



**WORD2VEC & SEMVUE**

**DÉTECTION ET VISUALISATION**

**DE TERMES PERTINENTS DE MAINTENANCE**

**EDF R&D**

**Laurent Pierre, Meryl Bothua**

**12 décembre 2017**



# SOMMAIRE

**1. CONTEXTE ET OBJECTIF**

**2. MÉTHODE WORD2VEC ET APPORT**

**3. RÉSULTATS OBTENUS : DÉMONSTRATION SEMVUE**

**4. PERSPECTIVES**

# 1. CONTEXTE ET OBJECTIF



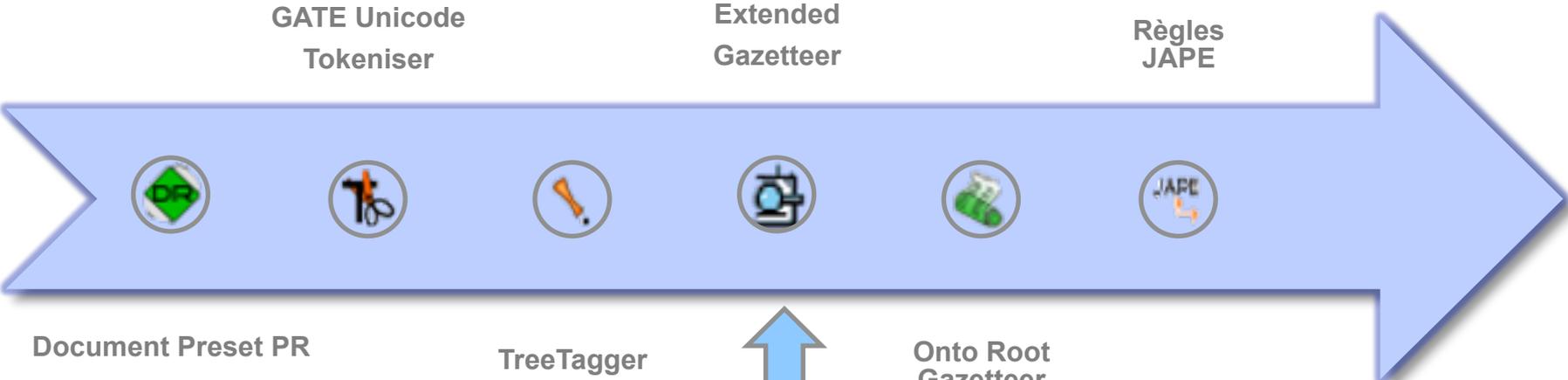
## Contexte général

- Programme de maintenance préventive
- Besoin d'extraction de connaissance sur des comptes rendus
- Instruire ce que la fouille de texte peut apporter

## Objectifs

- Aide à la création de ressources lexicales
- Identification automatique de termes similaires

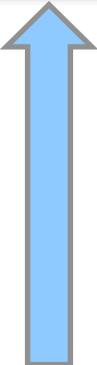
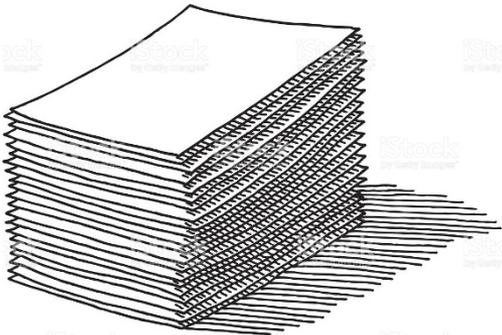
# APPORT DE W2V À UNE CHAÎNE GATE



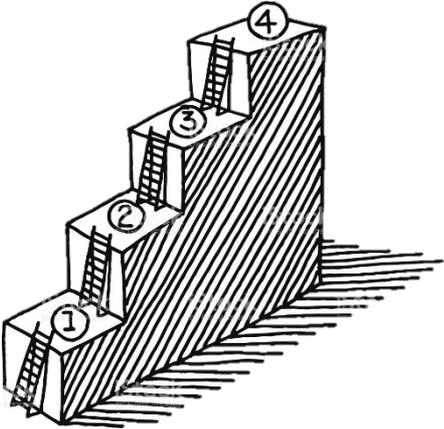
Document Preset PR

TreeTagger

Onto Root Gazetteer



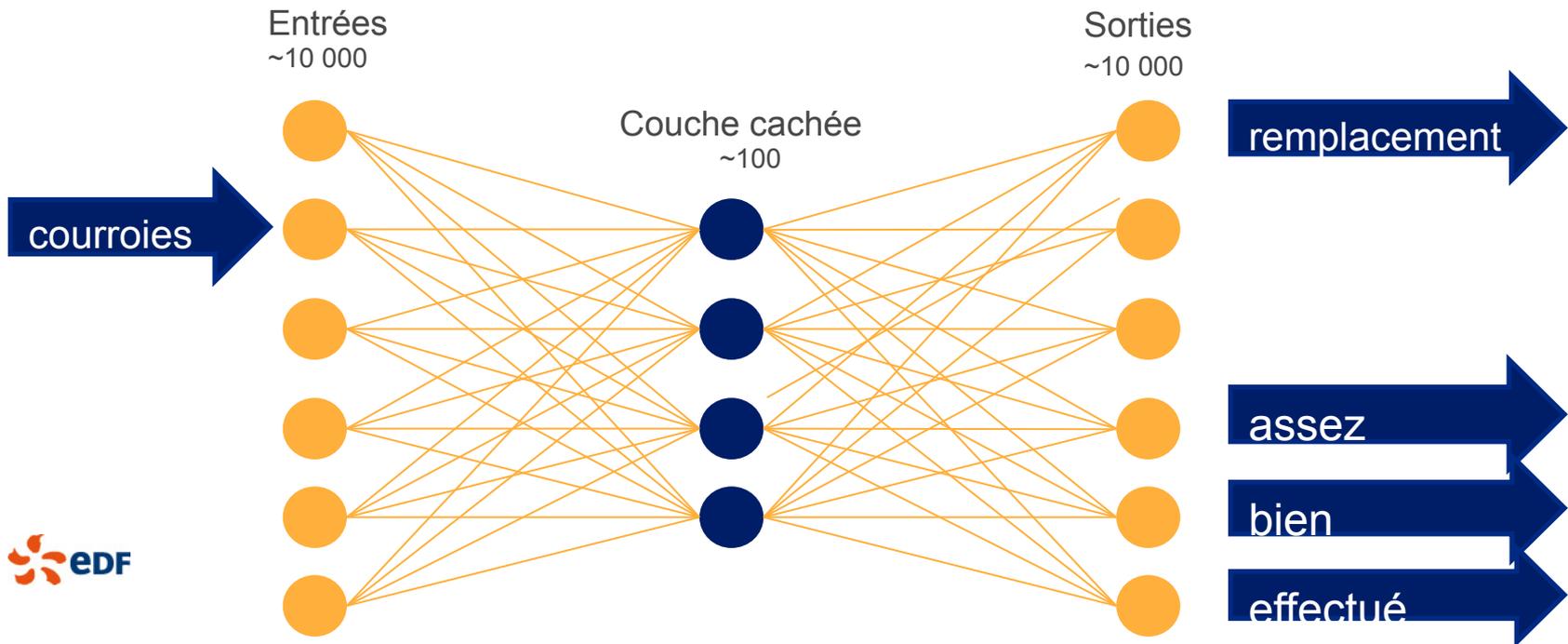
Word2Vec



# 2. LA MÉTHODE WORD2VEC

## Principe de l'algorithme

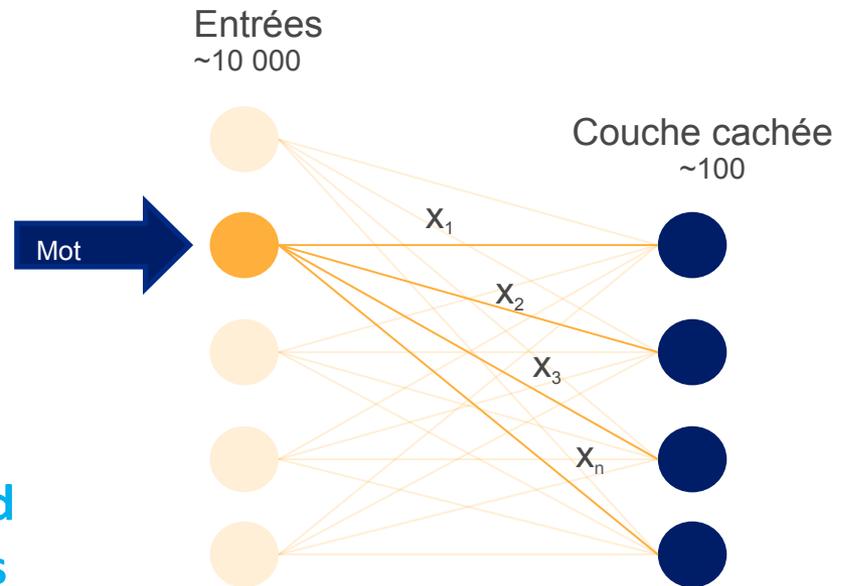
- Recherche à entrainer un réseau de neurones pour **prévoir le contexte** d'utilisation d'un mot
- Méthode non supervisée
- Nécessite des corpus plus ou moins importants
- Exemple avec : « ... remplacement **courroies** assez bien effectué ... »



# 2. LA MÉTHODE WORD2VEC

## ✓ Intérêt

- Permet de transformer un **mot** en **vecteur** (couche cachée)
- Représentation **continue**
- Ici, « mot » est transformé en un **vecteur**  $x_1, x_2, \dots, x_n$
- Méthode de plongement (« **word embeddings** ») ou **vecteurs mots**
- Permet de faire des sommes de mots, soustraction de mots, moyennes de mots...



## 2. PARAMÉTRAGE RETENU

- Approche plus pertinente : **CBOW**
- Dimension des vecteurs : **100**
- Fenêtre : **3**
- Nombre d'itérations : **800**
- Fréquence minimale des mots : **1**
- Prétraitement : **stopwords, lexiques**

# 3. RÉSULTATS OBTENUS

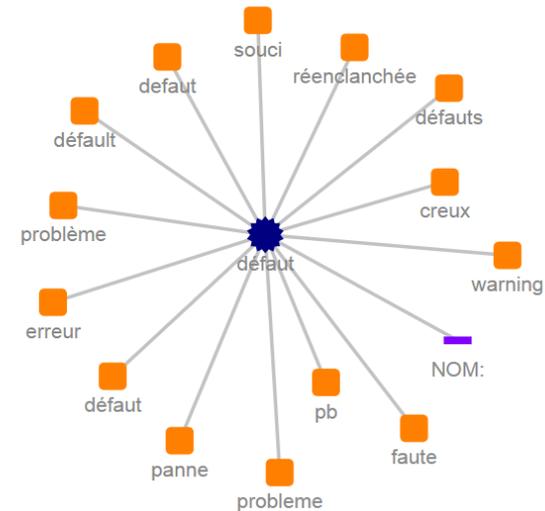
Détection automatique de :

- Synonymes ✓
  - Abréviations ✓
  - Fautes de frappe ou d' orthographe ✓
  - Multilinguisme ✓
- Taille du dictionnaire : ~ 5000 mots après prétraitements



# 3. VISUALISATION ET NAVIGATION DANS UN GRAPHE RDF : SEMVUE

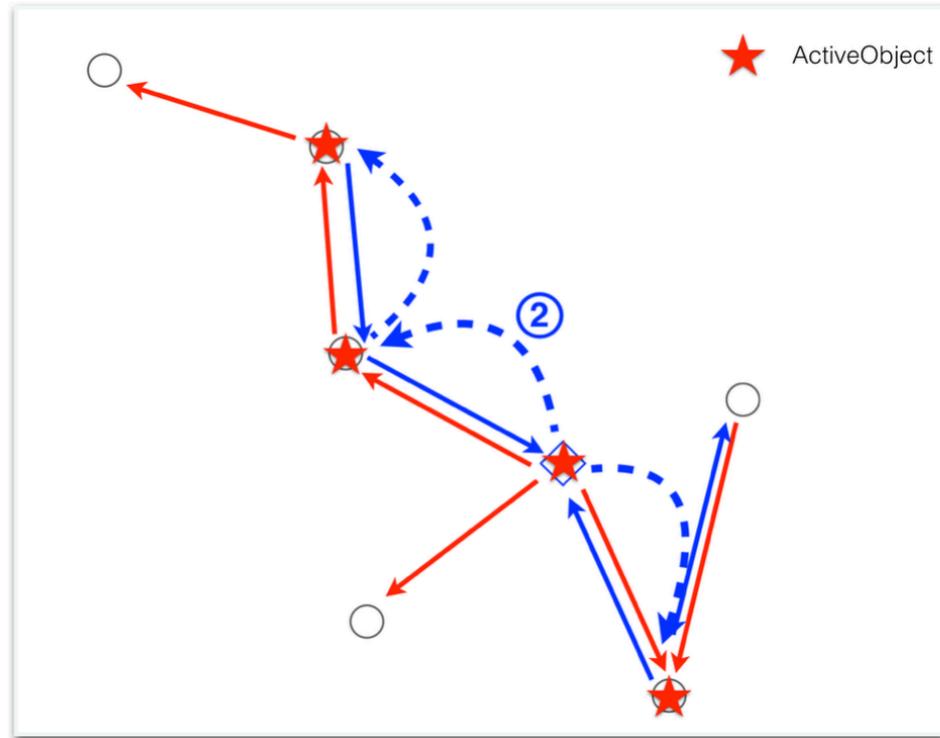
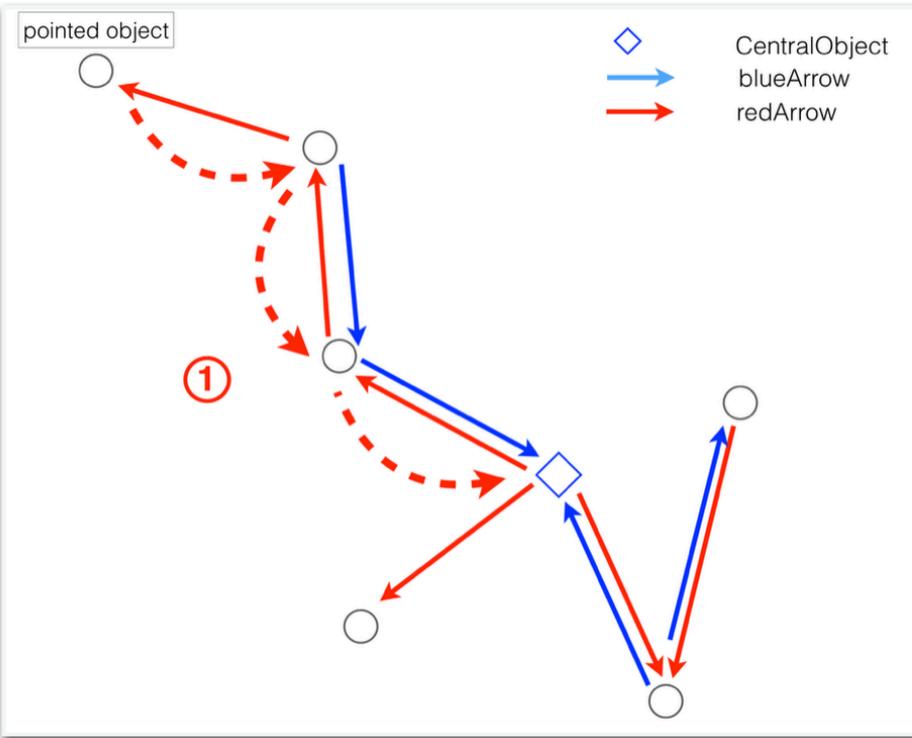
- **Afficher des types d'informations ciblées**
- **Déléguer à l'iconographie le typage**
- **Définir un périmètre minimal et suffisant autour des objets**
- **Elaguer le graphe**
- **Requêter et filtrer**



# 3. VISUALISATION SEMVUE

**Moyens :**

- **Description formelle de la politique de visualisation**
  - **Système de raisonnement**
- **Implémentation d'une ontologie dans un entrepôt RDF**



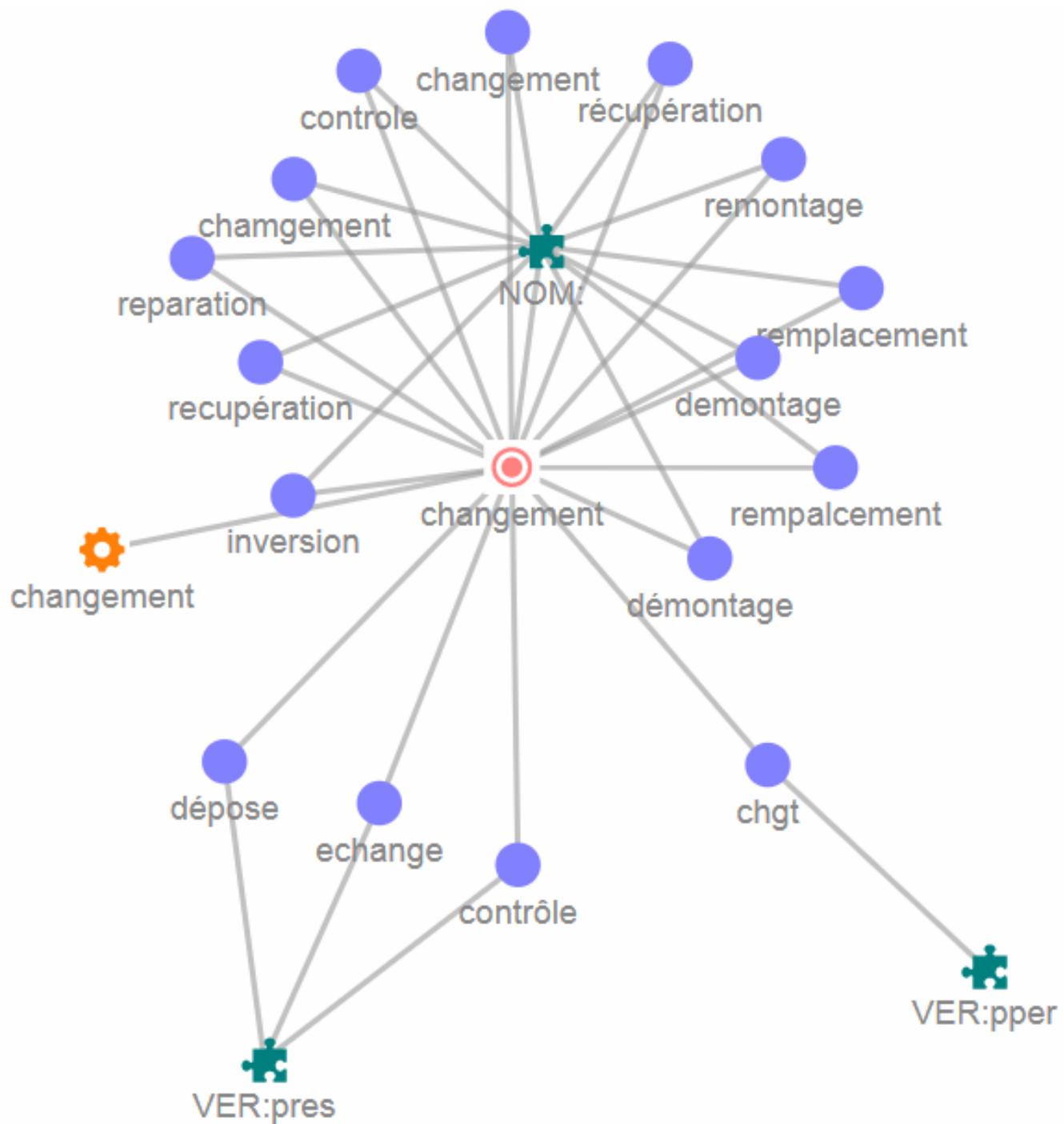
$\exists \text{ isPointed}\{\text{full}\} \sqsubseteq \text{IsSelectedFull}$

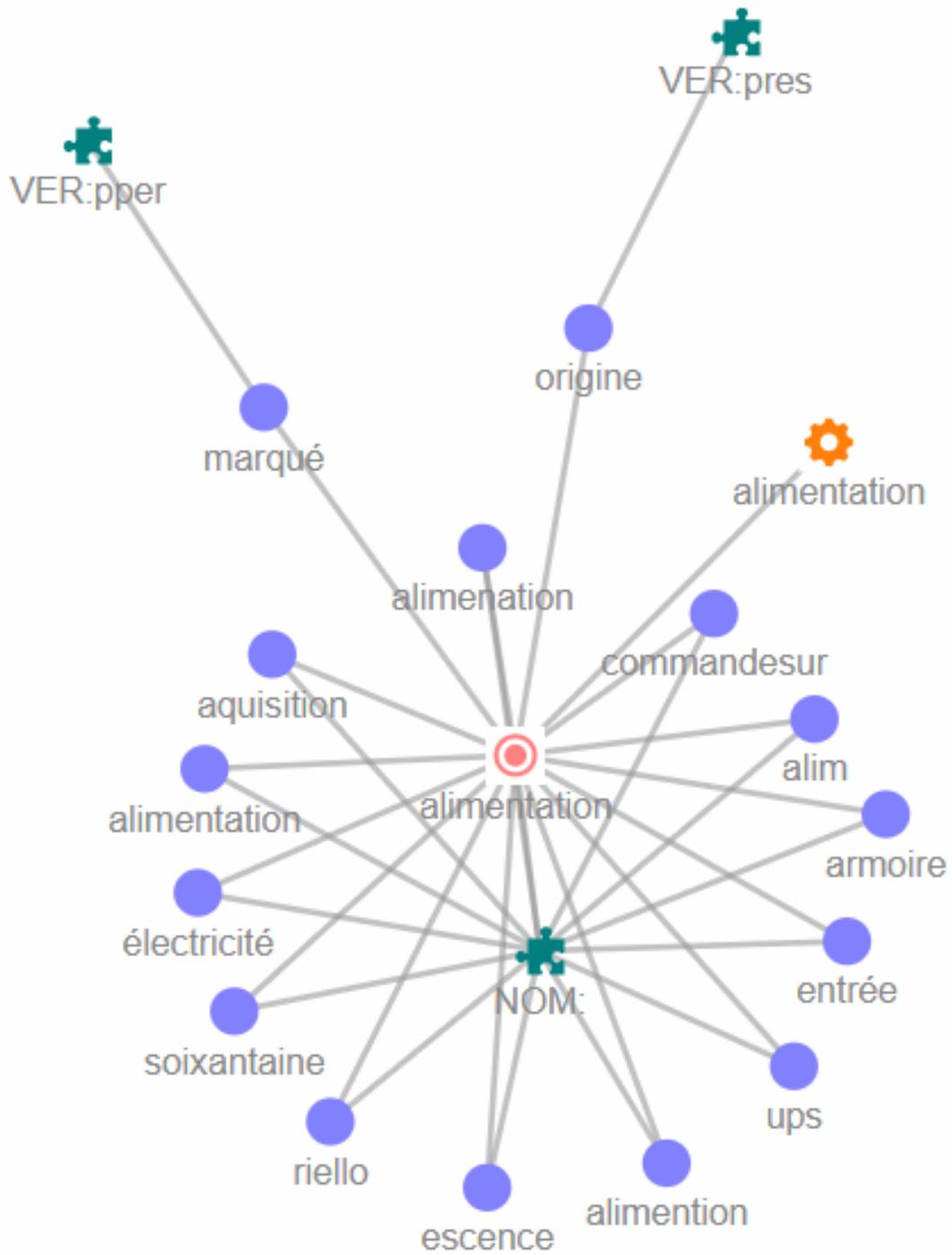
$\exists \text{ redArrow} . \text{IsSelectedFull} \sqsubseteq \text{IsSelectedFull}$

$\text{CentralObject} \sqcap \text{IsSelectedFull} \sqsubseteq \text{ActiveObject}$

$\exists \text{ blueArrow} . (\text{ActiveObject} \sqcap \text{IsModeFull})$   
 $\sqsubseteq \text{ActiveObject} \sqcap \text{IsModeFull}$







## 4. PERSPECTIVES W2V



- **Validation métier**
- **Réduction du bruit par injection de lexiques**
- **Correction orthographique**
- **Extraction d'actions sur des composants**
- **Gestion automatiquement les expressions multimots**
- **Mise en place d'une interface graphique sur W2V**

RECHERCHE	MOTS PROPOSES	TYPE	CLASSES	RELATION	
<input type="text" value="ROULEMENT"/> <input type="button" value="x"/> <input type="button" value="🔍"/>	<input checked="" type="checkbox"/> RLTS	Candidat	Composant	aPourCandidat	
	<input checked="" type="checkbox"/> ROULEMENTS	Candidat	Composant	aPourCandidat	
TYPE DU MOT : <input type="text" value="Representant"/>	<input checked="" type="checkbox"/> ROULTS	Candidat	Composant	aPourCandidat	
	<input checked="" type="checkbox"/> RL	Candidat	Composant	aPourCandidat	
	<input type="checkbox"/> POULIE				
CLASSE DU MOT : <input type="text" value="Composant"/>	<input checked="" type="checkbox"/> ROULT	Candidat	Composant	aPourCandidat	
	<input type="checkbox"/> TOURNEE				
	<input checked="" type="checkbox"/> REMPLT	Candidat	Action	porteSur	
	<input checked="" type="checkbox"/> ECHAUFFEMENT	Representant	Defaillance	aPourDefaillance	
CLASSES DE L'ONTOLOGIE : <input type="checkbox"/> owl:Thing <input checked="" type="checkbox"/> Action <input checked="" type="checkbox"/> Candidat <input checked="" type="checkbox"/> Composant <input checked="" type="checkbox"/> Defaillance <input checked="" type="checkbox"/> Representant	<input type="checkbox"/> VENTILO				
	<input type="checkbox"/> LEGER				
	<input type="checkbox"/> FLEXIBLE				
	<input checked="" type="checkbox"/> SERRAGE	Representant	Action	porteSur	
	<input type="checkbox"/> VENTI				
	<input type="checkbox"/> AEROSHELL				
	<input type="checkbox"/> RECALIF				
	<input type="checkbox"/> PIECE				
RELATIONS DE L'ONTOLOGIE <input type="checkbox"/> owl:topObjectProperty <input checked="" type="checkbox"/> aPourCandidat <input checked="" type="checkbox"/> aPourDefaillance <input checked="" type="checkbox"/> porteSur					
<input type="button" value="Editer dans Protégé"/>	<input type="button" value="Peuplement de l'ontologie"/>				