



**HAL**  
open science

# Génération d'explications textuelles en XAI: le cas de l'annotation sémantique

Jean-Philippe Poli, Wassila Ouerdane, Regis Pierrard

## ► To cite this version:

Jean-Philippe Poli, Wassila Ouerdane, Regis Pierrard. Génération d'explications textuelles en XAI: le cas de l'annotation sémantique. Rencontres francophones sur la logique floue et ses applications, Université de la Sorbonne, Oct 2021, Paris, France. pp.179-186. cea-03406887

**HAL Id: cea-03406887**

**<https://cea.hal.science/cea-03406887>**

Submitted on 28 Oct 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Génération d'explications textuelles en XAI : le cas de l'annotation sémantique

Jean-philippe Poli<sup>1</sup>

Wassila Ouerdane<sup>3</sup>

Régis Pierrard<sup>1</sup>

<sup>1</sup> Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France.

{jean-philippe.poli, regis.pierrard}@cea.fr

<sup>2</sup> Paris-Saclay, CentraleSupélec, MICS, 91190, Gif-sur-Yvette, France

wassila.ouerdane@centralesupelec.fr

## Résumé :

L'annotation sémantique d'images est un domaine d'une importance primordiale dans lequel l'apprentissage profond excelle. Cependant, certains domaines d'application, comme la sécurité ou la médecine, peuvent nécessiter une explication de cette annotation. L'intelligence artificielle explicable est une réponse à ce besoin. Dans ce travail, nous nous intéressons aux explications en langage naturel, dédiées aux utilisateurs humains pour leur fournir des indices sur le processus qui a conduit à l'attribution des étiquettes aux parties de l'image. Nous nous concentrons sur l'annotation sémantique d'images avec la logique floue qui s'est avérée être un cadre utile qui permet de capturer à la fois l'imprécision de la segmentation des images et celle de la connaissance spatiale humaine et du vocabulaire. Dans cet article, nous présentons un algorithme pour la génération d'explications textuelles de l'annotation sémantique de régions d'images.

## Mots-clés :

Explication, génération de langage naturel, annotation sémantique, problèmes de satisfaction de contraintes floues

## Abstract:

Semantic image annotation is an important field in which deep learning excels. However, an explanation of this annotation may be required in some application domains, like security or medicine. In the field of Explainable Artificial Intelligence, we study in particular an explanation which is a sentence in natural language that is dedicated to human users. It should provide them with clues about the process that leads to the decision : the labels assignment to image parts. We focus on semantic image annotation with fuzzy logic. It has proven to be useful when it comes to capture both image segmentation imprecision and the vagueness of human spatial knowledge. In this paper, we present an algorithm that generates textual explanation of the semantic annotation of image regions.

## Keywords:

Explanation, natural language generation, semantic annotation, fuzzy constraint satisfaction problems

## 1 Introduction

L'annotation sémantique d'images est la capacité pour un ordinateur d'étiqueter des images ou des régions d'images. C'est une tâche pri-

mordiale compte tenu de la production quotidienne d'images dans tous les domaines (ex : médecine, surveillance). Dans ce domaine, le deep learning a permis de construire des modèles capables de classer efficacement des images et de reconnaître des objets, parfois même au-delà des performances humaines [14].

Pour certaines applications critiques de l'Intelligence Artificielle (IA), la performance n'est pas le seul critère à optimiser [9] : de telles applications peuvent nécessiter une compréhension relative de la logique exécutée par l'IA. En d'autres termes, l'utilisateur final souhaite obtenir une réponse à la question "Pourquoi ?" [15]

Les problèmes de satisfaction de contraintes (CSP) ont été appliqués avec succès à l'annotation de figures géométriques [21] et à l'étiquetage de régions à partir d'un modèle [8]. Vanegas et al. ont étendu ces travaux aux problèmes de satisfaction de contraintes floues (FCSP) pour impliquer des relations spatiales floues et les ont appliqués avec succès à l'interprétation automatique d'images satellite [20]. Étant donné que les CSP et les FCSP sont des modèles interprétables et que le processus de résolution l'est également, ces approches sont de bonnes candidates pour l'annotation sémantique explicable des images. Dans ce cadre, Pierrard et al. [17] ont proposé des algorithmes pour extraire automatiquement des relations spatiales floues pertinentes à partir de quelques images d'apprentissage dont les régions sont segmentées et étiquetées. Les relations appropriées sont ensuite utilisées pour constituer un FCSP s'il s'agit d'annoter des

zones d'une image ou une base de règles pour classer l'image entière.

Ce travail s'intéresse à la génération d'une explication textuelle pour une annotation textuelle comme proposé par [17]. Étant donnée une solution d'un FCSP et le degré de satisfaction de chacune des contraintes impliquées, nous proposons et évaluons deux algorithmes qui extraient les éléments pertinents pour une explication.

L'article est structuré comme suit. La sect. 2 donne quelques éléments de compréhension sur les relations spatiales, les problèmes à satisfaction de contraintes floues et leur résolution. Ensuite, les sections 3 et 4 présentent respectivement les deux algorithmes de génération d'explications d'une annotation sémantique. Ils sont ensuite analysés et comparés dans la sect. 5.

## 2 Contexte

### 2.1 Relations spatiales floues

Le cadre de la logique floue présente de nombreux avantages pour une représentation des connaissances proche du langage naturel. En particulier, l'aspect linguistique des relations floues peut être utilisé directement dans l'explication [17].

De nombreuses relations spatiales floues ont été étudiées dans la littérature [2]. Par exemple, Vanegas considère trois types de relations spatiales : les relations topologiques, métriques et structurelles [20]. Les deux premiers types sont souvent utilisés en vision par ordinateur. On peut citer par exemple le cadre RCC8 qui définit les relations entre les régions, et leurs homologues flous qui ont été introduits dans [19]. Bloch a introduit un cadre basé sur la morphomathématique floue pour évaluer de telles relations [2]. En particulier, les relations directionnelles et métriques peuvent être exprimées sur la base de l'opérateur de dilatation floue.

Sans perte de généralité, dans la suite de cet article, nous utilisons spécifiquement des rela-

tions directionnelles, de distance et de symétrie. Les relations directionnelles et de distance [2] sont calculées comme un paysage flou et évaluées à l'aide d'une approche d'appariement de motifs flous [5]. La relation de symétrie [6] que nous utilisons recherche l'axe qui maximise une mesure de symétrie entre deux régions.

### 2.2 Problèmes de satisfaction de contraintes floues

Un problème de satisfaction de contraintes (CSP) consiste à affecter des valeurs à un ensemble de variables tout en respectant un ensemble de contraintes. Une extension des CSP à la logique floue afin de prendre en compte des contraintes flexibles est présenté dans [10]. C'est ce qu'on appelle un problème de satisfaction de contraintes floues (FCSP). Un FCSP est défini par :

- Un ensemble de variables  $X = \{x_1, \dots, x_n\}$ .
- Un ensemble de domaines  $D = \{D_1, \dots, D_n\}$  tel que  $D_i$  est la plage de valeurs qui peuvent être affectées à  $x_i$ .
- Un ensemble de contraintes flexibles  $C = \{c_1, \dots, c_p\}$ . Chaque contrainte  $c_k$  est définie par une relation floue  $R_k$  et par l'ensemble des variables  $V_k$  qui y sont impliquées.

L'algorithme de backtracking [10] est utilisé pour résoudre un FCSP. Il démarre par un ensemble vide d'affectations et sélectionne une variable  $x \in X$  à affecter. Ensuite, il trouve une valeur dans le domaine de  $X$  qui maintient la cohérence de l'instanciation courante par rapport à l'ensemble des contraintes  $C$ . Cette étape est répétée jusqu'à ce que toutes les variables soient affectées. Lorsqu'une variable  $x$  n'a plus de valeur à tester, l'algorithme revient en arrière et essaie la valeur suivante de la variable précédemment affectée.

Une affectation cohérente et complète est une solution. Une solution  $\gamma$  du FCSP est évaluée par son degré de cohérence [20], représenté par :

$$\text{cons}(\gamma) = \min_{c_k \in C} \mu_{R_k}(\gamma|_{V_k}) \quad (1)$$

où  $\gamma|_{V_k}$  est la projection de  $\gamma$  sur  $V_k$  et  $\mu_{R_k}$  la fonction d'appartenance représentant  $R_k$ .

Ce degré de cohérence permet également de comparer différentes solutions afin d'en extraire la meilleure.

### 2.3 Annotation d'images avec FCSP

Lorsqu'il s'agit d'annotation d'images, l'ensemble des variables  $X$  correspond aux objets que nous souhaitons annoter. Les variables partagent le même domaine  $D$  qui représente les régions de l'image, obtenues par segmentation de cette dernière. Ainsi,  $|X| \leq |D|$ . De plus, les contraintes dans  $C$  sont définies par des relations floues. Certaines d'entre elles peuvent également traiter des groupes d'objets [20]. Ainsi, il est possible de résoudre des problèmes d'annotation spécifiques dans lesquels les objets à annoter et les étiquettes sont connus (même s'ils sont détectés automatiquement). L'intuition derrière est qu'un tel problème d'annotation peut être combinatoire et les étiquettes sont affectées en conséquence les unes par rapport aux autres, par opposition aux méthodes classiques qui procèdent de manière plus individuelle.

Pierrard et al [17] appliquent cette approche à l'annotation d'organes dans des images médicales, en mettant l'accent sur la génération automatique du FCSP à partir de quelques données. Dans la suite de cet article, nous prendrons ce travail comme illustration avec un FCSP généré automatiquement.

Pour générer nos explications, nos algorithmes, Alg. 1 et Alg. 2, prennent en entrée une trace  $T = \langle P, s, \bar{C} \rangle$  de l'exécution des algorithmes de résolution.  $T$  est composée de :

- $P = \langle X, D, C \rangle$  est un FCSP.
- $s$ , une solution choisie parmi toutes les solutions de  $P$ , par exemple celle dont le degré de cohérence est maximal.  $s$  contient l'affectation de chaque variable dans  $X$ .
- $\bar{C}$ , l'ensemble des degrés de satisfaction de

chaque contrainte  $c \in C$ .

## 3 Génération d'explications textuelles complètes

Dans ce qui suit, nous présentons un premier algorithme pour la génération d'une explication en langage naturel, issu des travaux de [17].

### 3.1 Algorithme

L'Alg.1 transforme toutes les contraintes du FCSP en phrases en utilisant un réalisateur de surface (dans notre cas, SimpleNLG [11]). Le vocabulaire, i.e. l'ensemble des relations connues par le système, est constitué de : à gauche de, à droite de, au-dessous de, au-dessus de, près de, symétrique à et de forme allongée. Notons  $\bar{c}_x$  le complément de  $x$  dans le scope de  $c$ . Le modérateur est sélectionné parmi ceux cités dans le Tableau 1 en fonction du degré de satisfaction de  $c$ , comme dans les travaux de Budescu [4].

Modérateur	Degré de satisfaction
very high	de 0.9 à 1
high	de 0.7 à 0.9
average	de 0.4 à 0.7
moderate	de 0.2 à 0.4
low	de 0 à 0.2

Tableau 1 – Échelle de confiance simplifiée

### 3.2 Résultats

Dans ce travail, le FCSP est construit automatiquement, par apprentissage à partir de quelques images du jeu de données Visceral<sup>1</sup>. La figure 1 montre une des images et les différents organes.

La segmentation a été obtenue automatiquement et les régions ont été numérotées au fur et à mesure de leur identification, dans un ordre non maîtrisé (i.e. les différentes régions ne sont pas ordonnées). Cependant, afin de rendre plus compréhensible l'article, nous avons attribué un numéro aux régions, de gauche à droite, de haut en bas.

1. <http://www.visceral.eu/>

---

### Algorithm 1: Génération d'une explication complète

---

**Input:** la trace  $T = \langle P, s, \bar{C} \rangle$   
**Output:** Une explication textuelle complète

```

1 foreach variable non traitée  $x \in X$  do
2    $v \leftarrow$  valeur de  $x$  dans  $s$ 
3   Créer une phrase de la forme : "Region  $v$  is
   annotated as  $x$  with a moderator
   confidence because :"
```

4 **foreach** contrainte  $c \in C$  concernant la  
variable  $x$  **do**

5 **if**  $x$  est la première variable du scope  
de  $c$  **then**

6 | Générer une phrase de la forme :  
6 | "it is  $c \bar{c}_x$ "

7 **else**

8 | Générer une phrase de la forme :  
8 | " $\bar{c}_x$  is/are  $c x$ "

9 **end**

10 **end**

11 **end**

---

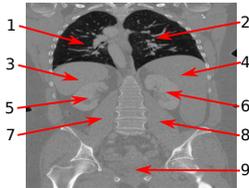


Figure 1 – Image IRM prise de dos avec les différentes régions à annoter

On considère la solution d'un tel FCSP ayant le degré maximal de consistance (autrement dit, de satisfiabilité). Le résultat de l'algorithme, est une explication longue, mais complète, de l'annotation (Figure 2).

Dans la section suivante, nous proposons de fournir des explications plus concises.

## 4 Génération d'explications textuelles concises

### 4.1 Considérations relatives aux sciences cognitives

Les sciences cognitives ont largement étudié la façon dont les humains représentent une scène ou balaiant des images. Il est donc naturel de considérer ces idées pour créer une explication.

Zwaan et *al.* présentent plus d'une décennie

Region 1 is annotated as the left lung with a high confidence **because** :

- it is *completely to the left of* region 2 (annotated as the right lung by the model),
- region 2 (right lung) is *completely to the right of* region 1,
- it is *above* region 3 (spleen),
- region 3 (spleen) is *completely below* region 1,
- it is *above* region 7 (left psoas),
- region 7 (left psoas) is *completely below* region 1,
- region 5 (left kidney) is *completely below* region 1.

Region 2 is annotated as the right lung with a very high confidence **because** :

- it is *completely to the right of* region 1 (left lung),
- region 1 (left lung) is *completely to the left of* region 2,
- region 3 (spleen) is *to the left of* region 2,
- region 4 (liver) is *below* region 2,
- it is *above* region 8 (right psoas),
- region 8 (right psoas) is *completely below* region 2,
- region 6 (right kidney) is *completely below* region 2,
- region 9 (bladder) is *below* region 2.

Figure 2 – Extrait d'explication d'une annotation avec l'approche complète. La version complète se trouve [17].

d'études sur le modèle de situation (représentation mentale) [22]. Ils mettent en évidence la difficulté de décrire correctement une scène spatiale avec le langage. Par exemple, si l'on décrit une pièce de manière circulaire, le premier et le dernier objets sont éloignés l'un de l'autre dans la description mais proches dans la pièce. Ils soulignent également l'importance de l'ordre dans lequel les parties de la scène doivent être décrites.

Cela nous amène aux études du balayage visuel d'images. L'attention des sujets est classiquement attirée par les points saillants. Les auteurs de [3] donnent un aperçu complet de leur détection automatique. Néanmoins, les sciences cognitives mettent en garde contre la difficulté de définir la saillance car elle peut être

dépendante du contexte, ou due à la singularité d'un objet, du but de l'utilisateur, etc. Les auteurs de [12] affirment que les sujets ont besoin de plus en plus de temps pour visualiser mentalement l'information au fur et à mesure qu'ils s'éloignent du point saillant de l'image. D'autres travaux étudient les difficultés des sujets à représenter une scène si la description est trop longue et si la description est trop précise [7]. Dans [18], les auteurs soulignent que le sens de balayage affecte la description d'une scène. Enfin, lorsqu'un même sujet regarde la même image, les chemins de balayage peuvent être différents [16] : ainsi, le chemin de balayage ne dépend pas seulement des objets dans l'image. Si plusieurs images similaires sont présentées, le chemin de numérisation peut également être de plus en plus efficace [16].

Finalement, les psychologues de la Gestalt [13] ont étudié les enjeux cognitifs de la perception visuelle, en particulier la forme des objets. Ils recommandent notamment de regrouper des objets similaires ou partageant des propriétés similaires.

Ce bref aperçu des sciences cognitives nous a aidé à concevoir notre stratégie d'explication.

## 4.2 Principes

Notre idée est d'améliorer le premier algorithme en prenant en compte les observations des sciences cognitives. Nous observons donc les principes suivants.

**Tri.** L'ordre des résultats a une importance : il est préférable de commencer par les régions des images qui sont saillantes, puis, d'utiliser des diagonales et des distances croissantes pour sélectionner les objets suivants.

**Saillance.** Il s'agit d'un concept difficile qui peut dépendre du contexte. A minima, le plus gros objet ou groupe d'objets peut être choisi comme point focal.

**Symétrie.** Un couple d'objets symétriques doit être traité ensemble.

**Priorité.** Il faut d'abord sélectionner les

contraintes les plus satisfaites.

**Associativité.** Il faut exploiter les propriétés des relations (par exemple "à gauche de") pour alléger les explications.

**Localité.** Si possible, nous utiliserons d'abord les contraintes avec les régions les plus proches dans l'image.

De plus, une explication doit d'une manière ou d'une autre indiquer comment la tâche a été accomplie. Dans notre cas, la résolution d'un FCSP est assez simple à expliquer puisque l'algorithme recherche les valeurs des variables telles que les contraintes sont satisfaites. Cependant, cela rend l'explication plus compliquée lorsque les contraintes ne sont pas toutes unaires, puisque ces affectations sont inter-dépendantes. Par exemple, une contrainte binaire forcera l'affectation de deux variables ensemble.

Enfin, nous nous limitons à un nombre maximal  $N_{\max}$  de contraintes pour chaque variable, choisies de sorte que la corrélation entre les contraintes sélectionnées soit minimale : par exemple, les valeurs de "à gauche de" et "très à gauche de" peuvent être corrélées et nous ne voulons donc pas les utiliser en même temps pour la même variable car elles sont redondantes. Nous utilisons l'information mutuelle comme critère pour détecter cette corrélation.

Dans ce qui suit, nous introduisons un algorithme qui considère ces différents principes.

## 4.3 Algorithme

L'explication produite par l'Alg 2 commence par une phrase générale qui indique la confiance globale concernant l'annotation en fonction du degré de cohérence de la solution (ligne 1). L'algorithme sélectionne ensuite la région de la segmentation qui est la plus saillante (ligne 2). Concernant cet objet, l'image est divisée en quatre quadrants. L'explication commencera par la région la plus saillante, puis par les autres objets du même quadrant, puis quadrant par quadrant, dans le sens des aiguilles d'une

montre. Cet ordre est matérialisé dans un ensemble ordonné  $X'$  (lignes 3-4).

---

**Algorithm 2:** Génération d'explications concises

---

**Input:** la trace  $T = \langle P, s, \bar{C} \rangle$   
**Output:** une explication textuelle concise

- 1 Écrire une phrase pour introduire le résultat et la confiance globale
- 2 Sélectionner  $f$  qui est le point focal dans l'image
- 3 A partir du centre de  $f$ , diviser l'image en 4 quadrants  $Q_1, \dots, Q_4$
- 4  $X' =$  ensemble de variables  $x \in s$  trié par quadrant
- 5 **while**  $X' \neq \emptyset$  **do**
- 6      $x \leftarrow \text{pop}(X')$
- 7      $S \leftarrow$  Sélectionner  $N_{max}$  contraintes  $c_i$  qui ne sont pas liées dans le graphe de connaissances et avec des degrés de satisfaction maximaux
- 8     Écrire la phrase “ $x$  est  $c_1, \dots$ , et est”
- 9     **if**  $x$  implique une contrainte de regroupement  $c$  **then**
- 10         Déplacer toutes les variables dans la portée de  $c$  au début de  $X'$
- 11     **end**
- 12 **end**

---

Pour chaque variable dans  $X'$ , l'algorithme doit sélectionner au plus  $N_{max}$  contraintes pour justifier l'explication. Les contraintes sont choisies en fonction non seulement de leur niveau de satisfaction (qui doit être le plus élevé possible pour ne pas surcharger le texte de modérateurs), mais aussi de leur lien mutuel et de la proximité avec les autres variables (lignes 5-12).

Le lien mutuel entre les relations est une partie délicate. Nous utilisons un graphe de connaissances sur les relations tel que proposé dans [17]. Un tel graphe met en évidence différents liens entre deux relations  $r_1$  et  $r_2$ , comme  $r_1 \Rightarrow r_2$ ,  $\neg r_1 \Rightarrow r_2$ , mais également la symétrie. Cette dernière est particulièrement importante pour éviter d'utiliser deux fois la même contrainte. Soit  $o_1$  et  $o_2$  deux objets dans l'image, et  $r$  une relation symétrique, si  $o_1 r o_2$  est utilisée dans une phrase, nous ne pouvons plus utiliser  $o_2 r o_1$ . Ensuite, l'algorithme recherche des contraintes de regroupement telles que “est symétrique à” qui constitue une paire de variables (ligne 9). En effet,

“This is the annotation of the given image (with a very high confidence). The right lung (region 2) is *symmetrical to* the left lung (region 1) and *above* the liver (region 4). The liver (region 4) is *at the right of* the right kidney (region 6) and *at the right of* the right psoas (region 8). The right psoas (region 8) is *above of* the bladder (region 9) and is *symmetrical to* the left psoas (region 7). The left psoas (region 7) is *below* the left kidney (region 5). The spleen (region 3) is *above* the left kidney (region 5) and is *below* the left lung (region 1).”

Figure 3 – Explication concise produite par Alg.2

la section précédente met en évidence que les groupes d'objets doivent être traités ensemble. Ainsi, les autres variables dans la portée de cette contrainte doivent être traitées juste après (ligne 10).

#### 4.4 Résultats

Dans ce travail, nous définissons le point focal comme le plus grand objet (en termes de surface). Nous définissons  $N_{max} = 2$ . Pour le même exemple (voir Figure 1), et la même solution  $s$ , le résultat est montré dans Figure 3.

La plupart des contraintes sont liées dans le graphe de connaissances, car nous avons utilisé principalement des relations directionnelles comme “à droite de” et “à gauche de”. Cela explique pourquoi nous atteignons rarement le nombre de contraintes  $N_{max}$ .

Le résultat est évidemment plus court et semble plus facile à lire. Le quadrant impose un ordre pour la description de chaque organe. L'explication semble moins redondante grâce à la sélection des contraintes.

### 5 Évaluation et discussion

Pour comparer les deux approches, nous avons utilisé le questionnaire présenté dans [1] : il est basé sur 17 questions organisées en 3 catégories : langage naturel, interaction homme-machine et contenu et forme. Chaque question est évaluée avec une échelle de Likert (de 1 “pas d'accord” à 5 “d'accord”). Notre panel est composé de 40 répondants, dont

20 membres du personnel médical (médecins, chirurgiens, infirmières, radiologues), l'autre moitié étant des informaticiens (6) et divers autres professionnels non médicaux (14). Afin de diminuer le temps que le personnel médical consacre au questionnaire, nous n'avons retenu que 12 questions sur les 17 initiales qui permettront de comparer les deux approches. Nous avons supprimé les questions sur la grammaire et celle qui indique si l'explication a fait changer d'avis un répondant. En raison du manque de place, la figure 4 met en évidence les réponses à quelques questions seulement.

Les deux explications sont comparables en termes de correction syntaxique (87% pour l'approche 1 et 95% pour l'approche 2), de compréhension du raisonnement (67,5% d'accord pour l'approche 1, 60% pour l'approche 2), et de la communication de l'incertitude (62,2% pour l'approche 1, 65% pour l'approche 2). La "compréhension du raisonnement" indique si les répondants peuvent déduire le processus de raisonnement lorsqu'ils lisent l'explication. Le critère de "communication d'incertitude" évalue la capacité de l'explication à dire à l'utilisateur à quel moment la décision est fiable. Dans notre cas, cela est obtenu par la traduction de la satisfaction des contraintes en parties de phrase comme "avec une confiance très élevée". Ces faits montrent que toutes les personnes n'ont pas compris comment l'algorithme annote les organes et pourquoi l'algorithme n'était pas sûr dans tous les cas.

Pour toutes les autres comparaisons, la deuxième approche surpasse la première approche. 19 personnes ont trouvé que la première explication était trop longue alors que seulement 1 répondant était préoccupé par la longueur de la deuxième explication. Les répondants ont trouvé la première explication répétitive (87,5%) et difficile à lire (72,5%), alors que seulement respectivement 22,5% et 10% du panel sont d'accord avec ces faits pour la seconde. Seuls 32,5% des répondants ont trouvé l'ordre des éléments dans l'explication

approprié pour l'explication 1 contre 72,5% pour la deuxième explication.

Les deux explications donnent confiance aux répondants vis à vis de l'étiquetage automatique (55% pour la première approche et 65% pour la seconde).

Ces résultats confirment les avantages du deuxième algorithme.

Il est important de noter que ces algorithmes ne sont pas spécifiques à un domaine. En effet, les relations sont génériques et elles pourraient être utilisées dans un autre domaine (comme l'annotation d'images satellite). De plus, ces algorithmes manipulent des régions d'images et n'ont aucune idée qu'elles représentent des organes. Cependant, les étiquettes qui sont utilisées sont des noms d'organes, car nous voulons une annotation sémantique. Nous n'utilisons pas de connaissance du domaine, par exemple pour remplacer le mot "région" par "organe" dans l'explication, ou pour utiliser un vocabulaire plus technique.

Les résultats montrent que l'ordre des éléments dans une explication est important pour les utilisateurs finaux. La concision semble également être un critère primordial.

#### Remerciements :

Nous remercions le panel pour sa participation, en particulier le personnel médical pour son temps en pleine crise sanitaire. Ce travail a été en partie financé par le programme d'innovation et de recherche Horizon 2020 de l'Union Européenne, dans le cadre du projet DeepHealth No 825111.

#### Références

- [1] I. Baaj and J. Poli. Natural language generation of explanations of fuzzy inference decisions. In *2019 IEEE International Conference on Fuzzy Systems*, pages 1–6, 2019.
- [2] I. Bloch. *Fuzzy Models of Spatial Relations, Application to Spatial Reasoning*, pages 51–58. Springer Berlin Heidelberg, 2013.
- [3] A. Borji, M.-M. Cheng, H. Jiang, and J. Li. Salient object detection : A survey. *Computational Visual Media*, 5 :117–150, 2014.
- [4] D. V. Budescu, H.-H. Por, and S. B. Broomell. Effective communication of uncertainty in the IPCC reports. *Climatic change*, 113(2) :181–200, 2012.

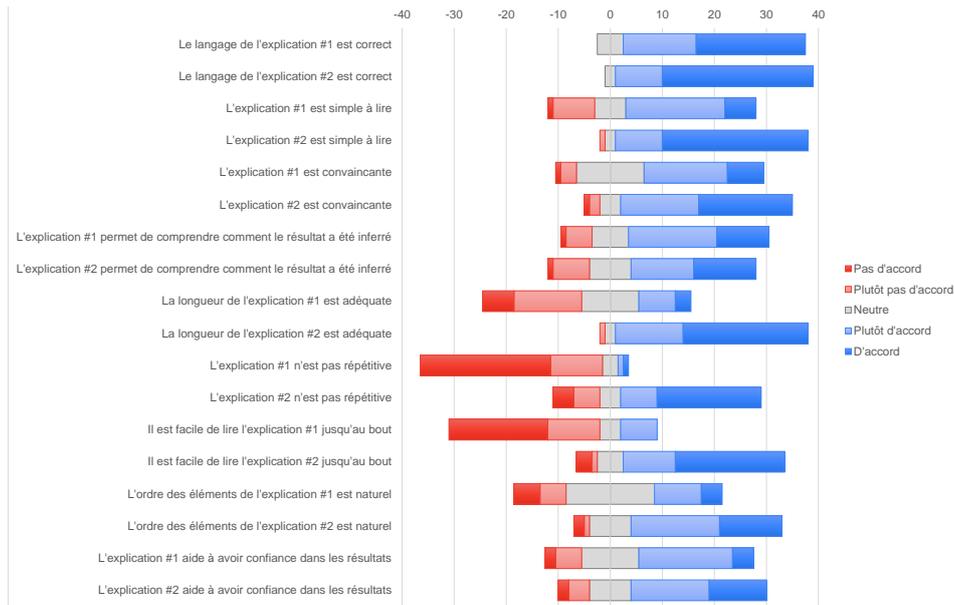


Figure 4 – Éléments marquants du sondage

- [5] M. Cayrol, H. Farreny, and H. Prade. Fuzzy pattern matching. *Kybernetes*, 11(2) :103–116, 1982.
- [6] O. Colliot. *Représentation, évaluation et utilisation de relations spatiales pour l'interprétation d'images. Application à la reconnaissance de structures anatomiques en imagerie médicale*. PhD thesis, Télécom ParisTech, 2003.
- [7] M. Denis, M.-R. Goncalves, and D. Memmi. Mental scanning of visual images generated from verbal descriptions : Towards a model of image accuracy. *Neuropsychologia*, 33(11) :1511 – 1530, 1995.
- [8] A. Deruyver, Y. Hodé, and L. Brun. Image interpretation with a conceptual graph : Labeling over-segmented images and detection of unexpected objects. *Artificial Intelligence*, 173(14) :1245 – 1265, 2009.
- [9] F. Doshi-Velez and B. Kim. Towards a rigorous science of interpretable machine learning. In *eprint arXiv :1702.08608*, 2017.
- [10] D. Dubois, H. Fargier, and H. Prade. Possibility theory in constraint satisfaction problems : Handling priority, preference and uncertainty. *Applied Intelligence*, 6(4) :287–309, 1996.
- [11] A. Gatt and E. Reiter. Simplenlg : A realisation engine for practical applications. In *Proceedings of the 12th European Workshop on Natural Language Generation*, pages 90–93, 2009.
- [12] S. Kosslyn, T. Ball, and B. Reiser. Visual images preserve metric spatial information : Evidence from studies of image scanning. *Journal of Experimental Psychology : Human Perception and Performance*, 4(1) :47–60, 1978.
- [13] R. Luccio. *Gestalt problems in cognitive psychology : Field theory, invariance and auto-organisation*, pages 1–19. Springer Berlin Heidelberg, 1993.
- [14] S. M. McKinney, M. Sieniek, V. Godbole, J. Godwin, N. Antropova, H. Ashrafian, T. Back, M. Chesus, G. C. Corrado, A. Darzi, M. Etemadi, F. Garcia-Vicente, F. J. Gilbert, M. Halling-Brown, D. Hassabis, S. Jansen, A. Karthikesalingam, C. J. Kelly, D. King, J. R. Ledsam, D. Melnick, H. Mostofi, L. Peng, J. J. Reicher, B. Romera-Paredes, R. Sidebottom, M. Suleyman, D. Tse, K. C. Young, J. D. Fauw, and S. Shetty. International evaluation of an AI system for breast cancer screening. *Nature*, 577(7788) :89–94, 2020.
- [15] T. Miller. Explanation in artificial intelligence : Insights from the social sciences. *Artificial Intelligence*, 2679 :1–38, 2019.
- [16] D. Noton and L. Stark. Scanpaths in eye movements during pattern perception. *Science*, 171(3968) :308–311, 1971.
- [17] R. Pierrard, J.-P. Poli, and C. Hudelot. Spatial relation learning for explainable image classification and annotation in critical applications. *Artificial Intelligence*, 292 :103434, 2021.
- [18] A. Román, A. Fathi, and J. Santiago. Spatial biases in understanding descriptions of static scenes : The role of reading and writing direction. *Memory & cognition*, 41, 2013.
- [19] S. Schockaert, M. D. Cock, C. Cornelis, and E. E. Kerre. Fuzzy region connection calculus : Representing vague topological information. *International Journal of Approximate Reasoning*, 48(1) :314 – 331, 2008.
- [20] M. C. Vanegas Orozco. *Spatial relations and spatial reasoning for the interpretation of Earth observation images using a structural model*. Theses, Télécom ParisTech, 2011.
- [21] D. Waltz. Understanding line drawings of scenes with shadows. In *The Psychology of Computer Vision*. McGraw-Hill, 1975.
- [22] R. A. Zwaan and G. A. Radvansky. Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2) :162, 1998.