1$  si  $w$  appartient aux mots annotant  $J$ , 0 sinon
- $P(b | J, \mathcal{A}) \cong P(b | J)$  et  $P(b | J) = \frac{\text{aire}(b)}{\sum_{b_p \in J} \text{aire}(b_p)}$
- $P(C_k | b, J, \mathcal{A}) = 1$  si  $C_k = C(b)$ , 0 sinon

## Associer des mots à une image

- Associer des mots à un blob :

$$\mathbf{P}(w|b, \mathcal{A}) = \mathbf{P}(w|C(b), \mathcal{A}) = \frac{\mathbf{P}(w, C(b)|\mathcal{A})}{\mathbf{P}(C(b)|\mathcal{A})}$$

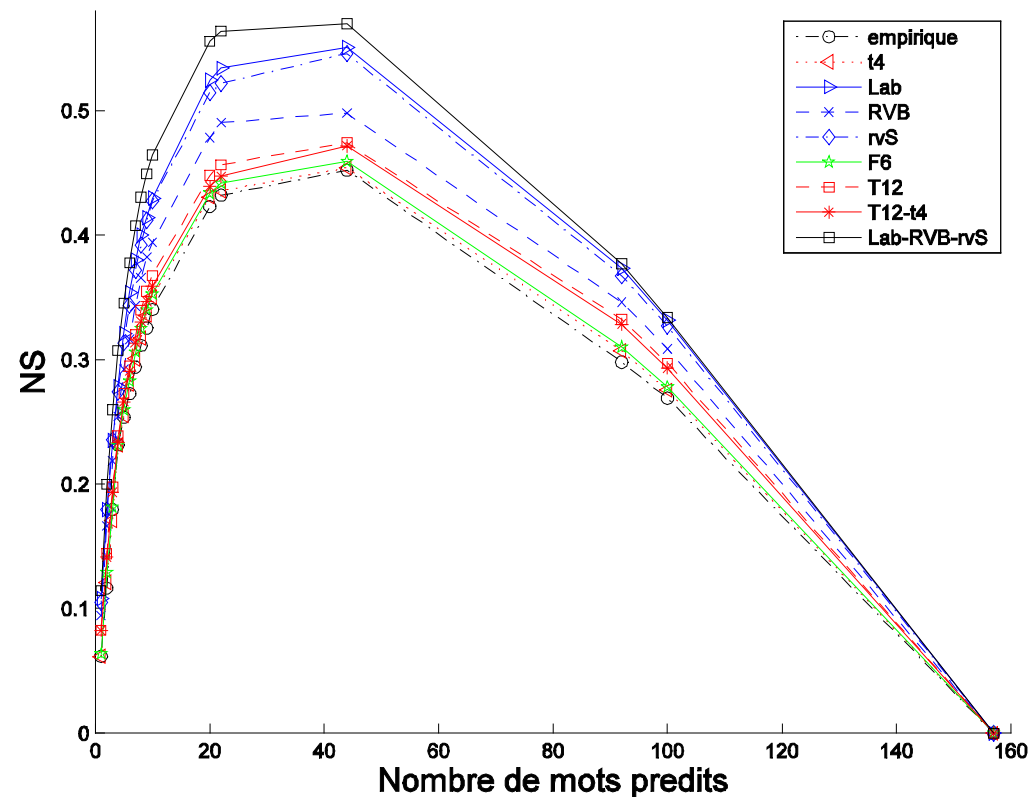
- Associer des mots à une image :

$$\mathbf{P}(w|I, \mathcal{A}) = \sum_{b \in I} \mathbf{P}(w|b, I, \mathcal{A}) \mathbf{P}(b|I, \mathcal{A})$$

où  $\mathbf{P}(b|I, \mathcal{A}) \simeq \mathbf{P}(b|I) = \frac{\text{aire}(b)}{\sum_{b_p \in I} \text{aire}(b_p)}$

# Expérimentations

- Corpus COREL : 7000 images d'apprentissage, 3000 images de test, lexique composé des 157 mots annotant au moins 20 images d'apprentissage





1: forest rabbit snow woodland  
 2: water sky tree snow  
 3: sky snow tree water  
 4: plane water jet snow  
 5: birds water fox rock  
 6: birds water fox rock



1: cat grass lion mane  
 2: water sky tree people  
 3: cat grass lion tree  
 4: tree rock lizard walls  
 5: closeup insect tree grass  
 6: grass ground shadows closeup



1: flower leaf plants  
 2: tree water flower  
 3: buildings people water  
 4: tree water street  
 5: birds rock nest  
 6: face ruins sculpture



1: fish reef water  
 2: water tree fish  
 3: people tree water  
 4: tree water buildings  
 5: flowers gardens shop  
 6: birds ground owl



1: sky sunset town  
 2: tree sky water  
 3: sky tree water  
 4: pattern people clouds  
 5: building church hills  
 6: building church hills



1: people tables tree  
 2: people flower tree  
 3: grass lion people tree  
 4: tree buildings lizard temple  
 5: birds grass nest rock  
 6: birds grass nest ground



1: boat horizon water  
 2: sky water people  
 3: mountains sky water  
 4: mountains water sky  
 5: dock sailboat building  
 6: dock sailboat closeup



1: dune grass sand valley  
 2: water sky people snow  
 3: people sky tree  
 4: people snow water  
 5: animal rocks people  
 6: building fence snow

1. Annotation manuelle

2. DIMATEX

3. PLSA-WORDS

4. PLSA-WORDSFEATURES

5. DIRECT

6. LSA (Monay & Gatica-Perez, 2004)

## Comparaison avec les modèles de l'état de l'art

Références	Modèles	NS	$\Delta$ NS	Gain NS
(Barnard et al., 2003)	empirique	0.425	-	-
	binary-D-2-region-cluster	0.604	0.179	+42%
	MoM-LDA	0.536	0.107	+25%
(Monay & Gatica-Perez, 2004)	empirique	0,427	-	-
	LSA	0.540	0.113	+26%
	PLSA-WORDS	0.571	0.144	+34%
DIMATEX (2005)	empirique	0.453	-	-
	Lab-RVB-rvS	0.583	0.132	+29%

## Complexité

- Le modèle *binary-D-2-region-cluster* (Barnard et al., 2003) nécessite pour être optimal en moyenne 10 itérations de l'algorithme EM pour un total de 511 nœuds. Chaque nœud nécessite l'apprentissage de plusieurs paramètres.
- Le modèle *PLSA-WORDS* (Monay & Gatica-Perez, 2004) nécessite l'apprentissage de plusieurs distributions de probabilités pour chacune des modalités, chacune nécessite plusieurs itérations de l'algorithme EM.

# Complexité de DIMATEX

- Le modèle DIMATEX
  - ne nécessite aucun apprentissage
  - ne possède aucun paramètre à optimiser
- Sa complexité moyenne est celle des VA-Files. C'est-à-dire :
  - $O(1)$  pour insérer un vecteur visuel dans la table
  - $O(1)$  pour annoter une image
- Une seule difficulté : la taille de la table de distributions jointes croît de manière exponentielle avec le nombre de dimensions visuelles.
  - De manière expérimentale, nous montrons que pour un nombre de dimensions supérieur à 15, les performances du système diminuent.

## Conclusion sur DIMATEX

- Le système DIMATEX obtient des scores similaires aux modèles de l'état de l'art.
- C'est un système d'annotation rapide à condition que le nombre de dimensions de l'espace visuel reste raisonnable
- Il a l'avantage d'être dynamique dans sa phase d'entraînement
  
- Le système DIMATEX :
  - ne permet pas de savoir quelles sont les caractéristiques visuelles d'un mot
  - n'utilise pas de critères pour sélectionner les dimensions visuelles les plus pertinentes



# Une méthode de recherche des caractéristiques visuelles d'un mot



Qu'est ce qui caractérise le mot « tiger » ?



tiger stone water

tiger water ground



tiger bengal grass

tiger forest

# Comment trouver les caractéristiques visuelles d'un mot ?

## ■ Principe :

- pour trouver les régions d'images similaires qui caractérisent un mot, utiliser une méthode classique d'apprentissage

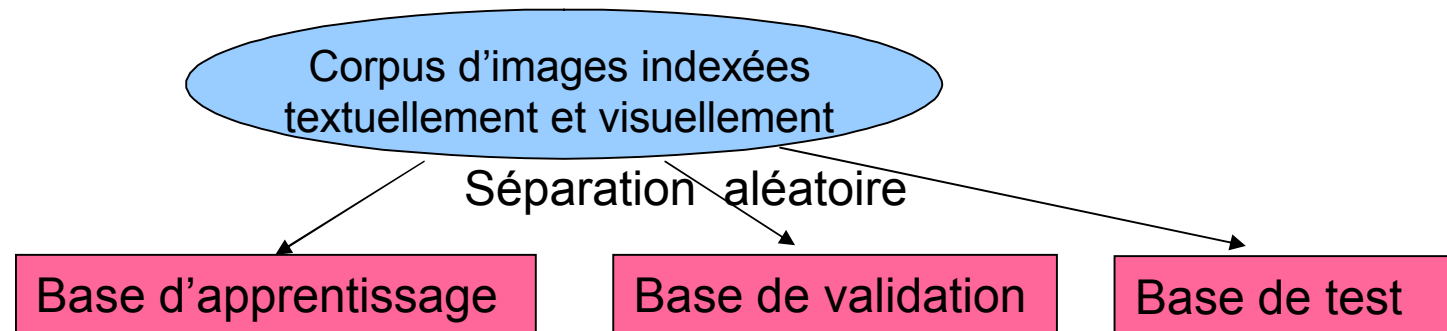
## ■ Proposition :

- la classification ascendante hiérarchique (CAH)
  - Avantages :
    - Maîtrise des paramètres
    - Clusters visuels interprétables

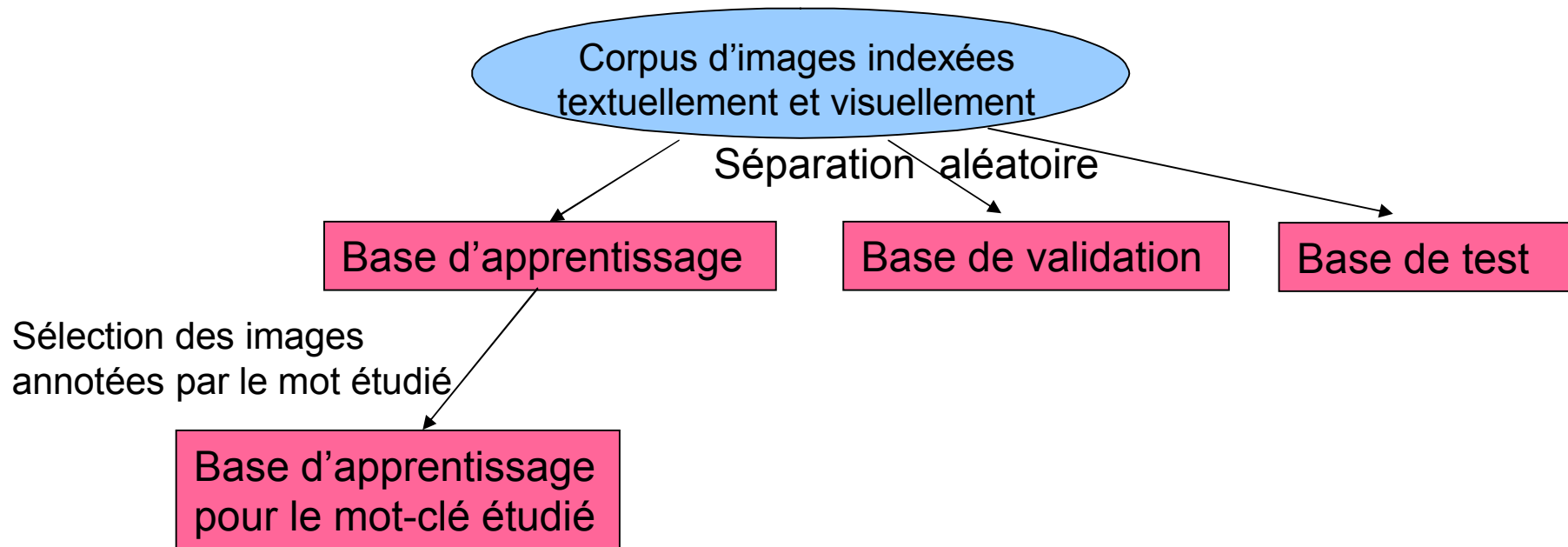
# Construction de clusters visuels par CAH

Corpus d'images indexées  
textuellement et visuellement

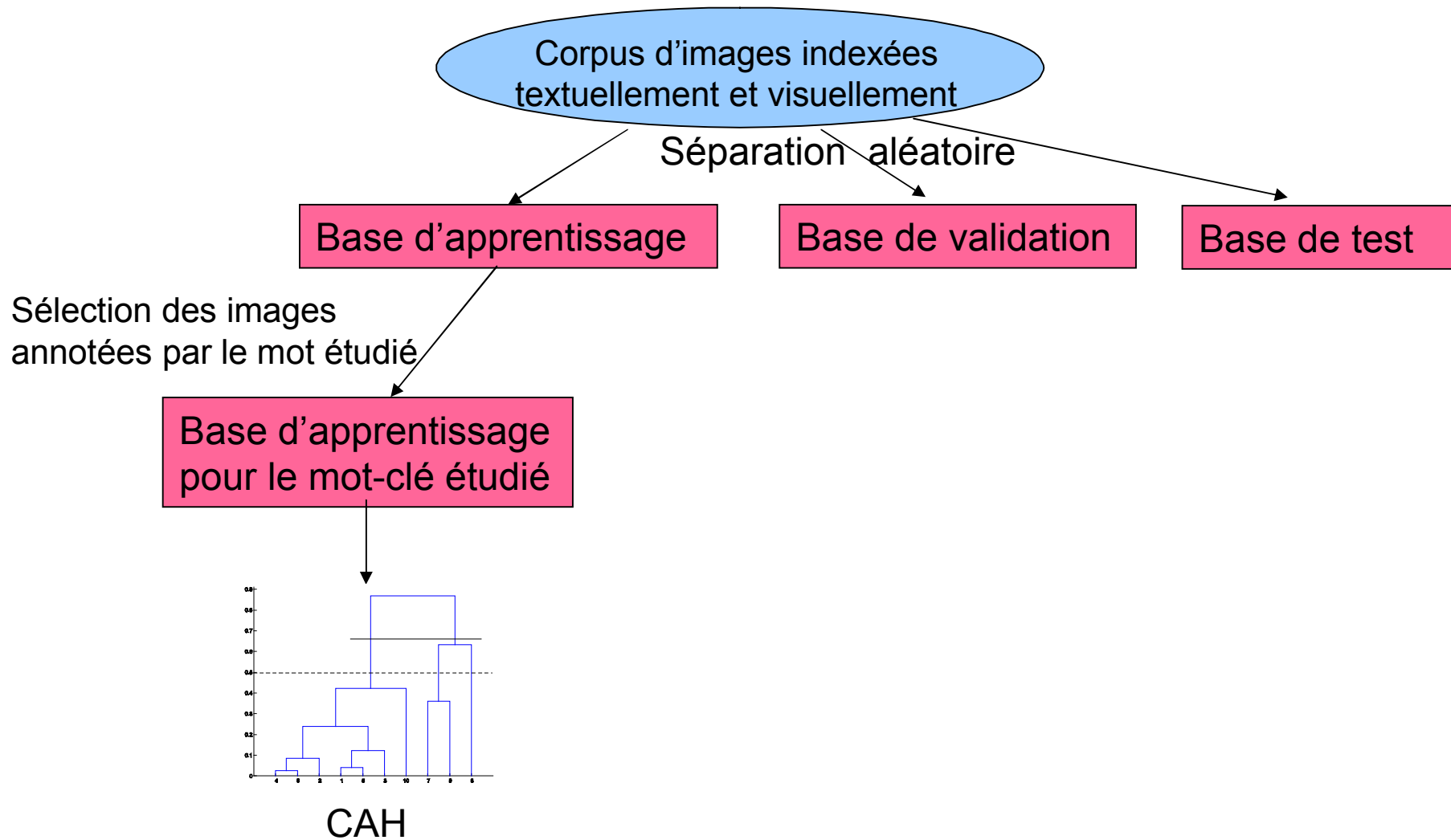
## Construction de clusters visuels par CAH



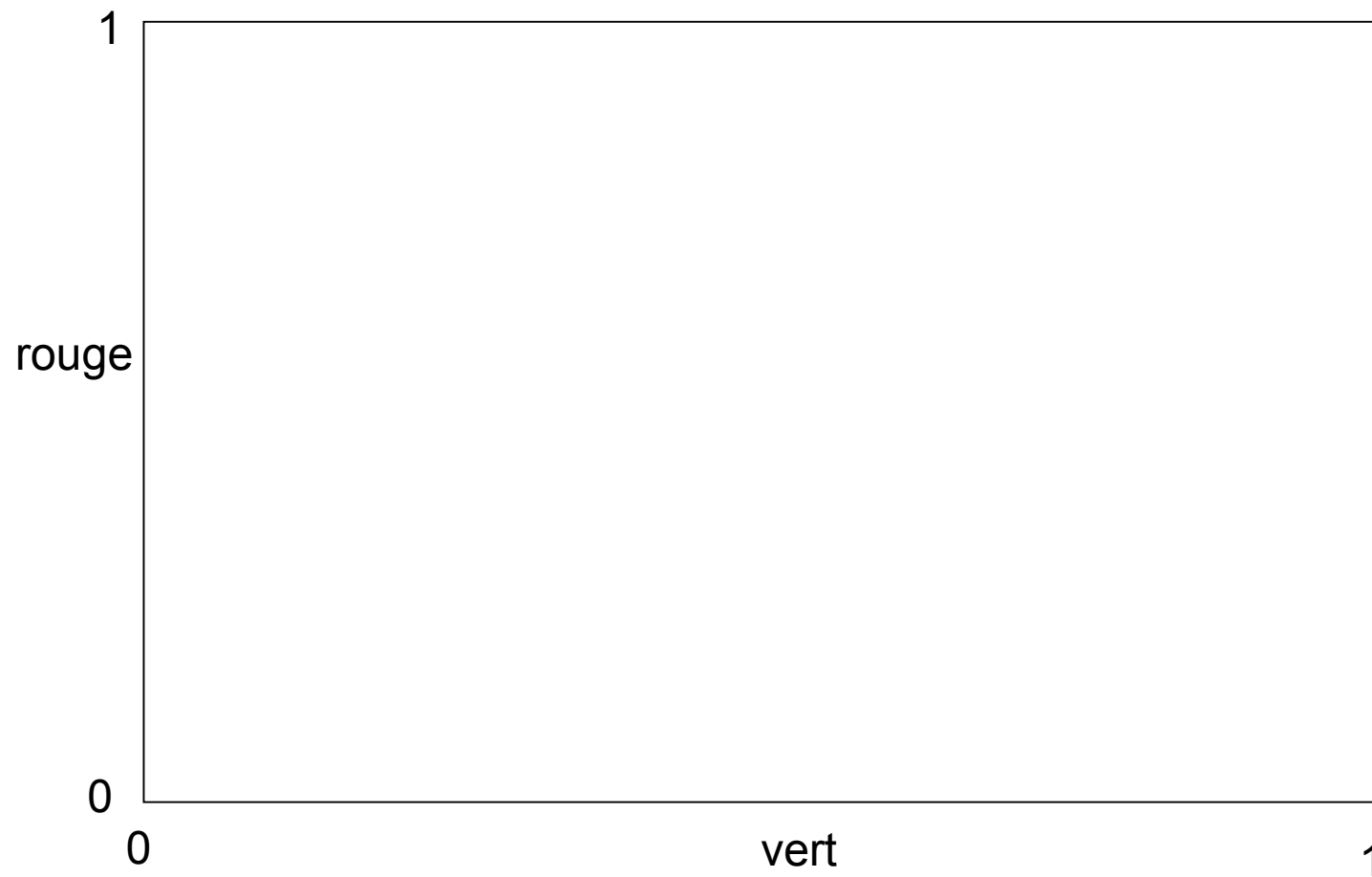
## Construction de clusters visuels par CAH



# Construction de clusters visuels par CAH

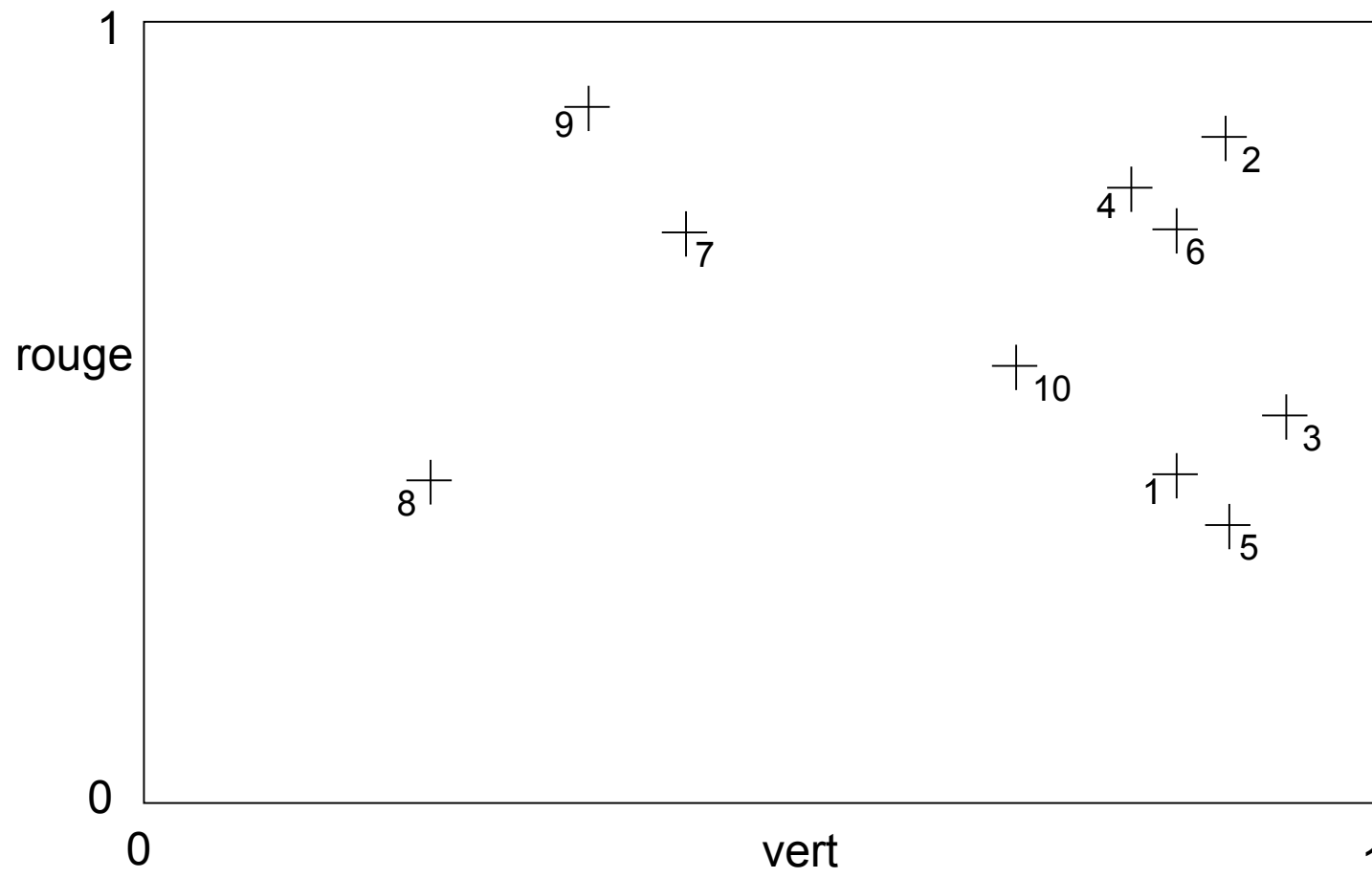


## Algorithme de la CAH : exemple

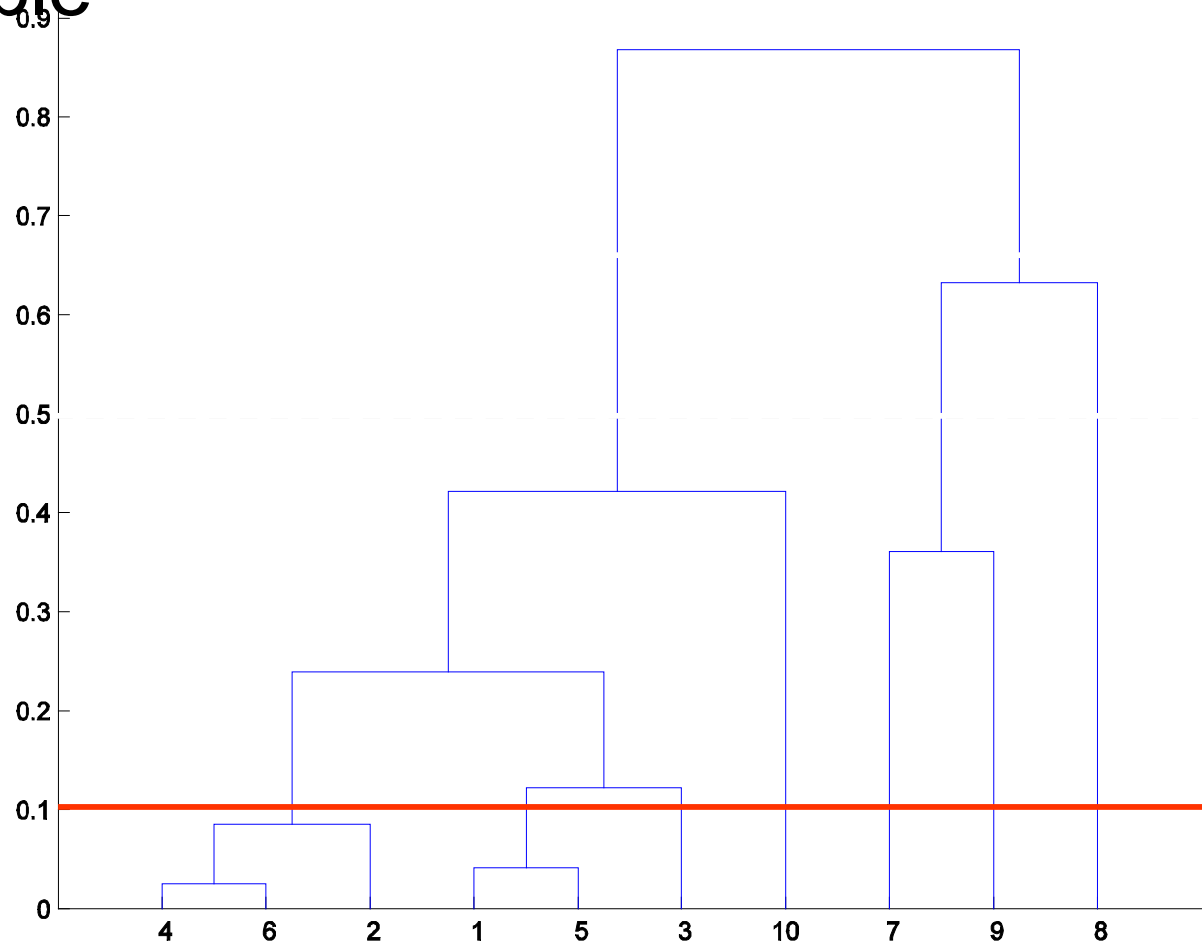


## Algorithme de la CAH : exemple

+ Vecteurs visuels des images d'apprentissage annotées par le mot « soleil »

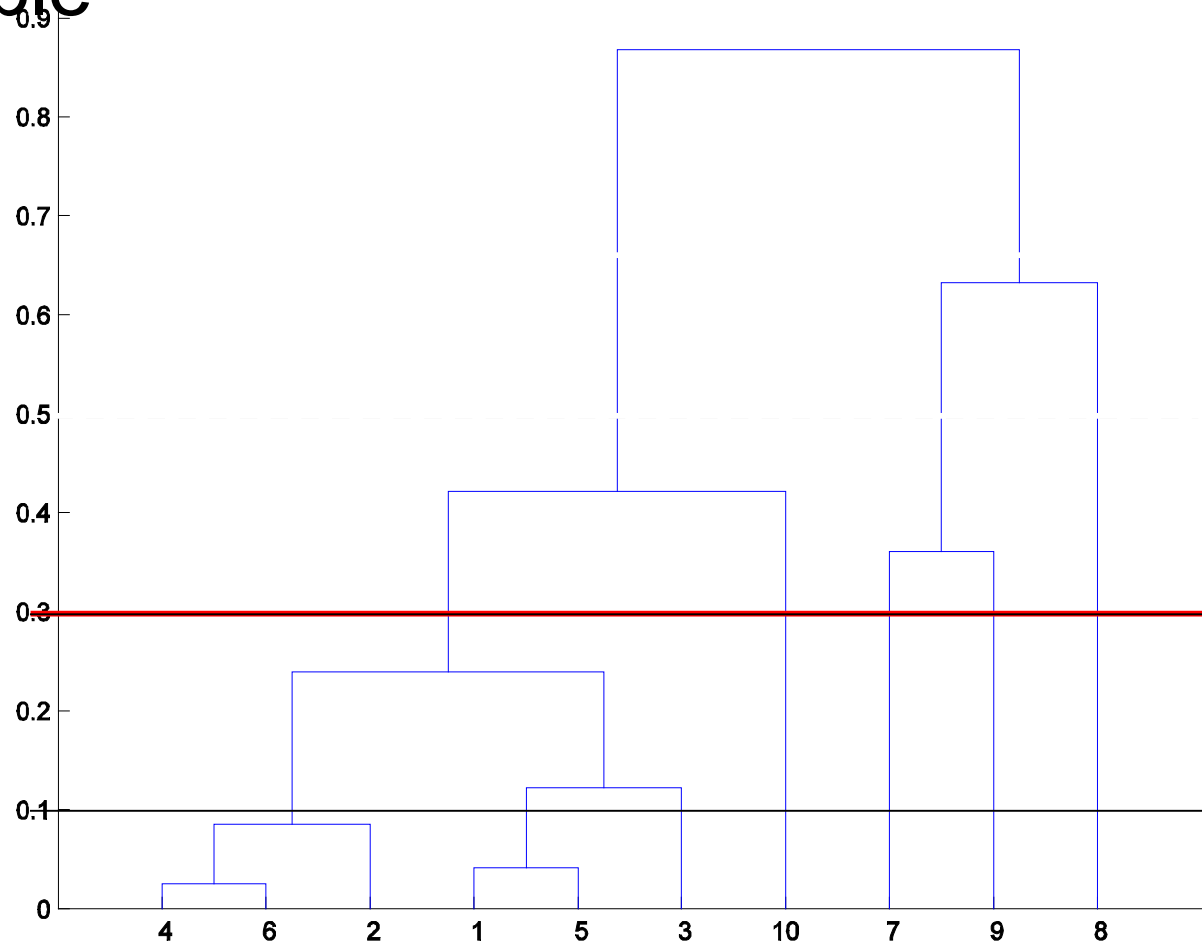


# Construction de clusters visuels par CAH : exemple



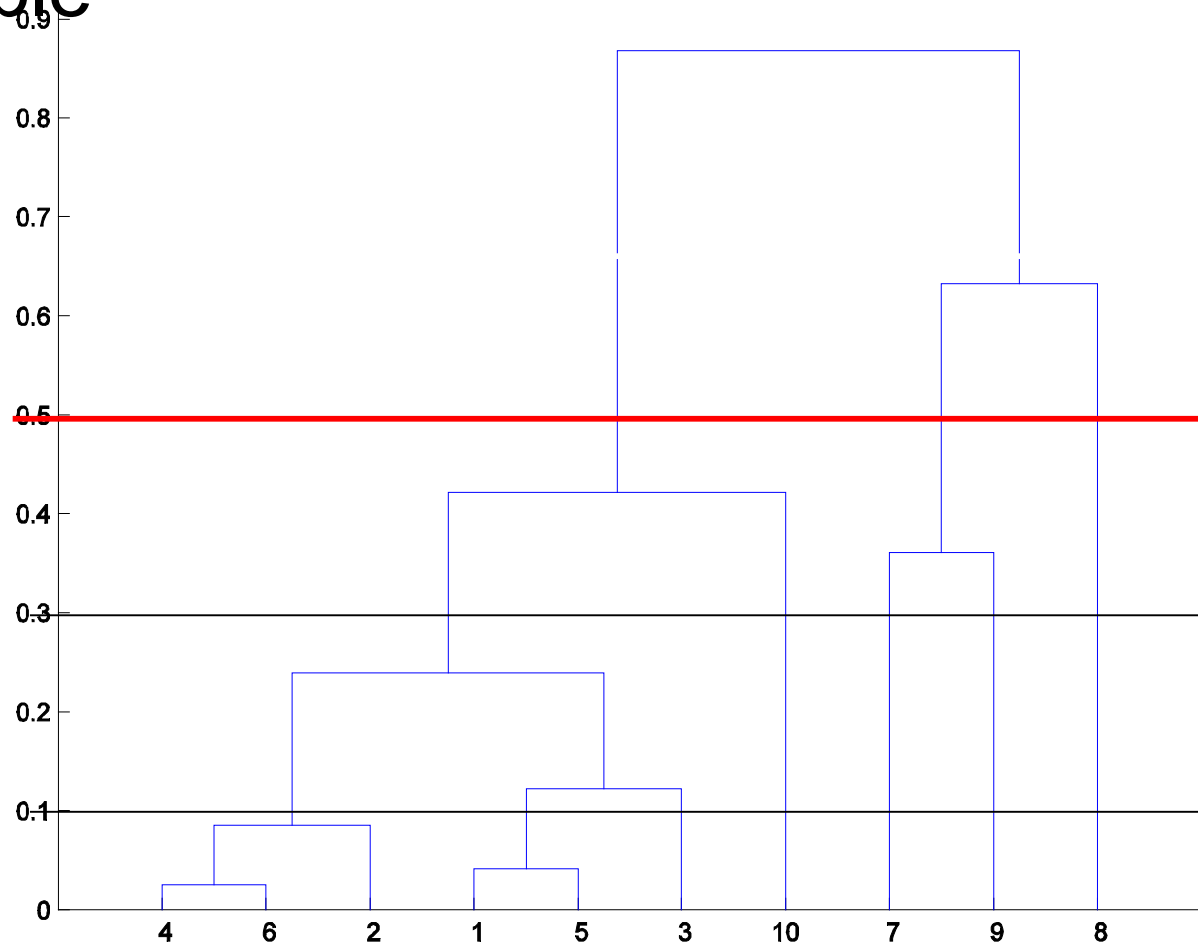
Vecteurs des images d'apprentissage du mot « Soleil »

# Construction de clusters visuels par CAH : exemple



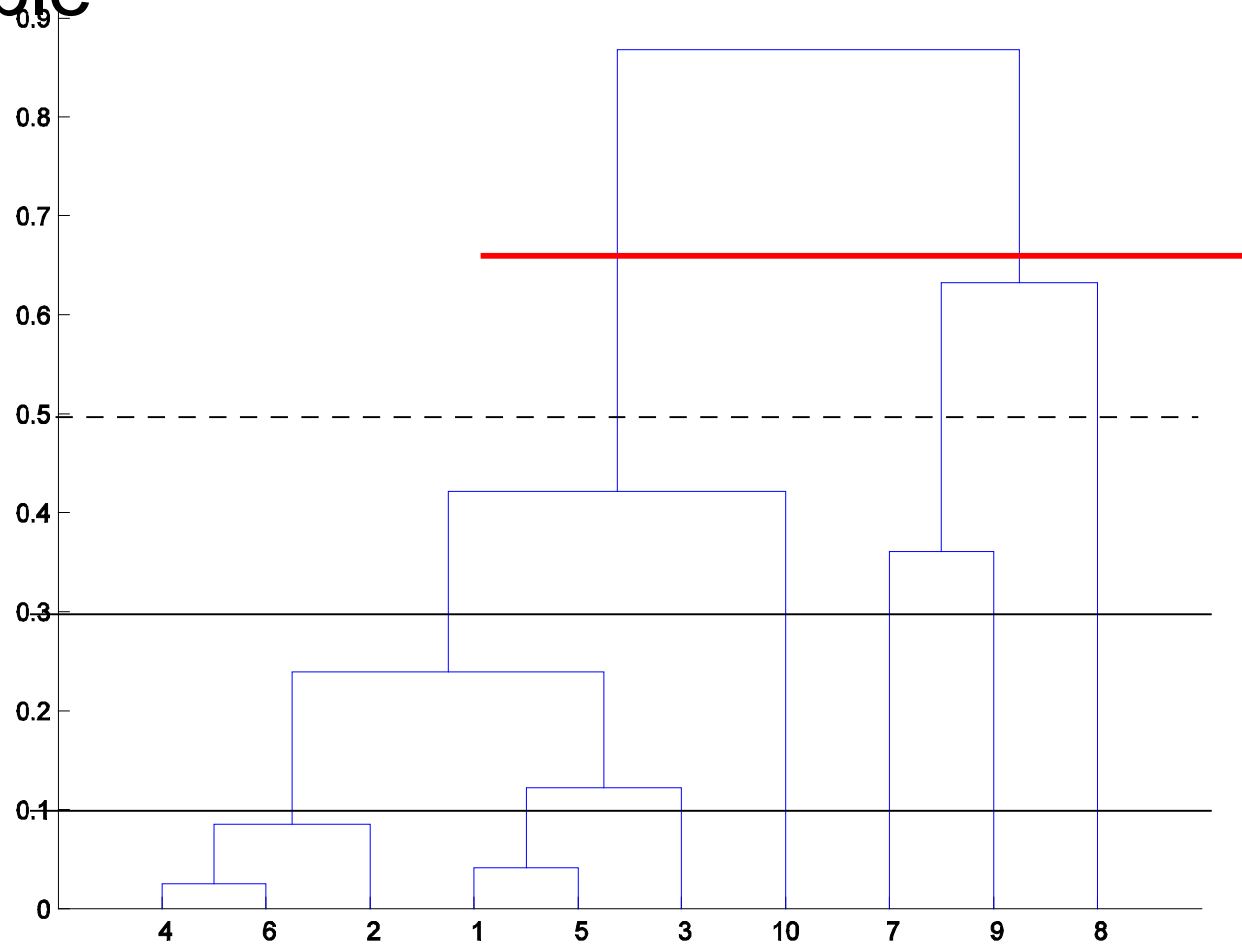
Vecteurs des images d'apprentissage du mot « soleil »

# Construction de clusters visuels par CAH : exemple



Vecteurs des images d'apprentissage du mot « soleil »

# Construction de clusters visuels par CAH : exemple



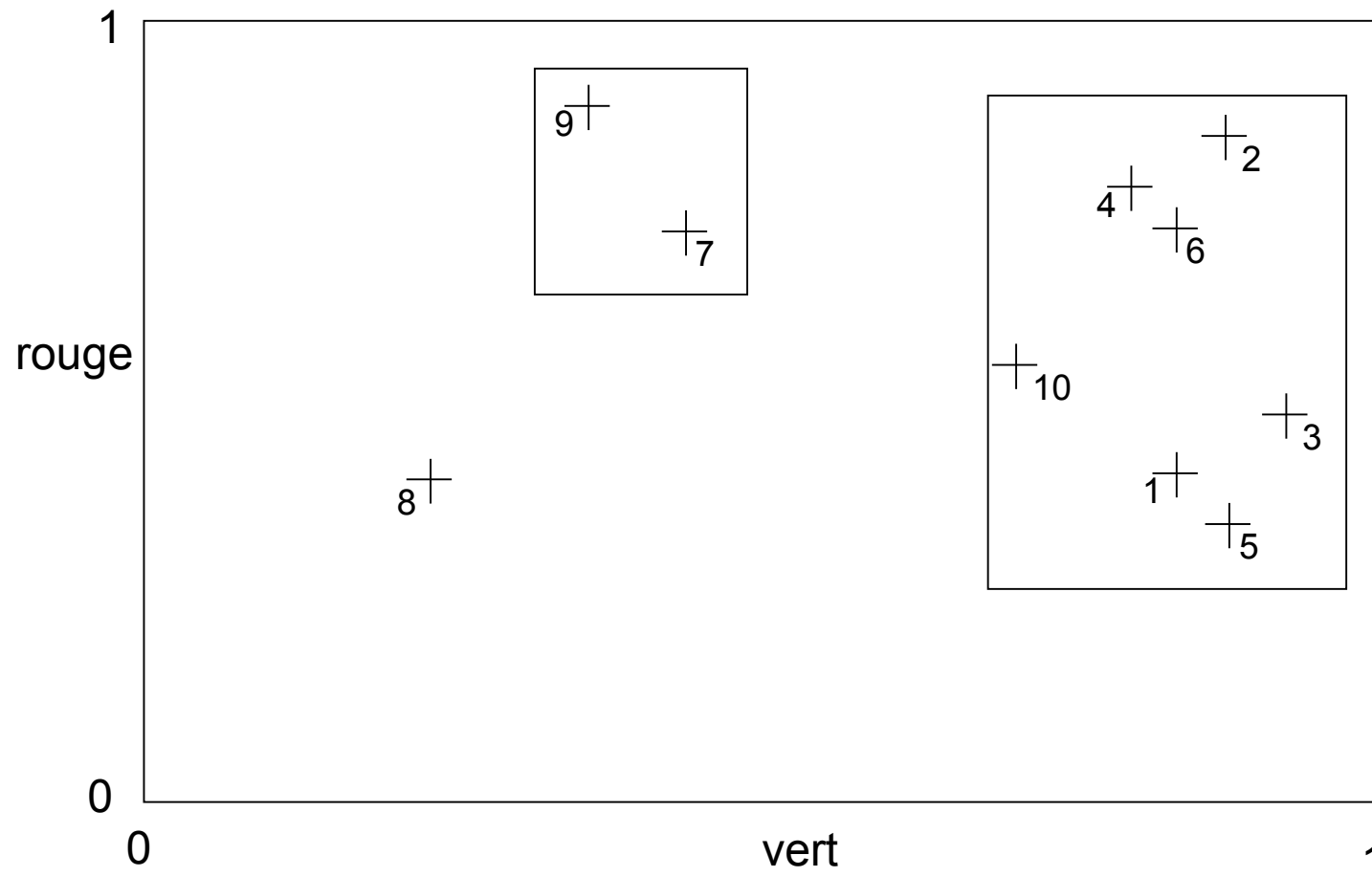
Vecteurs des images d'apprentissage du mot « soleil »

# Construction de clusters visuels par CAH : exemple



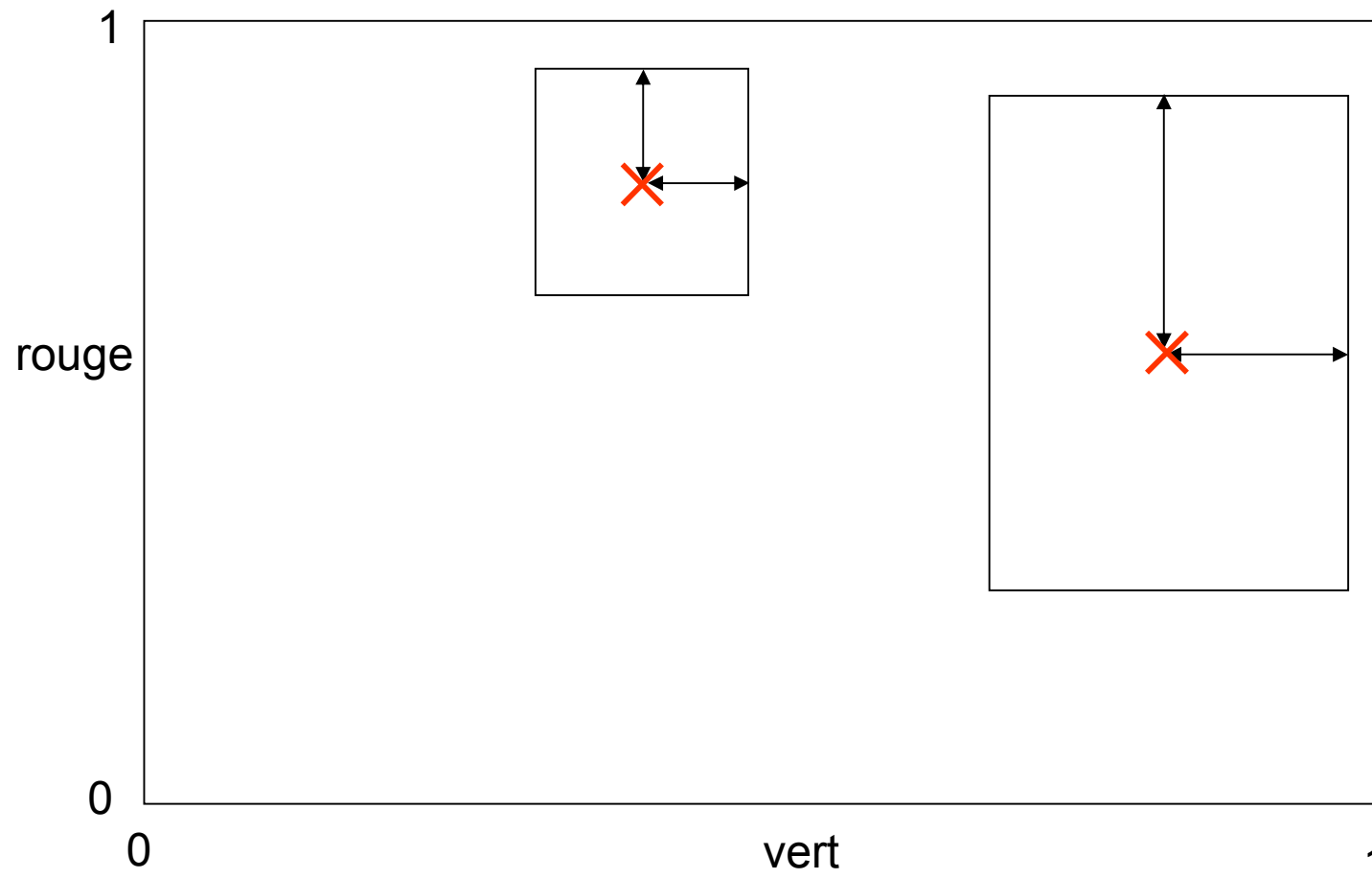
# Algorithme de la CAH : exemple

+ Vecteurs visuels des images d'apprentissage annotées par le mot « soleil »

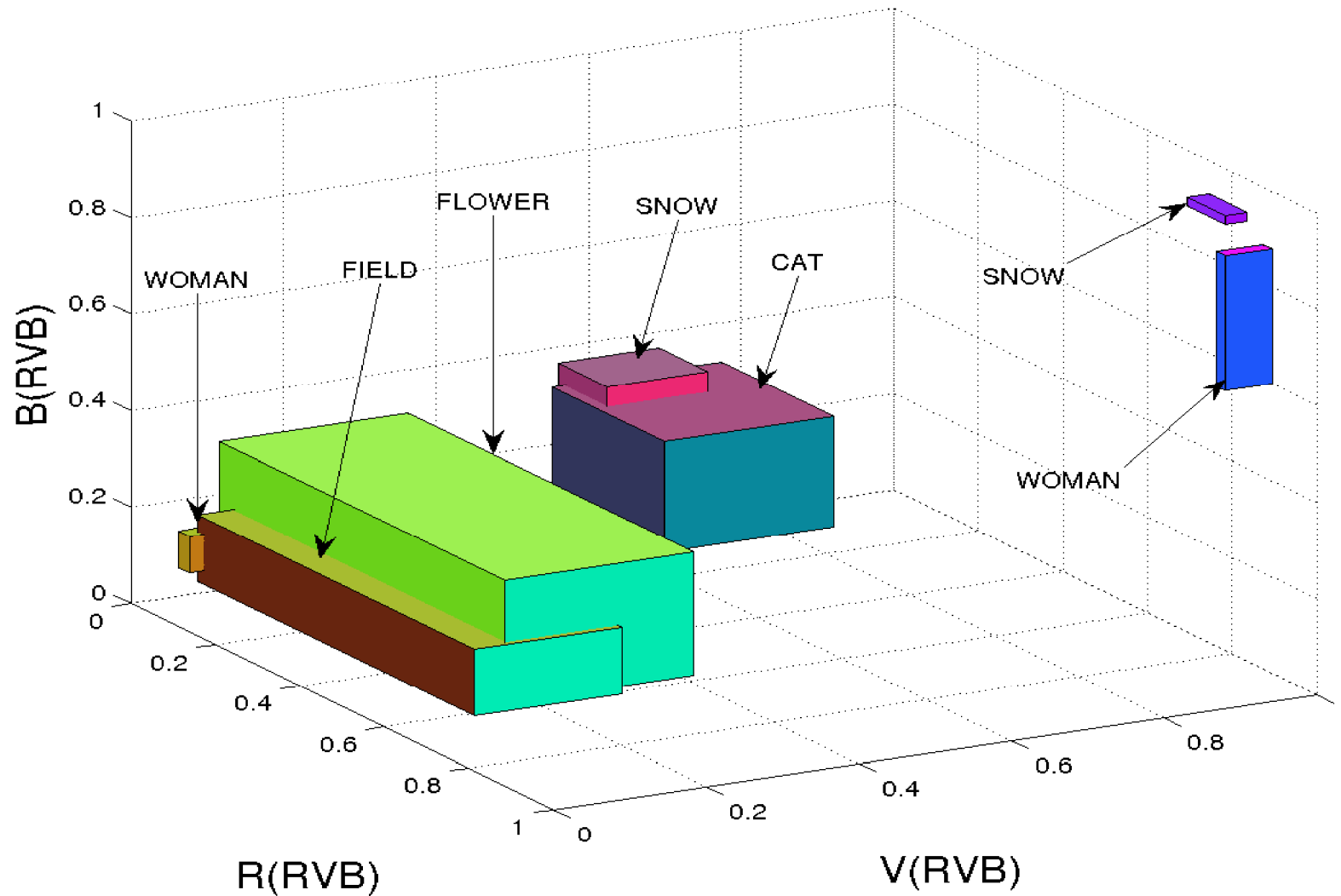


# Algorithme de la CAH : exemple

Clusters du mot « soleil » stockés en gardant uniquement la moyenne et l'écart-type pour chaque dimension



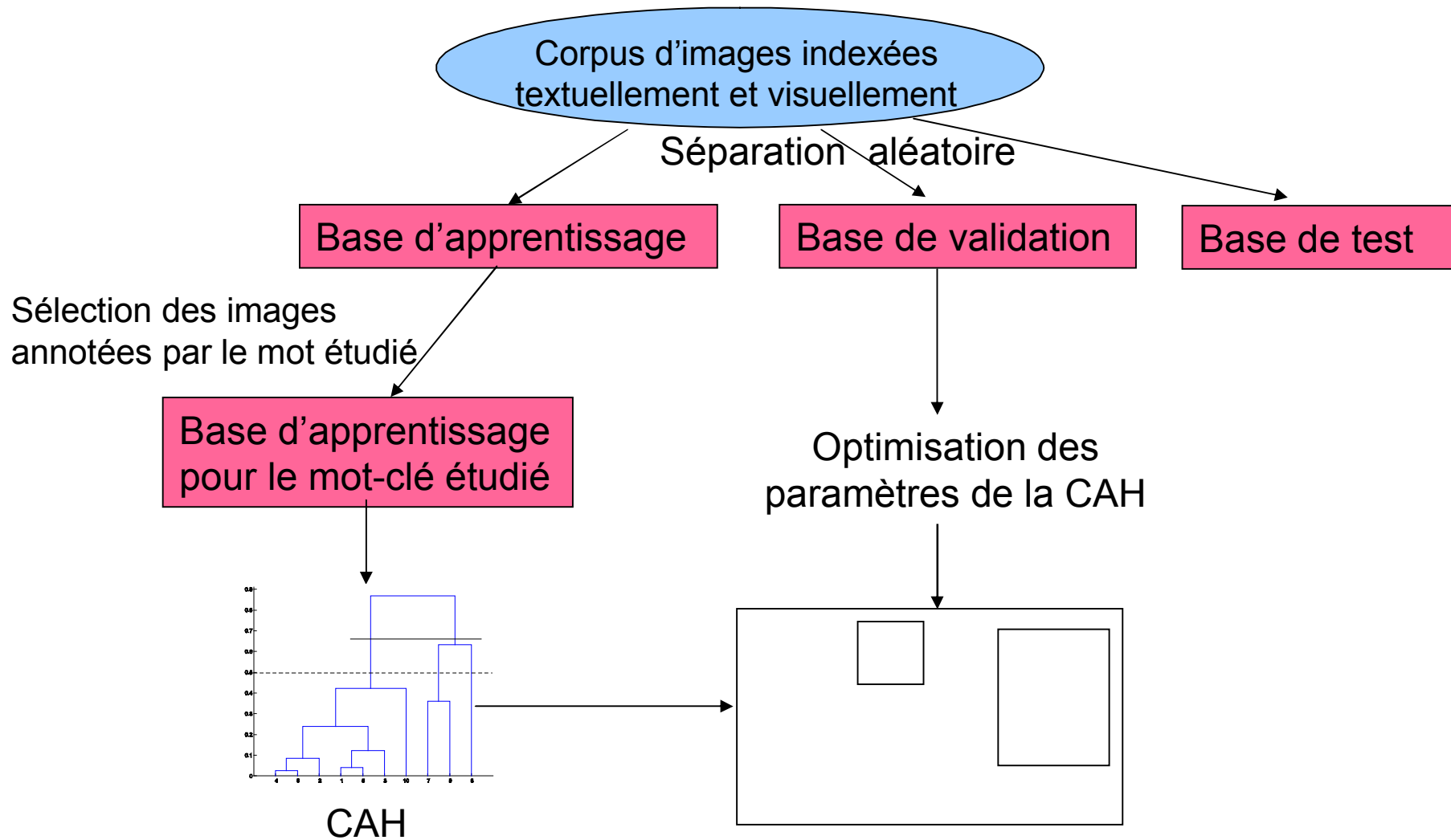
# Exemple de clusters visuels de plusieurs mots dans l'espace visuel



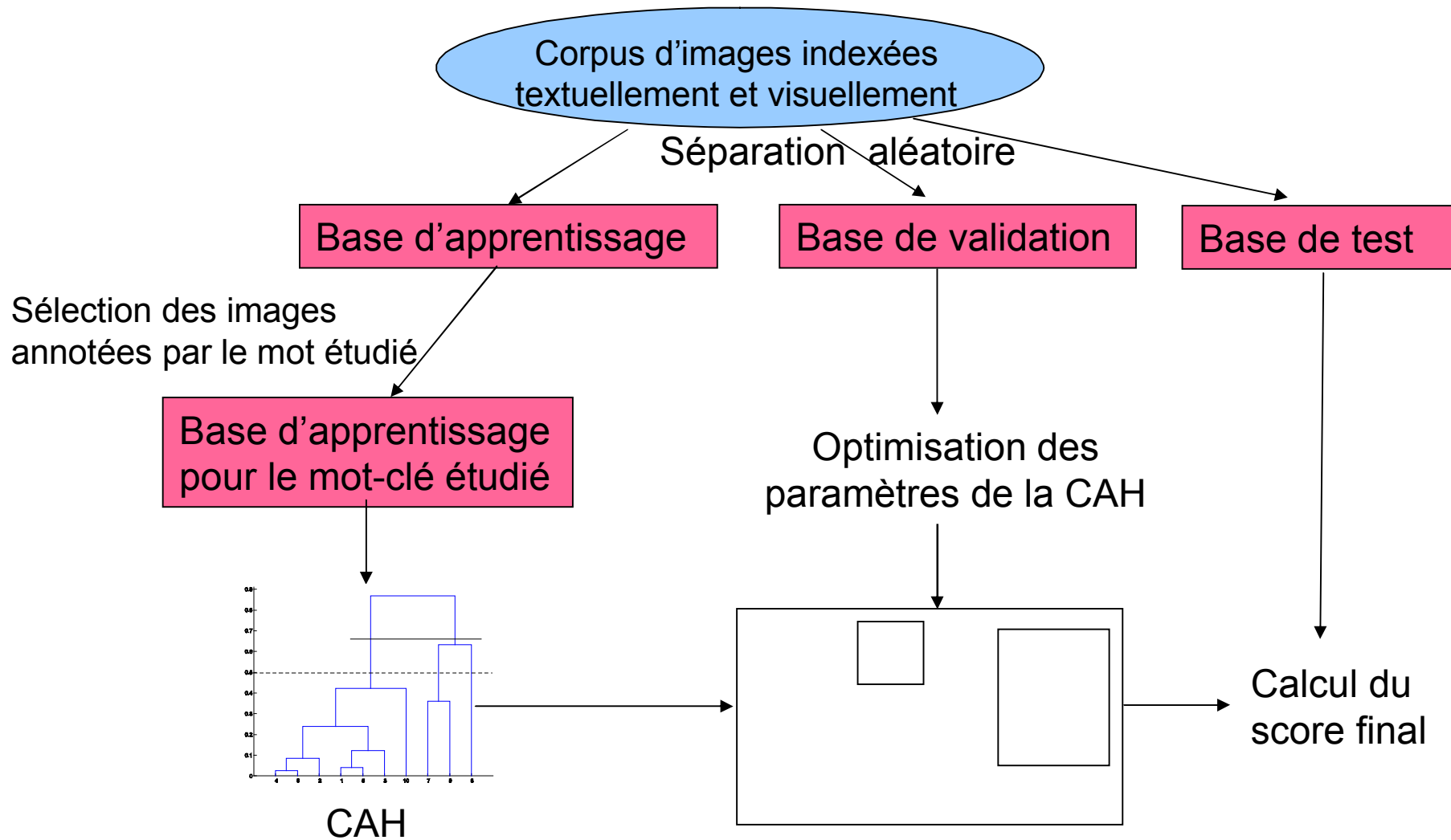
## Évaluation de la qualité des clusters obtenus

- Un blob est annoté par un mot s'il appartient à l'un des clusters de ce mot
- Une image est annotée par un mot si au moins  $B$  blobs de cette image sont annotés par ce mot

# Construction de clusters visuels par CAH



# Construction de clusters visuels par CAH

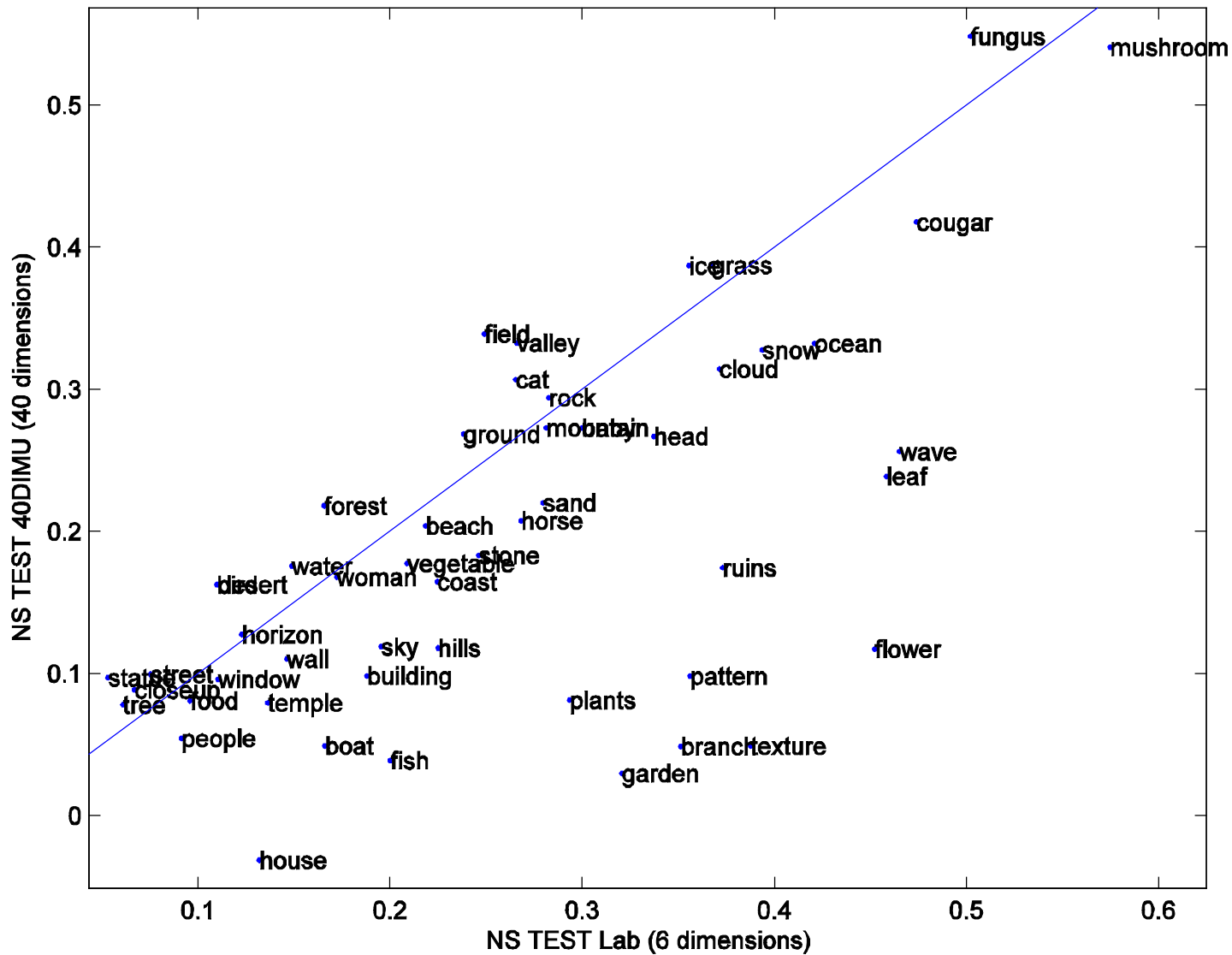


# Expérimentations

- Corpus COREL
  - 5000 images d'apprentissage
  - 2500 images de validation
  - 2500 images de test
- Espaces visuels :
  - Lab : 6 dimensions visuelles de couleurs
  - 40DIMU : 40 dimensions visuelles de l'espace U

	40DIMU	Lab
Nombre de dimensions	40	<b>6</b>
NS moyen sur 52 mots (validation)	0.236	<b>0.290</b>
NS moyen sur 52 mots (test)	0.192	<b>0.248</b>

# Lab versus 40DIMU



## Application : Filtrage de l'indexation textuelle d'images par le contenu visuel

- On suppose que tous les mots du lexique représentent le texte associé à l'image (ici 52 mots supposés extraits d'une page web)
- On filtre les mots avec les clusters visuels des mots obtenus par CAH
- On calcule le score NS à partir des mots associés initialement à l'image



Image 172052 (10 blobs)

### **Légende ( 3 sur 3 )**

water(OK) mountain(OK)  
coast(OK)

sensi=1.00 specif=0.65

preci=0.15 NS=0.65


### **20 mots associés par le système**

desert(7) water(6) sky(6) wave(6) hills(6)  
closeup(6) mountain(6) coast(6) tree(6)  
beach(6) boat(5) branch(5) temple(5) fish(4)  
sand(4) forest(4) cloud(4) people(4)  
horizon(3) valley(3)


### **32 mots non associés**

snow(2) statue(2) vegetable(1) rock(1) bird(1)  
wall(1) flower(1) head(1) building(1)  
window(1) woman(1) street(1) plants(1)  
field(1) cat cougar food fungus garden grass  
ground horse house ice leaf mushroom ocean  
pattern ruins stone texture

**Total: 52 mots**

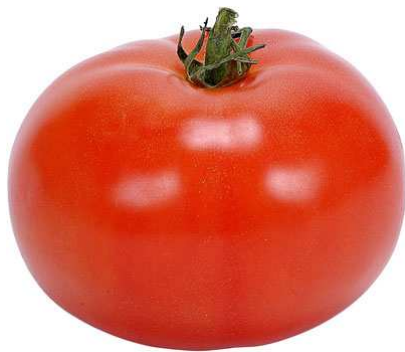


# Sélection des traits visuels en fonction du concept recherché à partir d'images mal annotées



## Problématique

- Quels traits visuels doit-on utiliser pour bien discriminer un certain concept (mot) ?
- Etant donnés :
  - le problème de la malédiction de la dimension
  - et l'existence de traits visuels non-informatifs



La couleur rouge est discriminante pour une tomate mûre, mais pas pour une fleur.

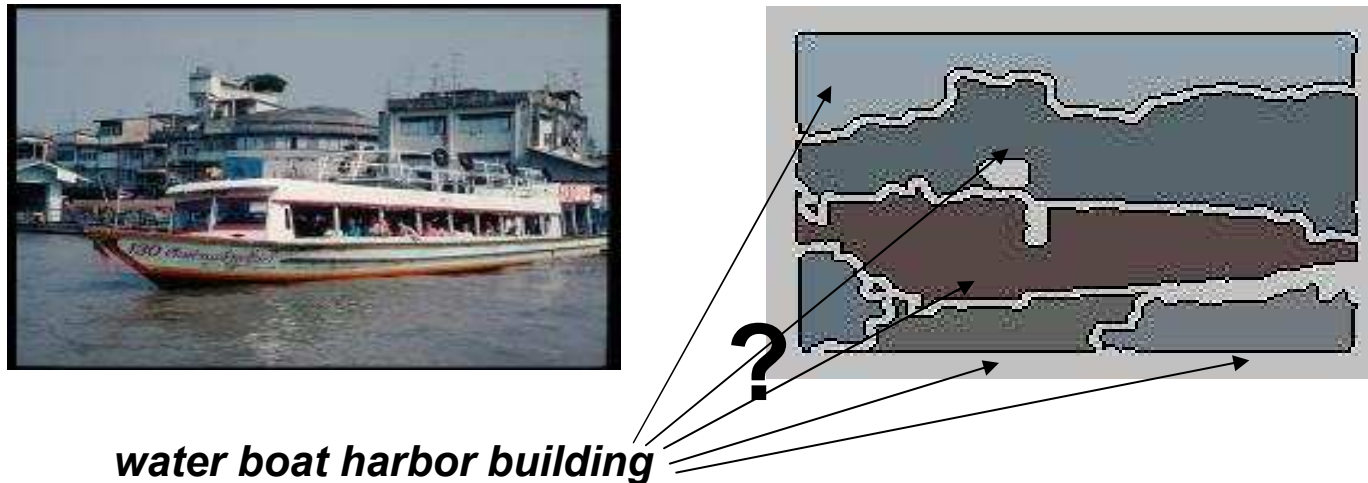


## La malédiction de la dimension

- Les espaces de grande dimension possèdent des propriétés particulières qui font que les intuitions géométriques peuvent se révéler fausses.
- Plus la dimension de l'espace est grande (Beyer et al., 1999) (Weber et al., 1998) :
  - Plus les vecteurs sont équidistants entre eux
  - Plus le partitionnement de l'espace devient exponentiel
  - Plus l'espace est creux
  - Plus la taille des bases de données augmente de façon exponentielle...
- L'apprentissage et la recherche par similarité sont donc plus difficiles sur des espaces de grande dimension

## Proposition

- Utiliser la méthode classique de l'Analyse Linéaire Discriminante (LDA) ?
  - Difficulté : Les bases d'images réelles (ex : Web, COREL) ne sont pas étiquetées par régions d'image

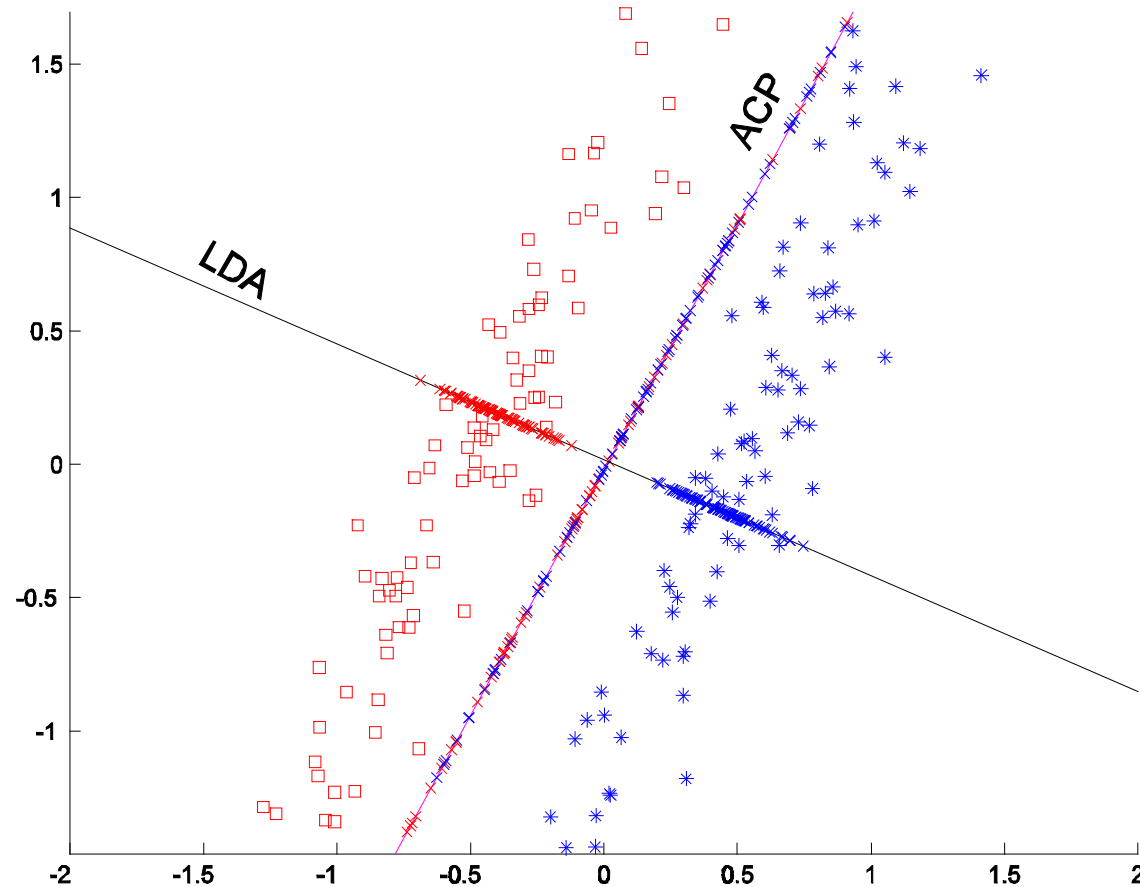


Par contre, les régions autour du « building » seront très changeantes...

- **Proposition : Approximer la LDA (ALDA)**

## Rappel : LDA versus ACP

- L'ACP recherche l'axe qui représente le mieux les données
- La LDA recherche l'axe qui sépare le mieux les classes



## Approximation des variances inter-/intra- classes pour estimer le pouvoir discriminant (LDA)

- Pour déterminer les traits visuels les plus discriminants en fonction du mot  $w_i$  considéré à partir de la base d'apprentissage
  - Séparation des images (et pas des régions) en 2 classes :
    - Celles qui sont annotées par ce mot
    - Celles qui ne sont pas annotées par ce mot
  - Calcul des variances interclasse (B) et intraclasse (W)
  - Pour chaque trait visuel  $v_j$  et pour chaque mot  $w_i$ , on calcule le pouvoir discriminant F:

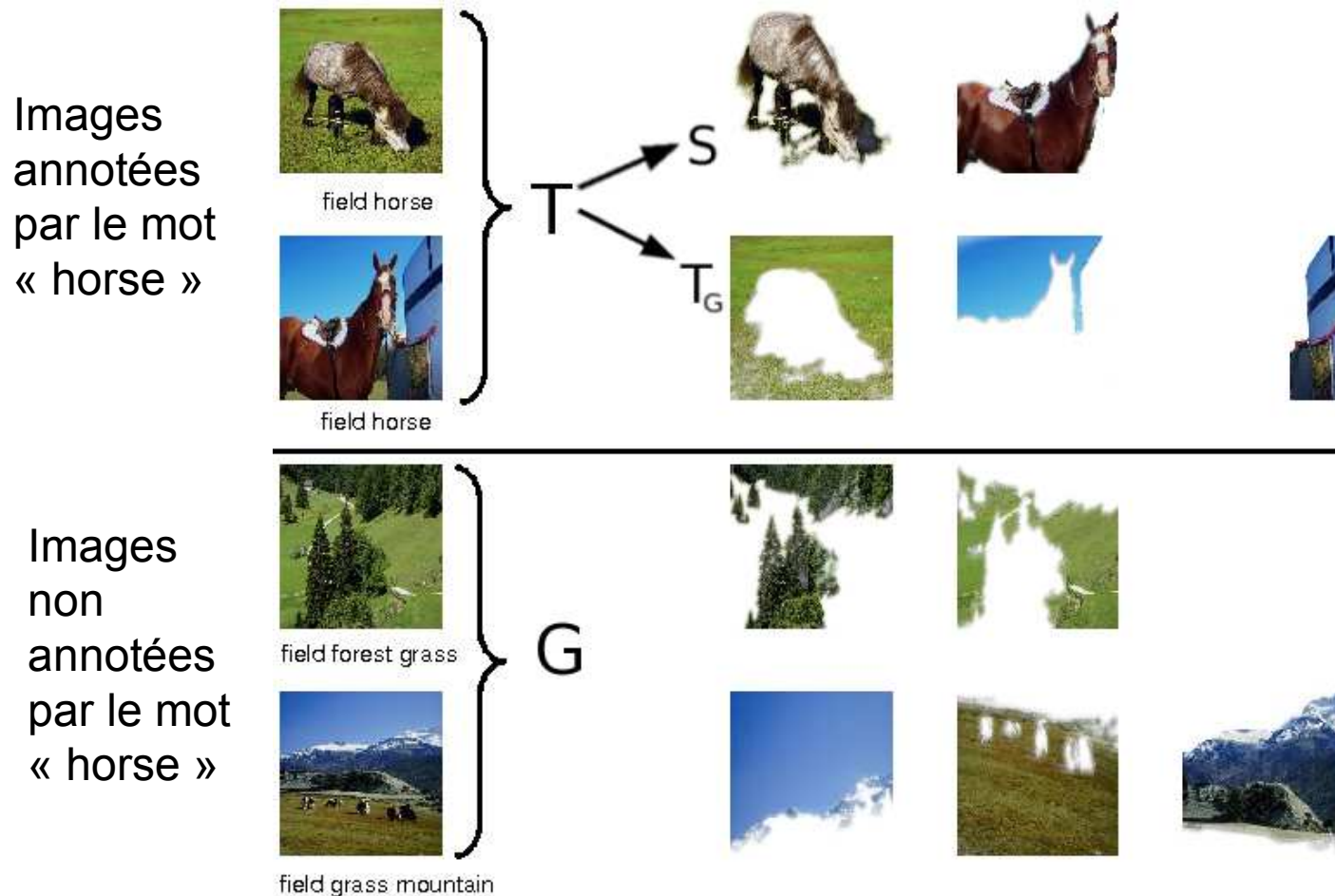
$$F(v_j; w_i) = \frac{B(v_j; w_i)}{B(v_j; w_i) + W(v_j; w_i)}$$

- Nous montrons théoriquement que **l'ordre des meilleurs F estimés par ALDA est le même que par LDA.**

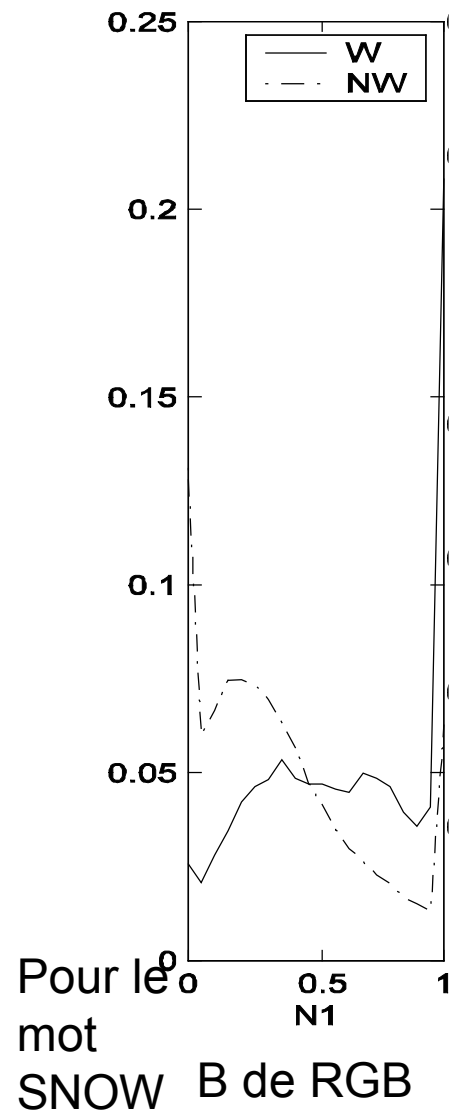
# Principe de démonstration de l'ALDA

Proposition : Le pouvoir discriminant sur les ensembles T et G peut être exprimé linéairement en fonction du pouvoir discriminant sur les ensembles S et G

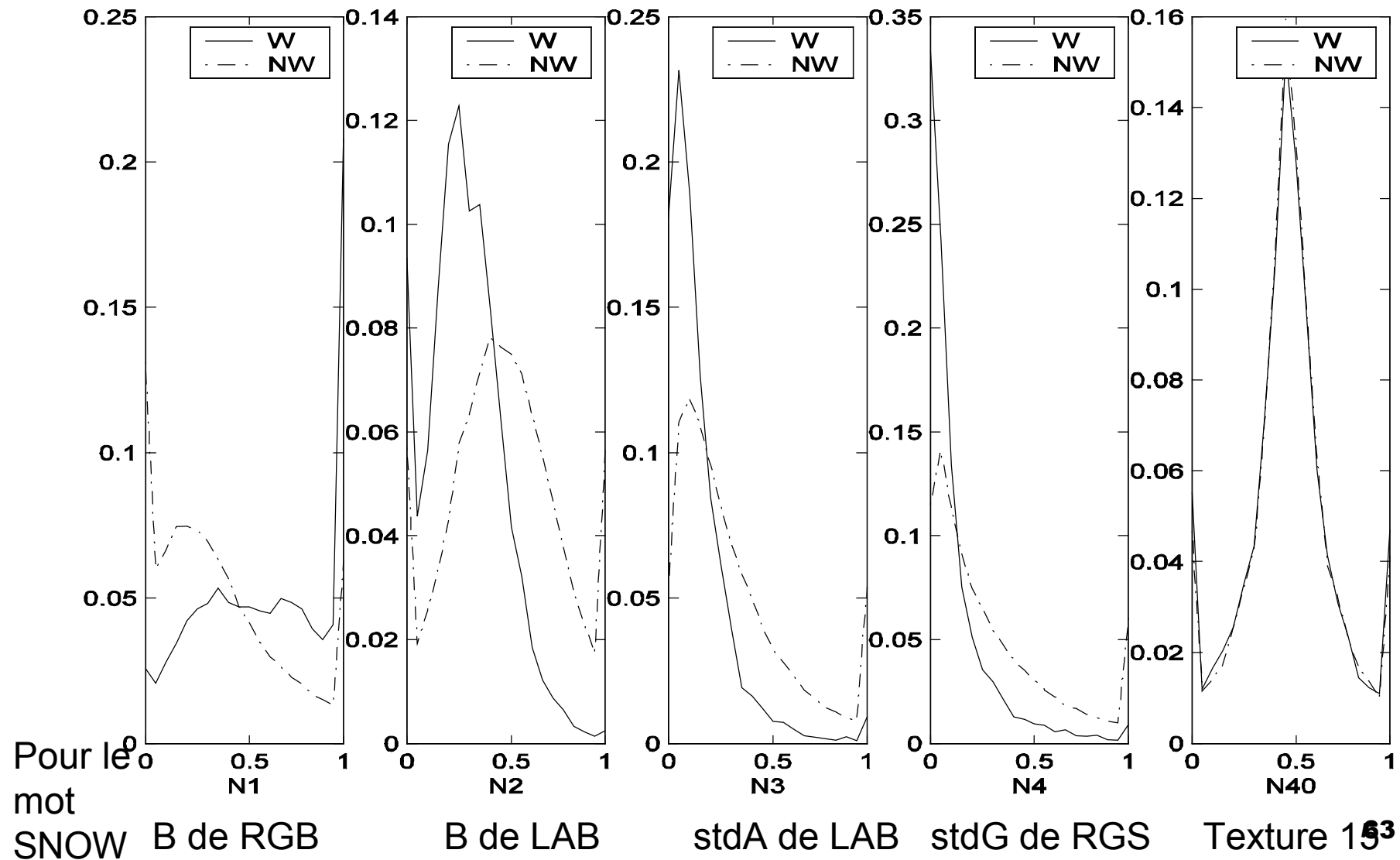
$$\hat{F}_{TG}(v_j; w_i) = A(w_i) \cdot F_{SG}(v_j; w_i) + B(w_i) \quad \text{avec } n_G \gg n_T$$



# Exemple : approximation de la LDA (ALDA)



## Exemple : approximation de la LDA (ALDA)



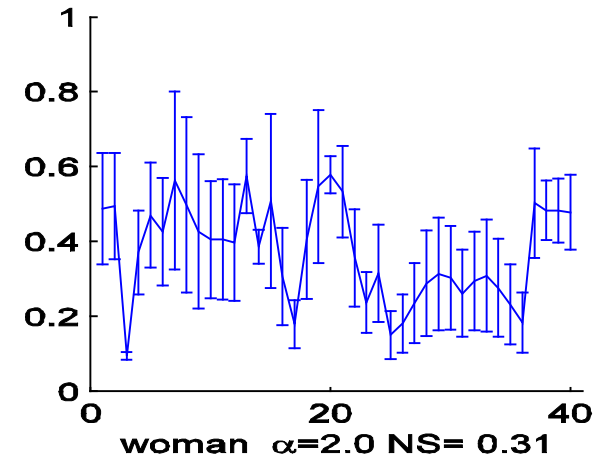
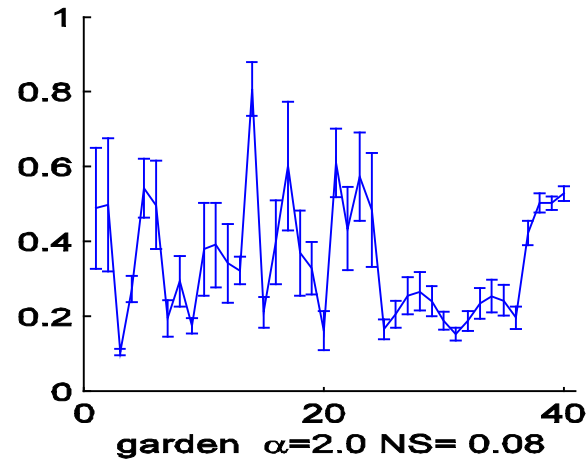
## Détermination du nombre de traits visuels à prendre en compte pour bien discriminer un mot

- Nous proposons de prendre les  $N$  dimensions visuelles qui cumulent  $\tau\%$  de la somme des pouvoirs discriminants de tous les traits.
  - Soient les traits visuels  $v_j$  ordonnés dans l'ordre décroissant des pouvoirs discriminants, nous choisissons  $N$  tel que :

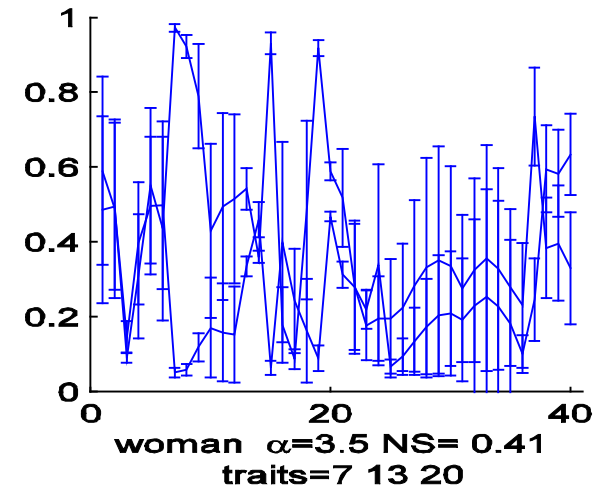
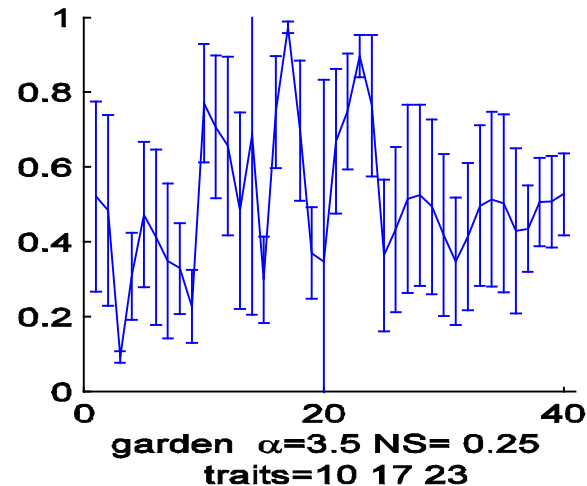
$$\sum_{j=1}^N F(v_j; w_i) = \tau \sum_{j=1}^{\delta} F(v_j; w_i)$$

# Exemples de clusters visuels dans l'espace usuel à 40 dimensions avec ou sans sélection par ALDA

Sans sélection



Avec sélection

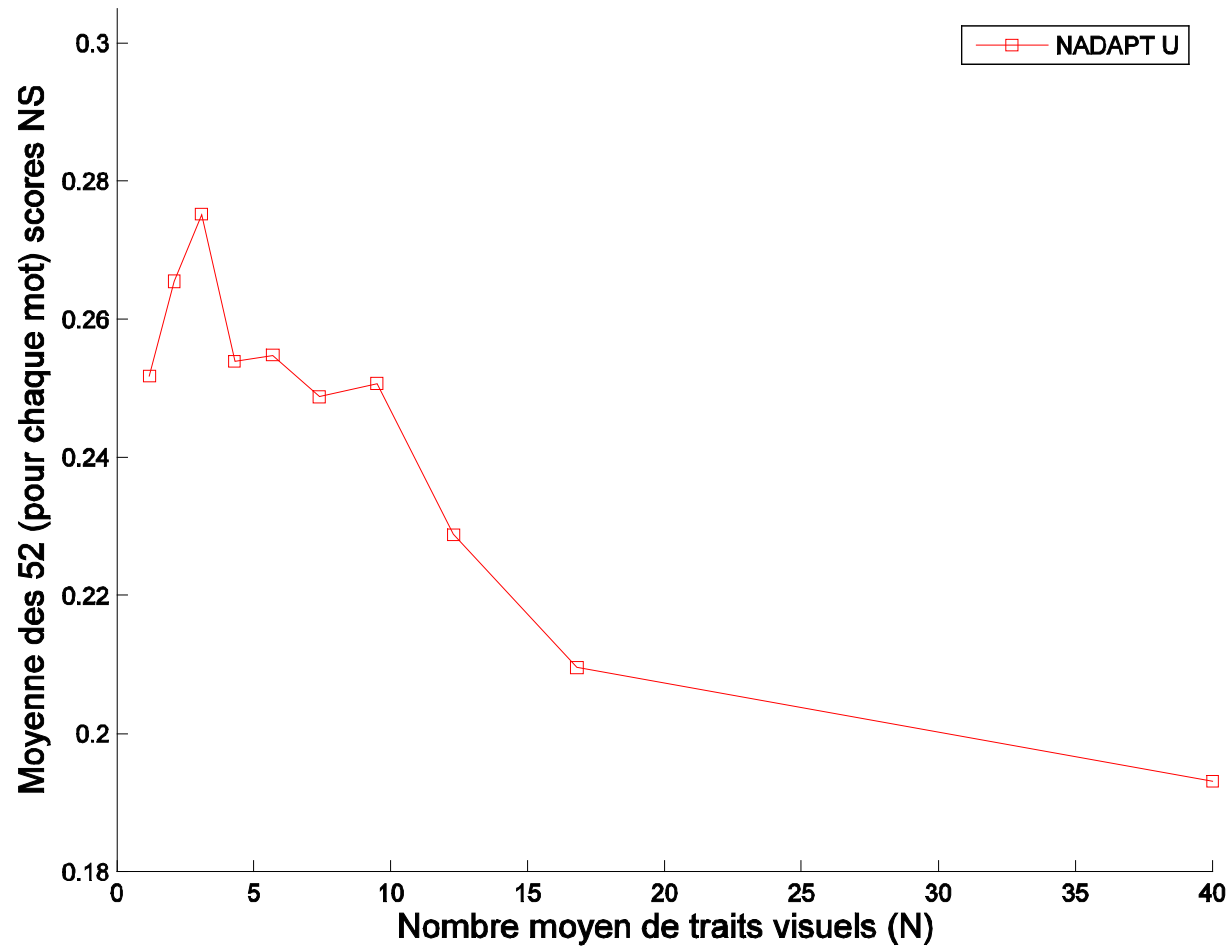


Écart-types rouge de RVB,  
vert de rvS et a de Lab

Rouge de RVb, rouge de  
rvS et a de Lab

# Scores moyens des classifications de 52 mots en fonction du nombre moyen de dimensions

## Expériences avec les traits visuels usuels (U)

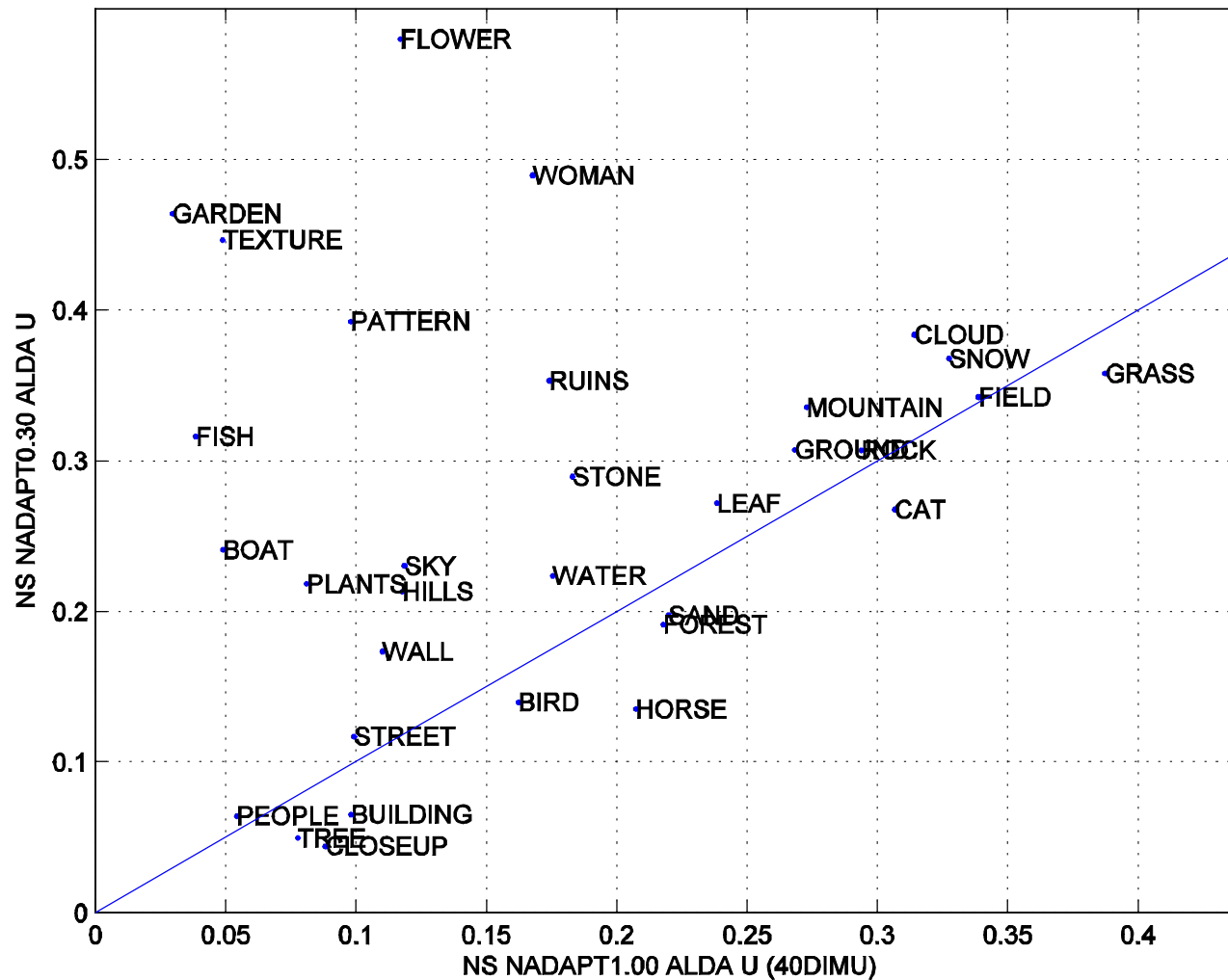


# Scores moyens des classifications de 52 mots en fonction du nombre moyen de dimensions

## Expériences avec les traits visuels usuels (U)



# Scores de classification de chaque mot sans/avec sélection de traits visuels



## Résumé des résultats

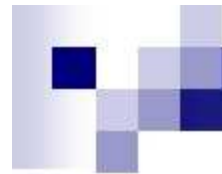
	<b>40DIMU</b>	<b>Lab</b>	<b>U avec sélection</b>	<b>ACP U</b>
<b><math>\tau</math></b>	<b>1.0</b>	<b>-</b>	<b>0.3</b>	<b>0.7</b>
Nombre moyen de dimensions	<b>40</b> -	<b>6</b> <b>(-85%)</b>	<b>3.1</b> <b>(-92%)</b>	<b>8</b> <b>(-80%)</b>
NS Moyen Classification	<b>0.192</b> -	<b>0.248</b> <b>(+29%)</b>	<b>0.275</b> <b>(+43%)</b>	<b>0.214</b> <b>(+11%)</b>

Nous montrons expérimentalement que d'autres descripteurs visuels, appelé hétérogénéité (Martinet et al., 2005) dérivés très rapidement des traits visuels U, apportent une information complémentaire que l'on peut utiliser pour améliorer les scores (fusion tardive jusqu'à +69%).

Nous obtenons expérimentalement des résultats similaires avec un autre critère de sélection : l'Approximation de la Diversité Marginale Maximale (AMMD).



# Conclusion



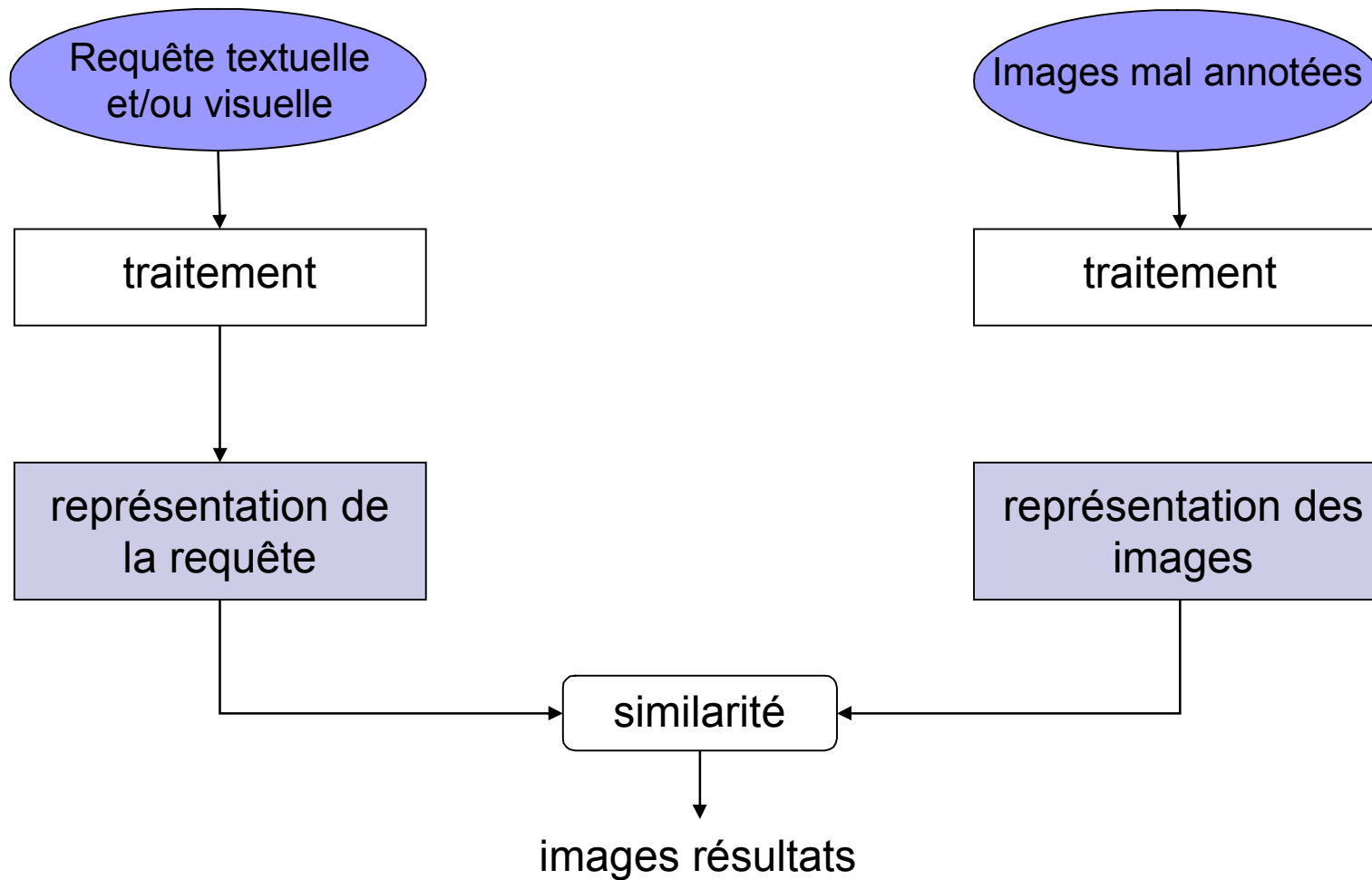


## Conclusion

- En résumé, les contributions de la thèse sont :
  - Un système d'auto-annotation rapide même sur un grand corpus d'images
  - Une méthode pour caractériser visuellement un mot-clé
  - Une méthode de sélection des dimensions sur des données mal-annotées

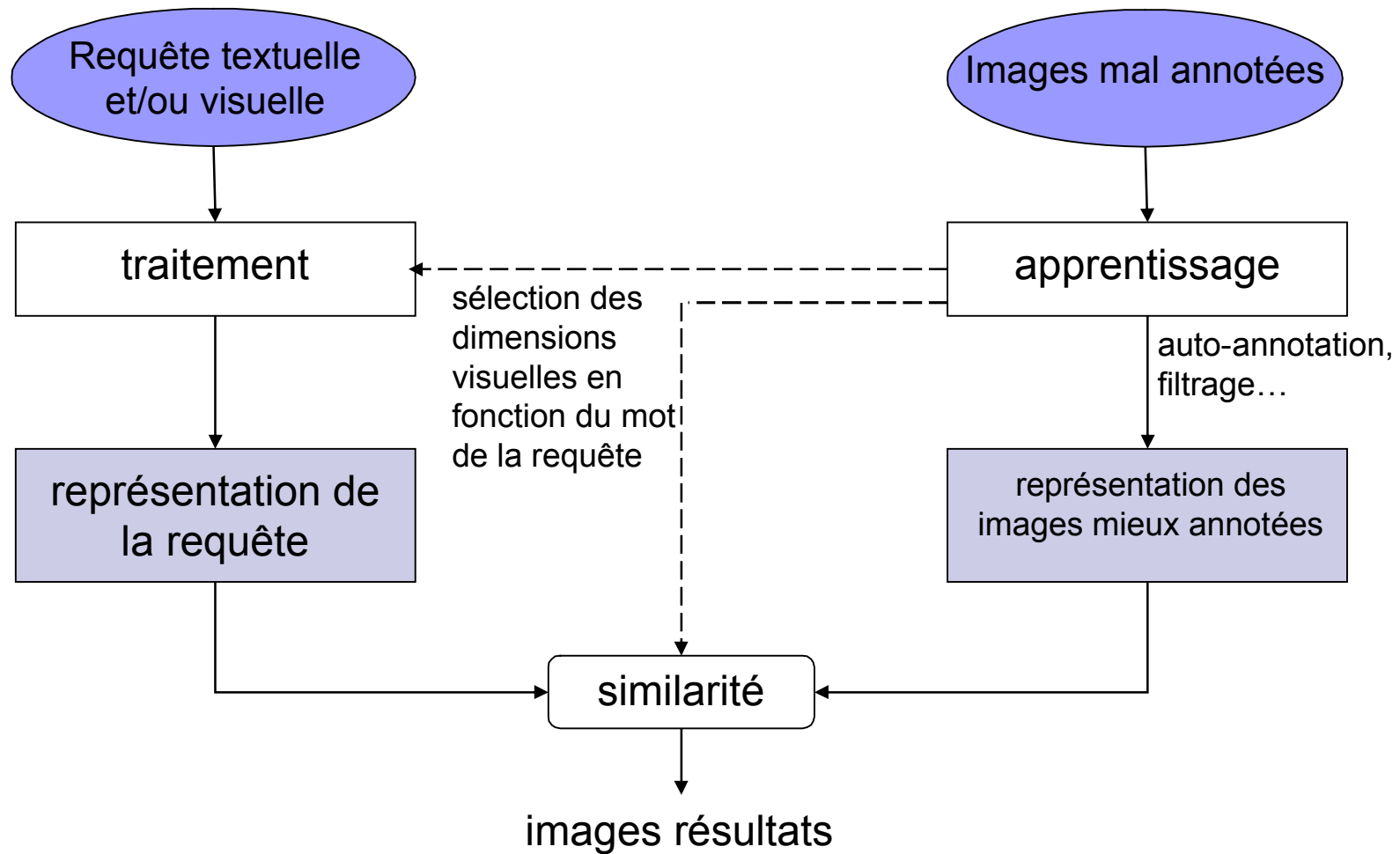
# Applications et perspectives

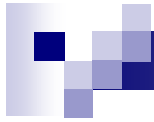
## Vers un moteur de recherche d'images sur le Web



# Applications et perspectives

## Vers un moteur de recherche d'images sur le Web





Merci de votre attention