

> Annotation des stades phénologiques et cultures dans les Bulletins de Santé du Végétal de la vigne

Atelier INOVIVE - Juillet 2023



Plan de la présentation

Corpus des BSV

Ressources

Caractérisation des ambiguïtés

Méthode de désambiguïstation



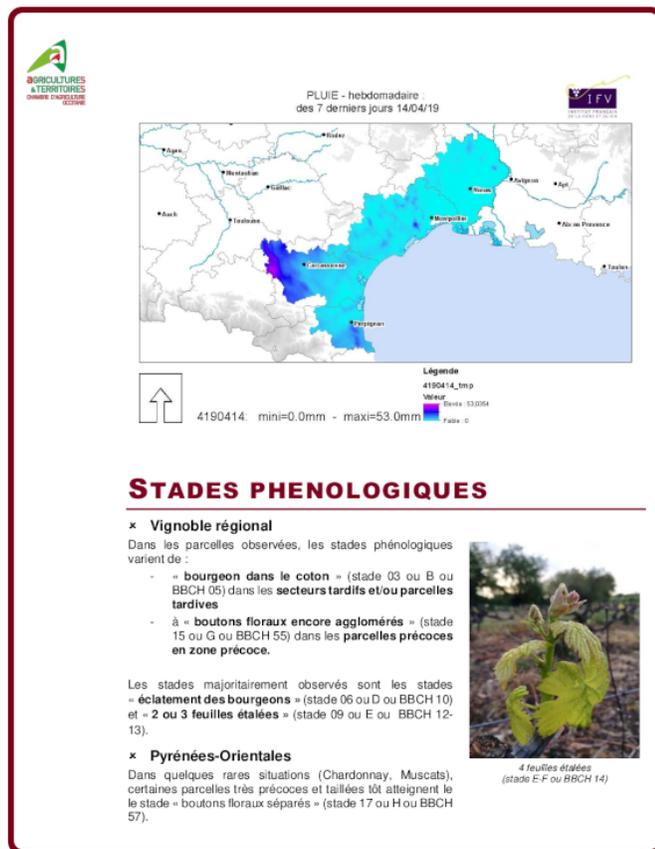
Corpus des BSV

Bulletin de Santé du Végétal (BSV)

C' est un **document d'information technique et réglementaire**, rédigé sous la responsabilité des Chambre Régionales d'Agriculture.

Il est destiné à fournir aux agriculteurs **des éléments de situation phytosanitaire** et d'**analyses de risque de qualité** pour les principales filières végétales de la région.

Bulletin de Santé du Végétal (BSV)



PDF

HTML

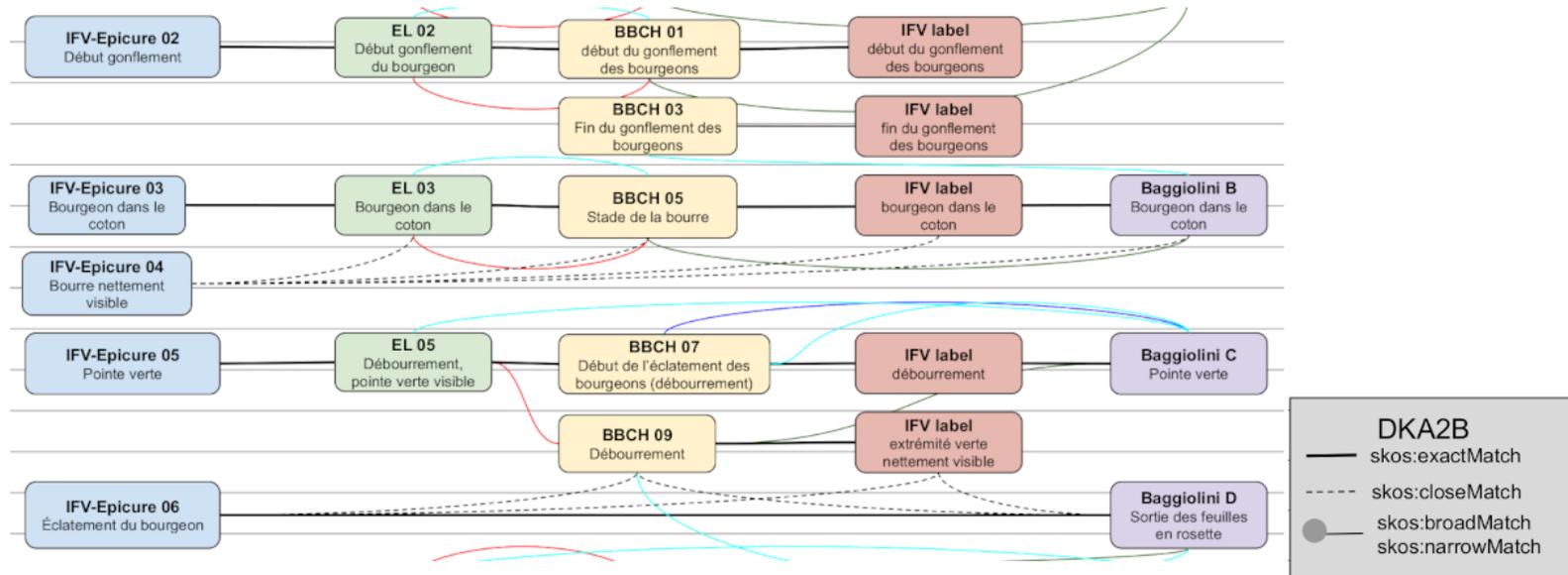
```
<h1 class="title 1" blk="32">STADES PHENOLOGIQUES</h1>
<h3 class="title 3" blk="33"> Vignoble régional</h3>
<p class="paragraph" blk="34">Dans les parcelles observées,
les stades phénologiques varient de :<br />-
" bourgeon dans le coton " (stade 03 ou B ou BBCH 05)
dans les secteurs tardifs et/ou parcelles tardives<br />-
à " boutons floraux encore agglomérés " (stade 15 ou G ou
BBCH 55) dans les parcelles précoces en zone précoce.</p>
<p class="paragraph" blk="35">Les stades majoritairement observés
<h3 class="title 3" blk="36"> Pyrénées-Orientales</h3>
```

Sous-corpus D2KAB Viticulture

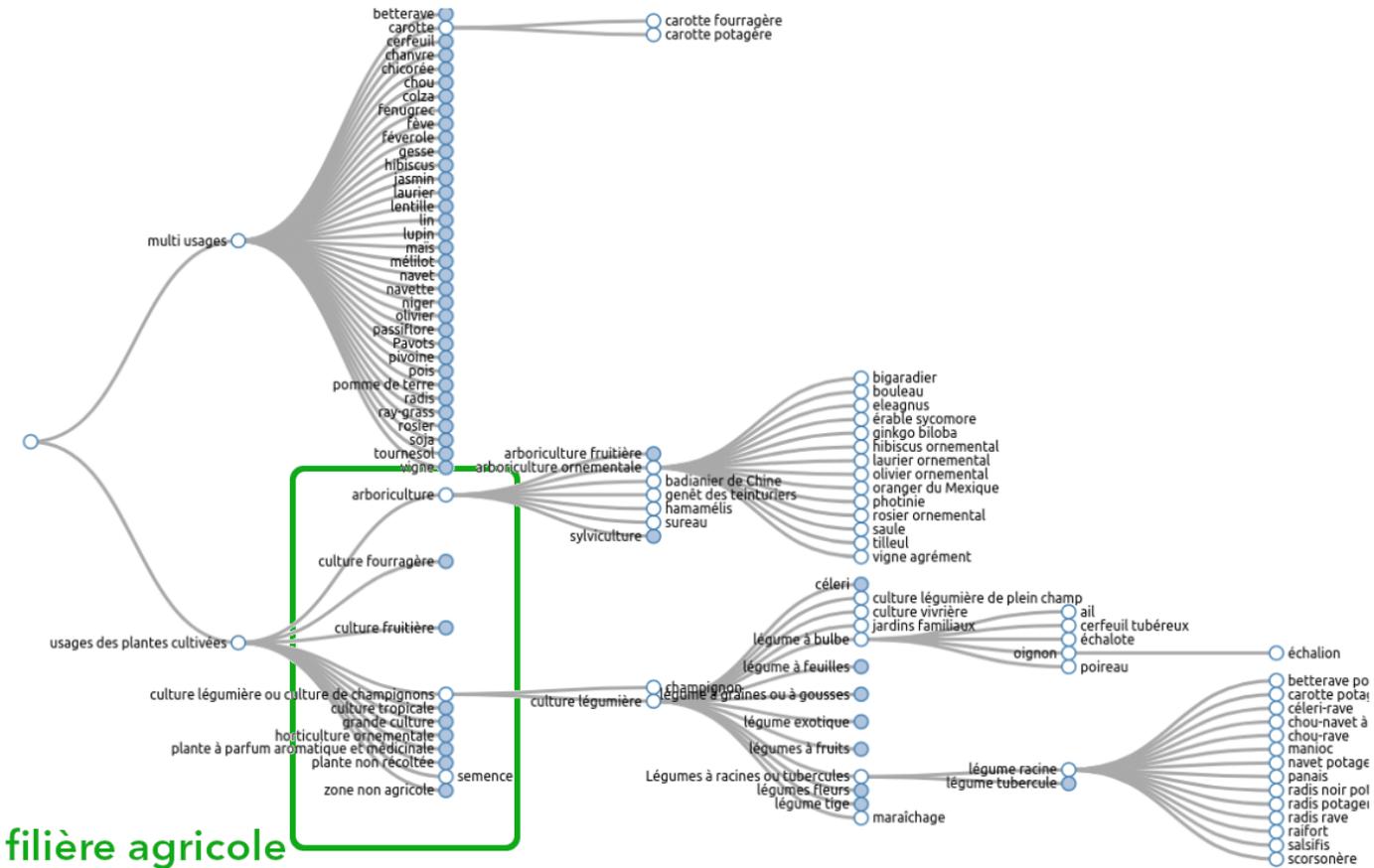
- 77 bulletins concernant la vigne
- provient du corpus de test D2KAB
- 2018-2019
- toutes les régions viticoles

Ressources

Ontologie et graphe de connaissance PPDO



Thésaurus FCU



Traitements

- PDF2block (PDF -> HTML)
- Segmentation, lemmatisation, POS tagging
- Chaîne de traitement AlvisNLP pour le repérage et la normalisation des entités

Caractérisation des ambiguïtés

Ambiguïté

Un segment du texte est identifié à tort comme appartenant à un certain type d'entité (Faux Positif).

*Le stade moyen actuel est **petits pois** -> Pas d'ambiguïté, on a bien affaire à un stade phénologique*

*Nous avons observé des attaques de mildiou sur des **petits pois** en particuliers dans des exploitations en agriculture biologique -> Ambiguïté, ici *petits pois* ne désigne pas un stade phénologique.*

Stades phénologiques

- stade phénologique, dans une échelle qui n'est pas celle de la vigne (*floraison*).
- culture (*petits pois*).
- stade de développement d'un bioagresseur (*maturité des larves*).
- symptôme lié à la présence d'un bioagresseur (*chute des feuilles*).

Cultures

- bioagresseurs (*champignon*),
- parties de plantes (*baies*)
- des symptômes liés à la présence d'un bioagresseur (*bois noir*)
- des descriptions météorologiques (*gel*)
- des noms de personnes (*véronique, olivier*).

Méthodes de désambiguïsation

Principe générale

Exploiter la structure de graphe entre les annotations d'un même bulletin pour identifier des entités qui pourraient être bruitées.

- Stades phénologiques : on s'intéresse aux voisinages (stades voisins intra ou inter-échelle)
- Cultures : on s'intéresse aux liens de parentés (notamment à la filière agricole)

Désambiguïsation des stades phénologiques

Hypothèse 1 : un stade phénologique est plus fiable si les stades proches sont également mentionnés dans le bulletin.

- Calculer la distance minimale entre chaque paire de stade
- Déterminer un voisinage dominant
- Retirer les mentions trop éloignées du voisinage dominant

Désambiguïisation des cultures

Hypothèse 2 : une mention de culture est plus fiable si la filière agricole à laquelle elle appartient est très présente dans le document.

- Aggrégation des mentions appartenant à une même filière agricole
- La fréquence conceptuelle de la filière agricole donne le score de confiance
- Retirer les mentions qui correspondent à des cultures dont la fréquence conceptuelle est trop faible

Résultats

	Mentions		Stades phénologiques	
	Corpus	Document	Corpus	Document
Annotation	4375	57	129	21,4
Désambiguïsé	4139	53,8	127	20,5

TABLE 1 – Annotation des stades phénologiques : nombres de mentions et stades phénologiques distincts annotés (dans le corpus et par bulletin), avant et après désambiguïstation.

	Mentions		Cultures	
	Corpus	Document	Corpus	Document
Annotation	1789	23.2	61	5.5
Désambiguïsé	1640	21.3	49	4.4

TABLE 2 – Annotation des cultures : nombres de mentions et cultures distinctes annotées (dans le corpus et par bulletin), avant et après désambiguïstation.

À venir

- Prendre en compte la sémantique des relations entre les noeuds (skos:exactMatch / ppdo:follows)
- Évaluer les résultats de désambiguïsation sur des entités validées manuellement

Merci de votre attention !

