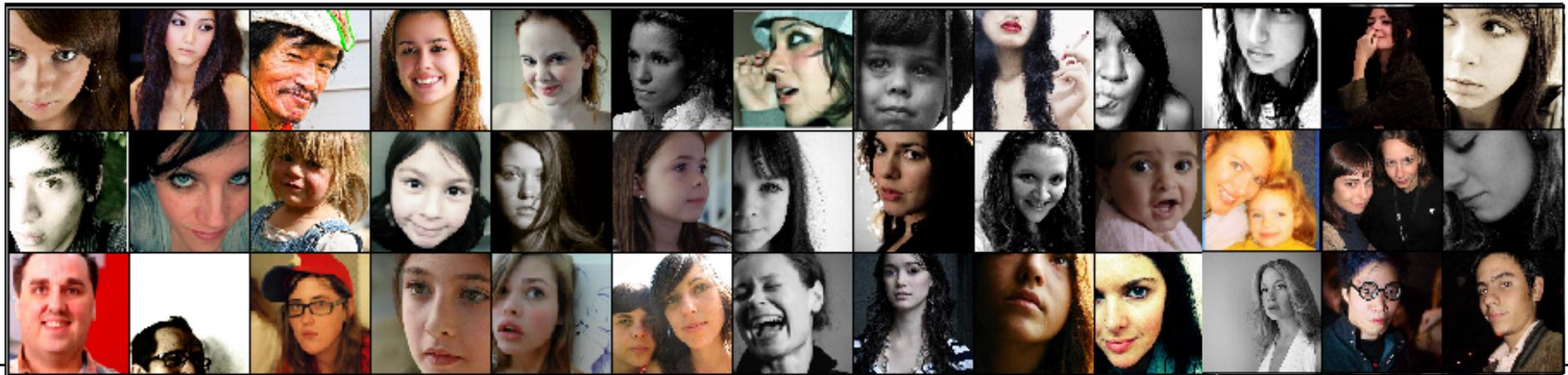


Classification multilingue et multimédia pour la recherche d'images dans le projet OMNIA

Atelier RISE

Inforsid

25 mai 2010



ANR



LIRIS

xerox



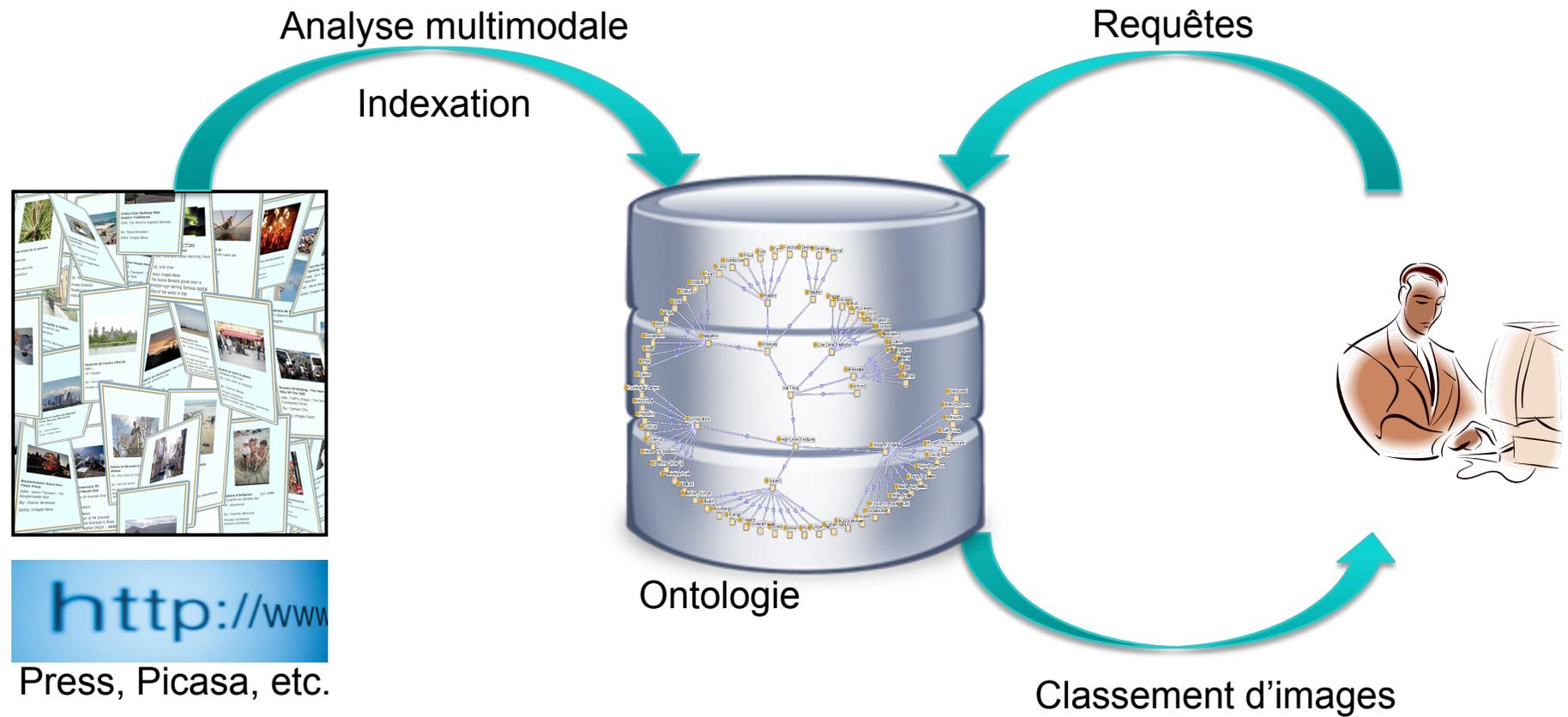
Buts du projet

OMNIA

“...explorer, développer et expérimenter de nouvelles techniques et de nouveaux outils pour **indexer, catégoriser, organiser** des documents, intégrant les aspects liés à **l'analyse du texte, des images** et du **contenu émotionnel**, ainsi que leur combinaison par une analyse multimodale” *

*ANNEXE 1 - Description scientifique du projet

Architecture générale



Les bases de données test

- **MIRFLICKR** : <http://press.liacs.nl/mirflickr/>
 - 25.000 images d'images créatives

- **Belga News**
 - 500.000 images de presse
 - des textes compagnons d'environ 50 mots

- Utilisées dans les compétitions ImageClef et Photo Annotation

MIRFLICKR



ANR



LIRIS

xerox



Belga News

-

1012—Australian Open champion Mary Pierce of France volleys the ball during the second-round match against Kyoko Nagatsuka of Japan in the Toray Pan Pacific Open women's tennis tournament in Tokyo 02 February. Pierce defeated Nagatsuka 6-4, 6-0.

-



1050730—AWA05 - 20020924 - BAGHDAD, IRAQ : Iraqi women sit under a portrait of Iraqi President Saddam Hussein in a waiting room in Baghdad's al-Mansur hospital 24 September 2002. Saddam Hussein is doggedly pursuing the development of weapons of mass destruction and will do his best to hide them from UN inspectors, the British government claimed in a 55-page dossier made public just hours before a special House of Commons debate on Iraq. Iraqi Culture Minister Hamad Yussef Hammadi called the British allegations "baseless." EPA PHOTO AFPI AWAD AWAD -

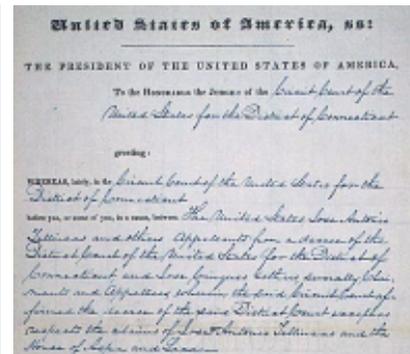
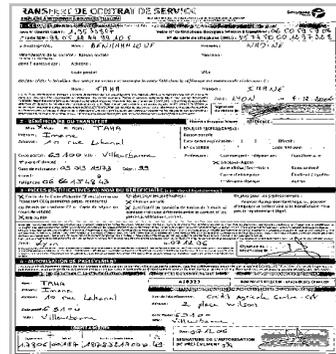


XRCE : Analyse des images pour la classification thématique et esthétique

- L'analyse contextuelle
- L'étape d'annotation

L'analyse contextuelle

Catégorisation *générique* et segmentation d'images

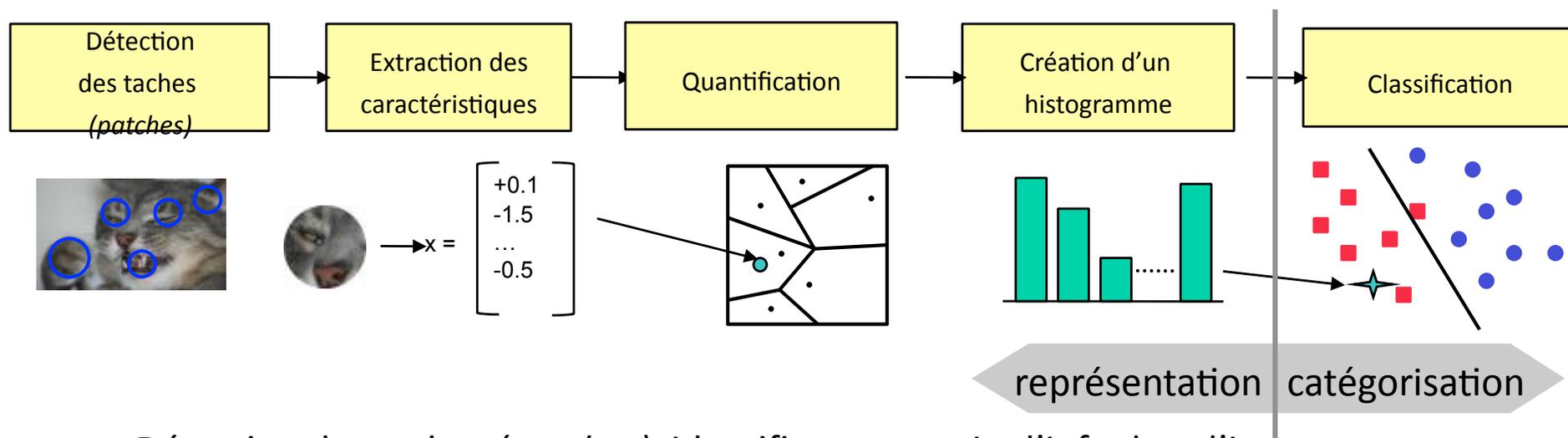


Générique parce que les mêmes techniques sont utilisées pour analyser plusieurs **objets**, **sujets** et **scènes** comme pour les **arts graphiques** (photos, peintures, dessins, documents, etc.)



Défi scientifique : **gérer la variabilité visuelle** inhérente aux différentes catégories ainsi que la variabilité en taille, forme, dimension, lumière, occlusion, etc.

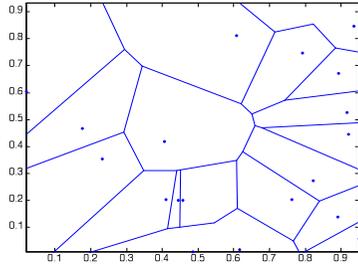
La chaîne *Sac de Mots*



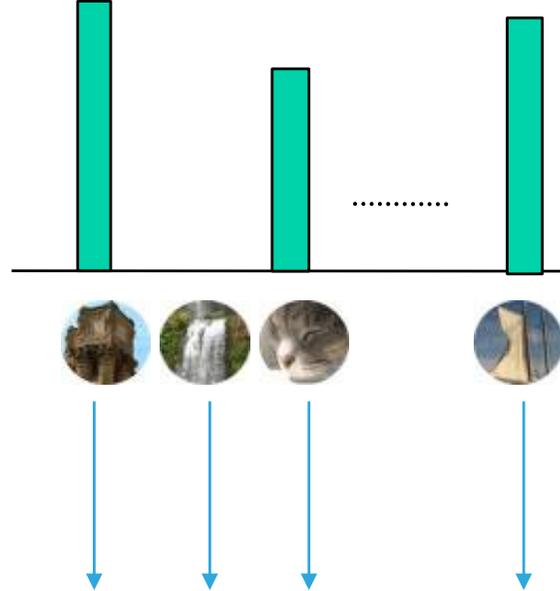
- Détection des taches (*patches*): identifier ou extraire l'info dans l'image
- Extraction des caractéristiques : évaluation quantitative des caractéristiques (de couleur et de texture)
- Quantification: mise en correspondance des caractéristiques locales avec les *mots visuels*
- Création d'un histogramme: évaluer une représentation globale de l'image
- Classification et apprentissage
 - Modèles discriminatifs : SVM (support vector machines)
 - Modèles génératifs: Bayes naïf, modèles bayésiens hiérarchiques

Vocabulaire visuel

K-moyennes



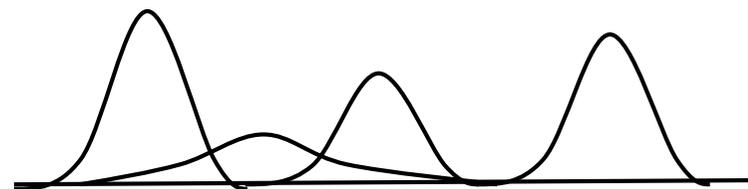
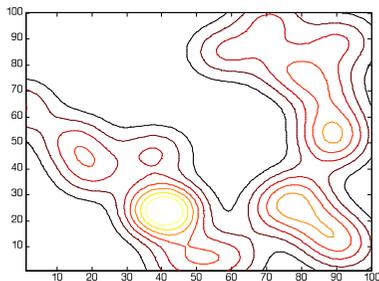
Affectation « dure »



$$[\#\{x_t \in C_1\}, \dots, \#\{x_t \in C_N\}]$$

x_t est un échantillon d'apprentissage et $C_1 \dots C_N$ indique les catégories de 1 à N .

GMM



Affectation « douce »

$$\sum_i [p(1 | x_t, \lambda), \dots, p(N | x_t, \lambda)]$$

le terme $p(i | x_t, \lambda)$ représente la probabilité que x_t appartienne à la gaussienne i du GMM

Classification et recherche d'images

1. Classificateur

Classificateurs « one-against-all »
$$K(f_I, f_J) = \sum_{t=1}^T \|\hat{f}_I - \hat{f}_t\|_1 \cdot \|\hat{f}_J - \hat{f}_t\|_1$$

où $\{f_1, f_2, \dots, f_T\}$ sont les vecteurs de caractéristiques des taches (patches) 1 à T

2. Recherche d'images

$$D(f_I, f_J) = \|\hat{f}_I - \hat{f}_J\|_1$$

où $\|\cdot\|_1$ est la norme L1, et \hat{f}_t, f_t sont normalisés pour avoir une norme L1 égale à 1.

Résultats de la classification

Résultat: mots-clés (concepts) avec des probabilités



mots-clés	probabilités d'existence
Indoor	0.00
Outdoor	1.00
Person	0.26
Day	0.94
Night	0.01
Water	0.01
Pathway	0.42
Vegetation	0.59
Tree	0.47
Mountains	0.08
Beach	0.02
Buildings	0.98
Sky	1.00
Sunny	0.65
Cloudy	0.20
Overcast	0.00
Animal	0.04

Étape d'annotation

Résultats de la campagne CLEF 2009

- Participation à deux tâches différentes: ***“Large Scale Visual Concept Detection and Annotation Task”*** et ***“Photo Retrieval Task.”***
- 5.000 images utilisées pour l'entraînement et 13.000 images pour tester.
- Les résultats ont été classés en fonction de deux types de mesure, un pour évaluer la performance par concept (**EER** et **AUC**), et un autre pour l'évaluation par photo (mesure hiérarchique **MH**).

LIRIS :

Reconnaissance de la sémantique émotionnelle portée par les images

Perception émotionnelle d'images ?

Positif - actif

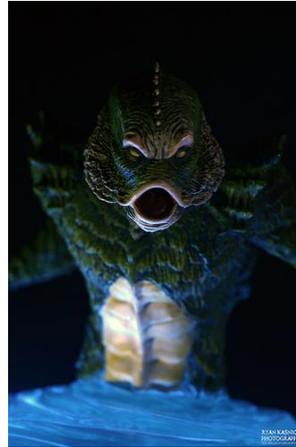
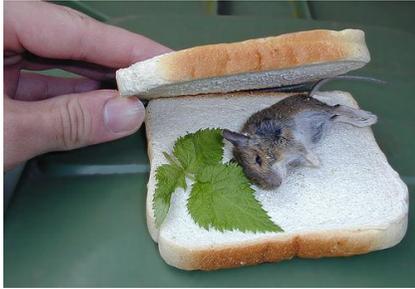


Positif - passif



Perception émotionnelle d'images ?

Négatif - actif



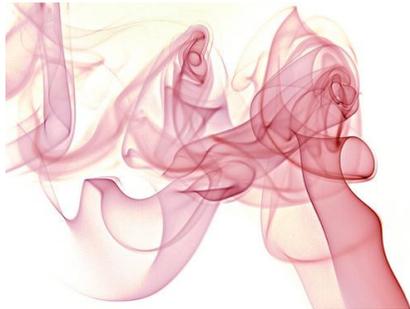
Négatif - passif



Caractère subjectif et ambigu : cas de conflits



Reconnaissance de la sémantique émotionnelle portée par les images



Positif - actif



Positif - passif



Négatif - actif

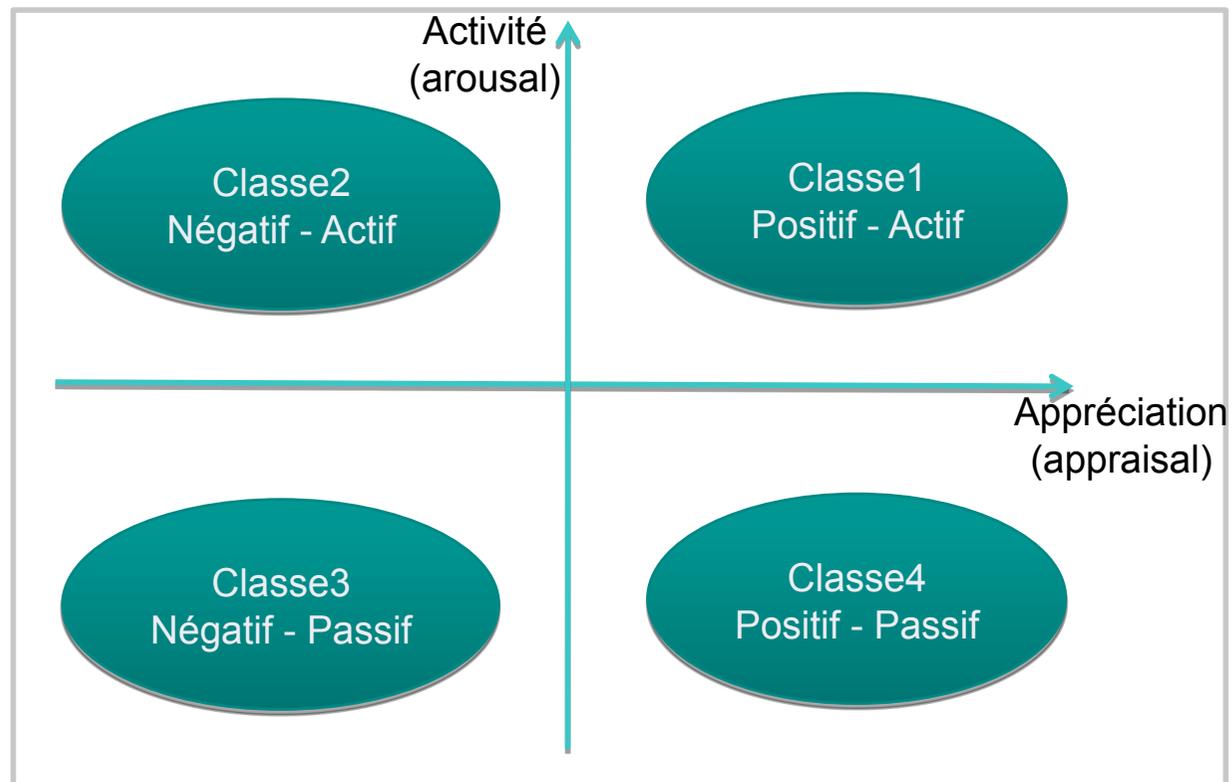


Négatif - passif

- Défis scientifiques :
 - Modéliser l'émotion
 - Franchir le fossé sémantique entre les informations de bas niveau (descripteurs visuels et/ou textuels) et les concepts de niveau sémantique élevé (émotions)
 - Fusionner des informations de nature différente pour la décision finale d'appartenance à une classe
 - Prendre en compte la nature complexe, ambiguë et conflictuelle des émotions

Représentation des émotions

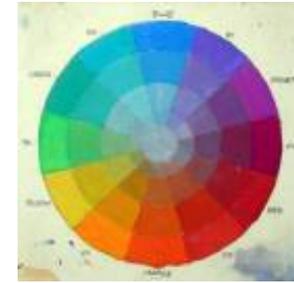
- Modèle discret / modèle dimensionnel



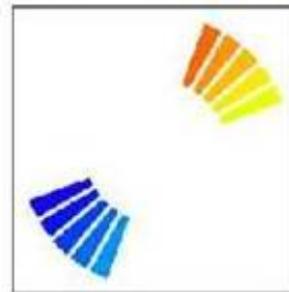
Descripteurs visuels utilisés

- Descripteurs de bas niveau :
 - Couleur
 - Moments de couleur
 - Histogramme de couleurs
 - Texture
 - Caractéristiques de Tamura
 - Matrice de cooccurrences
 - LBP (Local Binary Patterns)
 - Forme
 - Histogramme de contours
- Descripteurs de niveau intermédiaire :
 - Harmonie
 - Dynamisme

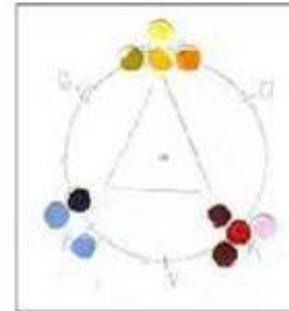
Descripteurs d'harmonie



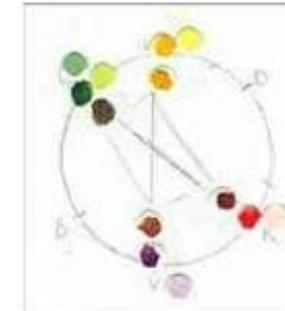
monochrome



bichrome



trichrome



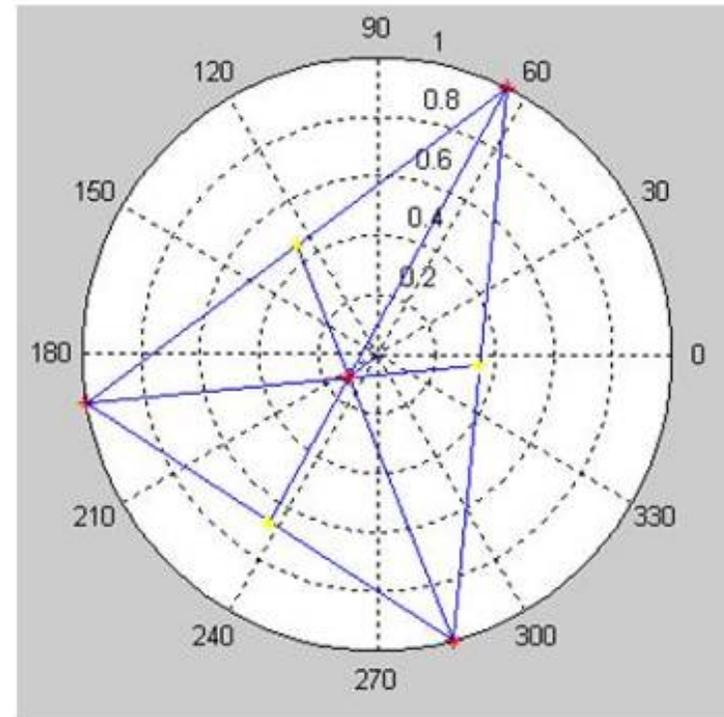
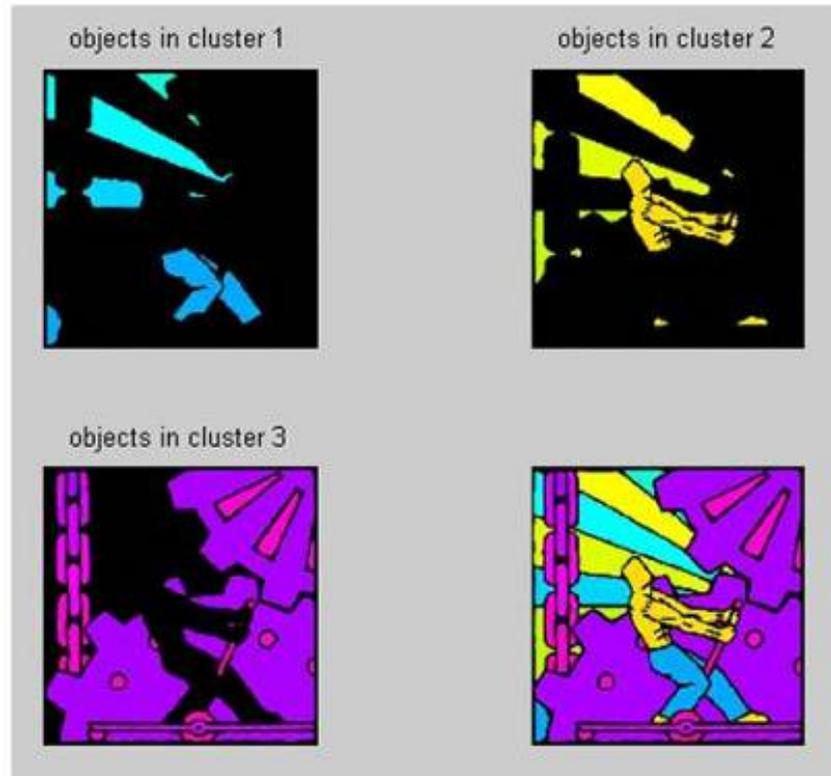
quadrichrome

D'après la théorie des couleurs d'Itten

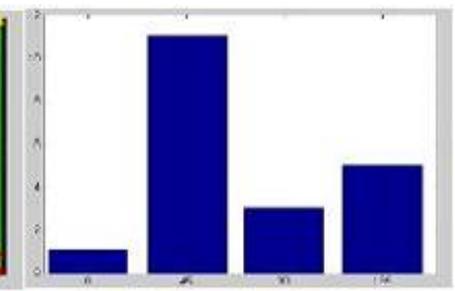
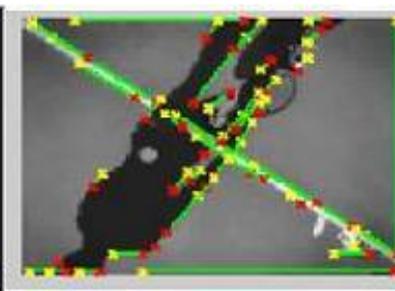
J. Itten, *Art of Color (Kunst der Farbe)*, Otto Maier Verlag, Ravensburg, Germany, 1961.



Descripteurs d'harmonie



Descripteurs de dynamisme



Utilisation de la théorie des fonctions de croyance (evidence theory) pour la fusion des descripteurs

- Elle a été introduite par [Dempster68], puis formalisée par [Smets98]
 - (modèle de croyance transférable TBM)
- Elle fournit un cadre pour raisonner sur des connaissances imparfaites :
 - incertaines
 - incomplètes
 - ambiguës
 - conflictuelles

Utilisation de la théorie des fonctions de croyance (evidence theory) pour la fusion des descripteurs

- Grâce aux fonctions de croyance, elle généralise :
 - les mesures de probabilité
 - les mesures de possibilité (sous-ensembles flous)
- Les fonctions de croyance (distributions de masses) permettent d'affecter un degré de croyance à l'association d'une ou plusieurs émotions à une image

La théorie des fonctions de croyance

- Espace de définition: Ω
 - n hypothèses H_i mutuellement exclusives
 - $\Omega = \{H_1, H_2, H_3, \dots, H_n\}$
- Cadre de discernement: 2^Ω (tous les sous-ensembles de Ω)
 - $2^\Omega = \{A \mid A \subseteq \Omega\}$
 - $2^\Omega = \{\emptyset, \{H_1\}, \{H_2\}, \dots, \{H_n\}, \{H_1, H_2\}, \dots, \Omega\}$ si Ω est fini
- Distribution de masses
 - $m^\Omega : 2^\Omega \rightarrow [0, 1]$
 - $A \rightarrow m^\Omega(A)$
- avec :
 - $m^\Omega(\emptyset) = 0$
 - $\sum_{A \subseteq \Omega} m^\Omega(A) = 1$

Règles de combinaison

- **Objectif**
 - Fusion de données issues de différents capteurs
- **Somme orthogonale conjonctive (TBM)**
 - 2 distributions de masse $m_{S_1}^\Omega, m_{S_2}^\Omega$ issues de S_1 et S_2 et indépendantes

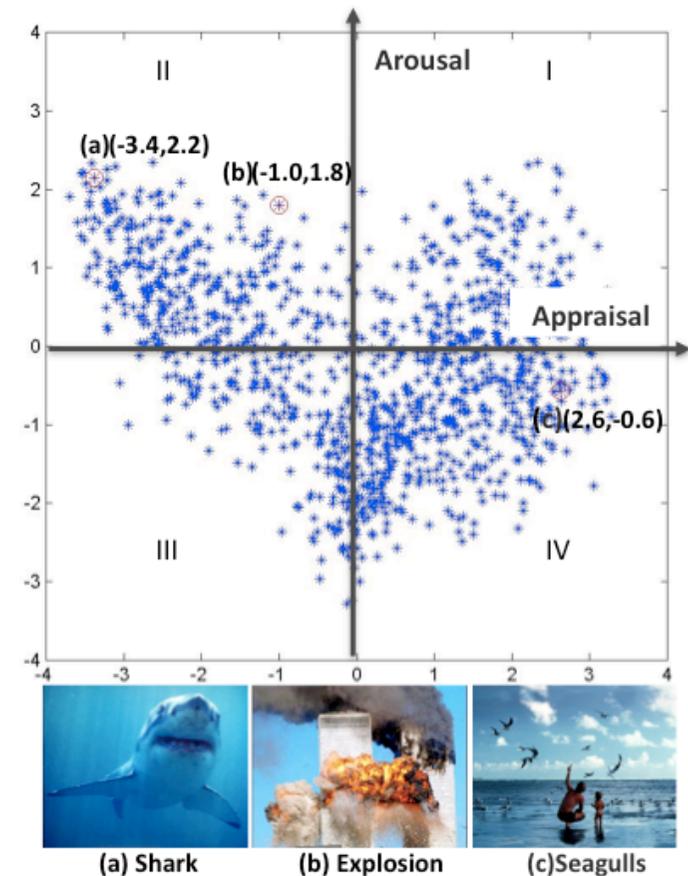
$$\forall A \subseteq \Omega, m_{S_1 \cap S_2}^\Omega(A) = \sum_{B \subseteq \Omega, C \subseteq \Omega, B \cap C = A} m_{S_1}^\Omega(B) \times m_{S_2}^\Omega(C)$$

- **Notion de conflit : $m_{S_1 \cap S_2}^\Omega(\emptyset) \neq 0$**
 - représente le fait que les sources ont amené à des observations contradictoires
 - permet de savoir si les mesures sont fiables et cohérentes
- **Grandeurs de décision**
 - Masse de croyance $\forall A \subseteq \Omega, bel(A) = \sum_{\emptyset \neq B \subseteq A} m^\Omega(B)$
 - Croyance (belief) $\forall A \subseteq \Omega, pl(A) = \sum_{A \cap B \neq \emptyset} m^\Omega(B)$
 - Plausibilité (plausibility)



Expérimentations : la base d'images

- **Utilisation de la base d'images IAPS**
 - Utilisée comme base de référence en psychologie
 - Caractérisation des images selon trois critères (arousal, appraisal, control)



P. J. Lang et al, International affective picture system (IAPS): Affective ratings of pictures and instruction manual, Technical Report A-8, University of Florida, Gainesville, 2008.

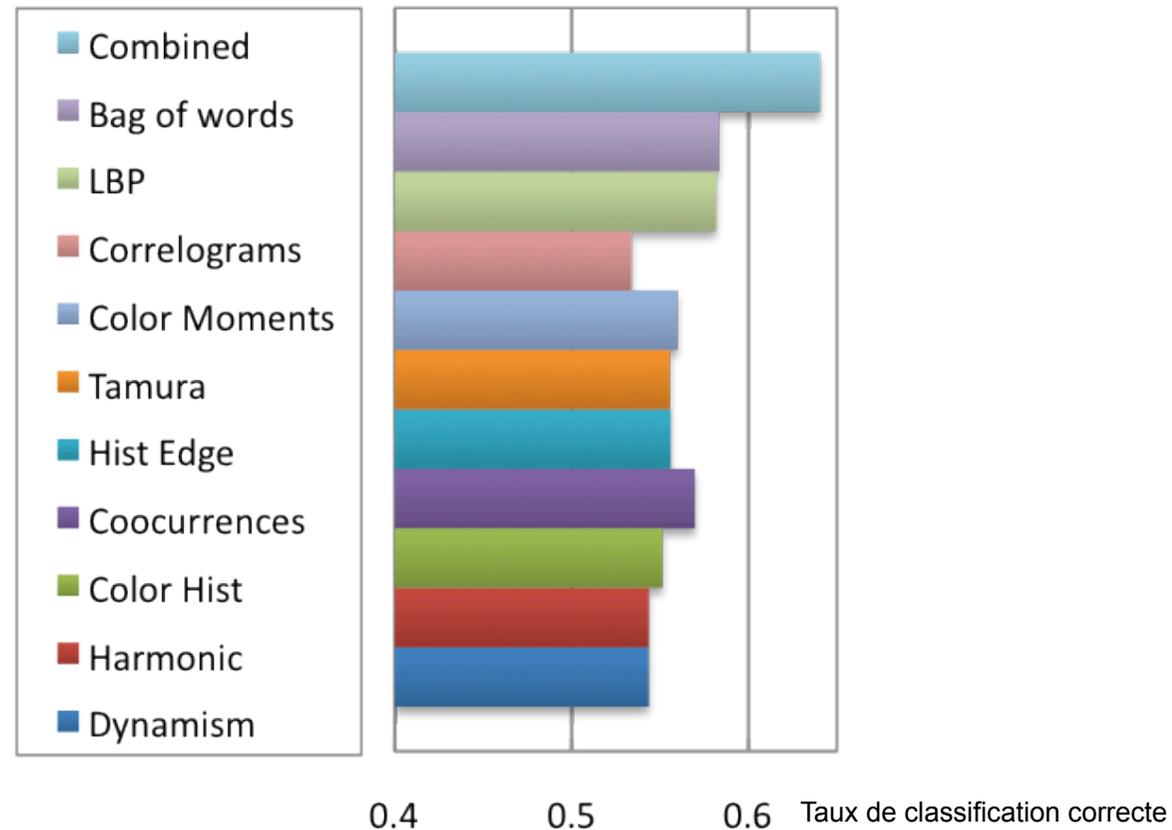
Expérimentations : évaluation des classificateurs

	<i>NN (%)</i>	<i>SVM (%)</i>	<i>Adaboost (%)</i>	<i>Knn (%)</i>
I	57, 21	61, 55	65,02	51, 33
II	63, 42	60, 34	62, 53	64,42
III	58, 21	62,61	61, 31	51, 52
IV	61, 72	65,75	64, 30	61, 71

Meilleurs pourcentages moyens de classification correcte :

- SVM : 62,6 %
- Adaboost : 63,3%

Expérimentations : évaluation des descripteurs



Meilleur pourcentage moyen de classification correcte :
Combined (fusion par la théorie des fonctions de croyance) : 64,6%

Expérimentations : théorie des fonctions de croyance

Prédit Réel	I	II	III	IV
I	63.32	12.25	11.23	11.15
II	11.05	61.42	12.27	11.82
III	16.21	12.53	66.19	10.52
IV	10.42	13.80	10.31	67.51
Total	100	100	100	100

Pourcentage moyen de classification correcte : 64,6%

Remarques

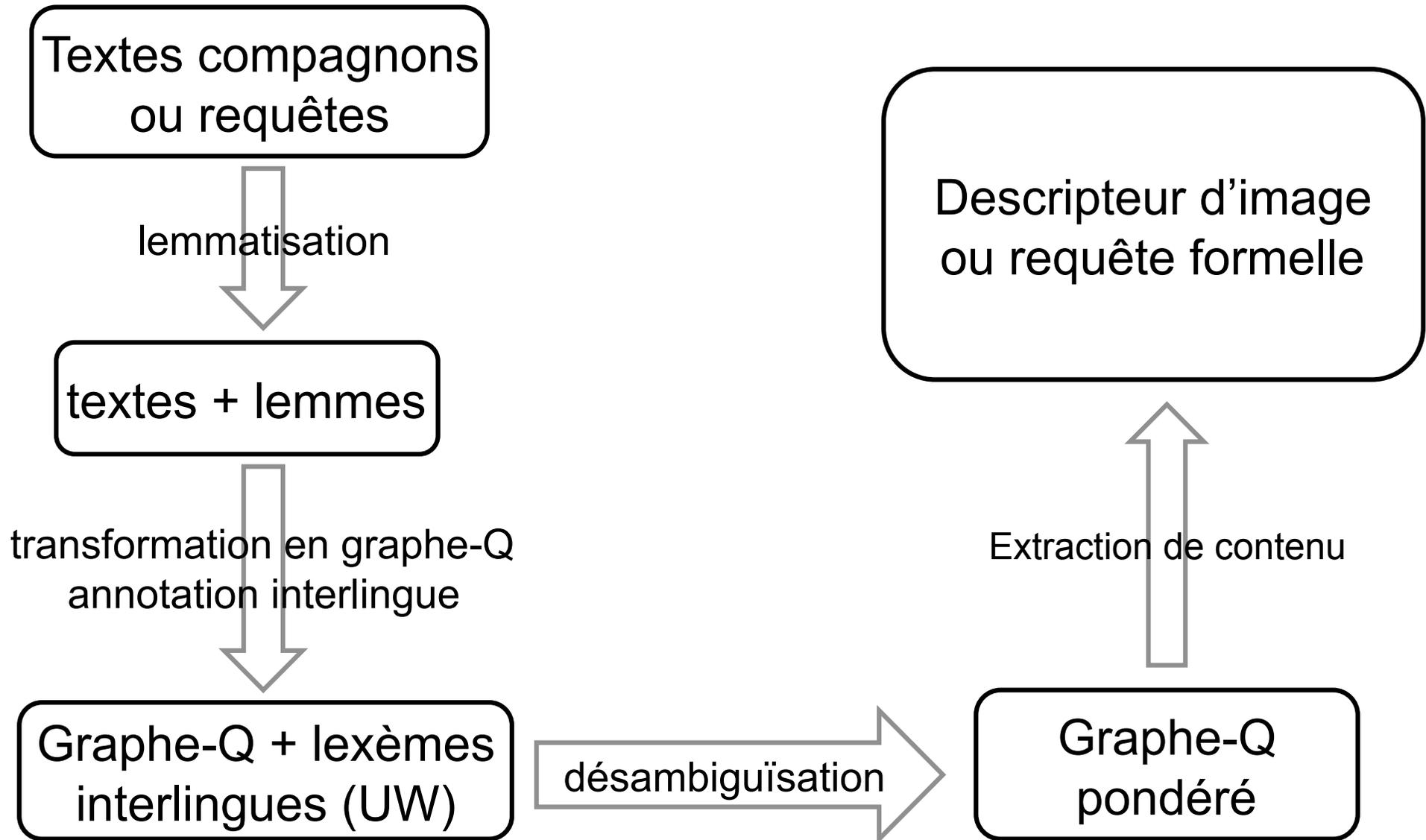
- Parmi les classificateurs traditionnels évalués, SVM et Adaboost sont les plus performants
- Parmi les descripteurs traditionnels évalués, LBP (texture) est le plus efficace
- Les descripteurs d'harmonie et de dynamisme portent des informations importantes liées aux émotions
- La combinaison des descripteurs par la théorie des fonctions de croyance permet d'obtenir de bons résultats et offre un cadre intéressant pour la fusion d'information de nature différente

Remarques

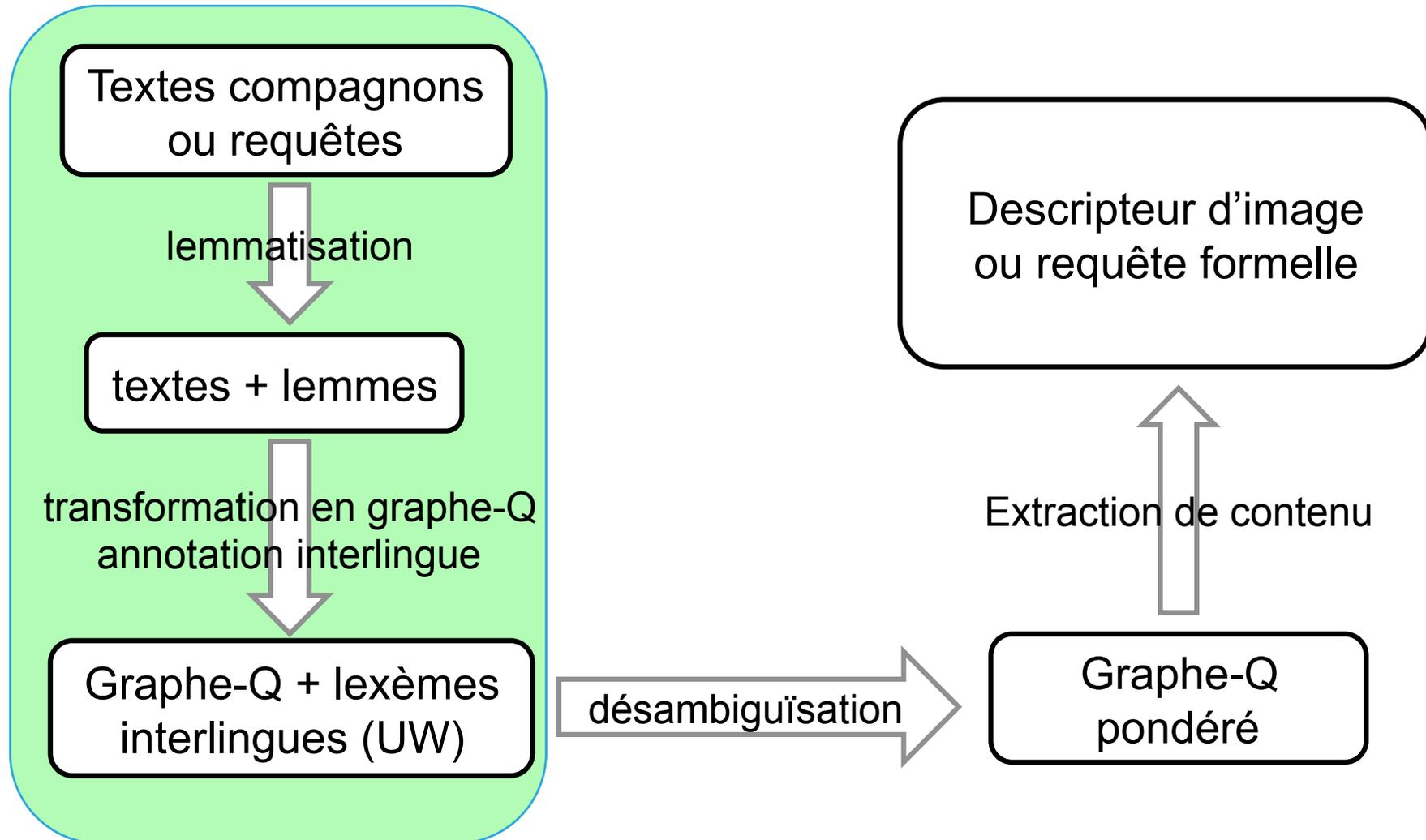
- La reconnaissance de la sémantique émotionnelle portée par des images contenant des objets à fort impact émotionnel ne peut se faire qu'avec l'information visuelle
 - Utilisation de l'information contenue dans les légendes (avec LIG)
 - Utilisation de l'identification des concepts dans l'image (avec Xerox)

LIG-GETALP :

**Extraction de contenu basée sur une
annotation interlingue des textes**



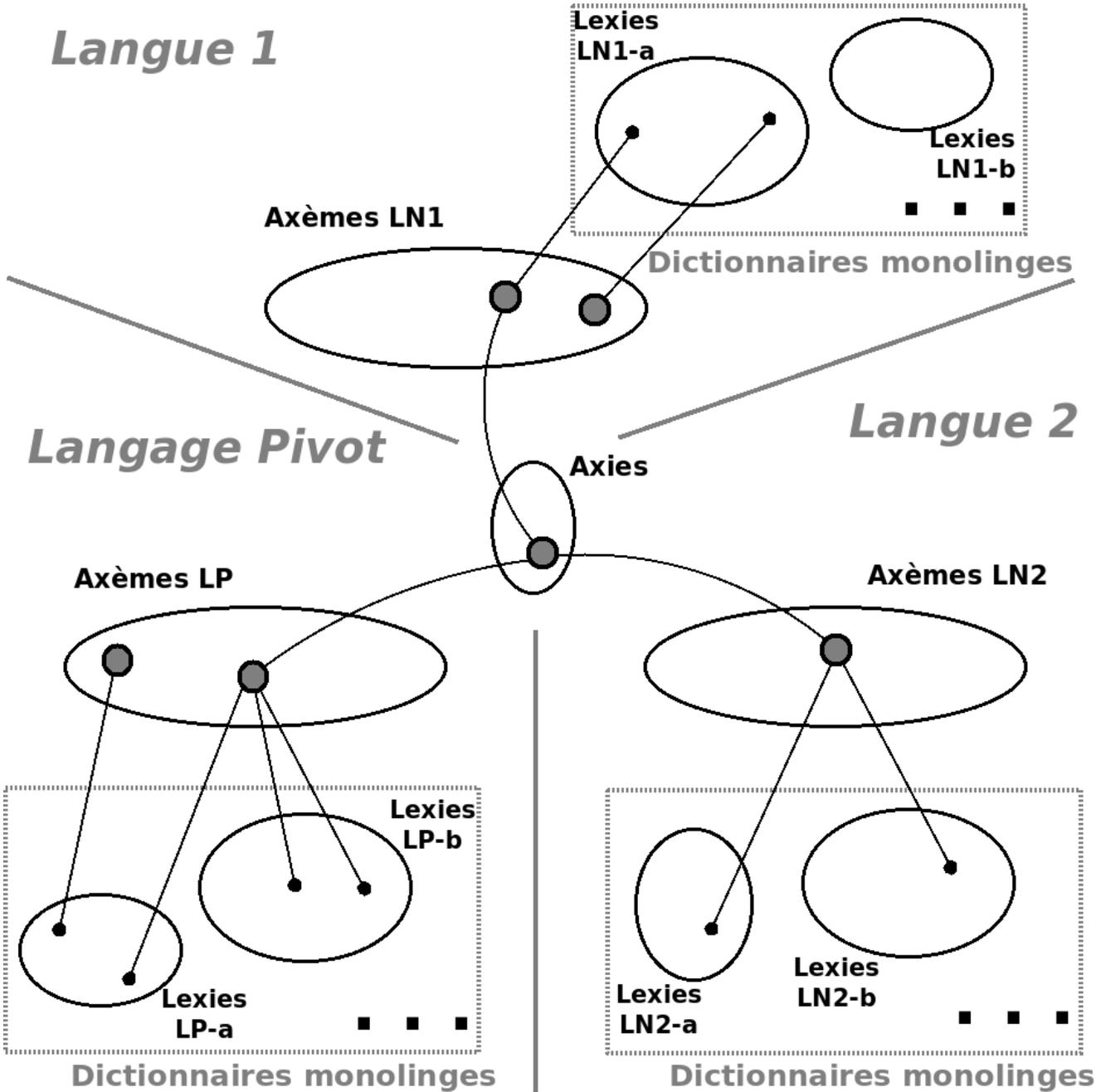
Annotation interlingue des textes



Lexèmes interlingues pour l'annotation des textes

- **Universal Networking Language (UNL)**
 - un projet international lancé en 1996 par l'Université des Nations Unies
 - un langage pivot abstrait (structure de graphe)
- **Universal Words (UW)**
 - mot vedette + restrictions
 - représente une acception sans ambiguïté
 - exemples :
 - book(icl>do,agt>human,obj>thing) ; book(icl>thing)
 - ikebana(icl>lower_arrangement)
 - 200.000 UW++ construites à partir des Synsets de WordNet

Plate-forme de gestion des dictionnaires : PIVAX



Structure de données pour l'annotation des textes : les graphes-Q

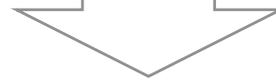
- Graphes orientés, étiquetés par des arbres simplement étiquetés (graphes-Q)
- Règles de réécriture
- 1 système-Q = n traitements-Q, 1 traitement-Q = {règles-Q}

- Intérêt :
 - possibilité de représenter les ambiguïtés d'un texte
 - permet à des non-informaticiens d'éditer un processus à l'aide de langages spécialisés

Texte lemmatisé sous forme de graphe-Q

in a waiting room

lemmatisation



\$OCC(\$FORM(in), ...

\$OCC(\$FORM(a), ...

\$OCC(\$FORM(waiting), ...

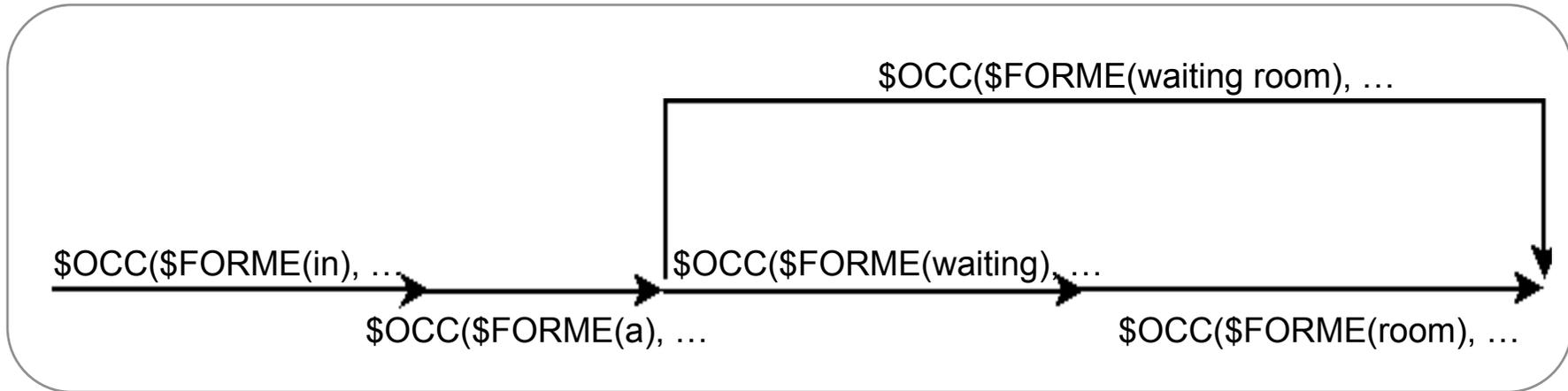
\$OCC(\$FORM(room), ...

\$OCC(\$FORM(waiting room), ...

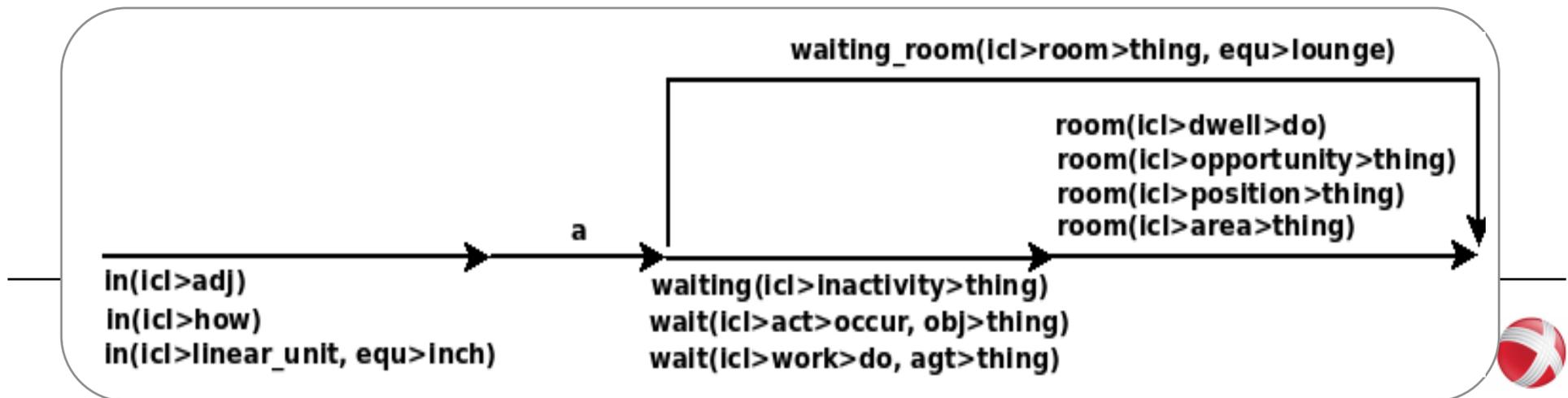
```
-01- $OCC($FORM(in), $LU(LEMMA(in), CAT(N)), $LU(LEMMA(in), CAT(A))) -02-  
-02- $OCC($FORM(a), $LU(LEMMA(a), CAT(UNK))) -03-  
-03- $OCC($FORM(waiting room), $LU(LEMMA(waiting room), CAT(N))) -04-  
-03- $OCC($FORM(waiting), $LU(LEMMA(waiting), CAT(N)), $LU(LEMMA(wait), CAT(V)))  
+ $OCC($FORM(room), $LU(LEMMA(room), CAT(N)), $LU(LEMMA(room), CAT(V))) -04-
```



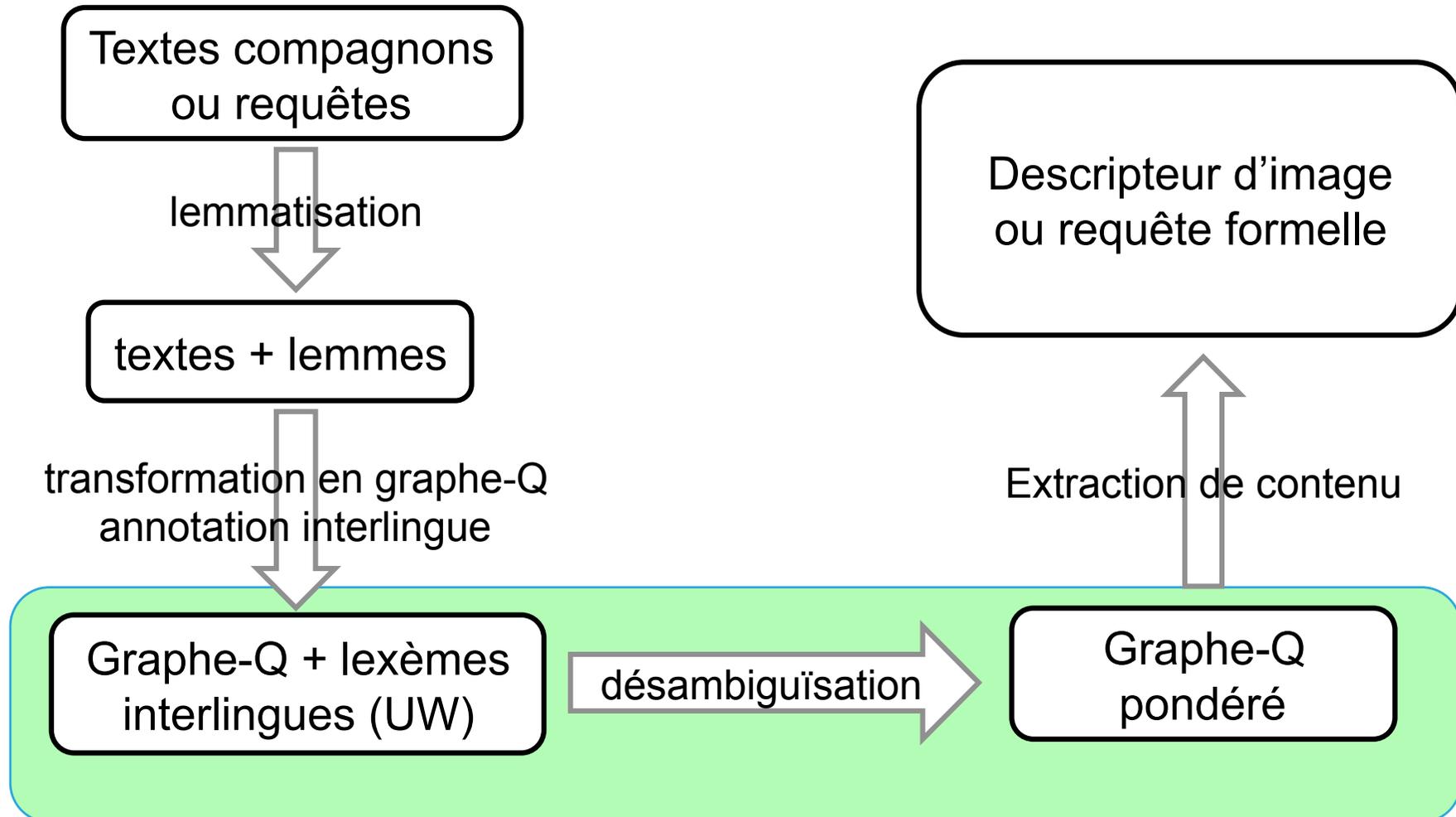
Ajout d'UW avec une règle-Q



$\$LU(LEMMA(waiting), CAT(N), U^*) ==$
 $\$LU(UWS(waiting(icl>inactivity>thing)), CAT(N)).$



Désambiguïisation basée sur des vecteurs conceptuels

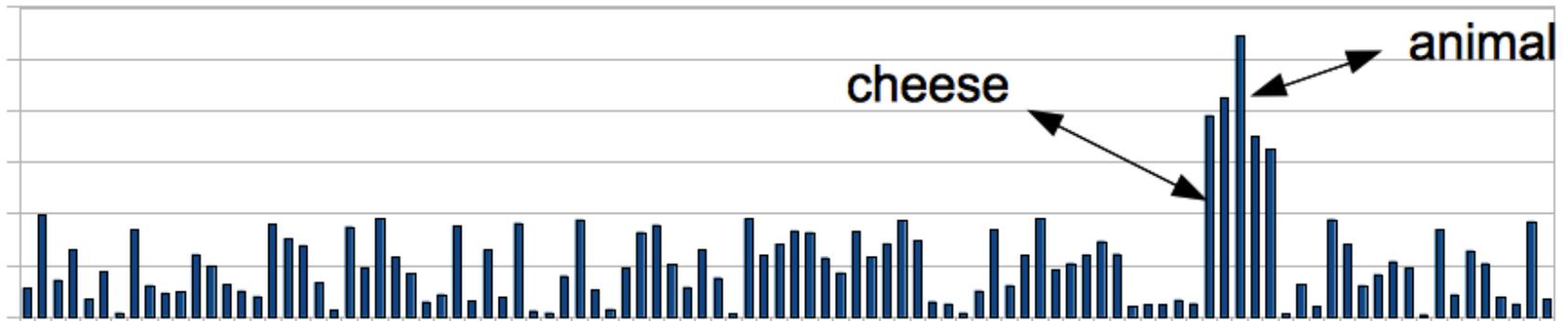


Les vecteurs conceptuels

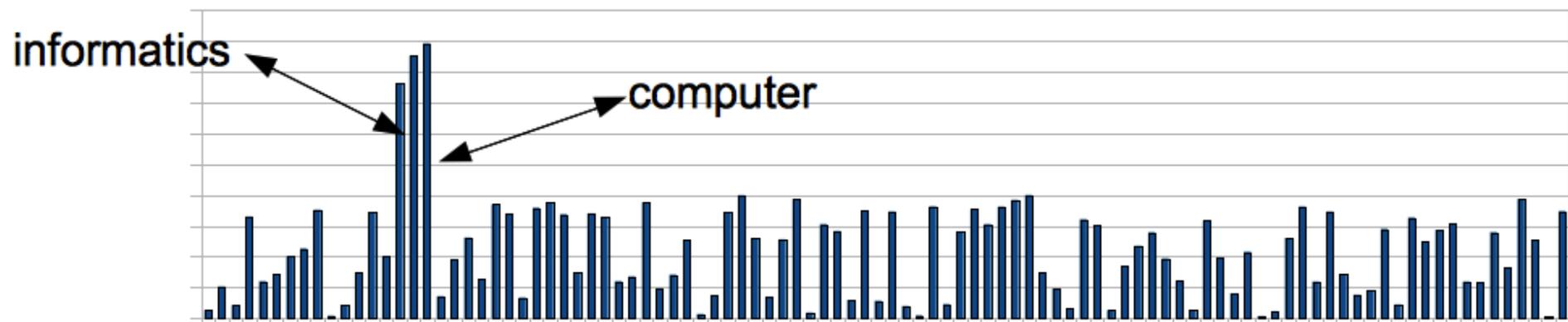
- Représentation thématique
 - Par exemple, 873 composantes (concepts issus du thésaurus Larousse)
 - (1) existence, (2) inexistence, (3) matérialité, ..., (516) liberté, ..., (872) jeux, (873) jouets
 - Une composante du vecteur correspond globalement à l'activation d'un concept
- Associable à :
 - toute donnée textuelle (mot, syntagme, paragraphe, ...)
 - tout autre document (image, vidéo, multimédia, ...)

Vecteurs conceptuels : exemples

- mouse(icl>rodent)

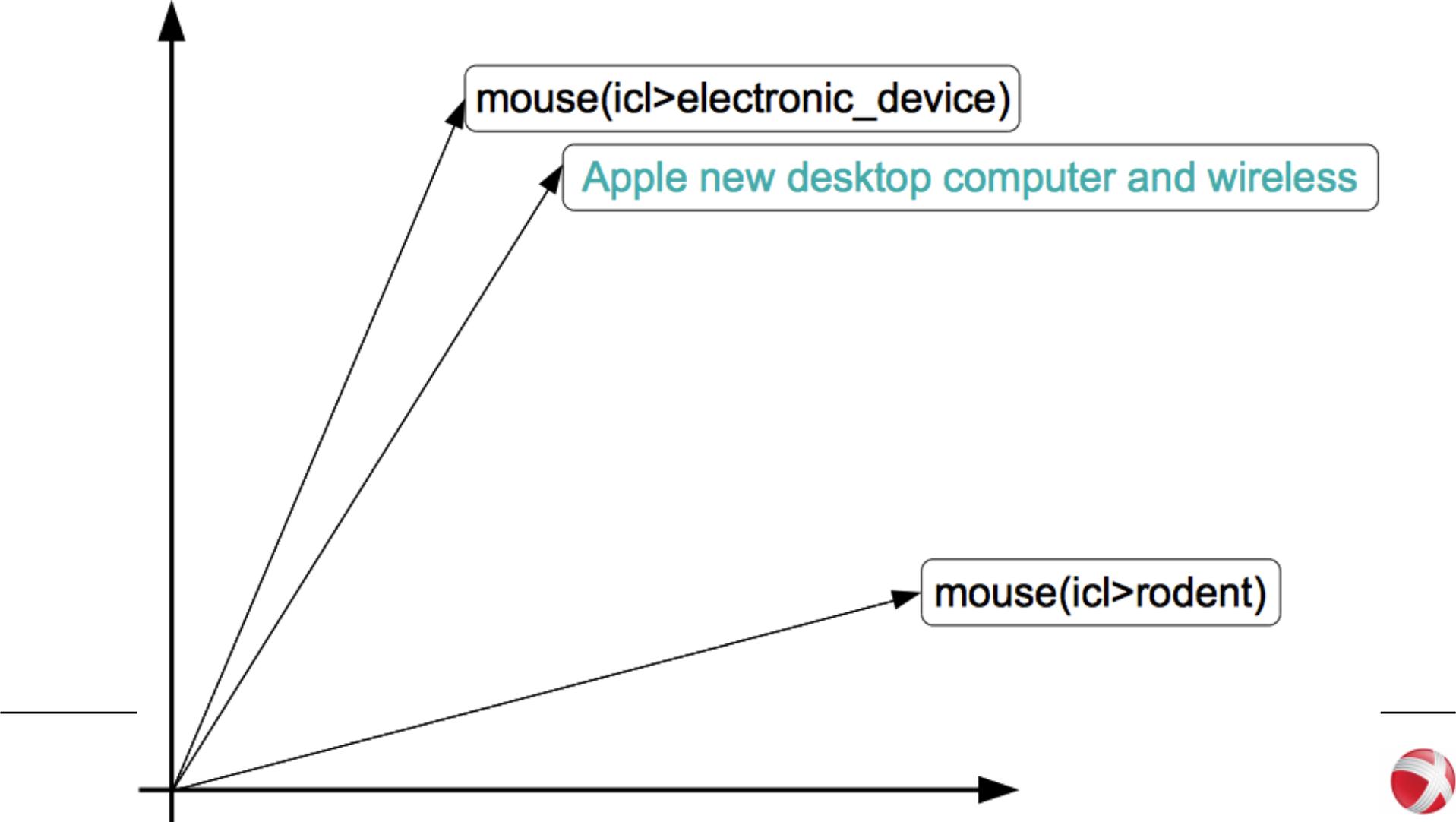


- mouse(icl>electronic_device)

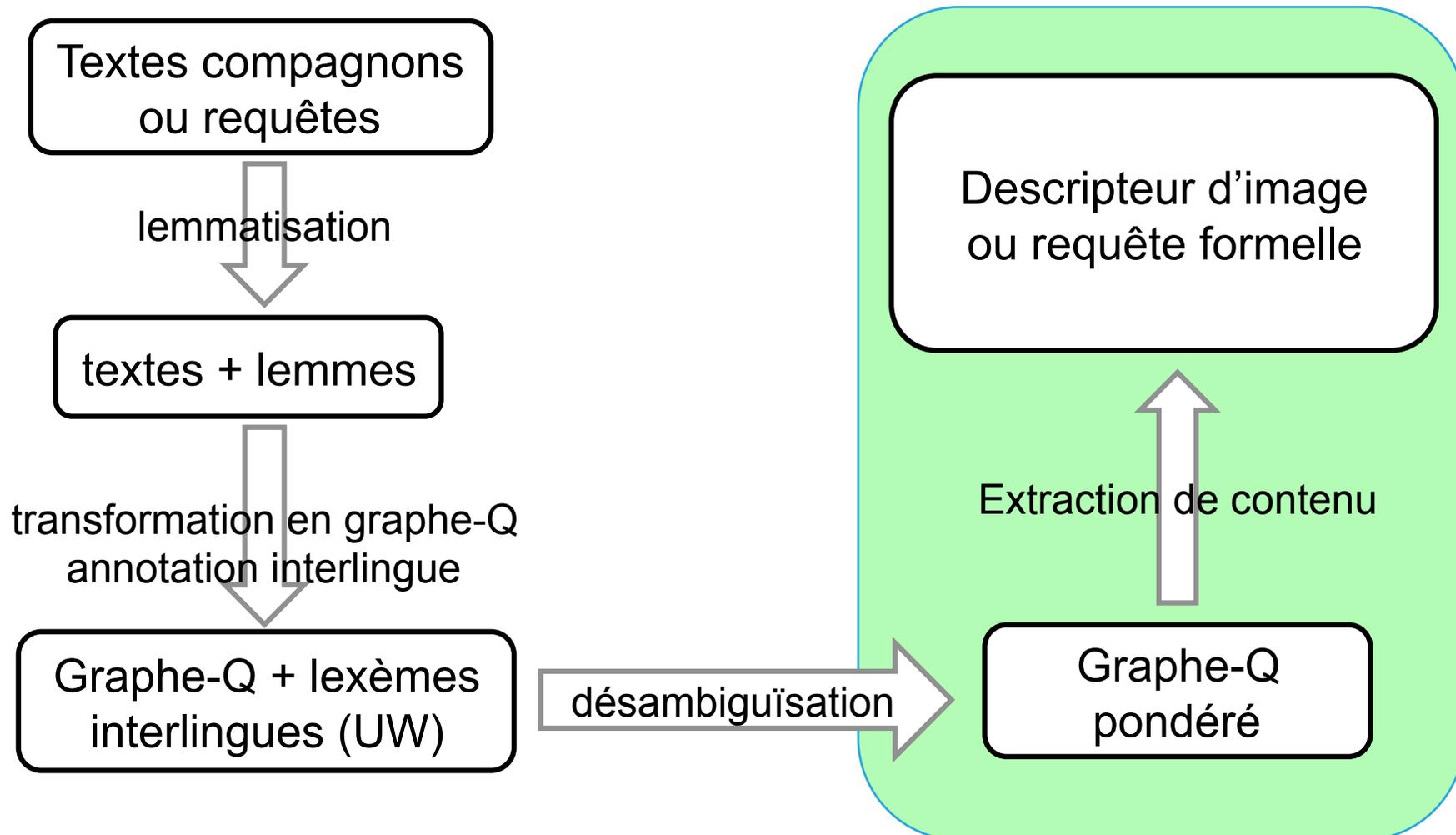


Principe de désambiguïsation

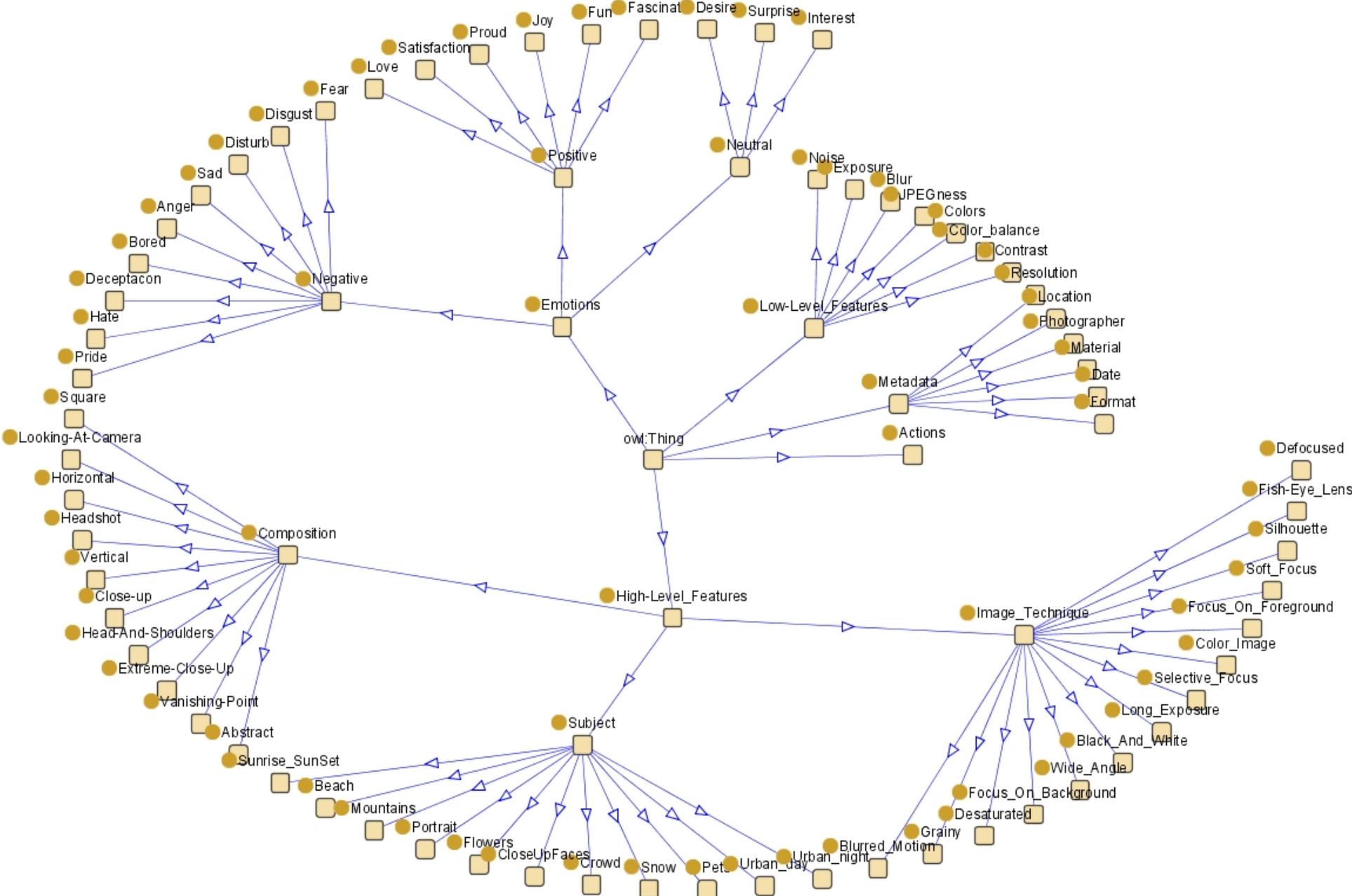
Apple new desktop computer and wireless mouse



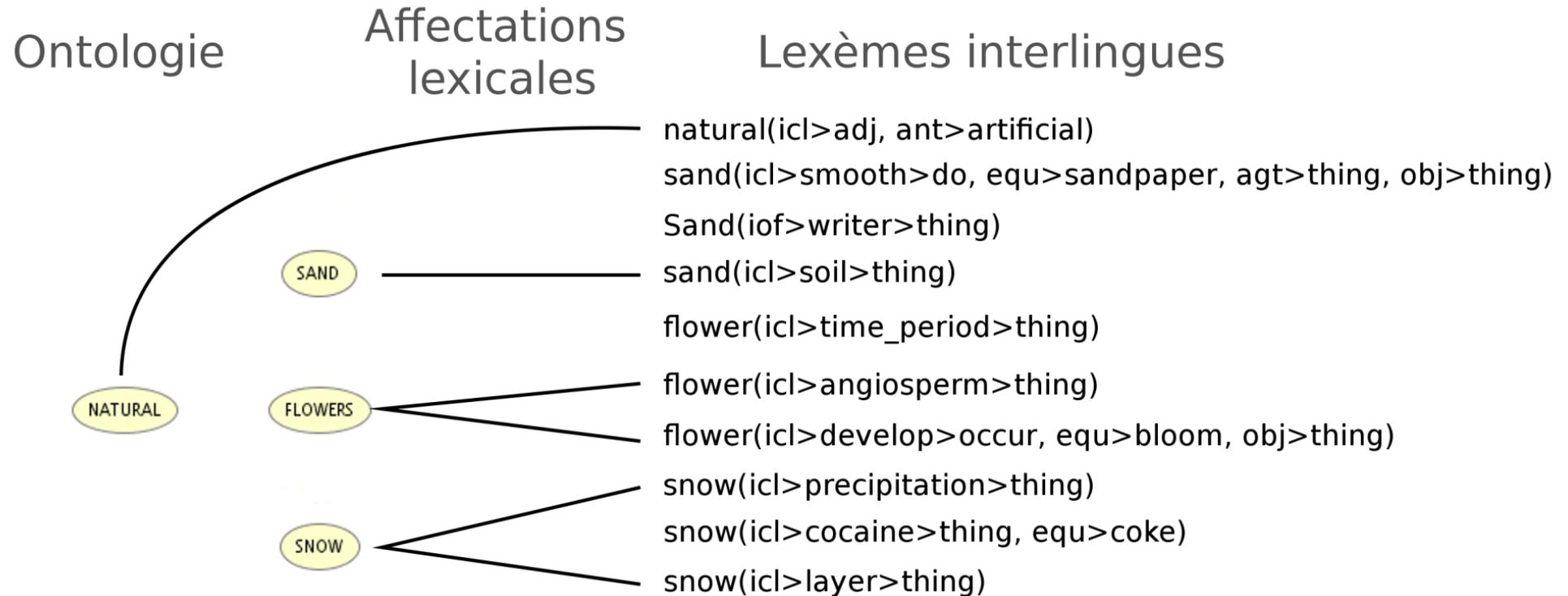
Extraction de contenu guidée par une ontologie



Ontologie pour la classification d'images



Correspondance ontologie ↔ UW



Défi à venir :

fusion des résultats des différentes analyses

Analyse des textes (LIG-GETALP)



Le nain de jardin voyageur du *Fabuleux Destin d'Amélie Poulain*, a lancé une nouvelle mode chez les touristes : faire poser leur mascotte en peluche favorite devant les plus célèbres monuments du monde. Ici, Bob le chien devant le Capitole de Washington DC.

City	0.023
Dog	0.023
Fashion	0.023
Garden	0.023
Horse	0.023
Monument	0.023
People	0.023
Place	0.023
Thing	0.186
Tourist	0.023
Traveler	0.046
Washington	0.023



Sad artistic photo in black and white. A single tree emerge from the fog in a grass field.

Black and White	0.056
Field	0.056
Fog	0.056
Grass	0.056
Marital Status	0.025
Feeling	0.056
Photo	0.056
Picture	0.056
Sad	0.056
Thing	0.444
Tree	0.056
Vegetal	0.112

Résultats issus du GVC (XRCE)



Neutral_Illumination	0,96
Outdoor	0,93
Landscape_Nature	0,9
No_Persons	0,82
Day	0,75
Sky	0,6
No_Blur	0,56
No_Visual_Season	0,52
Aesthetic_Impression	0,41
Partly_Blurred	0,36
Overall_Quality	0,32
Winter	0,32
Fancy	0,26
Water	0,25
Familiy_Friends	0,19
Sunset_Sunrise	0,19
Clouds	0,17



Neutral_Illumination	0,99
Day	0,85
Outdoor	0,85
No_Persons	0,84
No_Blur	0,72
Summer	0,59
Overall_Quality	0,58
Sunny	0,58
Plants	0,5
Fancy	0,37
No_Visual_Season	0,34
Partly_Blurred	0,34
Sky	0,3
Clouds	0,29
Citylife	0,26
Indoor	0,18

No_Visual_Season	0,83
Day	0,8
No_Persons	0,66
Sunny	0,56
Sky	0,47
Fancy	0,38
Overall_Quality	0,37
Single_Person	0,37
Animals	0,24
No_Visual_Time	0,24
Citylife	0,22
Clouds	0,2
Landscape_Nature	0,18
Building_Sights	0,14



Analyse affective (LIRIS)



$M(C1) = 0.5$
 $M(C1 \cup C4) = 0.4$
 $M(\Omega) = 0.1$



$M(C3) = 0.6$
 $M(C2) = 0.2$
 $M(C2 \cup C3) = 0.2$



$M(C2) = 0.95$
 $M(C2 \cup C3) = 0.05$

C1 : positif – actif
C2 : négatif – actif
C3 : négatif – passif
C4 : positif – passif

Les Ci peuvent également représenter des catégories d'émotion

- Tristesse
- Joie
- Colère
- ...

