

# **Estimer la prédictibilité (morpho) sémantique à l'aide de grands modèles de langage**

**Daniele Sanacore**

**Thématiques actuelles de la recherche en TAL  
9 décembre 2024**

# Contexte de recherche

Problématique et approche proposée

# Contexte de recherche

**Objet d'étude** : notion de **paradigme** en morphologie

**Problème** : trouver une méthode pour **identifier des paradigmes dérivationnels** en morphologie.

# Contexte de recherche

## Paradigmes en flexion :

L'organisation de la morphologie flexionnelle en paradigmes est relativement consensuelle (Carstairs–McCarthy, 1994; Ackerman et al., 2009; Blevins, 2016, Beniamine 2018, entre autres).

**Propriété centrale des paradigmes en flexion** : interprédictibilité sur le plan du contenu

	<b>SG</b>	<b>PL</b>
<b>M</b>	<i>vert</i>	<i>verts</i>
<b>F</b>	<i>verte</i>	<i>vertes</i>

# Contexte de recherche

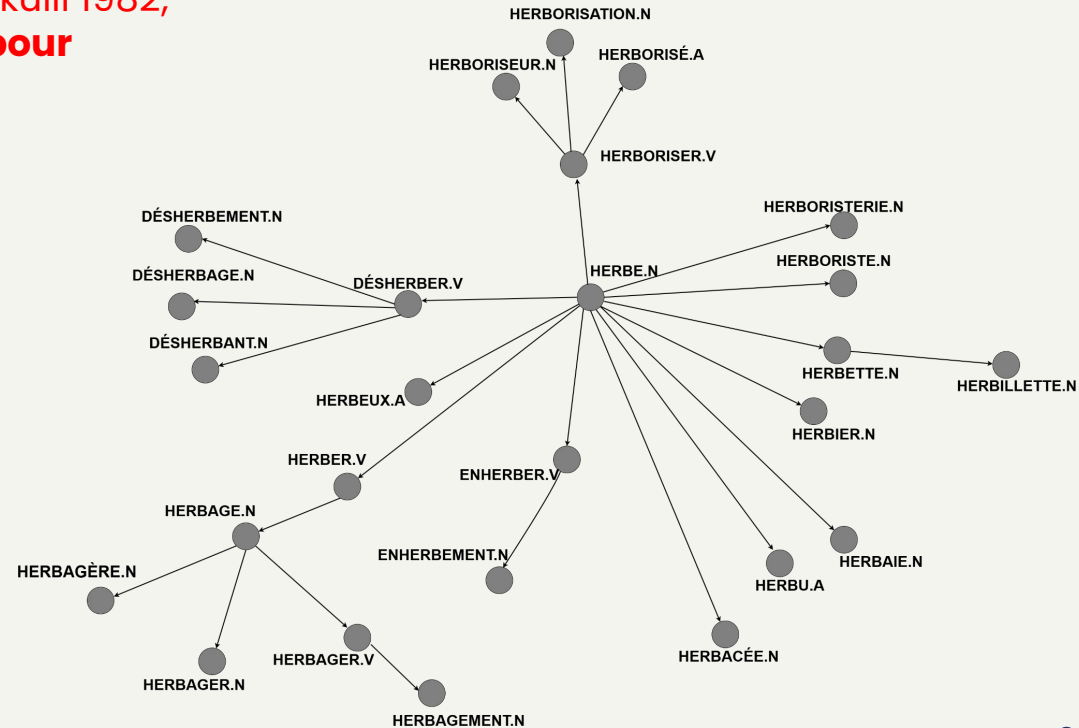
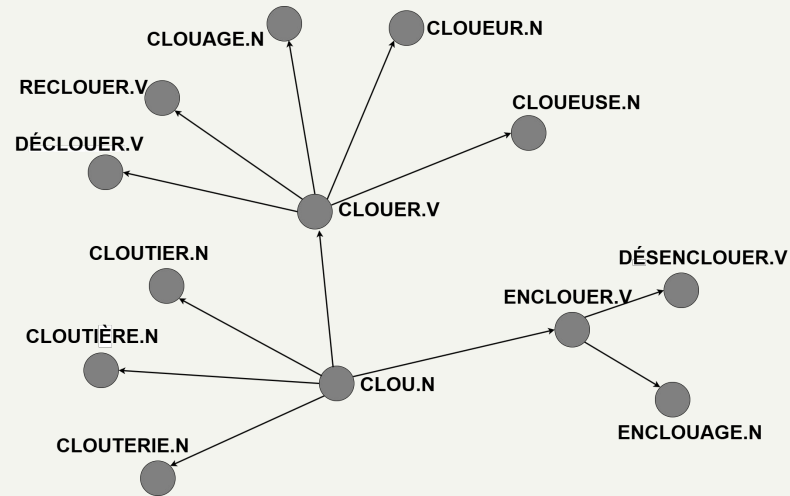
**Paradigmes en flexion** : La nature paradigmatique de la flexion est aujourd'hui **relativement consensuelle** (Carstairs–McCarthy, 1994; Ackerman et al., 2009; Blevins, 2016).

**Paradigmes en dérivation** : La nature paradigmatique de la dérivation est plus **débattue**, (Bauer, 1997; Štekauer, 2014; Bonami et Strnadová, 2019; Hathout et Namer, 2022; entre autres).

# Contexte de recherche

## À quoi ressemblent les paradigmes en dérivation ?

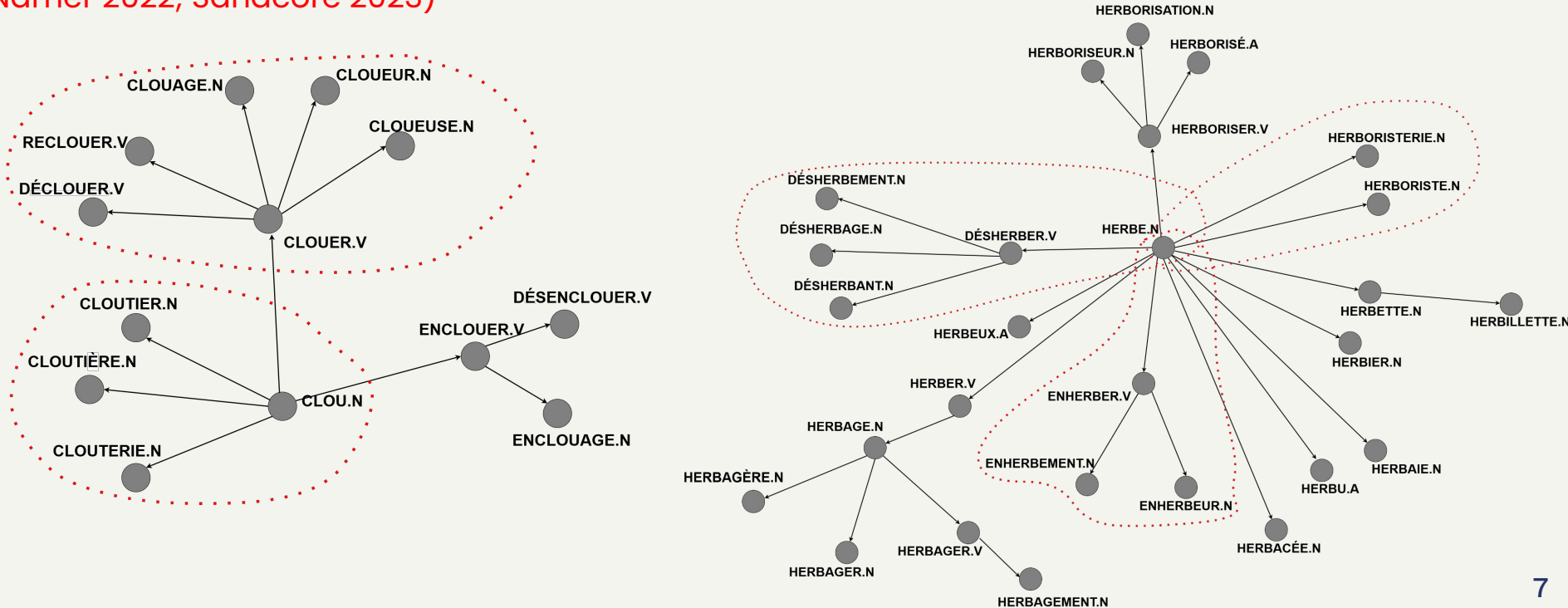
**Première possibilité :** les paradigmes dérivationnels correspondent à des familles de mots (Dokulil 1982; Stump 1991; Gaeta 2022), **problématique pour l'interprédictibilité**



# Contexte de recherche

## À quoi ressemblent les paradigmes en dérivation ?

**Deuxième possibilité :** les paradigmes dérivationnels correspondent à des familles de mots partielles qui expriment des concepts interprédictibles ( Bonami & Strnadová 2019; Hathout & Namer 2022; Sanacore 2023)



# Proposition méthodologique

**Problème** : nous ne savons pas à quoi ressemblent les paradigmes dérivationnels, quelles sont les “cases” disponibles et comment les identifier

**Approche proposée** : les relations sémantiques sont premières dans la structure des paradigmes dérivationnels (Roché 2017; Bonami & Strnadova 2019). Nous adoptons une approche basée sur la sémantique des frames (Petrucci, 1996; Fillmore, 2006), qui se sert de sorte de scénarios

**Mise en œuvre** : produire des scénarios sous forme d'histoires pour estimer l'interprédictibilité entre les concepts qui sont “réalisés” par les familles de mots



# Proposition méthodologique

## Exemple d'histoire construite sur le couple HERBE-HERBORISTE

*Dans le petit village de Gavot, il y avait une charmante **herboristerie** où les **habitants** allaient **acheter** les **remèdes naturels** pour soulager leurs maux. L'enseigne au-dessus de la porte indiquait clairement : "Herboristerie du Soleil". Au cœur de cette boutique, il y avait une grande sélection d'**herbes médicinales**, soigneusement rangées sur des étagères et dans des bocaux. Les clients venaient là pour trouver un traitement naturel contre les rhumatismes ou les douleurs articulaires. Un jour, la **propriétaire de l'herboristerie, Madame Dupont**, aida une vieille femme à choisir le bon remède pour soulager sa douleur au genou.*



**La co-occurrence systématique de certains éléments dans les histoires peut nous dire quelque chose sur la prédictibilité sémantique**

# Proposition méthodologique

## Pourquoi les histoires ?

- Les histoires permettent d'insérer les relations morphosémantiques que nous voulons étudier dans un contexte qui implique d'autres participants qui sont sémantiquement "proches" à la relation que nous prenons en compte
- La production des histoires, contrairement à l'utilisation de corpus, permet d'avoir des contextes qui dépendent uniquement de la relation que nous prenons en compte
- Le format narratif permet de bien identifier les participants et les rôles que ces participants jouent dans l'histoire

# Proposition méthodologique

L'approche basée sur les histoires s'inspire de la sémantique des frames (Fillmore 1976; Petruck 1996) et de la ressource FrameNet (Ruppenhofer et al. 2016)

**Scénario de vente :** Un **Acheteur** achète un **Bien** d'un **Vendeur** en lui payant une **Somme d'Argent**.

**Réalisation :** **Jean** a acheté **une nouvelle moto** chez un **revendeur**. Il a payé **5000 euros**.

**Acheteur :** L'acheteur a l'argent et veut acheter un bien

**Vendeur :** le vendeur reçoit de l'argent et cède un bien en échange

**Bien :** le bien est donné à l'acheteur par le vendeur en échange d'une somme d'argent

**Somme d'argent :** la somme d'argent est payée au vendeur par l'acheteur

# Proposition méthodologique

## Repérer les histoires

Deux possibilités :

- Interroger des locuteurs humains (compliqué et long)
- Produire les histoires automatiquement avec des outils de génération de textes/LLMs.

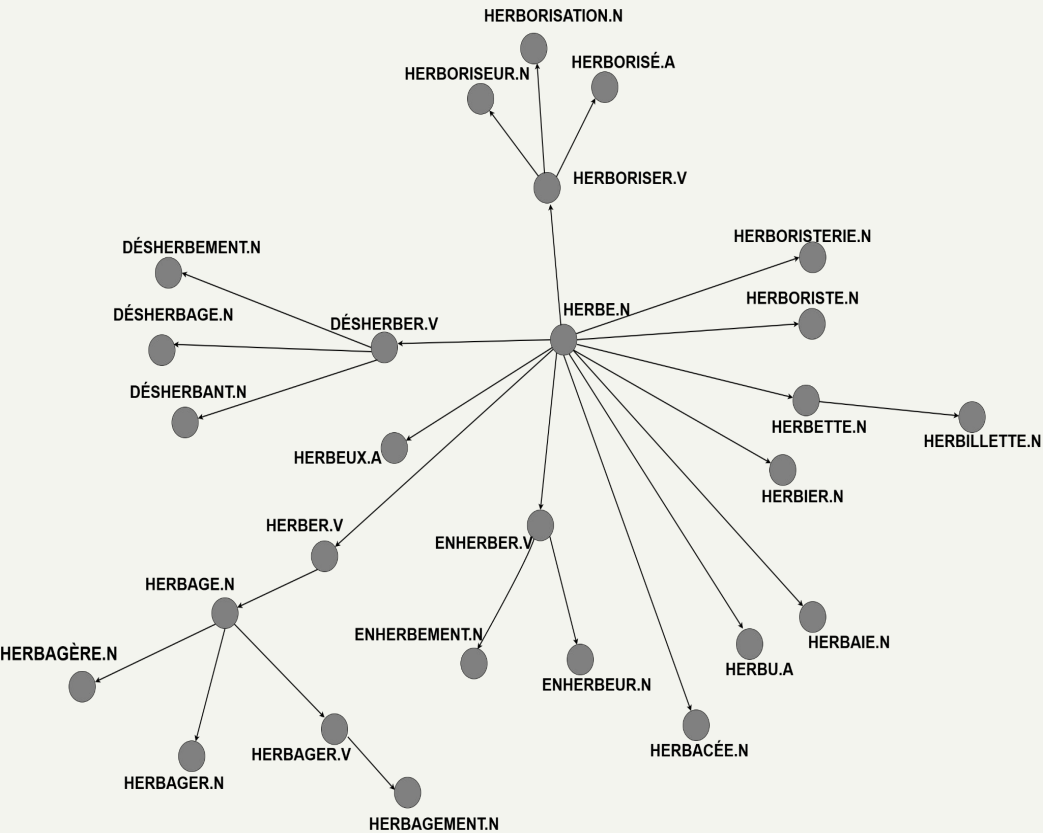
Pertinence des LLMs :

- Capacité de produire les histoires automatiquement et en grande quantité
- Capacité de produire des textes relativement cohérents d'un point de vue sémantique
- Capacité de produire de textes qui reflètent une certaine connaissance du monde

# Production des histoires

Utiliser les LLMs pour produire des histoires qui permettent d'estimer l'interprédictibilité

# LLMs pour générer des histoires



Nous prenons une “grosse” famille de mots :

- Nous produisons un jeu d’histoires construites sur des couples de mots extraits de cette famille
- Les histoires seront ensuite annotées (manuel et LLMs) pour estimer l’interprédictibilité des concepts qui sont réalisés dans cette famille

Hypothèse : dans la famille de HERBE, coexistent (au moins) trois paradigmes dérivationnels distincts et indépendants structurés par trois scénarios:

- Scénario de vente
- Élimination entité nuisible
- Cultivation

# LLMs pour générer des histoires

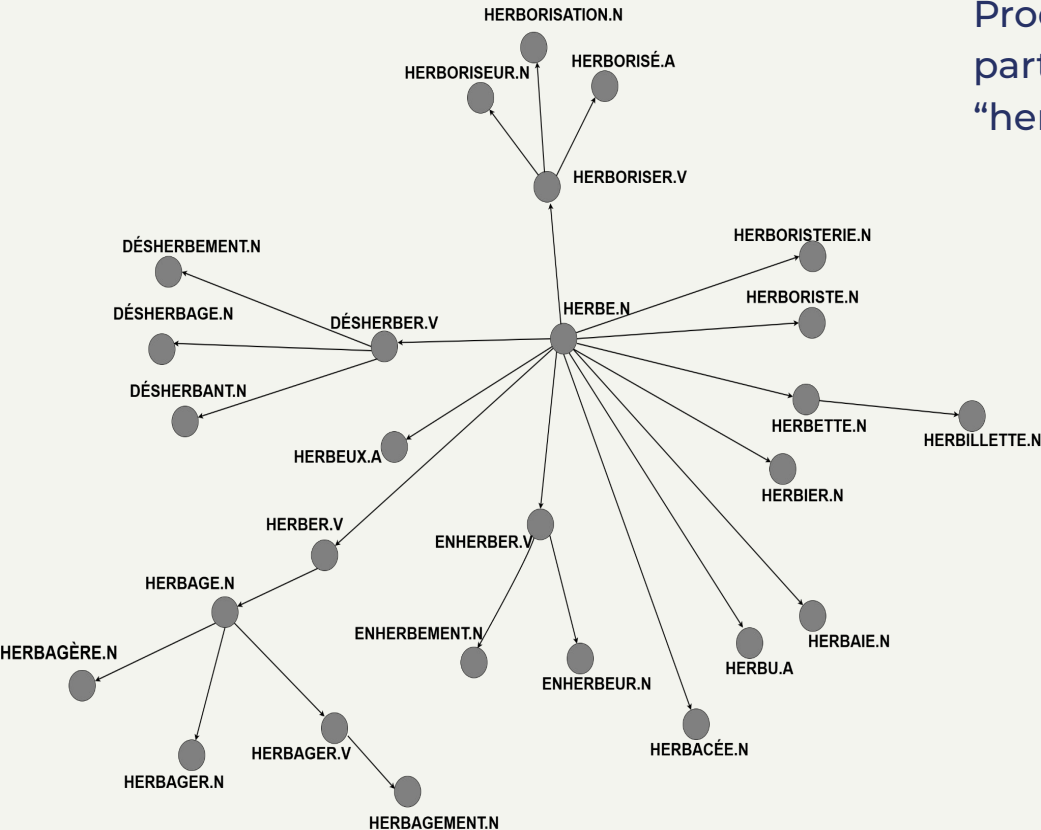


Ollama permet d'interroger une vaste gamme de LLMs "en local" sans passer par les interfaces web.

- Très facile à utiliser pour produire des textes, stocker les résultats dans des fichiers sur sa propre machine etc. (à condition d'avoir une machine assez puissante)
- Utilisable avec un notebook Python en local
- Grande variété de LLMs disponibles pour différentes tâches, langues, etc.

# LLMs pour générer des histoires

Production automatique de 600 histoires à partir de 12 couples extraits de la famille de “herbe” (50 histoires par couple):



- herbe - désherber
- désherber - désherbant
- herbe - désherbant
- herbe - herboriste
- herboriste - herboristerie
- herbe - herboristerie
- enherber - enherbement
- enherbement - herbe
- herbe - enherber
- désherber - herboriste
- désherbant - herboristerie
- désherbant - enherbement



# LLMs pour générer des histoires

À propos du prompt :

- Instructions données en anglais pour produire des données en français
- Jamais produire des histoires ou des traductions en anglais
- Produire des histoires qui font entre 75 et 100 mots
- Pas d'histoires en première personne
- Produire des histoires réalistes, qui reflètent la connaissance du monde réel (pas de fables)
- Pas d'humanisation d'animaux, si présents dans l'histoire

N.B: Le modèle est réinitialisé à chaque itération (une histoire=une itération) pour limiter la répétitivité de l'output

# LLMs pour générer des histoires

System prompt et création du prompt dans le corps de l'algorithme :

```
systemprompt = """ Your are a French native speaker.  
You will always answer in French, without using English.  
You will never provide an English translation of your answer.  
Your answers will be stories containing no less than 75 words and no more than 100 words.  
The stories that you will produce will be realistic, not fables.  
If the story involves animals, it will be realistic with respect to what animals do in the real world.  
You will never make stories in first person.  
"""
```

```
llm = Ollama(base_url='http://localhost:11434', model="llama3", system=systemprompt)
```

```
def create_prompt(row):  
    mot1 = row['mot1']  
    mot2 = row['mot2']  
    cat1 = 'noun' if row['cat1'] == 'N' else 'adjective' if row['cat1'] == 'A' else 'verb'  
    cat2 = 'noun' if row['cat2'] == 'N' else 'adjective' if row['cat2'] == 'A' else 'verb'  
    relation = row['relation']  
  
    if relation == '-':  
        return f"Tell me a story that contains the words '{mot1}' (which is a {cat1}) and '{mot2}' (which is a {cat2})."  
    else:  
        return f"Tell me a story that contains the words '{mot1}' (which is a {cat1}) and '{mot2}' (which is a {cat2}).  
        The story should reflect the following relationship between these words: {relation}."
```

# LLMs pour générer des histoires

Le modèle prend comme données d'input les couples de mots, leur catégorie syntaxiques et une glose qui illustre la relation entre les deux mots (lorsqu'elle est immédiate)

mot1	mot2	cat1	cat2	relation
herbe	désherber	N	V	désherber signifie débarasser un lieu de mauvaises herbes
désherber	désherbant	V	N	le désherbant est un produit utilisé pour désherber un terrain ou un jardin
désherbant	herbe	N	N	le désherbant est un produit utilisé pour éliminer les mauvaises herbes
herbe	herboriste	N	N	un herboriste est une personne qui vend des herbes et des graines médicinales
herboriste	herboristerie	N	N	une herboristerie est une boutique dans laquelle travaille un herboriste
herboristerie	herbe	N	N	une herboristerie est une boutique dans laquelle on vend des herbes médicinales
herbe	enherber	N	V	Enherber signifie planter de l'herbe dans un terrain
herbe	enherbement	N	N	l'enherbement est une pratique agricole qui consiste à planter de l'herbe dans un terrain
enherber	enherbement	V	N	l'enherbement est une pratique agricole qui consiste à enherber un terrain
désherbant	enherbement	N	N	-
désherber	herboriste	V	N	-
désherbant	herboristerie	N	N	-

**Fonction de la relation :** résoudre ambiguïtés sémantiques, guider la définition du scénario



La relation entre mot1 et mot2 est donc connue par le modèle. Ce qui n'est pas connu, ce sont tous les autres participants et relations qui coexistent fréquemment avec cette relation.

# LLMs pour générer des histoires

Modèle utilisé pour la production d'histoires : LLama 3 (8B)

## Exemples d'histoires produites

*Il était midi lorsque le jardinier arriva au parc pour entreprendre son travail. Sa tâche consistait à **désherber** les pelouses et les allées, libérant ainsi ces espaces des mauvaises **herbes** qui y avaient pris racine. Il prit son outil préféré, un désherbeur efficace, et se mit au travail. Les herbes envahissantes furent rapidement écartées, laissant apparaître le vert tendre de l'herbe bien cultivée.*

*Dans le petit village de Saint-Jean, il y avait un **herboriste** nommé Pierre qui passait son temps à explorer la nature. Il était particulièrement intéressé par les plantes médicinales et les fleurs rares qu'il trouvait dans les bois voisins. Un jour, alors qu'il allait collecter des feuilles pour faire une tisane, il remarqua que le terrain en face de lui était recouvert de mauvaises herbes. Il décida donc de se mettre au travail et de **désherber** ce terrain pour en faire un endroit où les gens pourraient se promener sans être gênés par la végétation abondante.*

# Annotation des histoires

Utiliser les LLMs pour annoter les histoires et estimer l'interprédictibilité

# LLMs pour annoter des histoires

Première étape de l'annotation : annotation manuelle des histoires pour avoir un premier "gold" et constitution d'un guide pour attribuer les rôles aux éléments syntaxiques de l'histoire.

SELLER: person selling a property to a buyer

SHOP: place where a commercial activity is realized, like a shop or the town market

MERCHANDISE: item sold from buyer to seller

CUSTOMER: person who buys a good from another by paying some money

COMMERCIAL ACTIVITY: activity in which one person sells a good to another

ELIMINATOR: person who removes an entity from an infested area

ELIMINATED ENTITY: entity removed from an infested area

ELIMINATING MEANS: product used to remove an entity from a given area (it is modified during the process, like a chemical product)

ELIMINATING INSTRUMENT: artifact used to remove an entity from a given area (it is not modified during the process, like scissors or lawn mowers)

LIBERATED PLACE: natural or artificial zone freed of a given harmful entity

ELIMINATION ACTIVITY: activity in which one person removes an entity from a given area

FARMER: person who grows a plant

CULTIVATION SITE: place where plants are cultivated

CULTIVATED PLANT: plant cultivated by one or more people \*on purpose\*; invasive weeds are not cultivated

FARMING ACTIVITY: activity in which plants are cultivated by one or more people



Les définitions des rôles se situent sur un niveau de granularité suffisamment général pour annoter n'importe quelle histoire

# LLMs pour annoter des histoires

```
# Configurer Le modèle Ollama
```

```
systemprompt = """
```

```
You are a semantic role annotator for French stories working as my assistant.
```

```
You will always answer in French, without providing an English translation of your answer.
```

```
You will base your decisions on the following semantic role definitions:
```

```
SELLER: person selling a property to a buyer
```

```
SHOP: place where a commercial activity is realized, like a shop or the town market
```

```
MERCHANDISE: item sold from buyer to seller
```

```
CUSTOMER: person who buys a good from another by paying some money
```

```
COMMERCIAL ACTIVITY: activity in which one person sells a good to another
```

```
ELIMINATOR: person who removes an entity from an infested area
```

```
ELIMINATED ENTITY: entity removed from an infested area
```

```
ELIMINATING MEANS: product used to remove an entity from a given area (it is modified during the process, like a chemical product)
```

```
ELIMINATING INSTRUMENT: artifact used to remove an entity from a given area (it is not modified during the process, like scissors or lawn mowers)
```

```
LIBERATED PLACE: natural or artificial zone freed of a given harmful entity
```

```
ELIMINATION ACTIVITY: activity in which one person removes an entity from a given area
```

```
FARMER: person who grows a plant
```

```
CULTIVATION SITE: place where plants are cultivated
```

```
CULTIVATED PLANT: plant cultivated by one or more people *on purpose*; invasive weeds are not cultivated
```

```
FARMING ACTIVITY: activity in which plants are cultivated by one or more people
```

```
"""
```

```
# Fonction pour annoter Les histoires
```

```
def annota_ruoli(story):
```

```
    prompt = f"""
```

```
Analyse les rôles sémantiques de l'histoire suivante:\n\n{story}\n\n
```

```
Répond uniquement avec un dictionnaire Python où les clés sont les rôles sémantiques et les valeurs sont les éléments syntaxiques de l'histoire qui r
```

```
Exemple de réponse attendue :
```

```
{{
```

```
    "SELLER": "-",
```

```
    "SHOP": "la boulangerie",
```

```
    "MERCHANDISE": "les croissants",
```

```
    "CUSTOMER": "un homme",
```

```
    "COMMERCIAL ACTIVITY": "l'achat",
```

```
    ...
```

```
}}
```

# LLMs pour annoter des histoires

## Exemple d'annotation

Dans le petit village de Gavot, il y avait une charmante *herboristerie* où les *habitants* allaient *acheter* les *remèdes naturels* pour soulager leurs maux. L'enseigne au-dessus de la porte indiquait clairement : "Herboristerie du Soleil". Au cœur de cette boutique, il y avait une grande sélection d'*herbes médicinales*, soigneusement rangées sur des étagères et dans des bocalux. Les clients venaient là pour trouver un traitement naturel contre les rhumatismes ou les douleurs articulaires. Un jour, la *propriétaire de l'herboristerie, Madame Dupont*, aida une vieille femme à choisir le bon remède pour soulager sa douleur au genou.

SHOP	<i>herboristerie</i>
CUSTOMER	<i>habitants; clients</i>
MERCHANDISE	<i>remèdes; herbes</i>
SELLER	<i>Mme Dupont</i>
COMMERCIAL ACTIVITY	<i>acheter</i>
...	-



# LLMs pour annoter des histoires

## Exemple d'annotation 2

Dans le petit village de Sérignac, il y avait une petite rue pavée où se trouvait la plus ancienne et la plus réputée **herboristerie** du pays. L'enseigne en bois portait le nom de "La Maison des Herbes" et les **villageois** y venaient pour **acheter** des **remèdes naturels** contre les maux courants. La boutique était dirigée par une femme sage et expérimentée, **Mme Martin**, qui connaissait à fond la médecine traditionnelle et **cultivait** elle-même dans son **jardin** les **herbes médicinales** que **vendait** en quantité limitée. Les clients venaient de loin pour acheter ses fameuses **décoctions et ses tisanes antibactériennes**, toutes faites à partir des herbes les plus précieuses de la région.

SHOP	<i>herboristerie</i>
CUSTOMER	<i>villageois</i>
MERCHANDISE	<i>remèdes naturels</i>
SELLER	<i>Mme Martin</i>
COMMERCIAL ACTIVITY	<i>acheter; vendre</i>
FARMER	<i>Mme Martin</i>
CULTIVATION SITE	<i>jardin</i>
CULTIVATED PLANT	<i>herbes</i>
FARMING ACTIVITY	<i>cultiver</i>
...	-

# LLMs pour annoter des histoires

**Probabilité conditionnelle** : la probabilité d'un rôle sémantique A d'être instancié dans une histoire étant donné la présence dans la même histoire d'un rôle sémantique B.

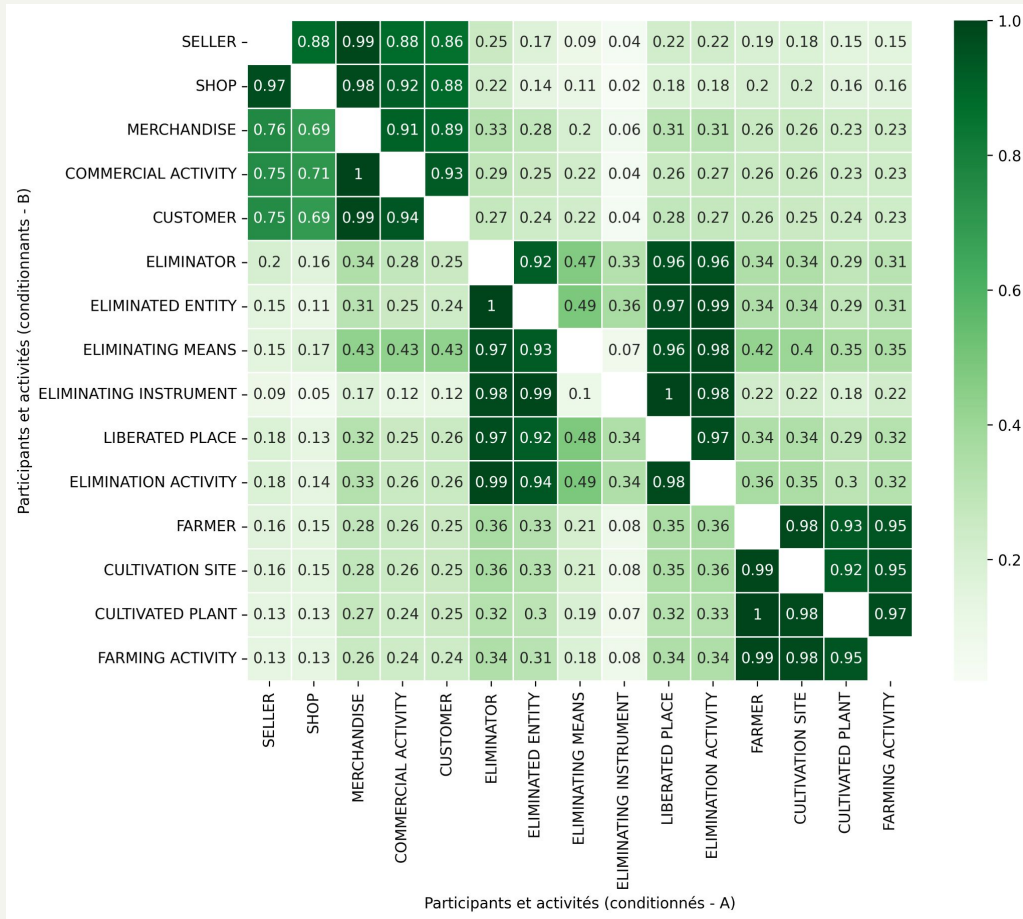
$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ when } P(B) > 0.$$

**Exemple** : la probabilité d'avoir un acheteur dans une histoire, étant donné la présence d'une boutique

# LLMs pour annoter des histoires

**Probabilités conditionnelles pour un rôle A d'être instancié dans une histoire étant donné la présence de B.**

- Trois zones d'interprédictibilité distinctes
- Moyen d'élimination et Instrument d'élimination en distribution presque complémentaire
- Vendeur et magasin pas systématiquement mentionnés dans les histoires



# LLMs pour annoter des histoires

## Parfois l'acheteur et la boutique ne sont pas mentionnés dans l'histoire

*"Les jardins de Marie-Rose, autrefois animés, étaient désormais déserts et abandonnés, laissant apparaître des flaques d'eau dans les fissures du sol craquelé. Déterminée à redonner vie à ce lieu délaissé, Marie-Rose décida de l'enherber. Elle acheta un sac de graines d'herbe fine et commença à semer des touffes dans les espaces dégagés. Au fil des jours, la verdure s'installa, recouvrant peu à peu le sol, et les flaques d'eau disparurent sous un épais tapis de feuilles nouvelles [...]."*

## L'instrument et le moyen d'élimination sont en distribution presque complémentaire

- [...] L'homme chargé de désherber les jardins, un vieux paysan nommé Pierre, est arrivé avec son outil préféré : une longue pelle en bois. Il a commencé à désherber le parc, enlevant soigneusement les herbes indésirables pour laisser place aux fleurs et aux arbres [...].
- [...] Il décida d'utiliser le désherbant qu'il avait acheté la semaine précédente. Il alla se placer au centre du jardin et versa le produit sur les parties infestées. Les mauvaises herbes commençaient à se recroqueviller et finissaient par mourir.

# LLMs pour annoter des histoires

3 LLMs de taille et “compétences” comparable utilisés pour l’annotation

- LLAMA 3 (8B)
- MISTRAL (8B)
- WIZARDLM (7B)

## À propos du prompt :

- Instructions en anglais
- Baser les décisions sémantiques sur les définitions des rôles et annoter uniquement par rapport à ces rôles
- Si un rôle sémantique est réalisé dans l’histoire, indiquer l’élément syntaxique qui l’instancie. Sinon insérer “-”
- Produire les annotations sous forme de dictionnaire Python et ensuite le transcrire dans la grille
- Deux séances d’annotation : (1) avec définitions (2) avec définitions et exemples (2 tours chacune)

# LLMs pour annoter des histoires

## Exemples fournis (3 de 7) :

```
{  
  "story": "Marc achète un désherbant dans un magasin de jardinage pour éliminer les mauvaises herbes dans le parc.",  
  "SELLER": "-",  
  "SHOP": "magasin de jardinage",  
  "MERCHANDISE": "un désherbant",  
  "CUSTOMER": "Marc",  
  "COMMERCIAL ACTIVITY": "achète",  
  "ELIMINATOR": "Marc",  
  "ELIMINATED ENTITY": "les mauvaises herbes",  
  "ELIMINATING MEANS": "un désherbant",  
  "ELIMINATING INSTRUMENT": "-",  
  "LIBERATED PLACE": "parc",  
  "ELIMINATION ACTIVITY": "éliminer",  
  "FARMER": "-",  
  "CULTIVATION SITE": "-",  
  "CULTIVATED PLANT": "-",  
  "FARMING ACTIVITY": "-"  
},
```

```
{  
  "story": "Le fermier utilise un pelle en bois pour éliminer les mauvaises herbes dans son champ de blé.",  
  "SELLER": "-",  
  "SHOP": "-",  
  "MERCHANDISE": "-",  
  "CUSTOMER": "-",  
  "COMMERCIAL ACTIVITY": "-",  
  "ELIMINATOR": "Le fermier",  
  "ELIMINATED ENTITY": "les mauvaises herbes",  
  "ELIMINATING MEANS": "-",  
  "ELIMINATING INSTRUMENT": "une longue pelle en bois",  
  "LIBERATED PLACE": "son champ de blé",  
  "ELIMINATION ACTIVITY": "éliminer les mauvaises herbes",  
  "WILD ANIMAL-MALE": "-",  
  "WILD ANIMAL-FEMALE": "-",  
  "WILD ANIMAL-OFFSPRING": "-",  
  "FARMER": "Le fermier",  
  "CULTIVATION SITE": "son champ de blé",  
  "CULTIVATED PLANT": "du blé",  
  "CULTIVATED FRUIT": "-",  
  "FARMING ACTIVITY": "-"  
},
```

```
{  
  "story": "Au marché, un fermier vend des paniers d'herbes qu'il cultive dans son jardin.",  
  "SELLER": "Un fermier",  
  "SHOP": "le marché",  
  "MERCHANDISE": "des paniers d'herbes",  
  "CUSTOMER": "-",  
  "COMMERCIAL ACTIVITY": "vend",  
  "ELIMINATOR": "-",  
  "ELIMINATED ENTITY": "-",  
  "ELIMINATING MEANS": "-",  
  "ELIMINATING INSTRUMENT": "-",  
  "LIBERATED PLACE": "-",  
  "ELIMINATION ACTIVITY": "-",  
  "FARMER": "Un fermier",  
  "CULTIVATION SITE": "son jardin",  
  "CULTIVATED PLANT": "herbes",  
  "FARMING ACTIVITY": "cultive"  
},
```

# LLMs pour annoter des histoires

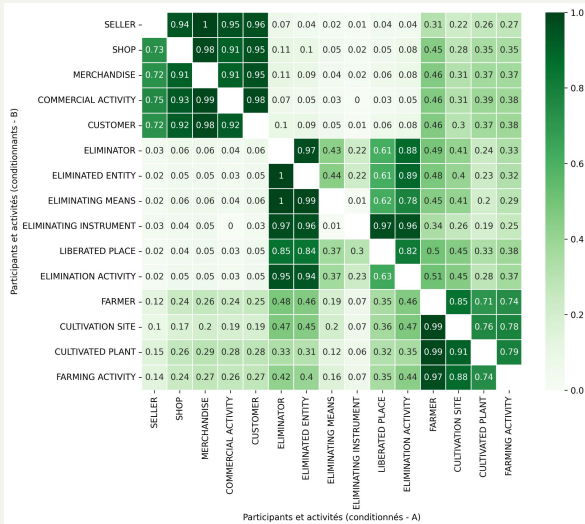
Résultats séance 1 : en utilisant seulement les définitions dans le prompt



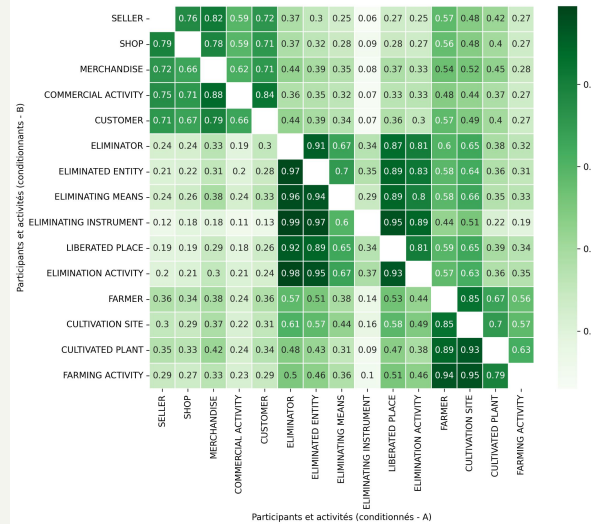
# LLMs pour annoter des histoires

Résultats séance 1 : en utilisant seulement les définitions dans le prompt

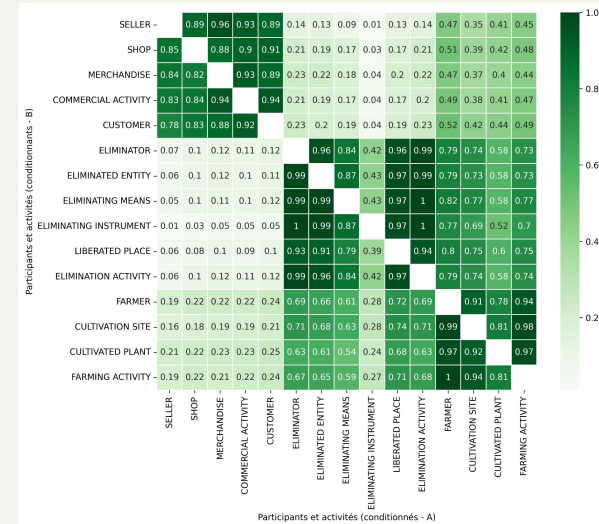
**MISTRAL (K=0.654)**



**LLAMA3 (K= 0.641)**



**WIZARDLM (K=0.619)**





# LLMs pour annoter des histoires

**Résultats séance 1:** tours d'annotations en utilisant les définitions et en donnant des exemples

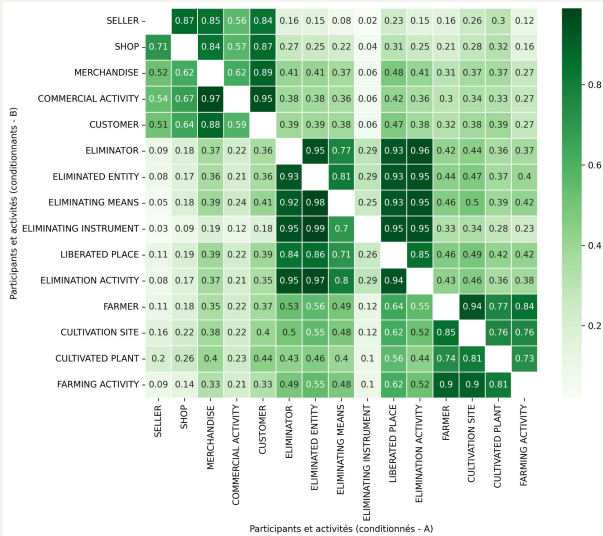


- Globalement, les scores augmentent légèrement (Llama et Mistral) ou sensiblement (Wizard)
- MAIS pour certains scores, on a une baisse (e.g. ELIMINATING MEANS pour Mistral)

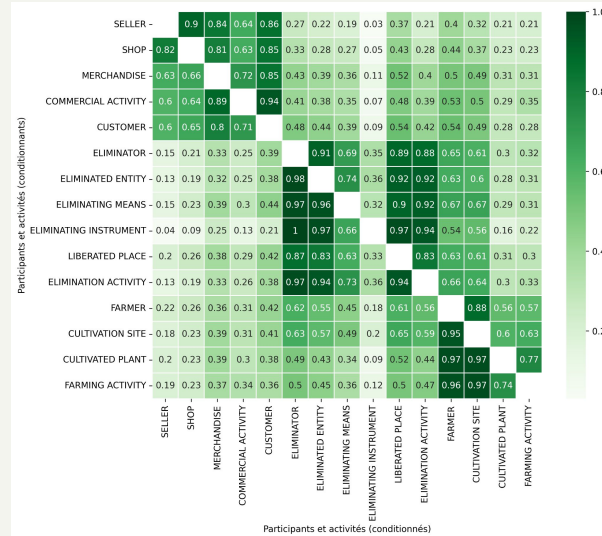
# LLMs pour annoter des histoires

Résultats séance 1: tours d'annotations en utilisant les définitions et en donnant des exemples

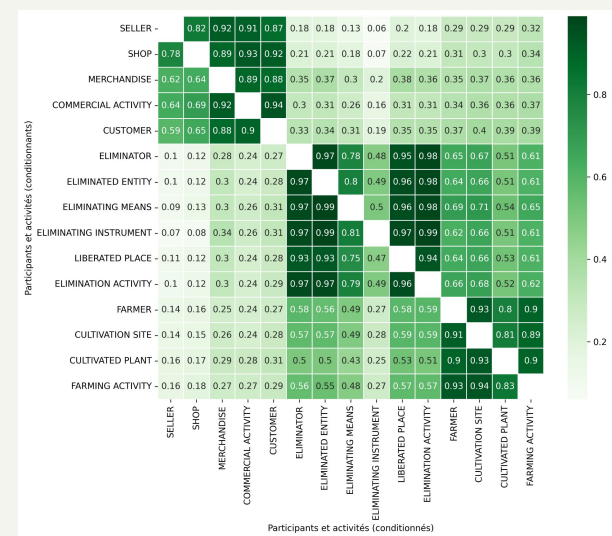
MISTRAL (K=0.709)



LLAMA 3 (K=0.662)



WIZARDLM (K=0.684)



# LLMs pour annoter des histoires

**Résultats séance 2:** tours d'annotation en utilisant les définitions dans le prompt

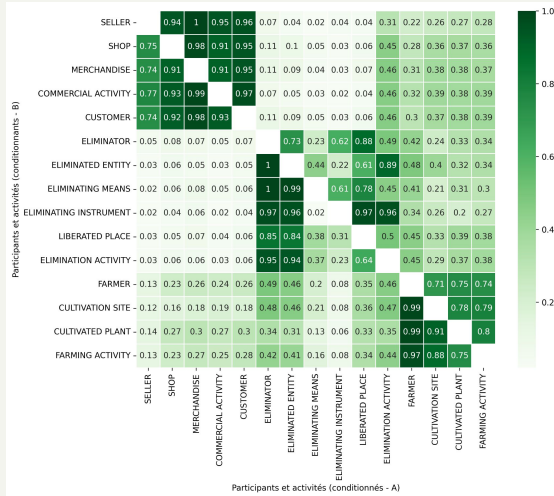


- Globalement, les scores sont similaires au premier tour sans exemples

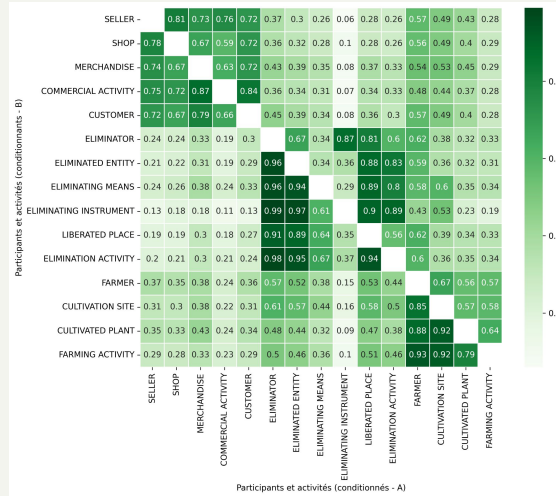
# LLMs pour annoter des histoires

Résultats séance 2 : tour d'annotations en utilisant les définitions dans le prompt

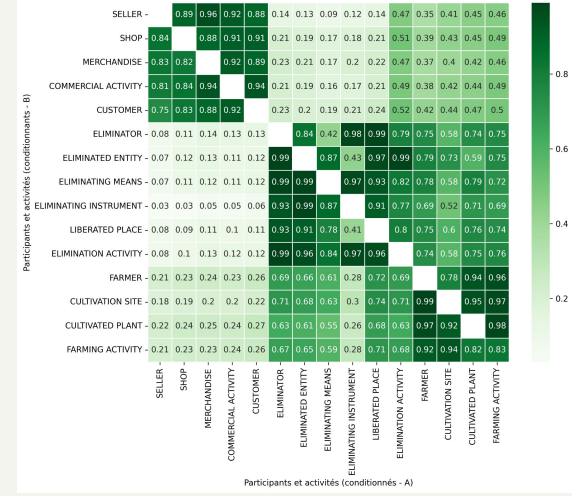
MISTRAL (K=0.674)



LLAMA3 (K=0.651)



WIZARDLM (K=0.631)



# LLMs pour annoter des histoires

**Résultats séance 2:** tours d'annotations en utilisant les définitions dans le prompt et les exemples

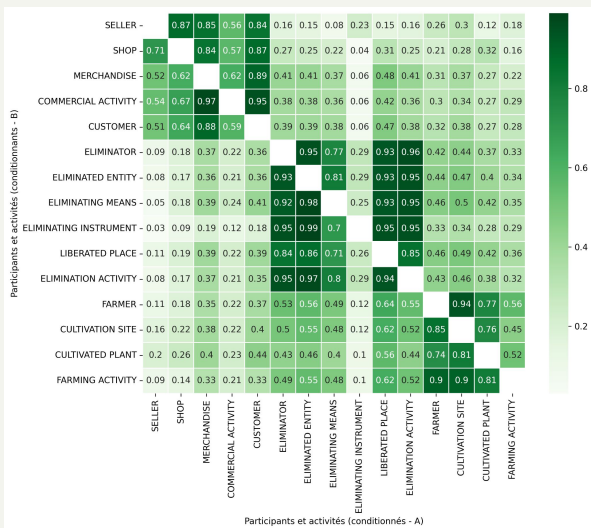


- Légère augmentation générale des scores, similairement au premier tour avec exemples
- Toujours avec exceptions (ex. ENDROIT LIBÉRÉ pour Mistral, -0.05)

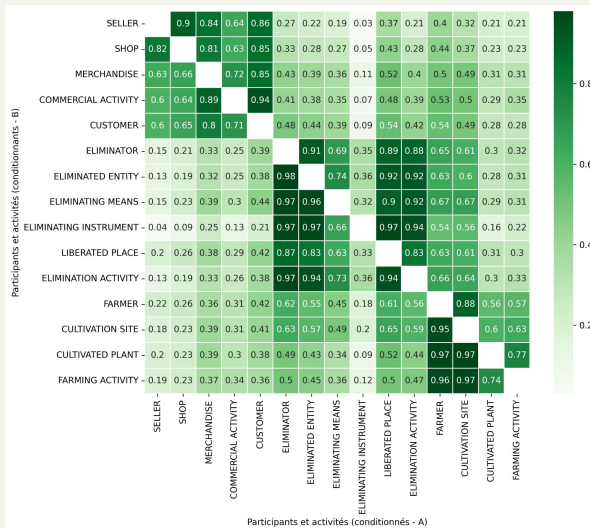
# LLMs pour annoter des histoires

Résultats séance 2 : tours d'annotations en utilisant les définitions dans le prompt et les exemples

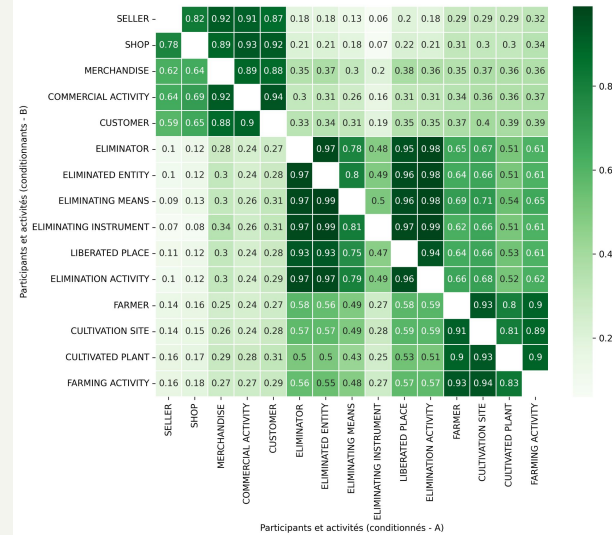
MISTRAL (K= 0.699)



LLAMA3 (K=0.669)



WIZARD (K=0.689)



# LLMs pour annoter des histoires

## Rôles problématiques : ELIMINATING INSTRUMENT, MEANS

[...] Mme Dupont décida donc de **désherber soigneusement le terrain** pour enlever ces herbes invasives. **Avec patience et méthode**, elle écarta les herbes inutiles [...]

**MEANS** : méthode et patience



[...] Elle avait finalement décidé de se lancer dans une opération de **désherbage** pour rétablir l'équilibre écologique. Un matin d'été, elle sortit son **désherbeur électrique** et se mit au travail pour désherber le jardin.

**MEANS** : désherbeur électrique



[...] Alors, Monsieur Martin avait décidé **d'utiliser un désherbant** pour éliminer ces plantes invasives. Il avait choisi un produit naturel et écologique, qui ne pourrait pas nuire aux herbes utiles. Le lendemain, il était sorti avec son **sac de désherbant** et s'était mis à traiter les mauvaises herbes. [...]

**INSTRUMENT** : son sac de désherbant



# LLMs pour annoter des histoires

## Rôles problématiques : FARMER, FARMING ACTIVITY

*Au cœur du bourg, une herboristerie offrait à ses habitants les dernières recettes de médicaments naturels et vendait également des désherbants. Le propriétaire, un **herboriste expérimenté**, allait chaque matin dans les bois pour **récolter les plantes sauvages les plus rares**.*

**FARMING ACTIVITY** : récolter les herbes



*Le fermier Jacques avait toujours été passionné par l'enherbement. Il estimait que **cette pratique agricole** était essentielle pour équilibrer les sols et favoriser la croissance des plantes à venir. Un jour, il décida de tester ses connaissances en **enherbant un terrain vaste et stérile** qui appartenait à son voisin. Il s'y rendit avec son outil préféré, un semoir électronique, et commença à planter une variété d'herbe sauvage.*

**FARMING ACTIVITY** : -





# LLMs pour annoter des histoires

**Rôles problématiques :** LIBERATED PLACE, ELIMINATED ENTITY

[...]Pierre avait **acquis un nouveau terrain** en friche, situé à la lisière du bois. Il avait alors entrepris de **l'enherber pour y réintroduire des espèces végétales résilientes**. Grâce à son expertise, le terrain était devenu un jardin verdoyant et fertile, prêt à produire des récoltes abondantes. L'enherbement avait ainsi permis l'émergence d'un nouveau champ, riche en biodiversité et en potentiel agricole.

**CLEARED PLACE :** terrain



La semaine dernière, une femme âgée est venue le trouver pour **demande de l'aide pour soigner sa mère qui souffrait d'une toux grave**. Jacques a recommandé un **remède fait à partir de thym et de menthe fraîche**, qu'il avait cueilli lui-même dans son jardin. La femme a été si impressionnée par la **rapidité de la guérison** que désormais, elle achète régulièrement des herbes médicinales chez Jacques pour ses amis et sa famille.

**CLEARED PLACE :** mère de la femme



**ELIMINATED ENTITY :** toux grave



# Conclusion et suite

# Conclusions (préliminaires)

- Les LLMs sont considérables comme pertinents pour une tâche d'estimation de l'interpédicibilité (production des histoires et annotation)
  - Toutes les annotations automatiques produites ont un accord satisfaisant ( $K > 0.60$ ) avec la première annotation manuelle effectuée
  - Une partie des annotations réalisées automatiquement identifie trois zones d'interpédicibilité distinctes (trois "paradigmes sémantiques")
- MAIS:**
- Certains rôles sémantiques posent plus problèmes que d'autres
  - Les mesures présentées prennent uniquement en compte la réalisation des rôles, pas les erreurs d'attribution
  - nécessite d'étudier davantage la stabilité de ces performances et le rôles des exemples
  - Nécessité d'étudier les performances sur plus de rôles et situations

# Conclusions (préliminaires)

L'estimation de la prédictibilité via les LLMs utilisés, surtout lorsque on fournit des exemples, permet d'identifier trois paradigmes sémantiques que l'on peut "projeter" sur la famille de départ pour la découper en trois familles paradigmatiques :

- HERBE, HERBORISTE, HERBORISTERIE
- HERBE, DÉSHERBER, DÉSHERBANT, DÉSHERBAGE, DÉSHERBEUR
- HERBE, ENHERBER, ENHERBEMENT, ENHERBEUR

**MAIS il s'agit d'un premier résultat qui doit être confirmé en reproduisant et améliorant la méthode**

# La suite

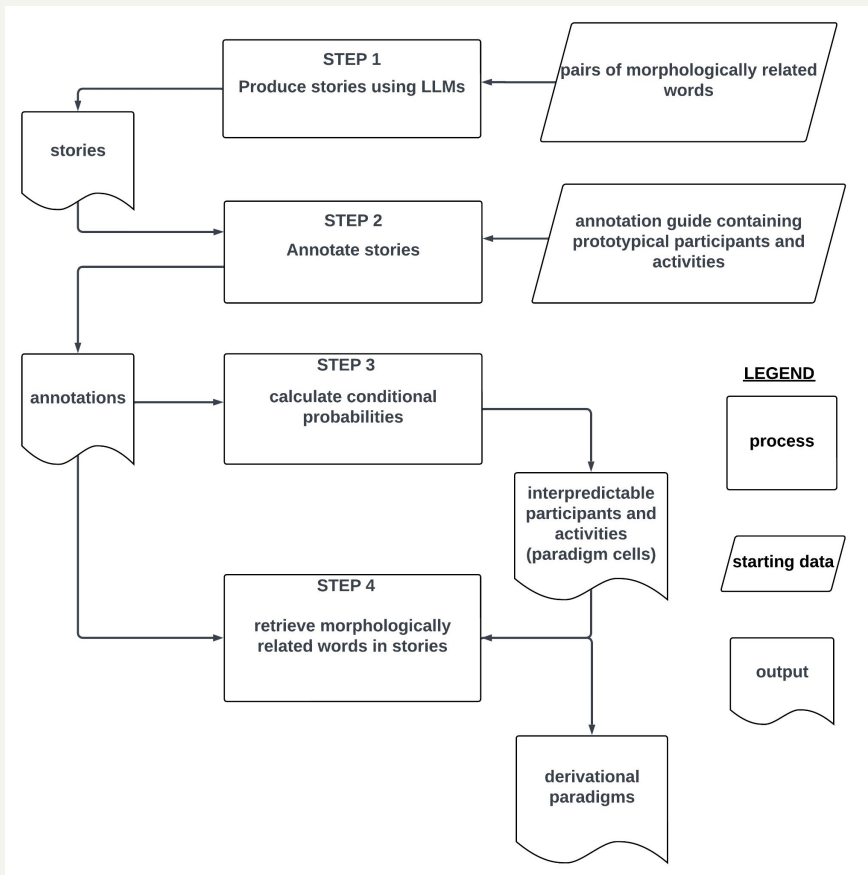
## Améliorer la procédure

- Tester différents prompts et formats d'exemples et faire plusieurs séances pour voir la stabilité des résultats
- Récolter des histoires produites par des humains pour étudier les similarités ou différences avec celles des LLMs

## Étendre la mise en œuvre

- Appliquer la méthode à un nombre important de familles de mots du lexique
- Étudier dans quelle mesure les histoires permettent d'aligner des familles de mots distinctes
- Étendre le guide d'annotation et introduire d'autres rôles et activités

# La suite - automatisation complète



- L'objectif à long terme est d'automatiser toutes les étapes de cette procédure
- La procédure est conçue pour être appliquées à toutes les familles du lexique
- Pour cela, il est fondamental d'étendre le guide d'annotation

**MERCI POUR VOTRE ATTENTION !**

# Références

- Baker, C. F., Fillmore, C. J., & Lowe, J. B. (1998). The Berkeley frameNet project. In *COLING 1998 Volume 1: The 17th International Conference on Computational Linguistics*.
- Bauer, L. (2019). Notions of paradigm and their value in word-formation. *Word structure*, 12(2), 153-175.
- Beniamine, S. (2018). Classifications flexionnelles. *Étude quantitative des structures de paradigmes* (Doctoral dissertation, Université Sorbonne Paris Cité–Université Paris Diderot (Paris 7)).
- Bonami, O., & Strnadová, J. (2019). Paradigm structure and predictability in derivational morphology. *Morphology*, 29(2), 167-197.
- Dokulil, M. (1982). Kotázce slovnědruhových převodů a přechodů, zvl. transpozice. *Slovo a slovesnost*, 43(4), 257-271.
- Fillmore, C. J. (1976). Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, 280(1), 20-32.
- Fradin, B. & G. Winterstein (2012). *Tuning agentivity and instrumentality: deverbal nouns in -oir revisited*. Paper delivered at *Décembrettes*, 8. 6-7
- Gaeta, L. (2022). Dangerous liaisons. An introduction to derivational paradigms. In *Paradigms in Word Formation*, 3-18. Amsterdam: John Benjamins.



# Références

- Hathout, N., & Namer, F. (2019). Paradigms in word formation: what are we up to?. *Morphology*, 29, 153-165.
- Hathout, N., & Namer, F. (2022). ParaDis: a family and paradigm model. *Morphology*, 32(2), 153-195.
- Namer, F., Hathout, N., Amiot, D., Barque, L., Bonami, O., Boyé, G., ... & Tribout, D. (2023). Démonette-2, a derivational database for French with broad lexical coverage and fine-grained morphological descriptions. *Lexique*, 33, 6-40.
- Petruck, M. R. (1996). Frame semantics. *Handbook of pragmatics*, 2.
- Roché, M. (2023). Les familles dérivationnelles: comment ça marche? *Lexique. Revue en Sciences du Langage*, (33).
- Ruppenhofer, J., M. Ellsworth, M. Schwarzer-Petruck, C.R. Johnson & J. Scheffczyk (2016). *Framenet II: Extended theory and practice*. Technical report, The Berkeley FrameNet Project.
- Sanacore, D. (2023). *Une histoire de famille: description morphosémantique des lexèmes construits et des relations dérivationnelles*. Ph.D. thesis, Université Toulouse le Mirail-Toulouse II.
- Štekauer, P. (2014). Derivational paradigms. In *The Oxford Handbook of Derivational Morphology*, 354±369. Oxford: Oxford University Press
- Stump, G.T. (1991). A paradigm-based theory of morphosemantic mismatches. *Language*. 675-725