

# Peuplement d'ontologie à partir de petites annonces immobilières

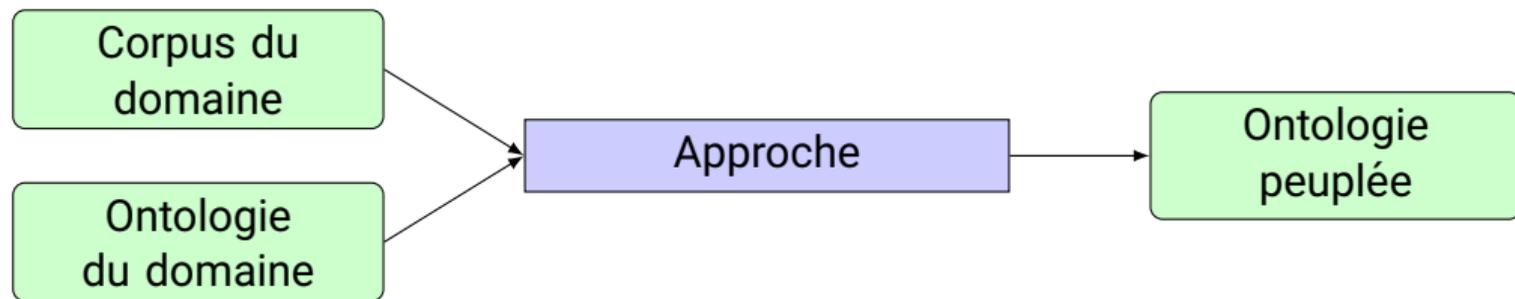
Céline Alec<sup>1</sup>

<sup>1</sup>Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, Caen, FRANCE

[celine.alec@unicaen.fr](mailto:celine.alec@unicaen.fr)



## Problématique étudiée : peuplement d'une ontologie de domaine



Corpus : chaque document est une description textuelle d'un objet du domaine

Exemple expérimental : petites annonces de ventes de maisons

## Peupler une ontologie

Ontologie :

*< Bien, type, Class >*

*< Maison, type, Class >*

*< Terrain, type, Class >*

*< surface, type, DataProperty >*

*< contient, type, ObjectProperty >*

...

### Exemple de peuplement

Un bien comprenant une maison de 110 m<sup>2</sup> et un terrain de 500 m<sup>2</sup>

*< bien1, isA, Bien >*

*< bien1, contient, maison1 >*

*< maison1, isA, Maison >*

*< maison1, surface, 110 >*

*< bien1, contient, terrain1 >*

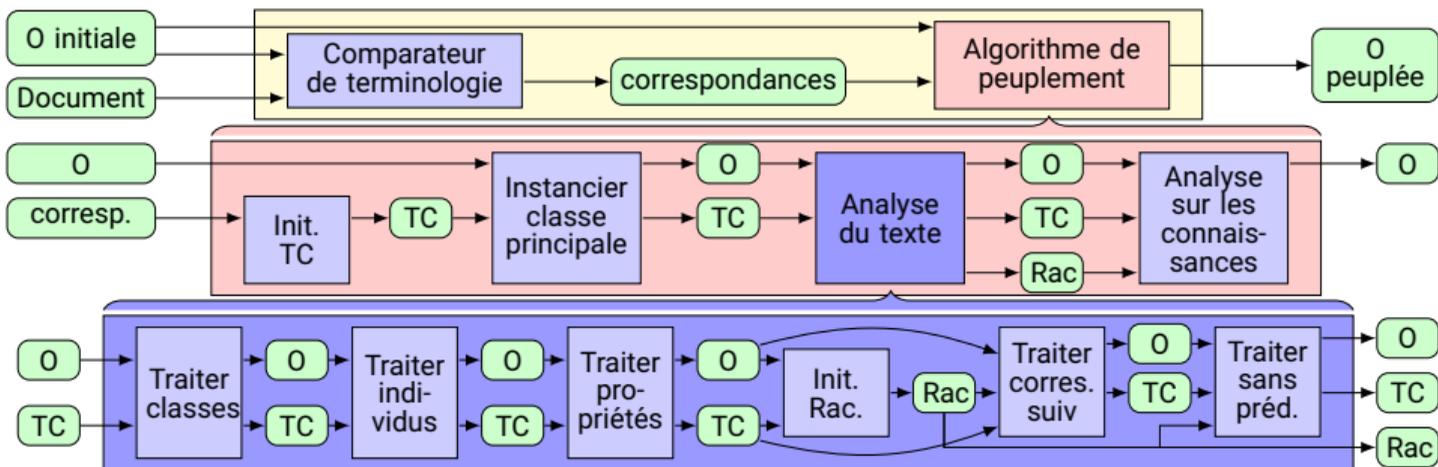
*< terrain1, isA, Terrain >*

*< terrain1, surface, 500 >*

## Positionnement par rapport à l'état de l'art

- Pas (ou peu) d'entités nommées
- Verbes peu parlants (« a », « possède », « contient », etc.), voire absents
- Propriétés typées à prendre en compte (ex : surfaces, prix, diagnostic énergétique, etc.)
- Propriétés non binaires (ex : un bien immobilier se situe à une certaine distance d'un point d'intérêt)

# Approche KOnPoTe (Knowledge graph/ONtology POpulation from TExts)



## Comparateur de terminologie

- Texte : découpage en phrases et lemmatisation
- Ontologie : lemmatisation de la terminologie

⇒ Création de correspondances **mention de texte** ↔ **entité ontologique**

« Cormelles-le-Royal : Magnifique maison à 15 min du centre-ville de Caen et à 3 min à pied des commerces et des écoles, 110 m<sup>2</sup> sur un terrain de 400 m<sup>2</sup>. Le rez-de-chaussée est composé d'une cuisine équipée, d'un salon avec cheminée, exposé sud-ouest, ainsi que d'une chambre de 15 m<sup>2</sup>. Premier étage : 2 chambres et 1 sdb. Proche des transports en commun. Terrain arboré et clos. Honoraires : 4%. »

# Pré-traitements

- Initialisation des traitements de correspondances

Texte	Entité	Type
Cormelles-le-Royal	Cormelles-le-Royal	individu
maison	Maison	classe
à	seSituéA	propriété objet
min	minVoiture	propriété typée
centre-ville	centre-ville	individu
...		

- Instanciation de la classe principale :  
<bien1, isA, Bien>

# Analyse sur le texte : traitement des classes/individus/propriétés

Idée : transformer les correspondances en individus/assertions

- Instanciation des classes :  $\langle \textit{individu}, \textit{isA}, \textit{classe} \rangle$
- Instanciation des propriétés :  $\langle \textit{sujet}, \textit{propriété}, \textit{objet} \rangle$

Quelques exemples :

Texte	Entité	Type	Individus	Assertions
Cormelles-le-Royal	Cormelles-le-Royal	individu	Cormelles-le-Royal	
maison	Maison	classe	maison1	$\langle \textit{maison1}, \textit{isA}, \textit{Maison} \rangle$
min	minVoiture	propriété typée	distance1	$\langle \textit{distance1}, \textit{isA}, \textit{Distance} \rangle$ $\langle \textit{distance1}, \textit{minVoiture}, 15 \rangle$

# Analyse sur le texte : initialisation des raccrochabilités

Contexte : verbes descriptifs peu parlants, voire absents

Intuition : manque d'assertions de propriété. Comment éviter des ajouts bruités ?

## Les raccrochabilités

= Un ensemble trié  $Rac(i)$  pour chaque individu  $i$

Contient des couples (*propriété, expression de classe*) t.q.  $i$  est « raccrochable » (en tant que sujet) à une instance de l'*expression de classe* via la *propriété*.

## Exemple

$Rac(bien1)$

- (*seSitueA, Commune*)
- (*estProcheDe, Lieu*)
- (*seSitueADistance, Distance*)
- (*contient, PartieDeBien*)

# Analyse sur le texte : traitement des correspondances suivantes

Intuition :

- Deux correspondances qui se suivent dans une même phrase ont un lien ?
- Gestion des propriétés n-aires :  $\langle\langle a \dots b \dots c \rangle\rangle \rightarrow \langle a, \dots, b \rangle$  et  $\langle a, \dots, c \rangle$

Exploitation des raccrochabilités pour trouver la meilleure propriété

Correspondance	Individus	Assertions	Individus précédents liés
...			
min	distance1	$\langle \text{distance1, isA, Distance} \rangle$ $\langle \text{distance1, minVoiture, 15} \rangle$	
centre-ville	centre-ville		
Caen	Caen		
...			

# Analyse sur le texte : traitement des correspondances suivantes

Intuition :

- Deux correspondances qui se suivent dans une même phrase ont un lien ?
- Gestion des propriétés n-aires :  $\langle\langle a \dots b \dots c \rangle\rangle \rightarrow \langle a, \dots, b \rangle$  et  $\langle a, \dots, c \rangle$

Exploitation des raccrochabilités pour trouver la meilleure propriété

Correspondance	Individus	Assertions	Individus précédents liés
...			
min	distance1	$\langle$ distance1, isA, Distance $\rangle$ $\langle$ distance1, minVoiture, 15 $\rangle$ $\langle$ distance1, distDuPI, centre-ville $\rangle$	
centre-ville	centre-ville		distance1
Caen	Caen		
...			

# Analyse sur le texte : traitement des correspondances suivantes

Intuition :

- Deux correspondances qui se suivent dans une même phrase ont un lien ?
- Gestion des propriétés n-aires :  $\ll a \dots b \dots c \gg \rightarrow \langle a, \dots, b \rangle$  et  $\langle a, \dots, c \rangle$

Exploitation des raccrochabilités pour trouver la meilleure propriété

Correspondance	Individus	Assertions	Individus précédents liés
...			
min	distance1	$\langle \text{distance1, isA, Distance} \rangle$ $\langle \text{distance1, minVoiture, 15} \rangle$ $\langle \text{distance1, distDuPI, centre-ville} \rangle$	
centre-ville	centre-ville		distance1
Caen	Caen		
...			

## Analyse sur le texte : traitement des correspondances suivantes

Intuition :

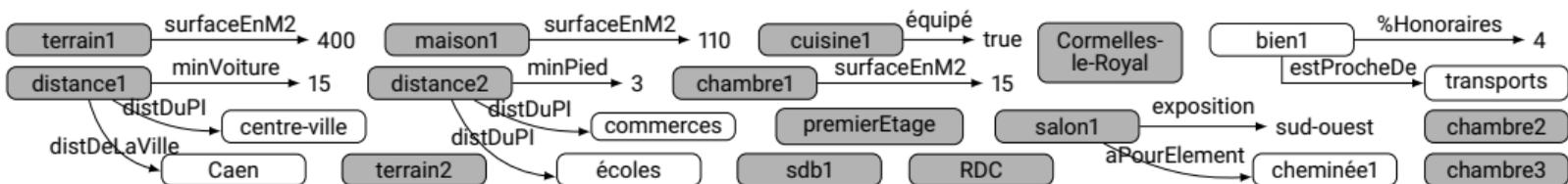
- Deux correspondances qui se suivent dans une même phrase ont un lien ?
- Gestion des propriétés n-aires :  $\langle\langle a \dots b \dots c \rangle\rangle \rightarrow \langle a, \dots, b \rangle$  et  $\langle a, \dots, c \rangle$

Exploitation des raccrochabilités pour trouver la meilleure propriété

Correspondance	Individus	Assertions	Individus précédents liés
...			
min	distance1	$\langle \text{distance1, isA, Distance} \rangle$ $\langle \text{distance1, minVoiture, 15} \rangle$ $\langle \text{distance1, distDuPI, centre-ville} \rangle$	
centre-ville	centre-ville	$\langle \text{distance1, distDeLaVille, Caen} \rangle$	distance1
Caen	Caen		distance1
...			

# Analyse sur le texte : traitement des individus sans prédécesseurs

Intuition : tout individu (sauf instance principale) doit être l'objet d'une assertion



⇒ Ajout de  $\langle \text{ sujet, propriété, individu\_sans\_prédécesseur} \rangle$  t.q.

- *sujet* et *individu\_sans\_prédécesseur* sont issus de la même phrase
- *propriété* est la plus adaptée au sens des raccrochabilités

## Exemple

Individus de la phrase 2 : RDC, cuisine1, salon1, chambre1

Assertions ajoutées :  $\langle \text{cuisine1, seTrouveAEtage, RDC} \rangle$ ,  $\langle \text{salon1, seTrouveAEtage, RDC} \rangle$ ,  $\langle \text{chambre1, seTrouveAEtage, RDC} \rangle$

# Analyse sur les connaissances

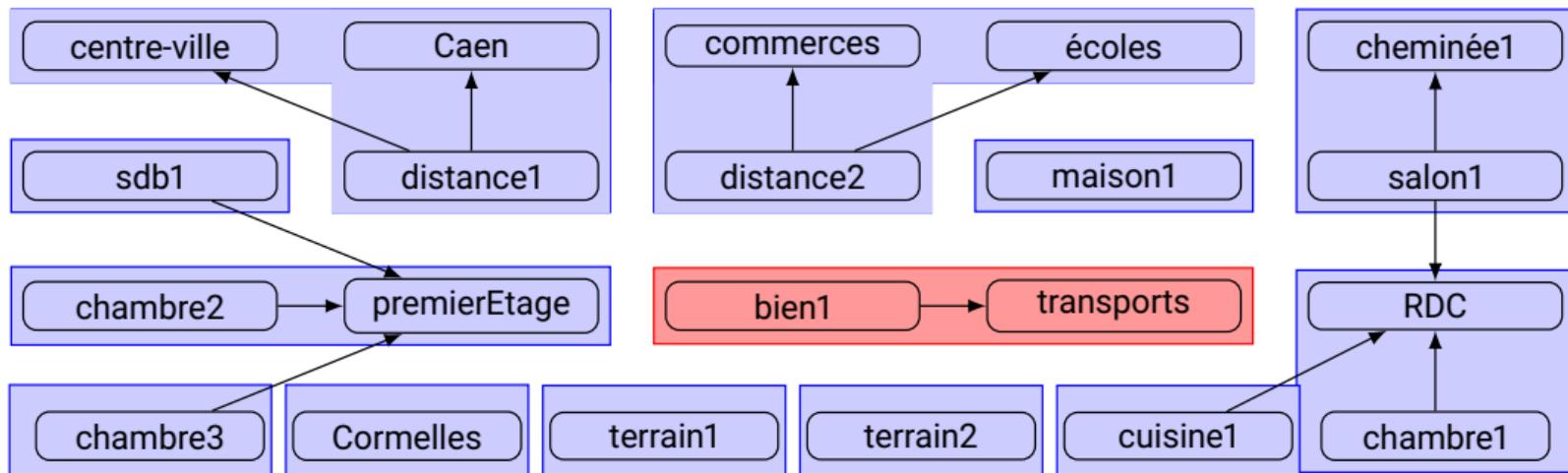
Intuition :

- L'instance principale doit mener à tout
- Les connaissances de l'ontologie peuvent aider

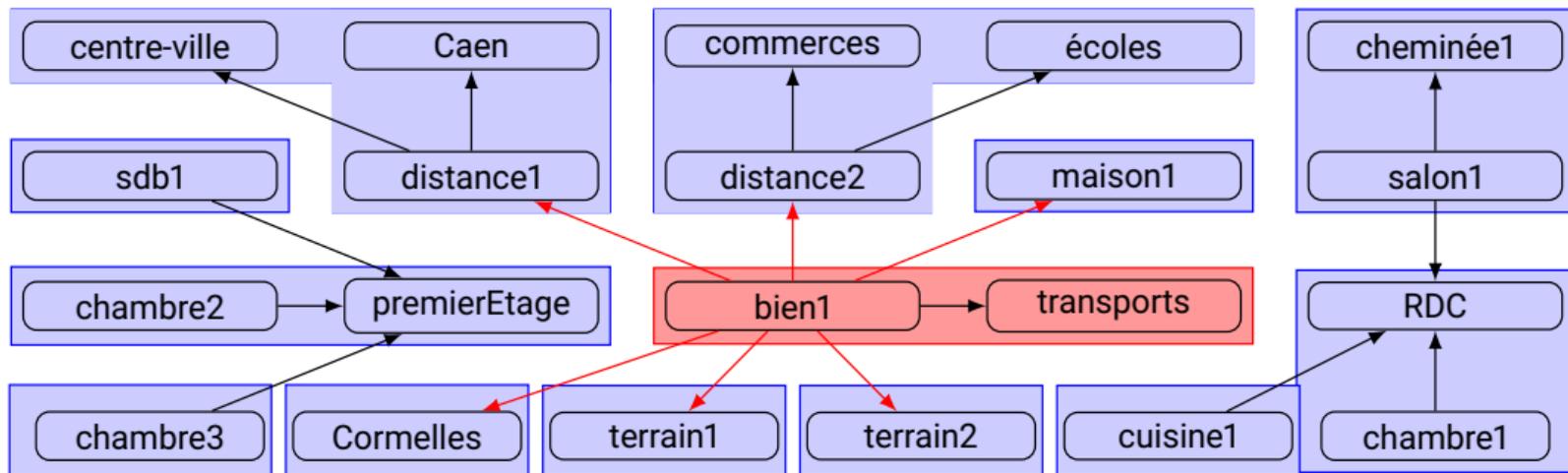
Idée :

Commencer par chercher ce qui est raccrochable à l'instance principale puis s'y distancer au fur et à mesure

# Analyse sur les connaissances : premier découpage en lots

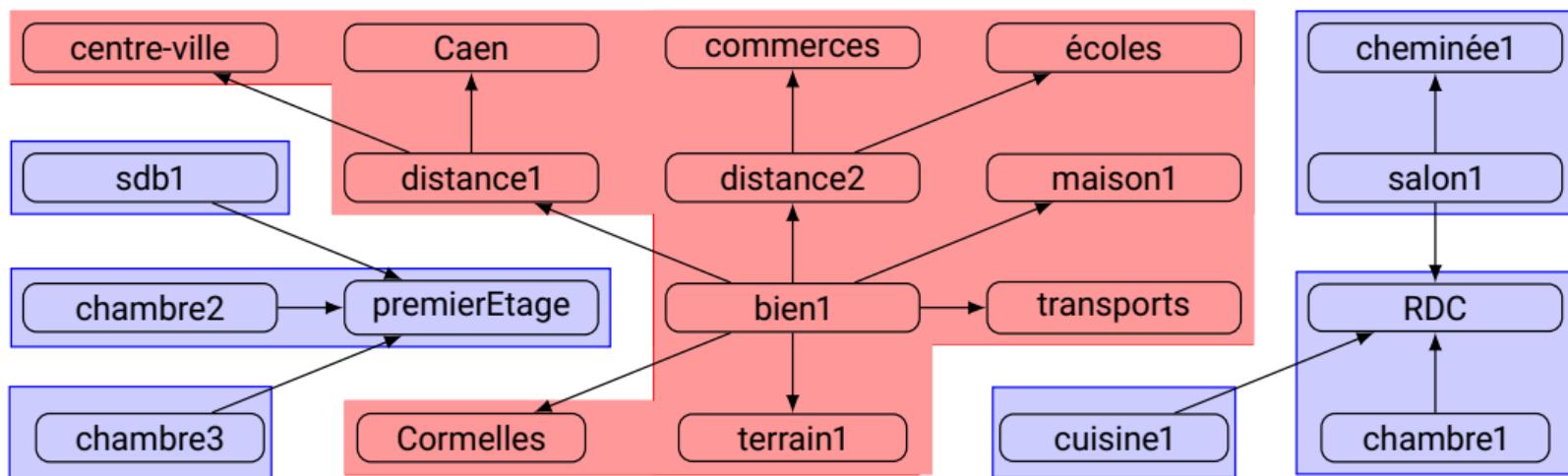


# Analyse sur les connaissances : premier découpage en lots



⇒ Fusion de terrain1 et terrain2

## Analyse sur les connaissances : deuxième découpage en lots



⇒ Ajout de :

<maison1, contient, chambre1>,    <maison1, contient, chambre2>,    <maison1, contient, chambre3>  
 <maison1, contient, cuisine1>,    <maison1, contient, salon1>,    <maison1, contient, sdb1>

⇒ Un seul lot!

## Expérimentations

- Corpus issu de 78 annonces + Ontologie du domaine
- Ontologie Gold Standard (GS) : Ontologie initiale manuellement peuplée avec les assertions représentant les descriptions du corpus
- Objectif : comparer l'ontologie obtenue par KOnPoTe avec le GS

$$Précision = \frac{VP}{VP + FP} \quad Rappel = \frac{VP}{VP + FN} \quad F\text{-mesure} = \frac{2 \times Précision \times Rappel}{Précision + Rappel}$$

# Résultats avec deux lemmatiseurs différents (*Aker* et *TT*)

Approche	Précision <sub>macro</sub>	Rappel <sub>macro</sub>	F-mesure <sub>macro</sub>	Précision <sub>micro</sub>	Rappel <sub>micro</sub>	F-mesure <sub>micro</sub>
KOnPoTe <sub><i>Aker</i></sub>	<b>0.9516</b>	<b>0.8740</b>	<b>0.9079</b>	<b>0.9465</b>	<b>0.8606</b>	<b>0.9015</b>
KOnPoTe <sub><i>TT</i></sub>	0.9496	0.8681	0.9039	0.9446	0.8545	0.8973
Analyse text. <sub><i>Aker</i></sub>	0.8989	0.4648	0.5994	0.8956	0.4726	0.6188
Analyse text. <sub><i>TT</i></sub>	0.8964	0.4579	0.5929	0.8937	0.4662	0.6127
Baseline+suiv <sub><i>Aker</i></sub>	0.8911	0.3138	0.4440	0.8741	0.3085	0.4561
Baseline+suiv <sub><i>TT</i></sub>	0.8879	0.3081	0.4377	0.8732	0.3036	0.4505
Baseline <sub><i>Aker</i></sub>	0.9234	0.1922	0.3099	0.9135	0.1926	0.3182
Baseline <sub><i>TT</i></sub>	0.9230	0.1888	0.3054	0.9138	0.1892	0.3135

## Conclusion et perspectives

- Approche de peuplement d'une ontologie de domaine à partir de descriptions textuelles d'objets de ce domaine
- Chaîne de traitements basée sur la problématique et non sur le domaine
- Résultats encourageants sur une première expérimentation
  
- Expérimentations sur d'autres domaines : coûteux
- Expérimentations sur le même domaine et corpus mais avec différentes ontologies : dégager des recommandations à suivre
- Exploiter KOnPoTe comme première étape pour corriger des erreurs d'annotation

# Merci pour votre attention

## Questions ?

# Contexte : Détection d'Erreurs et Correction d'Annotations

## Contexte général

Descriptions + annotations

Objectif à long terme : corriger  
les annotations erronées

⇒ via ontologie :  
peuplement + raisonnement

⇒ Généricité : descriptifs  
d'objets annotés



Maison 5 pièces 110 m<sup>2</sup>  
341 000 €

Caen (14000)

### Critères

Maison

Surface de 110 m<sup>2</sup>

Terrain de 700 m<sup>2</sup>

5 pièces

Classe énergie

A B C D E F G

GES

A B C D E F G

4 ch.

Honoraires : Oui

Référence : An1

Maison 5 pièces 110 m<sup>2</sup>

Exclusivité 610 immobilier , maison individuelle de 110 m<sup>2</sup> sur un terrain de 700 m<sup>2</sup> . La maison est au calme et à l'abris des regards. Sur la commune de Cairon au nord de Caen, avec commerces, école, et bus de ville à proximité.

Au rdc , entrée séjour salon , cuisine ouverte. Une chambre , un wc. A l'étage, 3 chambres , une salle d'eau et un wc.

Garage

Affaire suivie par Rouget-Briouze Cyril Agent Commercial immatriculé au RSAC de Caen sous le...

[Voir plus](#)

## Analyse sur le texte : traitement des classes/individus/propriétés

Idée : transformer les correspondances en individus/assertions

- Si correspondance de classe (sauf classe principale) :
  - ignorée si trace de négation (exemple : « Pas de garage »)
  - instanciation de la classe une ou plusieurs fois (exemple : « 2 chambres » → 2 instances de chambres)
- Si correspondance d'individus : on garde l'individu
- Si correspondance de propriété : on cherche les sujets et objets possibles permettant  $\langle \textit{sujet}, \textit{propriété}, \textit{objet} \rangle$ 
  - liste de sujets possibles en remontant les correspondances de la phrase tant qu'un sujet n'est pas dans le domaine (ou sinon nouvelle instance)
  - liste d'objets possibles : dépend du type de propriété et de correspondance
  - ajout si pas d'incohérence

# Analyse sur le texte : traitement des individus sans prédécesseurs

#	Individus de la phrase (objets recherchés en italique)	Assertions ajoutées (<instance_de_la_phrase, propriété, objet>)
1	<i>Cormelles-le-Royal, maison1, distance1, centre-ville, Caen, distance2, commerces, écoles, terrain1</i>	<distance1, distDeLaVille, <i>Cormelles-le-Royal</i> > <distance2, distDeLaVille, <i>Cormelles-le-Royal</i> > non ajoutées car incohérentes
2	<i>RDC, cuisine1, salon1, chambre1</i>	<cuisine1, seTrouveAEtage, RDC> <salon1, seTrouveAEtage, RDC> <chambre1, seTrouveAEtage, RDC>
3	<i>premierEtage, chambre2, chambre3, sdb1</i>	<chambre2, seTrouveAEtage, <i>premierEtage</i> > <chambre3, seTrouveAEtage, <i>premierEtage</i> > <sdb1, seTrouveAEtage, <i>premierEtage</i> >
4	bien1, transports	
5	<i>terrain2</i>	
6	bien1	